ARTICLES

erbB-2, p53, and Efficacy of Adjuvant Therapy in Lymph Node-Positive Breast Cancer

Ann D. Thor, Donald A. Berry, Daniel R. Budman, Hyman B. Muss, Timothy Kute, I. Craig Henderson, Maurice Barcos, Constance Cirrincione, Susan Edgerton, Craig Allred, Larry Norton, Edison T. Liu

Background: **We have previously reported that high expression of the erbB-2 gene (also known as HER-2/neu and ERBB2) in breast cancer is associated with patient response to dose-intensive treatment with cyclophosphamide, doxorubicin (Adriamycin), and 5-flurouracil (CAF) on the basis of short-term follow-up of 397 patients (set A) with axillary lymph node-positive tumors who were enrolled in Cancer and Leukemia Group B (CALGB) protocol 8541.** ***Methods:*** **To validate those findings, we conducted immunohistochemical analyses of erbB-2 and p53 protein expression in an additional cohort of 595 patients (set B) from CALGB 8541, as well as a molecular analysis of erbB-2 gene amplification in tumors from all patients (sets A and B). Marker data were compared with clinical, histologic, treatment, and outcome data.** ***Results:*** **Updated analyses of data from set A (median follow-up, 10.4 years) showed an even stronger interaction between erbB-2 expression and CAF dose, by use of either immunohistochemical or molecular data. A similar interaction between erbB-2 expression and CAF dose was observed in all 992 patients, analyzed as a single group. However, for set B alone (median follow-up, 8.2 years), results varied with the method of statistical analysis. By use of a proportional hazards model, the erbB-2 expression–CAF dose interaction was not significant for all patients. However, in the subgroups of patients randomly assigned to the high- or the moderate-dose arms, significance was achieved. When patient data were adjusted for differences by use of a prognostic index (to balance an apparent failure of randomization in the low-dose arm), the erbB-2 expression–CAF dose interaction was significant in all patients from the validation set B as well. An interaction was also observed between p53 immunopositivity and CAF dose.** ***Conclusions:*** **The hypothesis that patients whose breast tumors exhibit high erbB-2 expression benefit from dose-intensive CAF should be further validated before clinical implementation. Interactions between erbB-2 expression, p53 expression, and CAF dose underscore the complexities of predictive markers where multiple interactions may confound the outcome. [J Natl Cancer Inst 1998;90:1346–60]**

Randomized clinical trials have demonstrated modest survival improvement among breast cancer patients treated with adjuvant chemotherapy *(1–4)*. Several factors are related to outcome. They include combination chemotherapy *(4)*, dose and dose intensity *(5)*, sequencing of drug delivery agents and regimens *(6)*, and prognostic markers *(7)*. Reports of interactions between markers and treatment responsiveness or lack thereof have led to a separation of these factors into prognostic (independent of treatment) and predictive (interactive with treatment) categories *(8–11)*. The estrogen receptor (ER) is an example of a clinically important prognostic as well as predictive factor. Identification of molecular markers predictive of chemoresponsiveness allows for more selective and effective utilization of therapeutic agents.

The Cancer and Leukemia Group B (CALGB) 8541 trial *(5)* and a companion trial, 8869 *(10)*, demonstrated that patients assigned to a dose-intensive doxorubicin (Adriamycin)-based chemotherapy had significantly longer disease-free survival (DFS) and overall survival (OS) if their tumors exhibited high expression of the erbB-2 gene (also known as HER-2/neu and ERBB2).

This study was undertaken to explore further the hypothesis that there is a plausible interaction between erbB-2 expression and doxorubicin dose response. We analyzed additional tumors, included additional follow-up information, validated immunohistochemical analyses of erbB-2 protein expression with an independent assessment of immunohistochemical scoring and erbB-2 gene amplification, and assessed whether p53 (also known as TP53) gene expression, S-phase fraction, or DNA ploidy may contribute prognostic or predictive information.

This study consists of an analysis of the original 397 tumors previously reported *(10)* and 595 additional tumors. This analysis was not straightforward for a variety of reasons. Retrospective tumor accrual from primary and affiliate CALGB hospitals

Affiliations of authors: A. D. Thor, S. Edgerton, Evanston Hospital and Northwestern University, Evanston, IL; D. A. Berry, C. Cirrincione, Cancer and Leukemia Group B Statistical Office, Durham, NC; D. R. Budman, North Shore University Hospital, Manhasset, NY; H. B. Muss, Vermont Cancer Center, Burlington; T. Kute, Bowman Gray School of Medicine, Winston-Salem, NC; I. C. Henderson, University of California, San Francisco; M. Barcos, Roswell Park Cancer Center, Buffalo, NY; C. Allred, University of Texas at San Antonio; L. Norton, Memorial Sloan-Kettering Cancer Center, New York, NY; E. T. Liu, University of North Carolina at Chapel Hill.

Correspondence to: Ann D. Thor, M.D., Department of Pathology, Evanston Hospital, Evanston, IL 60201.

See ''Notes'' following ''References.''

was limited; tumor blocks from more than 500 patients could not be retrieved. Moreover, clinical management of breast cancer changed between 1985 and 1990; high-dose chemotherapy or transplant technologies were increasingly being used for high-risk patients, which reduced the numbers of high-risk patients in the later phase of the trial. An amendment to the CALGB 8541 protocol in 1988 recommended tamoxifen administration in a nonrandomized setting, which further complicated our analysis.

SUBJECTS AND METHODS

Patients

In CALGB trial 8541, 1572 women with stage II breast carcinoma were randomly assigned to receive one of three different regimens of adjuvant cyclophosphamide, doxorubicin (Adriamycin), and 5-fluorouracil (CAF): a high dose (cyclophosphamide at 600 mg/m^2, doxorubicin at 60 mg/m^2, and 5-fluorouracil at 600 mg/m^2 for four cycles), a moderate dose (cyclophosphamide at 400 mg/m^2, doxorubicin at 40 mg/m^2, and 5-fluorouracil at 400 mg/m^2 for six cycles), or a low dose (cyclophosphamide at 300 mg/m^2, doxorubicin at 30 mg/m^2, and 5-fluorouracil at 300 mg/m^2 for four cycles) on day 1 of a 28-day cycle, with the dose of 5-fluorouracil repeated on day 8 independent of hematologic values *(5).* For an update on the results of the clinical trial for the 1549 patients who completed treatment on the protocol (follow-up through December 1996), the reader is referred to the recent publication by Budman et al. *(12).* The patients in this study (CALGB 8869) were drawn from the larger trial of adjuvant chemotherapy (CALGB 8541). CALGB 8869 is a companion trial of translational studies. Follow-up for sets used in this study was extended through December 1997. (For set A, the median follow-up was 10.4 years; for set B, it was 8.2 years; for combined sets A and B, it was 9.3 years.)

Preliminary results published in 1994 *(10)* showed an association between erbB-2 expression and dose response to chemotherapy. That study included 397 patients enrolled early on in CALGB trial 8541 *(10).* In this analysis, these 397 patients are denoted as set A. Archived tumors from all other patients who completed therapy (n = 1152) were requested from participating institutions for validation of the earlier analysis. Only 616 additional tumor blocks could be retrieved; of these 616 blocks, 595 were assessable (denoted as set B in this analysis).

Tumor grading and histologic subtyping were performed by a reference pathologist using a modified nuclear grading schema of Black et al. *(13).* ER and progesterone receptor (PR) information was provided by participating institutions using their own reference laboratories (a combination of immunohistochemical and biochemical assays). Menopausal status was recorded from information provided by participating institutions.

Specimen Analysis

Blocks of 1013 primary breast cancers (combined sets A and B) were obtained. Each tissue block was sectioned for histologic analysis as previously described *(10,14);* residual invasive carcinoma was not identified in slides from 19 and 21 blocks submitted (requirement for inclusion in this analysis), resulting in 994 and 992 assessable samples, respectively, for erbB-2 and p53 analyses. This resulted in a retrospective tissue recovery rate of 64% (992 or 994 of 1549 patients, summarized in Table 1). Specifically excluded from entry in the clinical trial (CALGB 8541) were patients who had more than one tumor.

Immunohistochemical Analyses of erbB-2 and p53

Immunohistochemical analyses of erbB-2 and p53 were performed as described *(10).* Briefly, tissues were deparaffinized and rehydrated, and nonspecific reactivity was blocked with 10% normal horse serum (Vector Laboratories, Inc., Burlingame, CA), followed by addition of the anti-erbB-2 monoclonal antibody CB11 (BioGenex Laboratories, San Ramon, CA) diluted 1 : 900 in 0.01 *M* phosphate-buffered saline (PBS) (pH 7.4) overnight at 4 °C. After multiple PBS washes, tissue sections were sequentially incubated for 30-minute intervals at room temperature with biotinylated horse anti-mouse immunoglobulin G (1 : 500 dilution; Vector Laboratories, Inc.) and streptavidin–horseradish peroxidase (1 : 200; Zymed Laboratories, Inc., South San Francisco, CA) diluted in PBS. After multiple washes, peroxidase activity was visualized after reaction with diaminobenzidine (Sigma Chemical Co., St. Louis, MO), followed by counterstaining with hematoxylin, dehydration, and mounting under coverslips. The anti-erbB-2 monoclonal antibody CB11, which was reactive with the intracellular domain of erbB-2, was used for immunohistochemical staining of set B tumors because the polyclonal reagent used for set A was no longer available. After optimizing the assay with CB11, breast cancer cell lines MDA-MB-231 and MDA-MB-453 (American Type Culture Collection [ATCC], Manassas, VA), as well as fixed, embedded breast cancers from 20 non-CALGB patients (previously stained with polyclonal anti-erbB-2 antibody), were stained with the use of CB11. Slides with breast cancer were scored for percent positive tumor cells (from the monoclonal antibody assay) without knowledge of prior scores from the polyclonal antibody assay. Comparison of the staining showed a high correlation (correlation coefficient $r = .91$; $P<.001$; Thor A: unpublished data). Furthermore, data from the CALGB 8869 slides stained with either antibody were correlated similarly with the independently derived erbB-2 gene amplification data (*see* ''Results'' section). Control cell lines MDA-MB-231 and MDA-MB-453 were included with each assay. Stained slides were scored separately by two investigators, who each estimated the percentage of invasive tumor cells with membranous staining [as previously reported *(10)*].

Monoclonal anti-p53 antibody (PAb1801; Genesis Bio-Pharmaceuticals, Inc., Tenafly, NJ) was used as described *(10,14).* Previous studies by our laboratory *(14,15),* using this reagent and direct sequencing of p53 gene from breast or ovarian cancer tissues, have shown a high correlation between PAb1801 immunostaining and missense p53 mutations. The p53 immunopositivity was estimated as a percentage of tumor cells showing positive nuclear staining.

Differential Polymerase Chain Reaction for erbB-2 Gene Amplification

A hematoxylin–eosin-stained slide from each tumor block was mapped for cancer–stroma distribution *(10).* Thick sections (50 μm) were trimmed by use of this map to enhance the tumor-to-nontumor ratio for flow cytometry, and 10-μm sections were similarly microdissected for differential polymerase chain reaction (PCR) analysis. Molecular analyses for erbB-2 were successfully completed on 916 specimens by use of previously published approaches and algorithms *(16,17).* Two of the three reference genes used as controls (interferon alfa, interferon gamma, and N-ras) had to be judged unequivocally positive by two independent observers before a sample was deemed to be harboring gene amplification. Cell lines (e.g., SKBR3 from ATCC) with erbB-2 amplification were used as standards. Ratios of target-to-reference band intensities were compared. A greater than twofold increase in copy number by use of the normal tissue adjacent to the tumor as the control was considered to represent gene amplification and has been shown to be strongly correlated with high immunohistochemical expression [*(16); see* ''Results'' section).]

Technical Validation of Immunoassays

The relevance of a recent report of rapid p53 antigen degradation with storage *(18)* was of concern in the present study. So that we could address this issue, 24 of the 992 cases originally scored in 1991 and stored as tissue sections mounted on glass slides for 2–5 years were selected for restaining with anti-erbB-2 and anti-p53 antibodies. Correlation coefficients between the original and restained slides were very high: .71 for p53 and .92 for erbB-2. A significant reduction in immunostaining between the two time points was not discerned.

To estimate interobserver variability, a second pathologist (C. Allred) rescored 194 slides from the original 397 tumors (originally scored by A. D. Thor) in a blinded fashion. This pathologist used his own scoring system, which recorded the percent staining, the intensity of the staining, and the total score for each invasive breast cancer *(19).*

Flow Cytometry

Nine hundred eighteen tumors were assessable by flow cytometry for S-phase fraction and ploidy as described *(10).* Briefly, by use of prepared tumor maps, malignant and benign tissues were dissected apart and separately deparaffinized. The histogram of the benign tissue was compared with the histogram of the malignant tissue to correct for fixation and processing artifacts. The diploid DNA standard was obtained from the nonmalignant tissue from each tumor slide. The DNA index was obtained by a comparison of the ratios of the G_1 peaks of the malignant and benign tissue fractions. The S-phase fraction was obtained by use of the same rectangular fit model used previously *(10).*

Table 1. Patient and tumor characteristics of subjects enrolled in CALGB 8541 and 8869 clinical trials*

Characteristic	Companion trial 8869 (all patients)			Adjuvant trial 8541
	Low-dose arm	Moderate-dose arm	High-dose arm	
No. of subjects	327	340	346	1572
Age, y				
Median	50	50	50	50
Range	23–77	26–81	24–77	23–81
	No. (%)†			
Menopausal status				
Premenopausal	132 (40)	142 (42)	139 (40)	673 (43)
Postmenopausal	195 (60)	198 (58)	207 (60)	879 (56)
Unknown	—	—	—	20 (1)
Tumor size, cm				
⩽2	115 (35)	127 (37)	119 (34)	555 (35)
>2	211 (65)	212 (62)	224 (65)	994 (63)
Unknown	1 (<1)	1 (<1)	3 (1)	23 (2)
No. of positive lymph nodes				
1–3	196 (60)	201 (59)	194 (56)	927 (59)
4–9	96 (29)	104 (31)	110 (32)	456 (29)
⩾10	35 (11)	35 (10)	42 (12)	175 (11)
Unknown	—	—	—	14 (1)
Histologic type				
IDC	289 (88)	310 (91)	321 (93)	NA
LC	16 (5)	14 (4)	10 (3)	NA
Other	16 (5)	12 (4)	5 (1)	NA
Unknown‡	6 (2)	4 (1)	10 (3)	NA
Histologic grade				
Well differentiated	5 (2)	7 (2)	4 (1)	NA
Moderately differentiated	143 (44)	143 (42)	136 (39)	NA
Poorly differentiated	173 (53)	186 (55)	195 (56)	NA
Unknown	6 (2)	4 (1)	11 (3)	NA
Hormone receptors				
ER positive	210 (64)	219 (64)	232 (67)	1006 (64)
ER negative	114 (35)	116 (34)	107 (31)	527 (34)
Unknown	3 (1)	5 (1)	7 (2)	39 (2)
PR positive	179 (55)	194 (57)	180 (52)	844 (54)
PR negative	135 (41)	131 (39)	145 (42)	645 (41)
Unknown	13 (4)	15 (4)	21 (6)	83 (5)
Tamoxifen treatment				
Yes	113 (35)	110 (32)	122 (35)	502 (32)§
No	213 (65)	230 (68)	224 (65)	1047 (67)
Unknown	1 (<1)	0 (0)	0 (0)	23 (1)
DNA ploidy				
Diploid	117 (36)	111 (33)	122 (35)	NA
Aneuploid	175 (54)	200 (59)	193 (56)	NA
Unknown	35 (11)	29 (9)	31 (9)	NA
S-phase fraction				
⩽10%	92 (28)	97 (29)	94 (27)	NA
>10%	114 (35)	122 (36)	140 (41)	NA
Unknown	121 (37)	121 (35)	112 (32)	NA
Median, %	12	12	13	NA
Range, %	0–50	2–44	1–50	NA
erbB-2 immunohistochemistry by % cells positive				
0	125 (38)	138 (41)	137 (40)	NA
1–9	60 (18)	57 (17)	60 (17)	NA
10–49	49 (15)	45 (13)	49 (14)	NA
50–89	35 (11)	34 (10)	43 (12)	NA
⩾90	53 (16)	57 (17)	50 (14)	NA
Unknown	5 (2)	9 (3)	7 (2)	NA
erbB-2 amplification				
Absent	252 (77)	246 (72)	258 (75)	NA
Present	52 (16)	58 (17)	50 (14)	NA
Indeterminant	4 (1)	9 (3)	10 (3)	NA
Unknown	19 (6)	27 (8)	28 (8)	NA

Table 1 (continued). Patient and tumor characteristics of subjects enrolled in CALGB 8541 and 8869 clinical trials*

Characteristic	Companion trial 8869 (all patients)			Adjuvant trial 8541
	Low-dose arm	Moderate-dose arm	High-dose arm	
	No. (%)†			
p53 expression by % cells positive				
0	213 (65)	231 (68)	218 (63)	NA
1–9	58 (18)	39 (11)	56 (16)	NA
10–29	17 (5)	22 (6)	25 (7)	NA
30–49	6 (2)	6 (2)	7 (2)	NA
⩾50	29 (9)	34 (10)	33 (10)	NA
Unknown	4 (1)	8 (2)	7 (2)	NA

*Only 1549 patients completed treatment (*see* ''Subjects and Methods'' section). 8541 is the parent trial as described in the introduction; 8869 is the companion translational trial, which, for the purposes of this analysis, has been divided into sets A, B, and A and B combined. Significant differences were not observed between the various arms on the companion trial or in comparison to the adjuvant trial. Low, moderate, and high doses refer to the doses of CAF (i.e., cyclophosphamide, doxorubicin [Adriamycin], and 5-fluorouracil). CALGB = Cancer and Leukemia Group B; IDC = infiltrating ductal carcinoma; LC = lobular carcinoma; ER = estrogen receptor; PR = progesterone receptor; and NA = this characteristic was not available on the full study 8541 but only on the subset study 8869.

†Unless otherwise specified, values in columns = number of subjects (%).

‡Slides were unavailable for histologic review.

§Tamoxifen administration increased following a clinical amendment in 1988. The increased proportion of patients who received tamoxifen on CALGB 8541 described earlier by Muss et al. *(10)* reflects longer follow-up and more accurate documentation of tamoxifen administration.

Statistical Analysis

Survival interval was defined as the period between the study entry and death (for OS) or the period between study entry and documented relapse or death without relapse (for DFS). Patients without an event were censored at last follow-up. Survival curves were drawn with the use of the Kaplan–Meier product limit method *(20–22)*. The logrank test was used to compare two or more survival distributions. A proportional hazards model was used to relate the various covariables with outcome *(23)*. We chose the variables CAF dose, number of positive lymph nodes, tumor size, menopausal status, erbB-2 expression, and CAF dose interaction with erbB-2 expression for the multivariate proportional hazards models to be the same as those used in our previous publication *(10);* however, we added tamoxifen therapy and p53 expression and a p53 interaction with CAF dose because of their statistical significance. *P* values from univariate and multivariate proportional hazards models were derived from Wald's chi-squared statistics. We used chi-squared test or Fisher's exact test to compare categorical variables. The Kruskal–Wallis test was used to compare continuous variables across dose levels.

Statistical analyses included patients in the initial set (set A) and the subsequently accrued group (set B). Set A was the hypothesis-generating set, and set B was the validation set. Patients accrued into set B (n = 595) differed from those accrued into set A (n = 397) in many respects, including the following: a later date of trial entry (83% in the former and 53% in the latter that were entered after 1987; $P<.001$), a greater number of postmenopausal women (64% in the former and 53% in the latter; $P = .001$), a greater number of women who received tamoxifen (39% in the former and 28% in the latter; $P = .001$), a higher median S phase (14% in the former and 10% in the latter; $P<.001$), and somewhat fewer patients with 10 or more positive lymph nodes (14% in the former and 9% in the latter; $P = .041$). There were no significant differences between these two groups in ER or PR positivity, erbB-2 expression or amplification, DNA ploidy, or p53 positivity.

The principal analysis addressed whether set B confirmed the conclusion from set A *(10)*. The hypothesized interaction of erbB-2 expression and dose intensity of CAF was assessed by use of a multivariate proportional hazards model incorporating the following variables: number of positive lymph nodes (square root transformation used for better predictability and linearity required for Cox model analyses), tumor size, ER status, menopausal status, tamoxifen use, dose of CAF (coded 0 [low], 1 [moderate], and 2 [high]), erbB-2 expression, and an interaction term, the product of dose of CAF (coded number) and erbB-2 percent positivity. We first updated this model for set A to address whether our original hypothesis remained valid with an additional 7 years of follow-up. In all cases of comparison with set B, we considered only the most recently available clinical data from set A (median follow-up, 10.4 years). The primary analysis of set B used a multivariate model with the same variables as those used for set A, with a focus on the interaction of CAF dose and erbB-2 expression. As in our earlier publication *(10)*, we illustrated the interaction using Kaplan–Meier survival curves by dose for low erbB-2 expression (<50% cells showing positive expression) and for high erbB-2 expression (⩾50% cells showing positive expression) for set A , set B, and sets A and B combined. Kaplan–Meier survival curves for p53 are separated into no p53 expression (0% cells showing positive expression) and any p53 expression (⩾1% cells showing positive expression). *P* values from the logrank test are provided for comparison of survival curves. *P* values from the Wald statistic derived from the multivariate models, however, address the interaction question in a much better way.

Kaplan–Meier survival curves can be misleading because they do not account for differences that occur in prognosis among the various groups. Such occurrences are likely in subgroups of patients because the sample sizes within subgroups are much smaller than in the entire set. Therefore, in a secondary analysis to address the validation question, we adjusted for differences in prognoses. To effect this adjustment, we developed a prognostic index using a multivariate proportional hazards model that accounted for the following variables: square root of number of positive axillary lymph nodes (NPN), tumor size (TSIZE = 2 for >2 cm; TSIZE = 1 for ⩽2 cm), menopausal status (PRE = 1 for premenopausal; PRE = 0 for perimenopausal or postmenopausal), tamoxifen use (TAM = 1 for yes; TAM = 0 for no), and ER status (ERPOS = 1 for positive; ERPOS = 0 for negative). Specifically not considered in this index were erbB-2 positivity, dose of CAF, and time of entry into the trial because these factors were specifically being investigated. Fitting a proportional hazards regression model with the use of the above factors to the combined DFS data in sets A and B gave the following prognostic index:

$$\text{Index} = \exp[(0.439 \times \sqrt{\text{NPN}}) + (0.470 \times \text{TSIZE}) + (0.241 \times \text{PRE}) - (0.413 \times \text{TAM}) + (0.113 \times \text{ERPOS})].$$

Greater index values and larger or positive correlation coefficients correspond to poorer prognosis and increased risk of recurrence or death, whereas smaller index values and negative or smaller correlation coefficients indicate better prognosis and decreased risk of recurrence or death. ER positivity and the use of tamoxifen were correlated. In this trial, while ER positivity had a favorable prognostic implication, the sign of the coefficient of ER positivity (ERPOS) suggests otherwise. The reason for this apparent discrepancy is that the favorable ER prognosis was carried by the variable tamoxifen treatment (TAM). Both terms ERPOS and TAM were included in the index equation because, although they are correlated, they neither are collinear nor give identical information, since all patients with ER-positive tumors did not receive tamoxifen and some patients with ER-negative tumors did, in fact, receive the drug.

In our second analysis, we used this index to adjust DFS for the various subgroups as follows. We calculated an overall mean index for all 992 patients

(combining 397 patients from set A with 595 patients from set B) that represented the average risk for these patients. We next computed the mean index for patients by subgroup (set A versus set B by CAF dose schedule and erbB-2 expression [high versus low]). The overall mean index was then divided by each subgroup's mean index. This ratio was a measure of the relative risk of the entire set of patients compared with the mean index of the subgroup in question. Each subgroup's Kaplan–Meier DFS was then adjusted by raising it to the power of the calculated ratio. For example, if a subgroup had an average risk that was only 80% of the average risk for the entire patient cohort, then the ratio was 1.25. The subgroup's DFS would then be adjusted downward by raising the DFS by a power of 1.25. In this example, a 50% survival probability (survival proportion = 0.5) would become 42% (survival proportion = $0.5^{1.25}$). For a different hypothetical subgroup with a mean risk of 125% (compared with the entire patient cohort), the survival probability would be raised to the power 0.80. For this subgroup, a 50% survival probability (survival proportion = 0.5) would become 57% (survival proportion = $0.5^{0.8}$) with the ratio adjustment. In effect, this process adjusts prognosis of all groups so that they are comparable and rectifies imbalances that may occur. Logrank statistics are not defined for adjusted Kaplan–Meier survival curves. We compared these curves by using a chi-squared test and assuming exponential survival distributions, i.e., assuming constant hazard, with the exponential parameter adjusted as indicated above.

RESULTS

Patient Characteristics, erbB-2, and p53

Table 1 summarizes the clinical and biologic data for all 1013 patients enrolled in CALGB 8869 by treatment arm compared with all 1572 patients who participated in adjuvant trial CALGB 8541. Overall, patients in each treatment group were balanced by variables listed in Table 1 and were similar to those in the parent clinical trial. For either set A or set B, the DFS and OS of patients randomly assigned to the high- or moderate-dose arms were superior to those of patients on the low-dose arm. Patients accrued into set B differed from those accrued into set A in having a later date of trial entry ($P<.001$), a greater proportion of patients who received tamoxifen ($P = .001$), a greater proportion of postmenopausal patients ($P = .001$), a higher median S phase ($P<.001$), and a trend toward fewer positive lymph nodes (described in detail in the ''Subjects and Methods'' section).

Of the 992 tumors (combined sets A and B) immunostained for erbB-2, 60% exhibited some membranous reactivity (≥1%) and 27% exhibited nearly homogeneous staining with 50% or more of cells positive (Table 1). Twenty-nine percent of set A tumors and 27% of set B tumors exhibited this high-level reactivity. PCR analysis of erbB-2 gene amplification from 916 cases revealed amplification in 17% (21% in set A and 16% in set B). The Pearson product moment correlation between erbB-2 expression and gene amplification was significant ($P<.001$). PCR-negative cases showed low levels of immunopositivity (mean staining, 17%) compared with PCR-positive cases (mean staining, 71%). Associations between other prognostic variables and erbB-2 alterations (with the use of either immunohistochemical or molecular analysis for erbB-2) were generally similar between sets A and B. Data are presented for the group at large (sets A and B combined). The erbB-2 alterations (immunohistochemical or gene amplification) were significantly associated with steroid receptor negativity (ER, $P = .002$; PR, $P<.001$), no treatment with tamoxifen ($P<.001$), histologic tumor type ($P = .005$), higher tumor grade ($P<.001$), and a higher percentage of cells in S phase ($P = .018$). Correlations between erbB-2 expression and amplification persisted without significant differences for sets A and B ($r = .45$ and $P<.001$ for set A; $r = .51$ and $P<.001$ for set B).

p53 immunostaining was positive (≥1%) in 33% of 994 assessable tumors (Table 1) and was not significantly correlated with erbB-2 expression. p53 positivity was associated with tumor size ($P = .034$), higher tumor grade ($P<.001$), steroid receptor negativity (ER, $P<.001$; PR, $P<.001$), aneuploidy ($P = .001$), and a higher percentage of cells in the S phase ($P<.001$).

Technical Validation of Immunoassays

Correlation coefficients between the original and subsequently stained slides taken from long-term storage (*see* ''Subjects and Methods'' section) were very high: .71 for p53 and .92 for erbB-2. A significant reduction in immunostaining between the two time points was not discerned. For the 194 slides that were blindly rescored by a second pathologist (C. Allred), the percent staining, the intensity of the staining, and the total score (described earlier in ''Subjects and Methods'' section) were strongly correlated with the scoring by A. D. Thor. The percent staining was the parameter with the greatest correlation (adjusted $R^2 = 78\%$; $P<.001$); however, also correlated with the original scoring by A. D. Thor were the intensity of staining (adjusted $R^2 = 59\%$; $P<.001$) and the total score (adjusted $R^2 = 76\%$; $P<.001$). Moreover, the two independent assessments of erbB-2 expression showed similar correlations with other covariates, including interactions with dose in predicting DFS (by C. Allred, $P = .001$; by A. D. Thor, $P<.001$). These results substantiate the robustness of the erbB-2 data and suggest that staining intensity or a complex system that combines the intensity with the percent of cells staining may not provide superior predictive value.

Analysis of Outcomes for Combined Sets A and B

Characteristics including CAF regimen, age at enrollment, menopausal status, tumor size, number of positive lymph nodes, ER and PR contents, tamoxifen treatment, and erbB-2 gene amplification were significantly associated with DFS, as determined by univariate analysis (Table 2). The CAF regimen, tumor size, number of positive lymph nodes, histologic grade, ER and PR contents, tamoxifen treatment, erbB-2 amplification, and p53 expression were significantly associated with OS as well (Table 2). For all patients (Table 2), erbB-2 gene amplification (but not immunoexpression as a continuous variable) was associated with poorer prognosis ($P<.001$ for OS and $P<.001$ for DFS). p53 expression was associated with a shortened OS ($P = .025$) but not with a shortened DFS ($P = .45$).

Multivariate analyses of all data (sets A and B combined) demonstrated that various clinical and histologic factors, including the number of positive lymph nodes, tumor size, and tamoxifen therapy, were independent predictors of DFS and OS, whereas menopausal status was an independent predictor of DFS only [Table 3; *see also* Budman et al. *(12)* for full discussion of clinical issues]. The erbB-2 immunopositivity (as a continuous variable) showed an independent prognostic value for both DFS and OS ($P = .004$ and $P<.001$, respectively). The relationship between the erbB-2 immunohistochemical data and survival (outcomes) was stronger in the multivariate analysis than in the univariate analysis, since the interaction between erbB-2 expression and CAF dose was included in the multivariate model. The prognostic value of erbB-2 expression with the use of PCR-

Table 2. Univariate analysis of variables associated with disease-free and overall survival in all patients enrolled in the CALGB trial 8869*

Variable	No. of patients†	Disease-free survival‡			Overall survival¶		
		RR§	95% CI	Two-sided *P*‖	RR§	95% CI	Two-sided *P*‖
CAF regimen	1013	1.19	1.06–1.33	.003	1.15	1.01–1.30	.036
Age at enrollment#	1013	1.30	1.08–1.56	.003	1.11	0.90–1.33	.37
Menopausal status	1013	1.32	1.10–1.59	.003	1.11	0.90–1.36	.35
Tumor size	1008	1.73	1.41–2.13	<.001	1.90	1.50–2.41	<.001
No. of positive lymph nodes#	1013	2.45	2.07–2.90	<.001	2.61	2.17–3.13	<.001
Histologic type	993	1.03	0.72–1.46	.88	1.04	0.70–1.54	.84
Histologic grade	993	1.20	1.00–1.45	.054	1.51	1.22–1.87	<.001
ER content#	998	1.08	1.00–1.18	.070	1.17	1.06–1.30	<.002
PR content#	964	1.16	1.08–1.23	<.001	1.24	1.14–1.34	<.001
Tamoxifen treatment	1012	1.44	1.18–1.77	<.001	1.52	1.20–1.93	.001
DNA content (ploidy)	918	1.07	0.87–1.30	.53	0.97	0.77–1.21	.77
% S phase#	659	1.21	0.84–1.73	.300	1.36	0.91–2.02	.14
erbB-2 expression#	992	1.05	0.95–1.22	.31	1.11	1.00–1.28	.092
erbB-2 gene amplification	916	1.58	1.25–2.00	<.001	1.84	1.43–2.38	<.001
p53 expression#	994	1.06	0.94–1.16	.45	1.16	1.03–1.31	.025

*CALGB = Cancer and Leukemia Group B; RR = risk ratio; CI = confidence interval; CAF = cyclophosphamide, doxorubicin [Adriamycin], and 5-fluorouracil; ER = estrogen receptor; PR = progesterone receptor.

†Number of patients indicates number with complete data. The numbers of patients' tumors analyzed are not identical for all categories because of incomplete or unavailable data on some cases.

‡Favorable characteristics for disease-free survival were high CAF dose, older age, postmenopausal status, smaller tumor size, fewer positive lymph nodes, higher PR contents, tamoxifen treated, and no erbB-2 gene amplification.

§The RR compares two categories for each variable. For dichotomous variables, the RR compares the values for menopausal status (premenopausal versus postmenopausal), size (>2 cm versus ⩽2 cm), type (other versus infiltrating ductal), tamoxifen treatment (no treatment versus treated), ploidy (aneuploid versus diploid), and erbB-2 amplification (no amplification versus amplification). For continuous variables, we selected specific values to illustrate how to interpret the RR for the following variables: age (40 years versus 60 years), erbB-2 expression (50% versus 0%), p53 (30% versus 0%), number of positive lymph nodes (10 versus 1), S phase (15% versus 0%), ER (20% versus 0%), PR (20% versus 0%), and CAF dose (low versus moderate dose or moderate versus high dose).

‖*P* values are from the Cox proportional hazards model with the use of Wald's chi-squared test.

¶Favorable characteristics for overall survival included high CAF dose, smaller tumor size, fewer positive lymph nodes, lower histologic grade, higher ER and PR contents, tamoxifen treated, no erbB-2 gene amplification, and low p53 expression.

#These variables were analyzed on a continuous scale.

derived amplification data ($P<.001$ for DFS and $P<.001$ for OS; models not shown) was similar to protein expression (immunohistochemical) data. p53 was an independent marker of poor prognosis as well ($P = .035$ for DFS; $P = .023$ for OS). Although suggestive, the independent prognostic association between p53 expression and DFS had marginal significance because it was not maintained in all models (Table 3).

More importantly, the interaction between erbB-2 expression and CAF dose (predictive value of erbB-2) was significant for both DFS ($P = .001$ by immunohistochemistry; $P = .060$ by PCR) and OS ($P<.001$ by immunohistochemistry; $P = .010$ by PCR). The interactions between CAF dose and tamoxifen treatment and between erbB-2 expression and tamoxifen treatment were not significant in similar models (data not shown), illustrating a lack of interaction between tamoxifen treatment and erbB-2 expression or CAF dose in this study. p53 interacted with dose in predicting DFS ($P = .022$) but not OS ($P = .13$). Hence, for the combined set of patients (sets A and B), both erbB-2 and p53 expressions had independent prognostic and predictive values. Interactions between CAF dose and DNA ploidy or CAF dose and S phase were not observed (data not shown).

The data on the combined sets A and B are graphically displayed with the use of Kaplan–Meier plots using the same cut points as those used in our initial publication to enhance comparability (Fig. 1) *(10)*. Patients who were randomly assigned to the dose-intensive (high-dose) arm of the adjuvant trial and whose tumors expressed high levels of erbB-2 (⩾50% of positive cells) had a longer OS and DFS than similarly treated patients whose tumors expressed low levels of erbB-2 (OS = 78% versus 65% at 8 years; DFS = 69% versus 55% at 8 years).

Analysis of Outcomes of Sets A and B

As described in the ''Subjects and Methods'' section, there were two sets of patients in this study: Set A consisted of the 397 patients described previously *(10)*, and set B consisted of the newly added 595 patients. The updated multivariate analysis of DFS for set A with a median follow-up of 10.4 years is shown in Table 3, B. The effect of the various covariates is similar to that of the combined set (A and B). With additional follow-up data, the interaction between erbB-2 expression and CAF dose was even more significant than previously reported ($P<.001$). Multivariate analysis of DFS for the validation set (set B) is shown in Table 3, C. When set B was analyzed separately, the interaction between CAF dose and erbB-2 expression (measured as a continuous variable) was not statistically significant. However, the coefficient of the interaction term had the same sign as in set A, indicating that the interaction was in the same direction in both sets. (The correlation coefficient for DFS was −.011 and −.001 for sets A and B, respectively; for OS, it was −.011 and −.004 for sets A and B, respectively.)

The differences between the outcomes in set A and in set B were striking and raised concerns about the comparability of the two sets of patients. To address this possibility, we analyzed potential differences between sets A and B. Fig. 2 shows the Kaplan–Meier DFS curves for sets A and B by erbB-2 status. This figure uses a scheme to denote the 12 relevant subgroups of

Table 3. Multivariate Cox proportional analyses of clinical and biologic variables associated with disease-free survival and overall survival in patients enrolled in the CALGB trial 8869*

Variable	Disease-free survival			Overall survival		
	RR†	95% CI	Two-sided *P*‡	RR†	95% CI	Two-sided *P*‡
A) Sets A and B combined						
CAF	1.03	0.88–1.19	.73	1.04	0.88–1.24	.62
No. of positive LN§	2.62	2.20–3.13	<.001	2.67	2.20–3.25	<.001
Tumor size	1.65	1.33–2.03	<.001	1.76	1.38–2.25	<.001
Menopausal status	1.33	1.09–1.62	.006	1.02	0.82–1.28	.84
Tamoxifen therapy	1.43	1.15–1.78	.002	1.62	1.25–2.09	<.001
erbB-2 immuno§	1.35	1.11–1.57	.004	1.49	1.16–1.82	<.001
CAF × erbB-2§	—		.001	—		<.001
p53§	1.20	1.00–1.43	.035	1.23	1.03–1.47	.023
CAF × p53§	—		.022	—		.13
B) Set A						
CAF	1.14	0.91–1.44	.27	1.10	0.85–1.41	.48
No. of positive LN§	2.79	2.13–3.66	<.001	2.67	2.01–3.55	<.001
Tumor size	2.06	1.49–2.84	<.001	1.80	1.27–2.57	.001
Menopausal status	1.27	0.95–1.71	.105	1.01	0.73–1.39	.95
Tamoxifen therapy	1.16	0.83–1.62	.37	1.73	1.16–2.58	.007
erbB-2 immuno§	2.00	1.49–2.56	<.001	2.00	1.49–2.69	<.001
CAF × erbB-2§	—		<.001	—		<.001
p53§	1.52	1.16–1.98	.002	1.16	0.89–1.52	.28
CAF × p53§	—		.006	—		.43
C) Set B						
CAF	1.14	0.93–1.39	.21	1.01	0.80–1.27	.95
No. of positive LN§	2.52	1.96–3.24	<.001	2.67	2.03–3.15	<.001
Tumor size	1.54	1.16–2.04	.003	1.75	1.24–2.47	.001
Menopausal status	1.23	0.93–1.63	.15	1.03	0.74–1.43	.85
Tamoxifen therapy	1.69	1.25–2.29	.001	1.60	1.14–2.26	.007
erbB-2 immuno§	1.00	0.79–1.28	.99	1.11	0.82–1.49	.47
CAF × erbB-2§	—		.48	—		.12
p53§	1.09	0.86–1.39	.53	1.27	1.00–1.61	.065
CAF × p53§	—		.34	—		.21

*The results include erbB-2 as measured by immunohistochemistry (n = 984, patients with data on all factors in model). Results are similar with the use of erbB-2 as measured by polymerase chain reaction amplification. CALGB = Cancer and Leukemia Group B; RR = risk ratio; CI = confidence interval; CAF = cyclophosphamide, doxorubicin (Adriamycin), and 5-fluorouracil; LN = lymph nodes; erbB-2 immuno = percent erbB-2 cells positive by immunohistochemistry; CAF × erbB-2 = interactive term CAF dose 0, 1, 2 (low, moderate, high, respectively) multiplied by the percentage of erbB-2-positive cells determined by immunohistochemistry; CAF × p53 = interactive term CAF dose 0, 1, 2 (low, moderate, high, respectively) multiplied by the percentage of p53-positive cells determined by immunohistochemistry.

†The RR compares two categories for each variable. For dichotomous variables, the RR compares the values for menopausal status (premenopausal versus postmenopausal), size (>2 cm versus ⩽2 cm), type (other versus infiltrating ductal), tamoxifen treatment (no treatment versus treated), ploidy (aneuploid versus diploid), and erbB-2 amplification (no amplification versus amplification). For continuous variables, we selected specific values to illustrate how to interpret the RR for the following variables: age (40 years versus 60 years), erbB-2 expression (50% versus 0%), p53 (30% versus 0%), number of positive lymph nodes (10 versus 1), S phase (15% versus 0%), estrogen receptor (20% versus 0%), progesterone receptor (20% versus 0%), and CAF dose (low versus moderate dose or moderate versus high dose).

‡*P* values were determined by multivariate Cox proportional analysis. *P* values for erbB-2 and the interaction term CAF and erbB-2 are similar regardless of whether p53 is included or excluded from models.

§These variables were analyzed on a continuous scale.

patients—a triple code with the first letter indicating set (A or B), the second indicating dose schedule (high = H, moderate = M, or low = L), and the third indicating erbB-2 status (high [⩾50% tumor cells stained] versus low [<50% tumor cells stained]). Patients whose tumors stained low for erbB-2 had similar DFS characteristics when categorized by dose, regardless of whether they were in set A or set B (Fig. 2, A, C, and E). For patients with high erbB-2-expressing tumors, DFS times for the high- and moderate-dose groups only were nearly identical in sets A and B (Fig. 2, B and D). The 5-year DFS was 67% for set A and 69% for set B for the high-dose arm and 66% for set A and 60% for set B for the moderate-dose arm. When only the patients treated with the high and moderate doses were compared, sets A and B demonstrated similar interactions for erbB-2 expression by CAF dose, with the patients receiving the high dose exhibiting a longer DFS than the patients receiving the moderate dose among high erbB-2 expressors (logrank, $P = .006$ for set A [high dose] versus set A [moderate dose], and $P = .043$ for set B [high dose] versus set B [moderate dose]; Fig. 2, B and D). This is supportive evidence of an interaction between CAF dose intensity and erbB-2 expression.

In contrast, the behaviors of the low-dose groups from sets A and B were very different in the erbB-2-expressing patients (Fig. 2, F), with a 31% compared with a 63% 5-year DFS, respectively ($P = .002$). These differences have profound effects on the interactive term (interaction between erbB-2 expression and CAF dose) when sets A and B were independently analyzed: The lack of statistical significance of the interaction

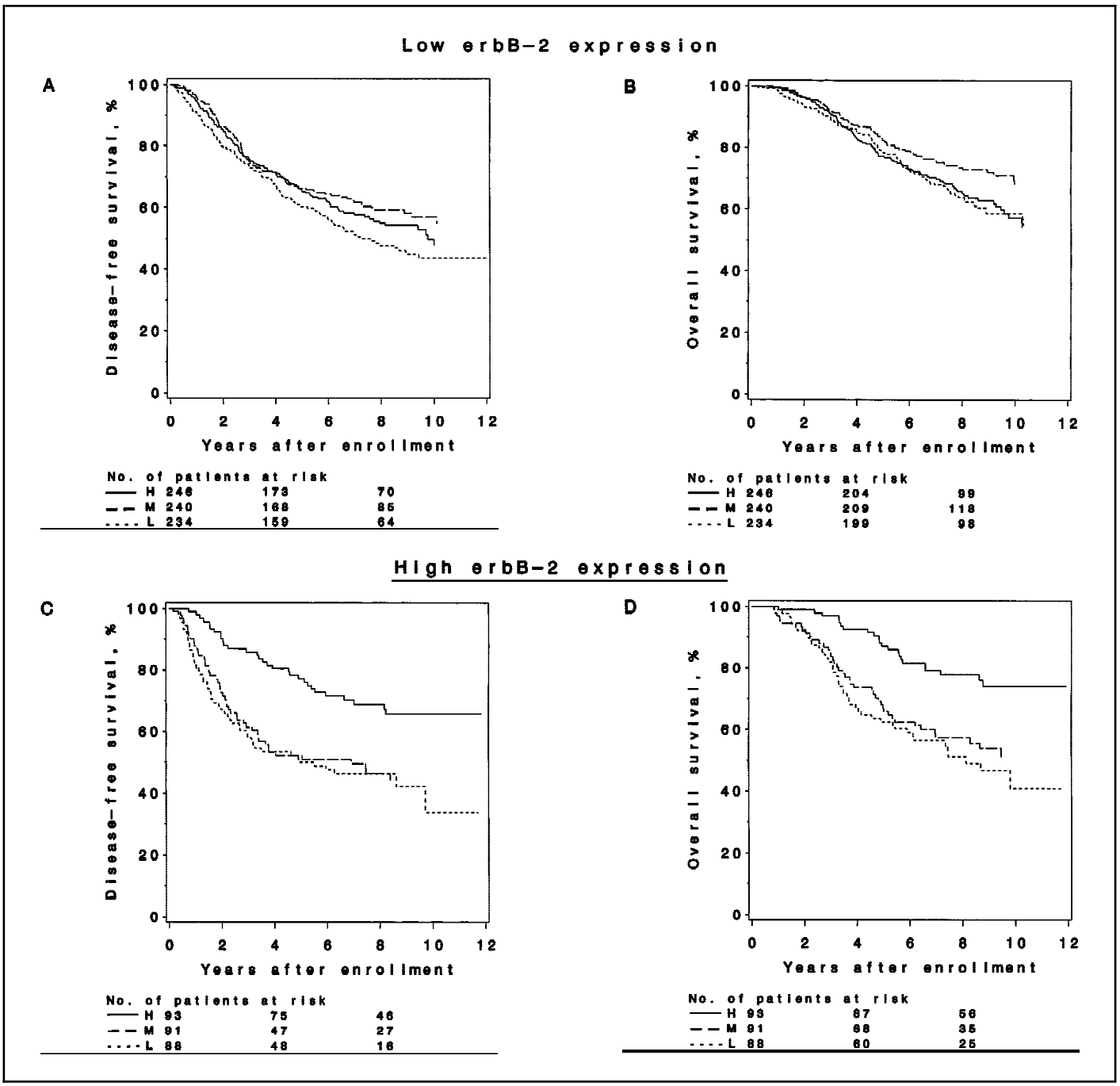


Fig. 1. Interaction of CAF (i.e., cyclophosphamide, doxorubicin [Adriamycin], and 5-fluorouracil) dose arm with erbB-2 expression in all patients. *P* values are derived from logrank tests. **A)** Disease-free survival (DFS) for low erbB-2 expression: erbB-2 <50%; n = 234, 240, and 246; 5-year DFS (95% confidence interval [CI]) = 60% (54%–66%), 66% (60%–72%), and 65% (59%–71%) for CAF dose that was low (L), moderate (M), or high (H), respectively; $P = .058$. **B)** Overall survival (OS) for low erbB-2 expression: 5-year OS (95% CI) = 78% (72%–83%), 82% (77%–87%), and 77% (71%–82%) for CAF dose that was L, M, or H, respectively; $P = .048$. **C)** DFS for high erbB-2 expression: erbB-2 ≥50%; n = 88, 91, and 93; 5-year DFS (95% CI) = 50% (40%–60%), 52% (42%–62%), and 71% (68%–85%) for CAF dose that was L, M, or H, respectively; $P<.001$. **D)** OS for high erbB-2 expression: 5-year OS (95% CI) = 63% (52%–72%), 66% (56%–75%), and 87% (79%–92%) for CAF dose that was L, M, or H, respectively; $P<.001$.

term (CAF dose and erbB-2 expression) in the multivariate analysis of set B was largely due to the good performance of patients in that group (set B, low CAF dose and high erbB-2 expression; data not shown).

To investigate the basis of this difference, the most important prognostic factors were compared across the 12 subgroups of patients (categorized by CAF dose, erbB-2 expression, and set A or set B; Table 4). These data revealed that patients with high erbB-2 expression from set A who were treated with low-dose CAF had significantly different prognostic characteristics than patients in set B with high erbB-2 expression treated with low-dose CAF. Specifically, the mean number of positive lymph nodes in the set B group with high erbB-2 expression treated with low-dose CAF was the least of all 12 subgroups (3.92). This was in contrast to set A patients with high erbB-2 expression who were treated with low-dose CAF; these patients had the

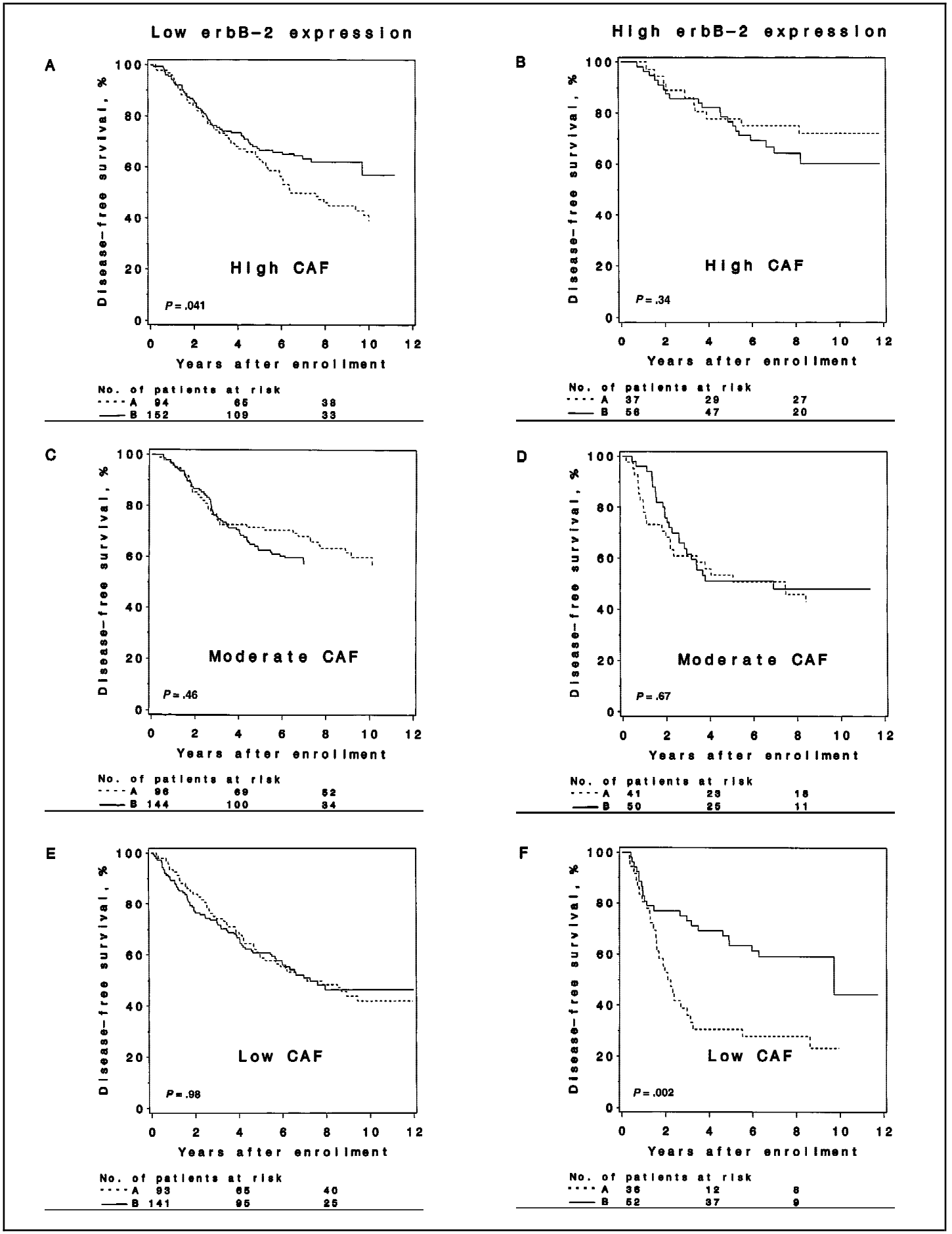


Fig. 2 ***(see previous page).*** Kaplan–Meier disease-free survival (DFS) comparison of interaction of sets A and B within CAF (i.e., cyclophosphamide, doxorubicin [Adriamycin], and 5-fluorouracil) arm and status of erbB-2 expression. **A)** High CAF dose and low erbB-2 expression: erbB-2 <50%; n = 94 and 152; 5-year DFS (95% confidence interval [CI]) = 63% (53%–72%) and 67% (58%–74%) for sets A and B, respectively. **B)** High CAF dose and high erbB-2 expression: erbB-2 ⩾50%; n = 37 and 56; 5-year DFS (95% CI) = 78% (62%–88%) and 77% (64%–86%) for sets A and B, respectively. **C)** Moderate CAF dose and low erbB-2 expression: n = 96 and 144; 5-year DFS (95% CI) = 71% (62%–80%) and 63% (54%–70%) for sets A and B, respectively. **D)** Moderate CAF dose and high erbB-2 expression: n = 41 and 50; 5-year DFS (95% CI) = 53% (39%–68%) and 51% (37%–65%) for sets A and B, respectively. **E)** Low CAF dose and low erbB-2 expression: n = 93 and 141; 5-year DFS (95% CI) = 59% (49%–68%) and 61% (53%–69%) for sets A and B, respectively. **F)** Low CAF dose and high erbB-2 expression: n = 36 and 52; 5-year DFS (95% CI) = 31% (18%–46%) and 63% (50%–75%) for sets A and B, respectively.

greatest mean number of positive lymph nodes (6.47). Set B patients with high erbB-2 expression who were treated with low-dose CAF were substantially different from their set A counterpart in other characteristics as well, including fewer premenopausal patients (27% versus 42%), larger tumor size (60% versus 50% T2 tumors), fewer ER-positive tumors (54% versus 69%), and greater tamoxifen use (37% versus 25%). These data illustrate the complex differences in prognostic variables between individual subgroups that may confound outcome comparisons.

To address the overall effect of the differences in prognostic variables, a prognostic index derived from the current data was then used to quantitate the differences between the subgroups in sets A and B (*see* ''Subjects and Methods'' section). The prognostic index is based on a mathematical formula that includes several factors associated with risk of recurrence or survival (e.g., number of positive lymph nodes). Patients with a higher index score had a worse outcome than those with a lower index score. The mean prognostic index for the entire set A was slightly larger than that for the entire set B (4.20 versus 3.62, Table 4); however, when the individual subgroups were compared, wide discrepancies were seen. Most notably, erbB-2-positive patients treated on the low-dose arm in set B had a prognostic index of 3.14 compared with the 4.77 in similarly treated patients in set A ($P = .059$, Table 4). These data indicate that the overall prognosis of patients in set B with high erbB-2 expression who were treated with low-dose CAF was substantially better than that of the other subgroups.

As indicated in the ''Subjects and Methods'' section, we adjusted for the effects of differing prognosis across the groups by modifying their Kaplan–Meier survival curves.

Table 4. Patient and tumor characteristics for all cases (sets A and B) of Cancer and Leukemia Group B trial 8869

				Lymph nodes		% of patients/tumors with¶				Prognostic index#	
Set*	CAF dose†	erbB-2‡	N§	Mean	SE‖	T2	Premeno	Tam	ER	Mean	SE
A	L	High	36	6.47	1.12	50	42	25	69	4.77	0.67
A	M	High	41	5.90	1.06	71	56	17	54	4.95	0.61
A	H	High	37	6.46	0.97	73	46	18	68	4.97	0.56
B	L	High	52	3.92	0.70	60	27	37	54	3.14	0.24
B	M	High	50	5.06	1.11	64	40	24	52	4.15	0.58
B	H	High	56	4.84	0.59	64	36	29	64	3.94	0.31
A	L	Low	93	4.55	0.52	69	44	30	65	4.00	0.28
A	M	Low	96	4.08	0.43	58	47	29	72	3.69	0.25
A	H	Low	94	4.99	0.55	67	48	35	77	4.07	0.28
B	L	Low	141	4.19	0.35	68	43	40	69	3.67	0.19
B	M	Low	144	4.35	0.38	61	33	43	70	3.49	0.19
B	H	Low	152	4.34	0.43	64	35	43	68	3.56	0.23
Summary											
A	All	All	397	5.03	0.27	65	47	28	68	4.20	0.15
B	All	All	595	4.38	0.21	64	36	39	66	3.62	0.11

*Set A = initial patient group (n = 397). Set B = validation patient group (n = 595).

†CAF = cyclophosphamide, doxorubicin (Adriamycin), and 5-fluorouracil; L = low dose; M = moderate dose; H = high dose.

‡erbB-2 expression: with the use of immunohistochemical data, high = ⩾50% tumor cells positive and low = <50% tumor cells positive.

§N = number of patients in category.

‖SE = standard error.

¶T2 [clinical stage based primarily on tumor size *(2,7)*] = % of patients with T2; Premeno = % of patients who are premenopausal; Tam = % of patients receiving tamoxifen; ER = % of tumors that are estrogen receptor positive.

#*See* ''Statistical analysis'' in ''Subjects and Methods'' section for the methodology used to calculate prognostic index values.

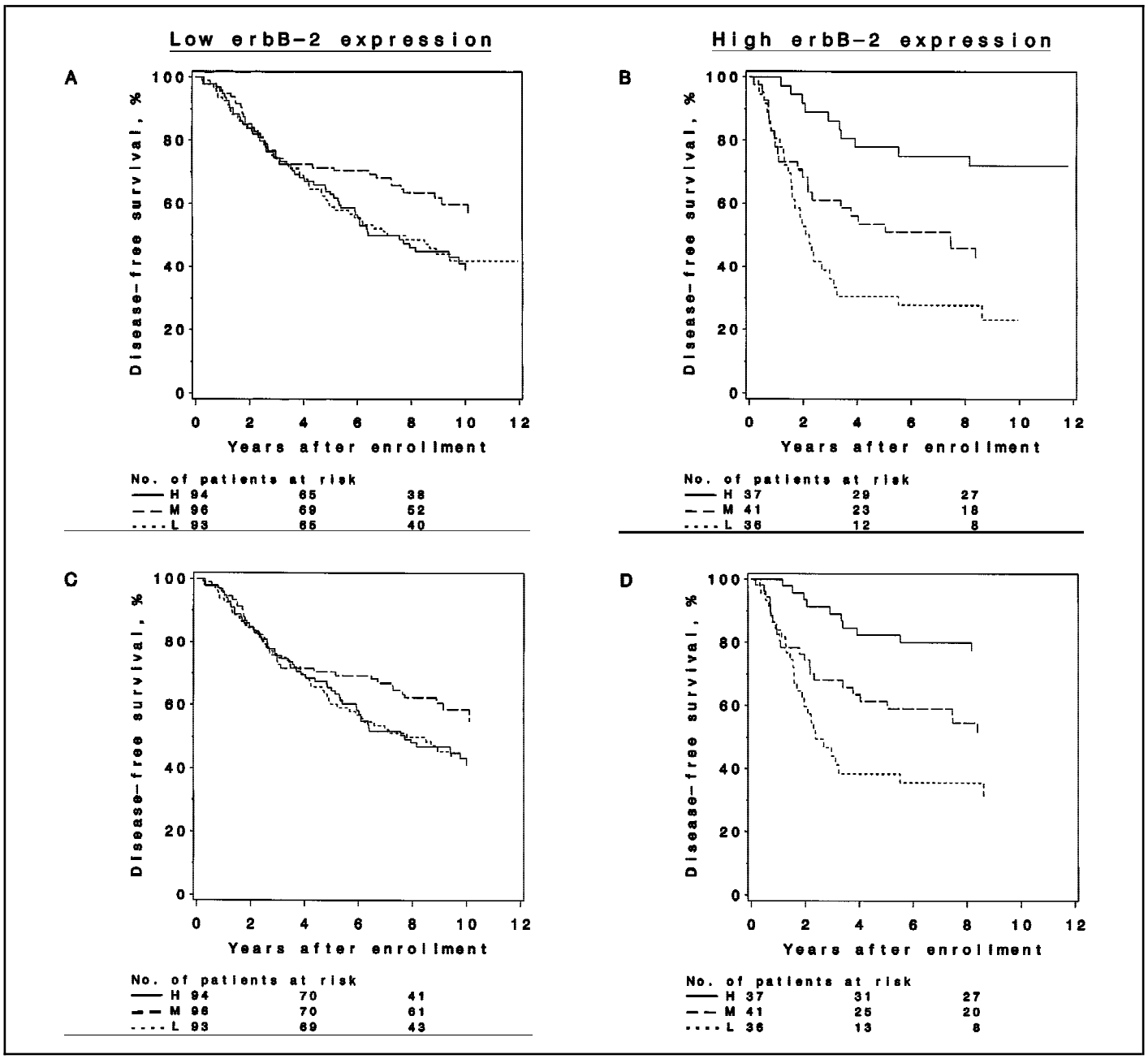


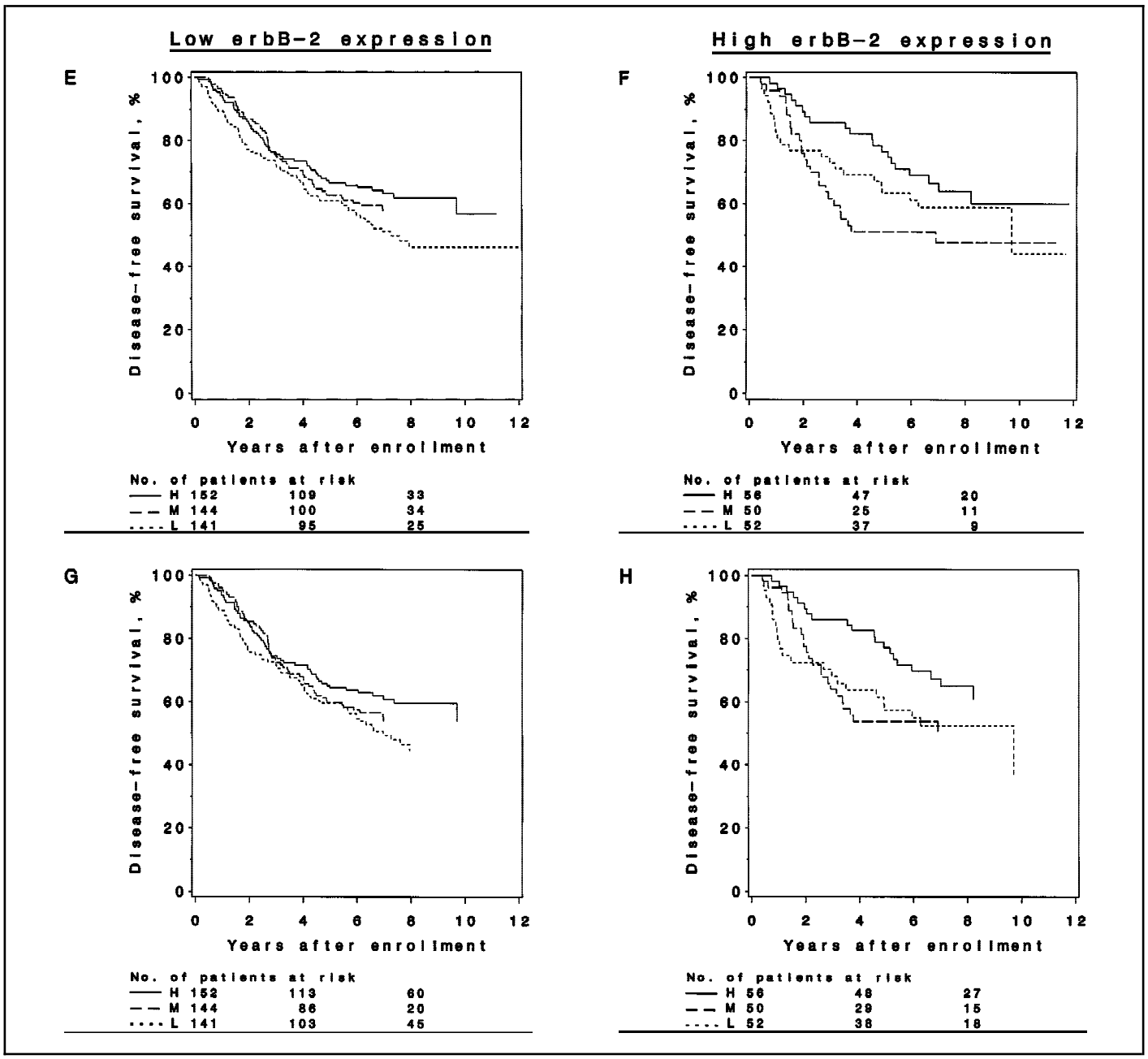
Fig. 3 ***(continues on facing page).*** Kaplan–Meier disease-free survival (DFS) curves showing interaction of CAF (i.e., cyclophosphamide, doxorubicin [Adriamycin], and 5-fluorouracil) arms with status of erbB-2 expression. The first four panels **(A, B, C, and D)** show set A, and the second four panels **(E, F, G, and H)** show set B. *P* values are derived from logrank tests for unadjusted curves only. Within each set, the first two panels **(A, B, E, and F)** show unadjusted survival curves. The second two **(C, D, G, and H)** show survival curves adjusted for prognosis as described in the text; since these are derived from adjusted survival probabilities, it is not possible to calculate logrank statistics or confidence intervals (CIs). **A)** Set A, low erbB-2 expression: erbB-2 <50%; n = 93, 96, and 94; 5-year DFS (95% CI) = 59% (49%–68%), 71% (62%–80%), and 63% (53%–72%) for CAF dose that was low (L), moderate (M), or high (H), respectively; $P = .053$. **B)** Set A, high erbB-2 expression: erbB-2 ⩾50%; n = 36, 41, and 37; 5-year DFS (95% CI) = 31% (18%–46%), 53% (39%–68%),

(Fig. 3 shows the unadjusted and adjusted curves.) As expected, the adjustment did not substantially change the configuration of the Kaplan–Meier curves for most of the subgroups with one major exception. After the adjustment for prognostic index, the set B, high-erbB-2-expressing groups then showed that the high-dose arm performed significantly better than either the moderate- or the low-dose arm. Thus, after we accounted for the prognostic index, the analysis of all subgroups supports the hypothesis that patients harboring erbB-2-positive tumors benefit from dose-intensive CAF adjuvant therapy and those harboring erbB-2-negative tumors do not.

p53 Interaction With CAF Dose

Like erbB-2 expression, in the combined sets A and B, p53 expression was an independent prognostic marker ($P = .035$ for DFS, and $P = .023$ for OS, Table 3). p53 expression also interacted with dose to predict treatment outcome with multivariate analysis, although the effect was less pronounced than that of erbB-2 expression (interaction between CAF dose and erbB-2 expression, $P = .001$; interaction between CAF dose and p53 expression, $P = .022$).

Further analysis suggests that the interactive effects on sur-



and 78% (62%–88%) for CAF dose that was L, M, or H, respectively; $P<.001$. **C)** Set A, low erbB-2 expression adjusted for prognosis; n = 93, 96, and 94. **D)** Set A, high erbB-2 expression adjusted for prognosis; n = 36, 41, and 37. **E)** Set B, low erbB-2 expression: n = 141, 144, and 152; 5-year DFS (95% CI) = 61% (53%–69%), 63% (54%–70%), and 67% (58%–74%) for CAF dose that was L, M, or H, respectively; $P = .12$. **F)** Set B, high erbB-2 expression: n = 52, 50, and 56; 5-year DFS (95% CI) = 63% (50%–75%), 51% (37%–65%), and 77% (64%–86%) for CAF dose that was L, M, or H, respectively; $P = .15$. **G)** Set B, low erbB-2 expression adjusted for prognosis. **H)** Set B, high erbB-2 expression adjusted for prognosis.

vival of erbB-2 expression and p53 expression may be additive (Fig. 4). Patients whose tumors showed low erbB-2 expression, regardless of p53 status, failed to derive benefit from high- or moderate-dose CAF (Fig. 4, A and B; DFS curves are similar but not shown). For those tumors with high erbB-2 expression, however, p53 seems to matter a great deal. Patients with high erbB-2 expression and p53-negative tumors showed differences in survival by treatment arm (OS, $P = .044$, Fig. 4, C; DFS, $P = .028$, data not shown), and patients on the high-dose arm demonstrated the best survival. Patients who had erbB-2-expressing tumors that also expressed p53 and who received the moderate-dose or especially the high-dose CAF (Fig. 4, D) survived substantially longer than those whose tumors were p53 negative. In particular, p53- and erbB-2-positive tumor patients who were treated on the high-dose arm (n = 34) had a greater than 90% OS at 10 years compared with a 39% 10-year survival if treated on the low-dose arm (n = 33). The number of patients in this set (high expression of erbB-2 and p53 positive) is limited (n = 101); hence, these results should be considered preliminary and hypothesis generating.

Discussion

Our initial correlative study led to the hypothesis that the only patients who benefit from the dose-intensive CAF regimen were

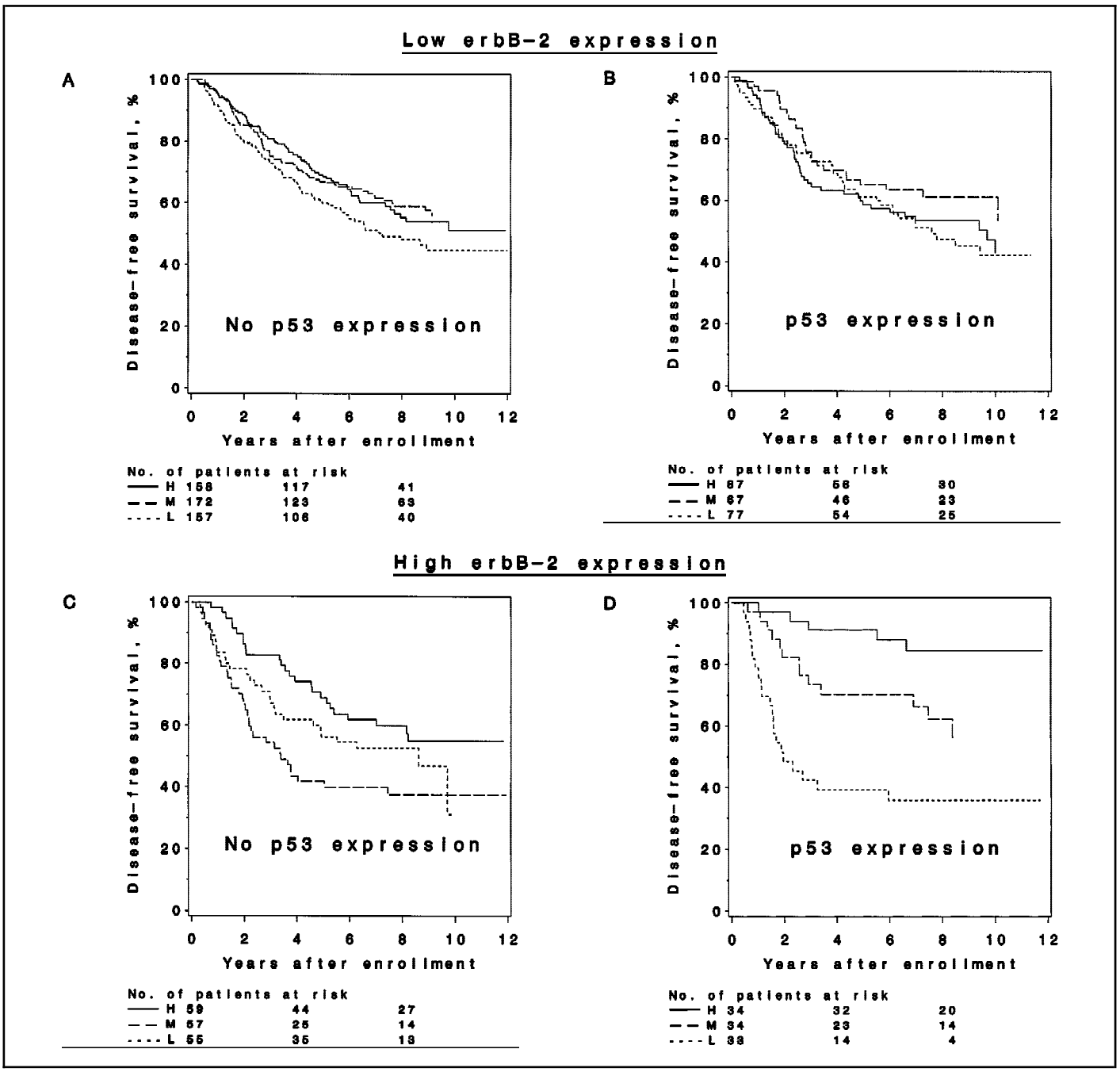


Fig. 4. Kaplan–Meier disease-free survival (DFS) curves showing interaction of CAF (i.e., cyclophosphamide, doxorubicin [Adriamycin], and 5-fluorouracil) dose with erbB-2 expression and p53 expression. *P* values are derived from logrank tests. **A)** Low erbB-2 expression (erbB-2 <50%) and no p53 expression (p53 = 0%); n = 157, 172, and 158; 5-year DFS (95% confidence interval [CI]) = 60% (52%–67%), 67% (59%–74%), and 68% (61%–75%) for CAF dose that was low (L), moderate (M), or high (H), respectively; $P = .11$. **B)** Low erbB-2 expression and any p53 expression (p53 ⩾1%); n = 77, 67, and 87; 5-year DFS (95% CI) = 61% (50%–71%), 65% (53%–76%), and 59% (48%–68%) for CAF dose that was L, M, or H, respectively; $P = .32$. **C)** High erbB-2 expression (erbB-2 ⩾50%) and no p53 expression; n = 55, 57, and 59; 5-year DFS (95% CI) = 56% (43%–68%), 42% (30%–55%), and 69% (56%–79%) for CAF dose that was L, M, or H, respectively; $P = .028$. **D)** High erbB-2 expression and any p53 expression; n = 33, 34, and 34; 5-year DFS (95% CI) = 39% (25%–56%), 70% (53%–83%), and 91% (77%–97%) for CAF dose that was L, M, or H, respectively; $P<.001$.

those whose tumors expressed erbB-2 *(9)*. This hypothesis raised the intriguing possibility that erbB-2 may be an important predictive marker for response to a specific adjuvant chemotherapy. On the basis of the data presented here, erbB-2 expression and perhaps p53 abnormalities may have value in predicting which patients are likely to respond to the dose-intensive CAF regimen. These data contradict the dogma that alterations of oncogenes always portend a worse outcome and suggest that such markers may even have a salutary prognostic or predictive value. These data may also explain, at least in part, the differing conclusions that have been reported for erbB-2 expression and p53 expression and clinical outcome when adjuvant treatment was not considered *(24)*.

Concerns were raised at the time of our initial report regarding the short-term follow-up and reproducibility of immunohistochemical assays for erbB-2 *(25)*. In this study, the median

follow-up of 9.3 years is considerably extended for all patients. This study includes the utilization of a different anti-erbB-2 antibody (also called p185) (with no major differences observed) as well as independently derived erbB-2 gene amplification data. Regardless of the antibody used, erbB-2 immunostaining was highly correlated with erbB-2 gene amplification by use of the Pearson product moment correlation ($r = .48$; $P<.001$) *(26)*. Either the immunohistochemical data or the amplification data, considered alone, had both prognostic and predictive value (gene amplification data not shown). Interobserver variance for quantifying erbB-2 expression on the same tissues was negligible—demonstrating reproducibility of immunohistochemical interpretation. The percent of positive cell staining provided superior prognostic and/or predictive information compared with data reflecting intensity of staining or a combined score (which takes into account percent of cells stained plus intensity of staining). Finally, we reassayed erbB-2 and p53 expression in sectioned, archived tissues stored for 2–5 years and found no significant deterioration of antigenicity. Thus, by multiple criteria, immunohistochemical erbB-2 data appear to be a *bona fide* reflection of the state of the erbB-2 expression in tumor tissues.

Prediction of response to treatment based on molecular markers is not novel. ER levels identify patients likely to respond to hormonal manipulation. The finding that markers can be used to predict adjuvant chemotherapy benefit has been explored only recently *(27)*. Earlier studies of breast cancers *(8,28)* suggested that erbB-2 expression was associated with poorer outcome in patients treated with non-doxorubicin-based adjuvant regimens, such as CMF (i.e., cyclophosphamide, methotrexate, and 5-fluorouracil). Comparison between our study (which used patients treated with CAF) and these previous investigations led us to speculate that a major factor in this difference is doxorubicin. To explore this hypothesis, we performed erbB-2 immunohistochemical analysis on 159 archived tissue blocks from patients enrolled in a previous CALGB adjuvant trial 8082 *(24)*. This randomized trial, initiated in 1980, compared CMFVP (i.e., cyclophosphamide, methotrexate, 5-fluorouracil, vincristine, and prednisone) with CMFVP plus VATH (i.e., vinblastine, doxorubicin [Adriamycin], thiotepa, and halotestin) in lymph node-positive patients *(29)*. With the use of similar methods, there were no differences in outcome between the VATH and the CMFVP arms in erbB-2-negative patients. VATH had an advantage over CMFVP in erbB-2-positive patients. Moreover, VATH improved DFS among erbB-2-positive patients to the level of erbB-2-negative patients *(24)*. These data support the hypothesis that doxorubicin is the key drug in the interaction that we observed between the CAF dose and erbB-2 expression. Similar interactions with erbB-2 and treatment have been reported for endometrial carcinoma *(30)*.

A major issue addressed in this article is whether the additional patients accrued to CALGB 8869 (set B) confirm the hypothesis of interaction between erbB-2 expression and CAF dose derived from the study of patients of set A. We have provided multiple analyses in this study, including traditional univariate and multivariate (Cox proportional hazards regression) models of combined sets A and B as well as set A and set B individually. Using the same multivariate model applied to set B as applied to set A, we were unable to identify a statistically significant interaction between CAF dose and erbB-2 expression. Therefore, using this method, we could not confirm our original hypothesis. However, outcome data (DFS and OS) from set B (particularly patients on the high- and moderate-dose arms) showed the same trend as those from set A. A comparison of prognostic variables and constructing prognostic indices showed that set B, particularly patients with high erbB-2 expression randomly assigned to the low-dose CAF arm, were not comparable to set A or combined sets A and B. This result reflects in part the difficulty of a retrospective correlative study where the major factor driving the scientific question (erbB-2 status) was not part of the randomization or stratification scheme. As alluded to in the ''Results'' section, multivariate models cannot fully account for imbalances, especially when the imbalance is itself multivariate. Given the complex drift of variables with time during the trial, comparisons between sets A and B were difficult without appropriate adjustments.

We accounted for higher order interactions among prognostic variables by using a prognostic index correction on the Kaplan–Meier survival curves. When applied to set B, an advantage to the high-dose arm in the high-erbB-2-expression group could then be discerned (Fig. 3), although, as is usually the case in confirmatory studies *(31)*, the differences between the dose arms are less dramatic than those seen in the hypothesis-generating set A. Although the adjusted results of set B appear consistent with our earlier analyses, they are not sufficiently conclusive to warrant unconditional acceptance of the original hypothesis for an interaction between CAF dose and erbB-2 expression.

Neither S-phase nor ploidy data showed any interaction with CAF dose. Interactions between CAF dose and p53 expression were observed, although they were less strong than the interaction observed between erbB-2 expression and CAF dose. Although Fig. 4 suggested that the interaction between p53 expression and CAF dose response was manifest only when erbB-2 was co-expressed, the interaction term for p53 was significant in multivariate models (Table 3, A) with or without erbB-2. Alterations of p53 or downstream effectors, such as $p21^{waf1/cip1}$, have now been shown to be associated with an improved chemoresponsiveness of breast cancer *in vitro (32)* and bladder cancer *in vivo (33)* to doxorubicin. On the basis of our findings, we hypothesize an interaction between p53 expression and CAF dose as well. The impact of a three-way interaction between erbB-2 immunohistochemical expression or gene amplification, high CAF dose, and p53 positivity in this study is provocative. Patients with abnormalities of both erbB-2 and p53 had a remarkable 10-year OS of 90%. To our knowledge, this is the first report to raise the possibility of an interaction between two molecular markers and a specific therapeutic regimen.

Taken together, the work described in this article and by others using CALGB 8541 resources *(10,24)* raises the possibilities 1) that dose-intensive CAF chemotherapy may improve survival of lymph node-positive breast cancer patients with erbB-2-positive tumors, 2) that intensive CAF chemotherapy may not benefit lymph node-positive patients with erbB-2-negative tumors, 3) that this erbB-2 expression–dose interaction may be specific to doxorubicin, and 4) that the erbB-2 expression–CAF dose interaction may be modulated by other molecular markers that interact with therapy, such as p53. Because the impact of erbB-2 and p53 on chemotherapeutic efficacy was not the primary end point for this study, these results should not be

generalized *(34)* or applied clinically until validated by other studies. Nevertheless, our study suggests that consideration of predictive biomarkers can have important implications in the treatment of lymph node-positive breast cancer. We may be able to identify a group of patients who may benefit from dose-intensive CAF, while other patients may be able to avoid the risk of CAF toxicity—perhaps the latter patients would derive greater benefit from another type of therapy.

REFERENCES

(1) Goldhirsch A, Gelber RD. Understanding adjuvant chemotherapy for breast cancer [editorial]. N Engl J Med 1994;330:1308–9.

(2) Bonadonna G. Karnofsky Memorial Lecture. Conceptual and practical advances in the management of breast cancer. J Clin Oncol 1989;7: 1380–97.

(3) Early Breast Cancer Trialists' Collaborative Group. Effects of adjuvant tamoxifen and cytotoxic therapy on mortality in early breast cancer: an overview of 61 randomized trials among 28,000 women. N Engl J Med 1988;319:1681–92.

(4) Early Breast Cancer Trialists' Collaborative Group. Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy: 133 randomized trials involving 31,000 recurrences and 24,000 deaths among 75,000 women. Lancet 1992;339:1–15, 71–85.

(5) Wood WC, Budman DR, Korzun AH, Cooper MR, Younger J, Hart RD, et al. Dose and dose intensity of adjuvant chemotherapy for stage II, node-positive breast carcinoma [published erratum appears in N Engl J Med 1994;331:139]. N Engl J Med 1994;330:1253–9.

(6) Recht A, Come SE, Henderson IC, Gelman RS, Silver B, Hayes DF, et al. The sequencing of chemotherapy and radiation therapy after conservative surgery for early-stage breast cancer. N Engl J Med 1996;334: 1356–61.

(7) National Institutes of Health Consensus Development Panel. Consensus statement: treatment of early-stage breast cancer. J Natl Cancer Inst Monogr 1992;11:1–5.

(8) Gusterson BA, Gelber RD, Goldhirsch A, Price KN, Save-Soderborgh J, Anbazhagan R, et al. Prognostic importance of c-erbB-2 expression in breast cancer. International (Ludwig) Breast Cancer Study Group. J Clin Oncol 1992;10:1049–56.

(9) Stal O, Sullivan S, Wingren S, Skoog L, Rutqvist LE, Carstensen JM, et al. c-erbB-2 expression and benefit from adjuvant chemotherapy and radiotherapy of breast cancer. Eur J Cancer 1995;31A:2185–90.

(10) Muss HB, Thor AD, Berry DA, Kute T, Liu ET, Koerner F, et al. c-erbB-2 expression and response to adjuvant therapy in women with node-positive early breast cancer. N Engl J Med 1994;330:1260–6.

(11) Clark GM. Prognostic and predictive factors. In: Harris JR, Lippman ME, Morrow M, Hellman S, editors. Diseases of the breast. Philadelphia: Lippincott-Raven; 1996. p. 461–85.

(12) Budman DR, Berry DA, Cirrincione CT, Henderson IC, Wood WC, Weiss RB, et al. Dose and dose intensity are determinants of outcome in the adjuvant treatment of breast cancer. J Natl Cancer Inst 1998;90:issue No. 16.

(13) Black MM, Opler SR, Speer FD. Survival in breast cancer cases in relation to the structure of the primary tumor and regional lymph nodes. Surg Gynecol Obstet 1955;100:543–51.

(14) Thor AD, Moore DH II, Edgerton SM, Kawasaki ES, Reihsaus ES, Lynch HT, et al. Accumulation of p53 tumor suppressor gene protein: an independent marker of prognosis in breast cancers. J Natl Cancer Inst 1992; 84:845–55.

(15) Kupryjanczyk J, Thor AD, Beauchamp R, Merritt V, Edgerton SM, Bell DA, et al. p53 gene mutations and protein accumulation in human ovarian cancer. Proc Natl Acad Sci U S A 1993;90: 4961–5.

(16) Liu E, Thor A, He M, Barcos M, Ljung BM, Benz C. The HER2 (c-erbB-2) oncogene is frequently amplified in *in situ* carcinomas of the breast. Oncogene 1992;7:1027–32.

(17) Neubauer A, Neubauer B, He M, Effert P, Iglehart D, Frye RA, et al. Analysis of gene amplification in archival tissue by differential polymerase chain reaction. Oncogene 1992;7:1019–25.

(18) Jacobs TW, Prioleau JE, Stillman IE, Schnitt SJ. Loss of tumor marker-immunostaining intensity on stored paraffin slides of breast cancer. J Natl Cancer Inst 1996;88:1054–9.

(19) Allred DC, Clark GM, Molina R, Tandon AK, Schnitt SJ, Gilchrist KW, et al. Overexpression of HER-2/neu and its relationship with other prognostic factors change during the progression of *in situ* to invasive breast cancer. Hum Pathol 1992;23:974–9.

(20) Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, et al. Design and analysis of randomized clinical trails requiring prolonged observation of each patient. I. Introduction and design. Br J Cancer 1976;34: 585–612.

(21) Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, et al. Design and analysis of randomized trials requiring prolonged observation of each patient. II. Analysis and examples. Br J Cancer 1977;35: 1–39.

(22) Zar JH. Biostatistical analysis. Englewood Cliffs (NJ): Prentice-Hall; 1974.

(23) Cox DR. Regression models and life tables (with discussion). J Royal Stat Soc 1972;34:187–220.

(24) Berry DA, Thor A, Cirrincione C, Edgerton S, Muss H, Marks J, et al. Scientific inference and predictions; multiplicities and convincing stories: a case study in breast cancer therapy. In: Bernardo JM, Berger JO, Dawid AP, Smith AF, editors. Baysian statistics 5. Oxford: Oxford University Press; 1996. p. 45–67.

(25) Press MF, Hung G, Godolphin W, Slamon DJ. Sensitivity of HER-2/neu antibodies in archival tissue samples: potential source of error in immunohistochemical studies of oncogene expression. Cancer Res 1994;54: 2771–7.

(26) Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science 1989;244:707–12.

(27) Ravdin PM, Chamness GC. The c-erbB-2 proto-oncogene as a prognostic and predictive marker in breast cancer: a paradigm for the development of other macromolecular markers—a review. Gene 1995;159:19–27.

(28) Allred DC, Clark GM, Tandon AK, Molina R, Tormey DC, Osborne CK, et al. HER-2/neu in node-negative breast cancer: prognostic significance of overexpression influenced by the presence of *in situ* carcinoma. J Clin Oncol 1992;10:599–605.

(29) Perloff M, Norton L, Korzun AH, Wood WC, Carey RW, Gottlieb A, et al. Postsurgical adjuvant chemotherapy of stage II breast carcinoma with or without crossover to a non-cross-resistant regimen: a Cancer and Leukemia Group B study. J Clin Oncol 1996;14:1589–98.

(30) Saffari B, Jones LA, el-Naggar A, Felix JC, George J, Press MF. Amplification and overexpression of HER-2/neu (c-erbB2) in endometrial cancers: correlation with overall survival. Cancer Res 1995;55:5693–8.

(31) Berry DA. When is a confirmatory randomized clinical trial needed? [editorial]. J Natl Cancer Inst 1996;88:1606–7

(32) Waldman T, Lengauer C, Kinzler KW, Vogelstein B. Uncoupling of S phase and mitosis induced by anticancer agents in cells lacking p21. Nature 1996;381:713–6.

(33) Lowe S, Jacks T. p53 and treatment of bladder cancer. Nature 1997;385: 123–4.

(34) Bitran JD, Samuels B, Trujillo Y, Klein L, Schroeder L, Martinee J. Her2/neu overexpression is associated with treatment failure in women with high-risk stage II and IIIA breast cancer treated with high-dose chemotherapy and autologous hematopoietic progenitor cell support following standard-dose adjuvant chemotherapy. Clin Cancer Res 1996;2: 1509–13.

NOTES

Present address: E. T. Liu, Division of Clinical Sciences, National Cancer Institute, National Institutes of Health, Bethesda, MD.

Supported by Public Health Service (PHS) CA44768 (A. D. Thor, S. Edgerton), CA33601 (D. A. Berry, C. Cirrincione), CA35279 (D. R. Budman), CA03927 (T. Kute), CA60138 (I. C. Henderson), CA59518 (M. Barcos), and CA47559 (E. T. Liu) from the National Cancer Institute (NCI), National Institutes of Health, Department of Health and Human Services. This research for CALGB 8869 (companion to 8541) was also supported by PHS grant CA31946 to the Cancer and Leukemia Group B (R. L. Schilsky, Chairman) and by PHS grant CA64061 from the NCI.

Manuscript received May 30, 1997; revised June 6, 1998; accepted July 16, 1998.