OXFORD

ARTICLE

Time-Varying Effects of Breast Cancer Adjuvant Systemic Therapy

Ismail Jatoi, Hanna Bandos, Jong-Hyeon Jeong, William F. Anderson, Edward H. Romond, Eleftherios P. Mamounas, Norman Wolmark

Affiliations of authors: NRG Oncology/ National Surgical Adjuvant Breast and Bowel Project, Pittsburgh, PA (IJ, EHR, EPM, NW); Department of Surgery University of Texas Health Science Center, San Antonio, TX (IJ); NRG Oncology Statistics & Data Management Center, and the Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA (HB, JHJ); NIH/NCI/Division of Cancer Epidemiology and Genetics, Bethesda, MD (WFA); Markey Cancer Center, University of Kentucky, Lexington, KY (EHR); UF Cancer Center at Orlando Health, Orlando, FL (EPM); Allegheny Cancer Center at Allegheny General Hospital, Pittsburgh, PA (NW).

Correspondence to: Ismail Jatoi, MD, PhD, FACS, Division of Surgical Oncology, Department of Surgery, University of Texas Health Science Center, 7703 Floyd Curl Drive, Mail Code 7738, San Antonio, TX 78229-3900 (e-mail: jatoi@uthscsa.edu).

Abstract

Background: The benefits of breast cancer adjuvant systemic treatments are generally assumed to be proportional (or constant) over time, but limited data suggest that some treatment effects may vary with time. We therefore systematically assessed the proportional hazards assumption across all 19 breast cancer adjuvant systemic therapy trials in the National Surgical Adjuvant Breast and Bowel Project (NSABP) database.

Methods: The NSABP breast cancer trials were tested for the proportionality of hazard rates between randomized treatment groups for five endpoints: overall survival, disease-free survival and recurrence, local-regional recurrence, or distant recurrence as first events. When the proportional hazards assumption did not hold, a "change point for the relative risk" technique was used to identify the temporal breakdown of the treatment effect.

Results: Time-varying treatment effects were observed in nearly half of the trials (nine of 19). In six (B-05, B-11, B-12, B-14, B-16, and B-20), novel treatment benefits diminished statistically significantly at specific time points following surgery. In B-09 and B-31, novel treatment benefits were delayed and emerged more than one year after surgery (1.57 and 1.32 years correspondingly), but the benefit in B-09 reversed after the third year of follow-up. In one trial (B-23), the initial advantage and subsequent disadvantage of one of the regimens was evident.

Conclusions: Breast cancer adjuvant systemic therapy can have statistically significant time-varying effects, which should be considered in the design, analysis, reporting, and translation of clinical trials. These time-dependent effects will have greater relevance as the number of long-term breast cancer survivors increases.

Following surgery for primary breast cancer, most patients receive adjuvant systemic therapy, and the benefits of these treatments are generally assumed to be proportional (or constant) over time (1). However, the proportional hazards assumption has never been systematically assessed in a large clinical trial dataset and time-varying treatment effects never fully elucidated. For instance, an early effect of breast cancer adjuvant therapy may potentially diminish over time after cessation of therapy, or there may exist a lag time before a treatment effect becomes evident.

In randomized clinical trials, time to occurrence of a clinically important event (eg, death, disease progression, etc.) is a primary outcome measure, and the Cox proportional hazards regression model is often used to evaluate the effect of treatment on this time-to-event outcome (2). Under the assumption of proportional hazards, the estimate of the hazard ratio (HR) for an endpoint (ie, the relative risk of having the event for a treated patient compared with an individual in the control group) provides summary results for the duration of the

ARTICLE

clinical trial. These summary estimates assume proportional risks and/or benefits (constant HR) irrespective of length of follow-up, even though the hazard rates may vary over time. For example, if a treatment reduces the average risk of recurrence by 50% during the first three years after breast cancer diagnosis (ie, HR of 0.5), then risk of recurrence should decrease from 8% to 4% at year three, 6% to 3% at year four, 2% to 1% at year six, and so forth. Thus, the absolute benefit might vary but the proportional (or relative) benefit would remain constant over time.

However, some reports now suggest that the proportional hazards assumption of the Cox model does not hold and that the relative effects of breast cancer adjuvant systemic therapy may indeed vary over time (3–6). Since 1971, the National Surgical Adjuvant Breast and Bowel Project (NSABP) has completed 19 breast cancer adjuvant therapy trials (7–23). More than ten years of average follow-up are now available for most of these trials, with some having more than 20 years of average follow-up. The NSABP dataset therefore provided a unique opportunity to assess the long-term effects of adjuvant systemic therapy in a large set of breast cancer trials. We have analyzed each of those trials to discern potential changes in the optimal efficacy of adjuvant systemic therapy at specific time points following breast cancer diagnosis and surgery.

Methods

NSABP Trial Data Synthesis

These trials were approved by local human investigations committees or institutional review boards in accordance with assurances filed with and approved by the Department of Health and Human Services. Written informed consent was required for participation in each trial.

The assumption of proportional hazards between treatment groups for overall survival (OS), disease-free survival (DFS), recurrence (R), local-regional recurrence (LRR), and distant recurrence (DR) as first events was tested for all completed breast cancer adjuvant NSABP trials, ie, B-05, B-07 to B-16, B-19, B-20, B-22, B-23, B-25, B-28, B-30, and B-31 (7–23). A brief summary of these trials is presented in Table 1. All time-to-event endpoints were measured from the date of random assignment to the date of diagnosis of the specified event. Patients otherwise event-free were censored at their last follow-up. DFS events included all local, regional, and distant recurrences, second primary cancers, and deaths from any cause. All of the analyses were based on the original treatment assignment.

Statistical Analysis

The standard Cox proportional hazards model was employed for the trials in which receptor status was not collected and trials with only either receptor-positive or receptor-negative populations. The Cox proportional hazards model, stratified by receptor status, was employed for the trials that included both receptor-positive and receptor-negative patients, ie, B-09, B-15, B-16, B-22, B-25, B-28, B-30, and B-31. Patients with estrogen receptor (ER)–positive and/or PgR-positive tumors were identified as receptor positive; otherwise the patients were considered to be receptor negative. Women with both ER and PgR status unknown were excluded from the analyses of these trials. Only eligible patients with follow-up were considered in the present work.

The formal test for lack of proportionality assumption for treatment indicator was performed by artificially creating a

Table 1. Summary of the NSABP breast cancer adjuvant systemic therapy clinical trials

Trial	Trial period*	Treatment	No. of patients†	Receptor status	Nodal status	No. of events‡				
						OS	DFS	R	LRR	DR
B-05	1972–1996	P vs placebo	348	Not recorded	+	256	276	214	76	136
B-07	1975–1996	PF vs P	671	Not recorded	+	446	517	392	130	262
B-08	1976–1996	PMF vs PF	686	Not recorded	+	444	506	404	131	266
B-09	1977–2004	PF+TAM vs PF	1545	ER±/PR±	+	1090	1222	949	324	616
B-10	1977–1996	PF+CP vs PF	254	Not recorded	+	154	181	159	54	105
B-11	1981–1996	PAF vs PF	683	ER- and /or PR-	+	378	441	371	162	209
B-12	1981–2004	PAF+TAM vs PF+TAM	1073	ER+	+	667	762	511	134	373
B-13	1981–2004	Surgery vs surgery + M→F	731	ER-	–	240	351	204	80	116
B-14	1982–2007	Placebo vs TAM	2817	ER+	–	1187	1652	796	312	477
B-15	1984–2004	AC vs AC+CMF(x3) vs CMF	2295	ER±/PR±	+	1182	1461	1208	425	761
B-16	1984–2004	AC+TAM vs TAM§	833	ER±/PR±	+	467	568	382	117	260
B-19	1988–2006	M→F vs CMF	1074	ER-	–	266	428	228	96	132
B-20	1988–2006	TAM vs M→F+TAM vs CMF+TAM	2299	ER+	–	439	774	380	148	231
B-22	1989–2006	AC vs AC(i) vs AC(ii)	2255	ER±/PR±	+	1053	1367	1048	364	681
B-23	1991–2006	AC±TAM vs CMF±TAM	1952	ER-	–	335	545	316	126	186
B-25	1992–2006	AC(ii) vs AC(ii-i) vs AC(ii-ii)	1977	ER±/PR±	+	794	1069	854	294	558
B-28	1995–2009	AC vs AC→T₁	3036	ER±/PR±	+	862	1266	919	288	626
B-30	1999–2012	AC→T₂ vs AT₂ vs AT₂C	5240	ER±/PR±	+	1169	1769	1218	262	948
B-31	2000-open	AC→T₁ vs AC→T₁+H	2050	ER±/PR±; HER2+	+	395	642	487	120	367

* Trial period: year opened to accrual – year closed to follow-up. A = adriamycin; C = cyclophosphamide; CP = corynebacterium parvum; DFS = disease-free survival; DR = distant recurrence; ER = estrogen receptor; F = 5-fluorouracil; H = herceptin; i = intensified; ii = intensified and increased; LRR = local-regional recurrence; M = methotrexate; OS = overall survival; P = L-Phenylalanine mustard; PR = progesterone receptor; R = recurrence; TAM = tamoxifen; T_1 = taxol; T_2 = taxotere.
† Eligible with follow-up and known receptor status.
‡ The location for some of the recurrences was unknown, and could not be classified as either LRR or DR.
§ B-16 is a three-arm trial, the primary comparison relates to these two regimens.

time-dependent covariate $Z(t) = TRT \times \ln(t)$ and testing the statistical significance of this interaction term at .05 level in the Cox regression model

$$h(t) = h_0(t)\exp\{\beta_1 \times TRT + \beta_2 \times Z(t)\}. \qquad (1)$$

The rejection of the null hypothesis $H_0 : \beta_2 = 0$ indicated the violation of the proportional hazards assumption. This test, originally proposed by Cox, has the advantage of simplicity and has been shown to have equally good power for detection of nonproportionality as some other commonly used tests (2,24). When proportional hazards assumption was not satisfied, we used a "change point" for the relative risk technique to identify the form of the time-dependent covariate (25,26). The time interval was split at point τ into two epochs ($t \le \tau$ and $t > \tau$), and the Cox proportional hazards model (1) was fit with the corresponding time-dependent function $Z(t)$ defined as

$$Z(t) = TRT \times I(t > \tau) = \begin{cases} TRT & \text{if } t > \tau \\ 0 & \text{if } t \le \tau \end{cases} \quad (2).$$ In general, when indicator

function is used to describe the time-dependent effect of treatment, the interpretation of the coefficients β_1 and β_2 becomes quite straightforward with $\exp(\beta_1)$ and $\exp(\beta_1 + \beta_2)$, reflecting the hazard ratio between two treatment groups before and after time point τ correspondingly. To identify the optimal value for τ, the model was fit for a set of τ values equal to each of the event times and the value that maximized the log partial likelihood was chosen. Proportional hazards assumption was then tested on each side of the change point τ, and the process was repeated if it was violated on either side of τ. As a secondary analysis to identifying the empirical change point, the time axis was divided at five years and the treatment effects were reported over the two intervals of five years or less and more than five years.

The standard estimator proposed by Kaplan and Meier was used to plot the estimates of the survival curves (27). The kernel-smoothed estimates of the hazards function were obtained from a standard univariate product-limit estimate. A plot of the logarithm of the cumulative hazard rates by treatment group vs time was used as an example of a graphical technique to check proportionality of the hazards assumption (25). The corresponding curves should be approximately parallel to each other if the assumption is not violated. All reported P values were two-sided, and all analyses were performed in SAS (Version 9.4, Cary, NC).

Results

The assumption of proportionality was violated for one or more endpoints in nine of the 19 NSABP trials: B-05, B-09, B-11, B-12, B-14, B-16, B-20, B-23, and B-31 (Table 2) (7,9,11,13,15,17,19,23). It did not hold for one endpoint in B-09 (DFS), B-12 (survival), B-20 (LR recurrence), and B-31 (distant recurrence). It did not hold for two endpoints in B-05 (DFS and recurrence) and B-16 (survival and DFS) and for three endpoints in B-11, B-14, and B-23 (DFS, recurrence, and LR recurrence).

Change points for the relative risks ranged from 1.1 to 9.1 years following initial surgery (Table 3). However, the change in treatment effect for the majority of the trials occurred between one and four years. Change points ranged between one to two years for B-05, B-23, and B-31, between two to three years in B-12, and between three to four years in B-11 (recurrence and LR recurrence), B-14 (DFS and recurrence), B-16, and B-20. A change point of approximately 4.5 years was identified in B-11 (DFS) and about nine years in B-14 (LR recurrence). Two change points for the relative risk of DFS events were identified for the B-09 protocol (Table 4).

In B-05, patients were randomly assigned to L-Phenylalanine mustard (P) vs placebo, and the entire benefit of the

Table 2. Results of the test of proportionality of the hazards* for different treatment groups by dataset and endpoint

			P		
Trial	Survival	Disease-free survival	Recurrence	Local-regional recurrence	Distant recurrence
B-05	.86	.05	.01	.06	.16
B-07	.17	.52	.35	.29	.10
B-08	.55	.40	.23	.08	.73
B-09†	.67	.03	.11	.16	.47
B-10	.23	.32	.33	.61	.10
B-11	1.00	<.01	.04	.04	.49
B-12	.02	.12	.07	.13	.33
B-13	.83	.34	.87	.32	.33
B-14	.23	<.01	.01	<.01	.12
B-15†	.78	.96	.51	.70	.43
B-16†	.04	.02	.09	.25	.25
B-19	.31	.10	.34	.54	.21
B-20	.77	.21	.18	.03	.47
B-22†	.89	.81	.98	.55	.63
B-23	.17	<.01	.01	<.01	.42
B-25†	.28	.89	.20	.08	.80
B-28†	.36	.71	.84	.72	.60
B-30†	.38	.30	.15	.71	.13
B-31†	.85	.53	.38	.19	.03

* Tested the statistical significance of the presence of interaction between treatment and ln(time). Two-sided P values less than or equal to .05 were considered statistically significant.
† Using the stratified-by-receptor-status Cox regression model.

Table 3. Summary of the findings for the trials in which proportionality of the hazards was not satisfied*

Trial	Outcome	Median follow-up time, y	Treatment group	HR (95% CI)†	P‡	Change point, y, τ	First interval, ≤ τ HR (95% CI)	P	Second interval, > τ HR (95% CI)	P
B-05	DFS	21.0	Placebo	Ref						
			P	0.83 (0.66 to 1.05)	.004	1.14	0.45 (0.27 to 0.73)	.001	1.02 (0.78 to 1.35)	.88
	Recurrence		P	0.78 (0.60 to 1.02)	.002	1.14	0.39 (0.23 to 0.66)	<.001	1.03 (0.75 to 1.43)	.84
B-11	DFS	12.7	PF	Ref						
			PAF	0.81 (0.67 to 0.97)	.002	4.47	0.68 (0.55 to 0.85)	<.001	1.44 (0.96 to 2.18)	.08
	Recurrence		PAF	0.75 (0.61 to 0.92)	.01	3.79	0.65 (0.52 to 0.83)	<.001	1.23 (0.78 to 1.92)	.37
	LR Recurrence		PAF	0.59 (0.43 to 0.81)	.006	3.81	0.48 (0.33 to 0.68)	<.001	1.62 (0.73 to 3.61)	.24
B-12	OS	20.3	PF+TAM	Ref						
			PAF+TAM	1.02 (0.88 to 1.19)	.002	2.79	0.54 (0.35 to 0.83)	.005	1.13 (0.96 to 1.33)	.15
B-14	DFS	20.4	Placebo	Ref						
			TAM	0.74 (0.67 to 0.82)	<.001	3.42	0.53 (0.44 to 0.65)	<.001	0.84 (0.75 to 0.94)	.002
	Recurrence		TAM	0.59 (0.51 to 0.68)	.006	3.47	0.46 (0.37 to 0.58)	<.001	0.70 (0.58 to 0.84)	<.001
	LR recurrence		TAM	0.47 (0.37 to 0.60)	.001	9.08	0.35 (0.26 to 0.48)	<.001	0.79 (0.54 to 1.16)	.23
B-16	OS	16.1	TAM	Ref						
			AC+TAM	0.89 (0.74 to 1.07)	.002	3.47	0.51 (0.34 to 0.76)	.001	1.03 (0.84 to 1.27)	.74
	DFS		AC+TAM	0.84 (0.71 to 0.99)	.003	3.85	0.63 (0.49 to 0.81)	<.001	1.05 (0.84 to 1.31)	.69
B-20	LR recurrence	14.6	TAM	Ref						
			M→F+TAM	0.76 (0.53 to 1.08)	.02	3.8	0.59 (0.34 to 1.03)	.07	0.91 (0.57 to 1.45)	.68
			CMF+TAM	0.35 (0.22 to 0.55)			0.09 (0.03 to 0.29)	<.001	0.58 (0.34 to 0.99)	.04
B-23	DFS	11.0	CMF±TAM	Ref						
			AC±TAM	1.13 (0.96 to 1.34)	.001	1.5	0.66 (0.46 to 0.96)	.03	1.31 (1.08 to 1.59)	.005
	Recurrence		AC±TAM	1.06 (0.85 to 1.32)	.006	1.5	0.65 (0.43 to 0.98)	.04	1.30 (1.00 to 1.70)	.05
	LR recurrence		AC±TAM	1.14 (0.81 to 1.62)	<.001	1.93	0.50 (0.28 to 0.89)	.02	2.03 (1.26 to 3.28)	.004
B-31	Distant recurrence	9.0	AC→T_1	Ref						
			AC→T_1+H	0.51 (0.42 to 0.64)	.02	1.32	0.81 (0.53 to 1.25)	.35	0.45 (0.35 to 0.57)	<.001

* All statistical tests were two-sided. A = adriamycin; C = cyclophosphamide; CI = confidence interval; DFS = disease-free survival; F = 5-fluorouracil; H = herceptin; HR = hazard ratio; LR = local-regional; M = methotrexate; OS = overall survival; P = L-Phenylalanine mustard; TAM = tamoxifen; T_1 = taxol; y = years.
† Assuming proportionality of hazards.
‡ For statistical significance of time-dependent treatment effect (P value for β_2 in [1] with Z(t), as defined in [2]).

chemotherapeutic agent occurred during the first 1.14 years following surgery. In B-11, the addition of doxorubicin to P and 5-fluorouracil (F) statistically significantly improved DFS only during the first 4.47 years. In B-12, the addition of doxorubicin to P, F, and tamoxifen (TAM) improved OS during the first 2.79 years, with no benefit after that period. In B-14, adjuvant TAM had its greatest effect in improving DFS and reducing overall recurrences during the first 3.42 and 3.47 years, respectively, but it continued to have a statistically significant (though diminished) benefit thereafter. Moreover, TAM had a statistically significant benefit in reducing LR recurrences for up to 9.08 years (Table 3).

In B-16, a short course of chemotherapy (doxorubicin and cyclophosphamide) with TAM was found to have a superior benefit on OS than TAM alone in postmenopausal patients with node-positive, primarily receptor-positive tumors. However, that effect on OS was only evident during the initial 3.47 years following surgery, with no benefit thereafter (Table 3). In B-20, chemotherapy and TAM was compared with TAM alone in patients with ER-positive, node-negative tumors, and the addition of chemotherapy (cyclophosphamide, methotrexate, and F [CMF]) was primarily beneficial in reducing the risk of LR recurrence during the initial 3.8 years following surgery. After that period, CMF continued to have a benefit, although the magnitude of that benefit was statistically significantly reduced (Table 3).

In B-23, patients randomly assigned to doxorubicin and cyclophosphamide (+/-TAM) had better prognosis initially

Table 4. Summary of the findings for the NSABP B-09 protocol*

Trial	Outcome	Median follow-up time, y	Treatment group	HR (95% CI) †	P‡	Change point, y, τ	First interval, ≤ 1.57 HR (95% CI)	P	Second interval, > 1.57, ≤ 3.32 HR (95% CI)	P	Third interval, > 3.32 HR (95% CI)	P
B-09	DFS	23.7	PF	Ref	.003	1.57						
			PF+TAM	0.99 (0.88 to 1.10)	<.001	3.32	1.01 (0.81 to 1.27)	.91	0.62 (0.50 to 0.78)	<.001	1.24 (1.05 to 1.45)	.01

* All statistical tests were two-sided. CI = confidence interval; DFS = disease-free survival; F = 5-fluorouracil; HR = hazard ratio; NSABP = National Surgical Adjuvant Breast and Bowel Project; P = L-Phenylalanine mustard; TAM = tamoxifen.
† Assuming proportionality of hazards.
‡ For statistical significance of time-dependent treatment effects.

with respect to DFS (until 1.5 years) and LR recurrences (until 1.93 years) compared with patients randomly assigned to cyclophosphamide, methotrexate, and F (+/- TAM), but they experienced a delayed detriment later on.

Delayed treatment effect was evident in B-31. A statistically significant benefit of adjuvant trastuzumab on distant recurrences emerged 1.32 years after surgery and remained essentially constant thereafter (HR = 0.45, 95% confidence interval [CI] = 0.35 to 0.57, $P < .001$) (Table 3). However, patients in the study arm of the B-31 trial received one year of trastuzumab, and it may have potentially taken this long to realize the full benefits of therapy.

In B-09 (PF+TAM vs PF), the benefit of adding TAM to the chemotherapy regimen did not emerge until 1.57 years after surgery (HR = 0.62, 95% CI = 0.50 to 0.78, $P < .001$), but after 3.32 years the effect reversed (HR = 1.24, 95% CI = 1.05 to 1.45, $P = .01$) (Table 4 and Figures 1 and 2). When an unstratified analysis of this trial was performed by including all of the patients regardless of their receptor status, the same two change points for the relative risk (1.57 and 3.32 years) were detected with no treatment effect on the first interval, statistically significant benefit of PF+TAM on the second time interval, and the reversed treatment effect on the last time interval. However, the beneficial effect of PF after 3.32 years was not statistically significant (HR = 1.15, 95% CI = 0.99 to 1.33, $P = .07$).

Supplementary Table 1 (a table similar to Table 3, available online) presents the results of our secondary analyses in which a fixed change point of five years was employed. As expected, the results were similar among the trials, in which an empirically identified change point was close to five years (B-11, B-14, B-16, and B-20). However, the point estimates of the hazard ratios and their statistical significance were quite different for the rest of the trials. In addition, the nonproportionality of the hazards could no longer be detected in the majority of the studies.

Discussion

The hazard ratio point estimates indicate average treatment effects during the window of a clinical trial, but clinicians often utilize these summary measures to predict long-term treatment effects under the assumption that such effects are proportional (or constant) over time. Yet, our study suggests that this assumption may no longer hold for long-term follow-up. Indeed, nine of the 19 NSABP breast cancer adjuvant systemic therapy trials demonstrated nonproportional (time-varying) treatment effects. In six of these trials (B-05, B-11, B-12, B-14, B-16, and B-20), the benefits of adjuvant therapy diminished statistically significantly at specific time points following
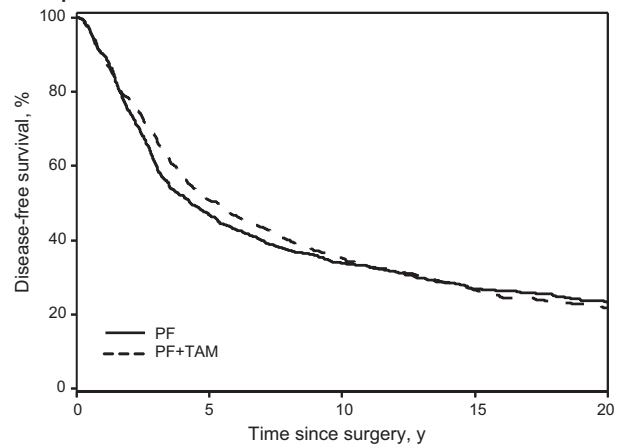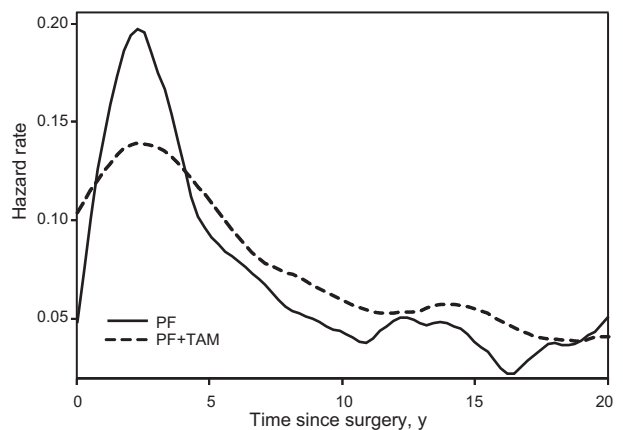

Figure 1. Kaplan-Meier estimator **(A)** and smoothed hazard rates **(B)** of disease-free survival for National Surgical Adjuvant Breast and Bowel Project (NSABP) B-09, PF + TAM vs PF. **A)** The Kaplan-Meier estimator for disease-free survival function is a 'cumulative' measure that describes the fraction of patients surviving disease-free up to a given time point (following the successful intervention for the initial breast cancer diagnosis). **B)** The hazard rate for disease-free survival is a 'conditional' measure that describes the instantaneous rate of predefined events among patients who are disease free at a given time point (following the successful intervention for the initial breast cancer diagnosis). F = 5-fluorouracil; P = L-Phenylalanine mustard; TAM = tamoxifen.

surgery, generally before the fourth year. In B-23, novel treatment was associated with an initial benefit and then a delayed increased risk of recurrence. In two other trials (B-09 and

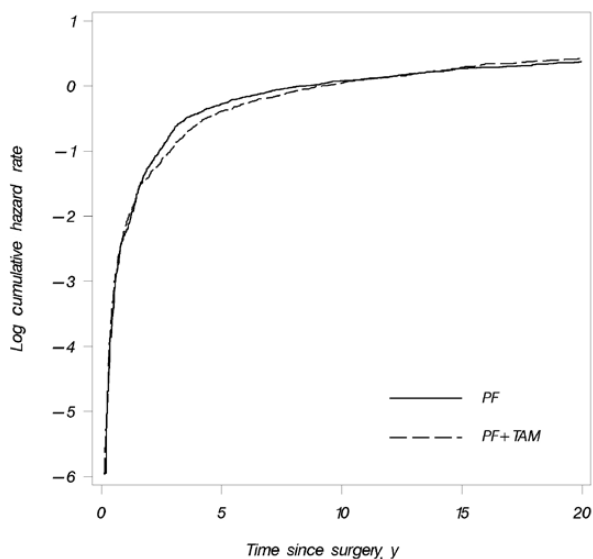Graphical check of the PH assumption: NSABP B-09

Figure 2. Plot of the logarithm of the cumulative hazard rates vs time for PF + TAM and PF treatment groups for National Surgical Adjuvant Breast and Bowel Project (NSABP) B-09. F = 5-fluorouracil; P = L-Phenylalanine mustard; TAM = tamoxifen.

B-31), treatment benefits were delayed, although the benefit in B-09 was transient, ie, lasting between approximately 1.5 and 3.5 years following initial surgery, with a reversal of the effect after 3.5 years.

In the B-09 trial, 17% of patients had unknown receptor status and were excluded from the current analyses. Among patients with known receptor status, there was a 70/30 mixture of patients with receptor-positive/receptor-negative tumors, who were equally split between PF+TAM versus PF, with TAM given for only two years. The large fraction of receptor-negative tumors might partially explain the initial delay in effect with the subsequent short interval of benefit, possibly because of the short, two-year course of TAM. For statistical, not clinical, verification, an unstratified sensitivity analysis of this trial was performed by including all of the patients regardless of their receptor status. The same two change points for the relative risk were detected, and the results, similar to the stratified analysis, were obtained in terms of treatment effects on each interval. A more detailed analysis of this trial is beyond the scope of this manuscript. Other trials have shown a superiority to more extended regimens, ie, five years of TAM superior to two years and 10 years superior to five years (28).

Prior to this report, a few other studies had suggested that the benefits of many adjuvant systemic therapies were nonproportional and often limited to the initial years following surgery. Analysis of clinical trial data from the National Cancer Institute of Milan comparing hazard functions among patients treated with CMF vs the untreated control group found that essentially all the benefit of CMF occurred during the first four years following surgery (4). Additionally, analysis of the Cancer and Leukemia Group B (CALGB) and US Breast Cancer Intergroup data revealed that high-dose vs low-dose adjuvant cyclophosphamide, doxorubicin, and F reduced the risk of breast cancer recurrence or death by 55% in the first year following surgery and 30% in the second year, with no advantage thereafter (5). Moreover, inspection of recurrence hazard curves for node-negative breast cancers in the NSABP trials seemed to suggest that

much of the benefit of adjuvant therapy was evident during the initial years following surgery (6).

Although we have dwelled on the nonproportional effects of breast cancer adjuvant systemic therapy, it is widely acknowledged that standard breast cancer prognostic and predictive factors (ER expression, tumor size, S-phase fraction, etc.) also have nonproportional effects (1,29,30). For example, the ER-positive to ER-negative hazard ratio for breast cancer death varies over time and at its extreme is less than 1.0 during the initial eight years following surgery, and thereafter the hazard ratio is greater than 1.0 (1). The nonproportional effects of adjuvant systemic therapy are consistent with these observations. Indeed, if predictive biomarkers (such as ER expression) show nonproportional effects, then the biological mechanisms responsible for early and late breast cancer–specific events may vary, and one should also consider the possibility of time-varying treatment effects.

The optimal graphic representation of time-dependent treatment effects merits some consideration. The Kaplan-Meier plots of the survival curves by treatment group are usually chosen to present the results of a clinical trial. It is a cumulative measure that depicts the fraction of patients surviving event-free up to a given time point over the course of a clinical trial (Figure 1A) (27). However, breast cancer is a chronic and heterogeneous disease that may recur many years after initial diagnosis and treatment. Given this clinical uncertainty, outcomes might be better visualized with the hazard curves, especially when the proportional hazards assumption is not satisfied (6,31). The hazard function is a conditional measure that depicts the instantaneous rate or "force" of failure (Figure 1B). For example, in B-09, the hump-shaped DFS hazard rate curves of the two treatment groups cross at two time points, indicating an initial delayed benefit of the experimental treatment that was reversed later (Figure 1B) (9). Yet, this time-varying treatment effect is not immediately evident from the Kaplan-Meier plot, in which the reversal of the experimental treatment effect becomes visually apparent only when the DFS curves converge around eight years and then eventually cross at 12 years (Figure 1A). Although B-09 is not a recent trial, it demonstrates that curves of hazard rates over time are potentially as important as Kaplan-Meier plots. In general, kernel-based methods are often used to produce smoothed plots of the hazard function estimates (25). However, because the decreasing number of patients at risk generally increases the bias and the variance of the estimator, one must be careful in interpreting these estimates, especially in later follow-up (so called "boundary effect" of smoothed hazard function estimates). In addition, special attention should be paid to the choice of the bandwidth, which determines the degree of smoothness of the kernel-smoothed estimator, because more smoothness also generally leads to increased bias, even though it might result in a lower variability (32).

In the presence of a time-dependent treatment effect, the baseline hazard function (for the standard therapy group) provides important information (33). As is evident from Figure 1B, the baseline risk of DFS events in the control group of B-09 is quite high in the first five years after surgery. Therefore, the majority of DFS events are observed during this time frame and it dominates the overall treatment effect. When interpreting results similar to B-05, for example, it might be inaccurate to simply suggest that the experimental treatment no longer works in the longer follow-up (7). Rather, it worked initially, but failure rates in surviving patients then became similar in both treatment groups (Table 3).

A limitation of our study relates to the issue of multiple comparisons. Multiple endpoints for multiple trials were screened based on the P values. However, because one of the main purposes

of this study was to examine the departures from proportionality in the long-term follow-up trials, the statistical significance level of .05 was chosen in order to increase the likelihood of detection. Once the nonproportionality of hazards between the treatment groups was detected, the "change point" τ for the relative risk was identified. As estimates of the change points are derived based on the data in hand, their independent validation would be desirable. In addition, knowledge of the distributional properties of τ would be useful for the construction of the confidence intervals and are currently being investigated (26). The discovery of the change-points for the trials in our study was purely from the statistical standpoint. Their interpretation from the clinical point of view might be difficult. Similar to our secondary analyses, other authors have suggested using a fixed time point during the follow-up, such as five or seven years, and consider early and late differences between treatment groups (6).

In summary, the goal of adjuvant therapy is to positively impact both short-term and long-term outcomes and hopefully to cure patients with breast cancer. Nonproportional treatment effects should therefore be considered in the design, analysis, and reporting of breast cancer adjuvant systemic therapy trials. The proportionality of hazards is often a reasonable assumption for short-term duration clinical trials. However its violation is an important issue that frequently arises in long-term follow-up (33,34). Even though no formal statistical test of nonproportionality was employed for the primary analysis of the earlier trials, the authors commented on the diminished treatment effect over time for B-05 and B-09 (7–9). The remainder of the trials did not report on the nonproportionality of the hazards at the time of their primary analyses. However, we were able to detect the violation of this assumption in more than half of the trials as long-term follow-up data became available. Even though the hazard ratio point estimates, which represent average treatment effects over the duration of a clinical trial, could still be a reasonable summary measure, provided that the individual hazards are not strongly nonproportional (6), the results of our study highlight its potential shortcomings. The check for proportional hazard assumption is routinely done at the analysis stage of the trial, and several approaches exist to handle the data revealing the violation of this assumption, one of which was presented in the current work. At the same time, the proportionality of hazards is often assumed at the design stage of the clinical trial. Even though it is widely accepted that it is hard to design a trial under nonproportionality assumption, careful consideration should be given to the potential time-dependent treatment effect (34–36). Many adjuvant treatments only have early benefits, and longer follow-up may dilute the beneficial point estimates of those effects. Conversely, the early stoppage of some trials may potentially obscure important late effects. Thus, careful consideration should be given to developing appropriate timelines for the primary analysis of breast cancer adjuvant therapy trials, with particular attention paid to disease-specific endpoints and any scientific or statistical evidence suggesting time-varying treatment effects. Moreover, with recent improvements in breast cancer therapy, greater numbers of breast cancer patients will now survive longer. Some of these long-term survivors are at increased risk for delayed recurrences, and additional extended adjuvant therapy trials should be developed to assess novel long-term treatment strategies.

Funding

Notes

References

1. Jatoi I, Anderson WF, Jeong JH, et al. Breast cancer adjuvant therapy: time to consider its time-dependent effects. *J Clin Oncol.* 2011;29(17):2301–2304.
2. Cox DR. Regression Models and Life-Tables. *J Royal Stat Soc. Series B (Methodol).* 1972;34(2):187–220.
3. Jeong JH, Jung SH, Wieand S. A parametric model for long-term follow-up data from phase III breast cancer clinical trials. *Stat Med.* 2003;22(3):339–352.
4. Demicheli R, Miceli R, Moliterni A, et al. Breast cancer recurrence dynamics following adjuvant CMF is consistent with tumor dormancy and mastectomy-driven acceleration of the metastatic process. *Ann Oncol.* 2005;16(9):1449–1457.
5. Berry DA, Cirrincione C, Henderson IC, et al. Estrogen-receptor status and outcomes of modern chemotherapy for patients with node-positive breast cancer. *JAMA.* 2006;295(14):1658–1667. Erratum in *JAMA.* 2006;295(20):2356.
6. Dignam JJ, Dukic V, Anderson SJ, et al. Hazard of recurrence and adjuvant treatment effects over time in lymph node-negative breast cancer. *Breast Cancer Res Treat.* 2009;116(3):595–602.
7. Fisher B, Fisher ER, Redmond C. Ten-year results from the National Surgical Adjuvant Breast and Bowel Project (NSABP) clinical trial evaluating the use of L-phenylalanine mustard (L-PAM) in the management of primary breast cancer. *J Clin Oncol.* 1986;4(6):929–941.
8. Fisher B, Redmond C, Wolmark N, et al. Disease-free survival at intervals during and following completion of adjuvant chemotherapy: The NSABP experience from three breast cancer protocols. *Cancer.* 1981;48(6):1273–1280.
9. Fisher B, Redmond C, Brown A, et al. Adjuvant chemotherapy with and without tamoxifen in the treatment of primary breast cancer: 5-year results from the National Surgical Adjuvant Breast and Bowel Project Trial. *J Clin Oncol.* 1986;4(4):459–471.
10. Fisher B, Brown A, Wolmark N, et al. Evaluation of the worth of corynebacterium parvum in conjunction with chemotherapy as adjuvant treatment for primary breast cancer. Eight-year results from the National Surgical Adjuvant Breast and Bowel Project B-10. *Cancer.* 1990;66(2):220–227.
11. Fisher B, Redmond C, Wickerham DL, et al. Doxorubicin-containing regimens for the treatment of stage II breast cancer: The National Surgical Adjuvant Breast and Bowel Project experience. *J Clin Oncol.* 1989;7(5):572–582.
12. Fisher B, Redmond C, Dimitrov NV, et al. A randomized clinical trial evaluating sequential methotrexate and fluorouracil in the treatment of patients with node-negative breast cancer who have estrogen-receptor-negative tumors. *N Engl J Med.* 1989;320(8):473–478.
13. Fisher B, Costantino J, Redmond C, et al. A randomized clinical trial evaluating tamoxifen in the treatment of patients with node-negative

breast cancer who have estrogen-receptor-positive tumors. *N Engl J Med.* 1989;320(8):479–484.

14. Fisher B, Brown AM, Dimitrov NV, et al. Two months of doxorubicin-cyclo-phosphamide with and without interval reinduction therapy compared with 6 months of cyclophosphamide, methotrexate, and fluorouracil in positive-node breast cancer patients with tamoxifen-nonresponsive tumors: results from the National Surgical Adjuvant Breast and Bowel Project B-15. *J Clin Oncol.* 1990;8(9):1483–1496.

15. Fisher B, Redmond C, Legault-Poisson S, et al. Postoperative chemotherapy and tamoxifen compared with tamoxifen alone in the treatment of positive-node breast cancer patients aged 50 years and older with tumors responsive to tamoxifen: results from the National Surgical Adjuvant Breast and Bowel Project B-16. *J Clin Oncol.* 1990;8(6):1005–1018.

16. Fisher B, Dignam J, Mamounas EP, et al. Sequential methotrexate and fluo-rouracil for the treatment of node-negative breast cancer patients with estrogen receptor-negative tumors: eight-year results from National Sur-gical Adjuvant Breast and Bowel Project (NSABP) B-13 and first report of findings from NSABP B-19 comparing methotrexate and fluorouracil with conventional cyclophosphamide, methotrexate, and fluorouracil. *J Clin Oncol.* 1996;14(7):1982–1992.

17. Fisher B, Dignam J, Wolmark N, et al. Tamoxifen and chemotherapy for lymph node-negative, estrogen receptor-positive breast cancer. *J Natl Cancer Inst.* 1997;89(22):1673–1682.

18. Fisher B, Anderson S, Wickerham DL, et al. Increased intensification and total dose of cyclophosphamide in a doxorubicin-cyclophosphamide regimen for the treatment of primary breast cancer: findings from National Surgical Adjuvant Breast and Bowel Project B-22. *J Clin Oncol.* 1997;15(5):1858–1869.

19. Fisher B, Anderson S, Tan-Chiu E, et al. Tamoxifen and chemotherapy for axillary node-negative, estrogen receptor-negative breast cancer: Findings from National Surgical Adjuvant Breast and Bowel Project B-23. *J Clin Oncol.* 2001;19(4):931–942.

20. Fisher B, Anderson S, DeCillis A, et al. Further evaluation of intensified and increased total dose of cyclophosphamide for the treatment of primary breast cancer: findings from National Surgical Adjuvant Breast and Bowel Project B-25. *J Clin Oncol.* 1999;17(11):3374–3388.

21. Mamounas EP, Bryant J, Lembersky B, et al. Paclitaxel after doxorubicin plus cyclophosphamide as adjuvant chemotherapy for node-positive breast can-cer: results from NSABP B-28. *J Clin Oncol.* 2005;23(16):3686–3696.

22. Swain SM, Jeong JH, Geyer CE Jr, et al. Longer therapy, iatrogenic amenorrhea, and survival in early breast cancer. *N Engl J Med.* 2010;362(22):2053–2065.

23. Romond EH, Perez EA, Bryant J, et al. Trastuzumab plus adjuvant chemother-apy for operable HER2-positive breast cancer. *N Engl J Med.* 2005;353(16):1673–1684.

24. Ng'andu NH. An empirical comparison of statistical tests for assessing the proportional hazards assumption of Cox's model. *Stat Med.* 1997;16(6):611–626.

25. Klein JP, Moeschberger ML. *Survival Analysis: Techniques for Censored and Trun-cated Data.* Statistics for Biology and Health; 2003.

26. Liang KY, Self SG, Liu XH. The Cox proportional hazards model with change point: an epidemiologic application. *Biometrics.* 1990;46(3):783–793.

27. Kaplan EL, Meier P. Nonparametric-Estimation from Incomplete Observa-tions. *J Am Stat Assoc.* 1958;53(282):457–481.

28. Burstein HJ, Temin S, Anderson H, et al. Adjuvant endocrine therapy for women with hormone receptor-positive breast cancer: American Society of Clinical Oncology clinical practice guideline focused update. *J Clin Oncol.* 2014;32(21):2255–2269.

29. Hilsenbeck SG, Ravdin PM, de Moor CA, Chamness GC, Osborne CK, Clark GM. Time-dependence of hazard ratios for prognostic factors in primary breast cancer. *Breast Cancer Res Treat.* 1998;52(3):227–237.

30. Bryant J, Fisher B, Gündüz N, Costantino JP, Emir B. S-phase fraction com-bined with other patient and tumor characteristics for the prognosis of node-negative, estrogen-receptor-positive breast cancer. *Breast Cancer Res Treat.* 1998;51(3):239–253.

31. Jatoi I, Tsimelzon A, Weiss H, et al. Hazard rates of recurrence following diag-nosis of primary breast cancer. *Breast Cancer Res Treat.* 2005;89:173–178.

32. Hess KR, Serachitopol DM, Brown BW. Hazard Function Estimators: A Simu-lation Study. *Stat Med.* 1999;18(2):3075–3088.

33. Putter H, Sasako M, Hartgrink HH, et al. Long-term survival with non-pro-portional hazards: results from the Dutch Gastric Cancer Trial. *Stat Med.* 2005;24(18):2807–2821.

34. Lu J, Pajak TF. Statistical power for a long-term survival trial with a time-dependent treatment effect. *Controlled Clin Trials.* 2000;21(6):561–573.

35. Lakatos E, Lan KKG. A comparison of sample size methods for the logrank statistic. *Stat Med.* 1992;11:179–191.

36. Royston P, Parmar MKB. An approach To trial design and analysis in the era of non-proportional hazards of the treatment effect. *Trials.* 2014;15:314.