Statistical Interaction In The Survival Analysis Of Early Breast Cancer Using Registry Data

A Danzel

Citation

A Danzel. Statistical Interaction In The Survival Analysis Of Early Breast Cancer Using Registry Data. The Internet Journal of Oncology. 2004 Volume 2 Number 2.

Abstract

Purpose: An interaction analysis is used to investigate subgroup effects between breast conserving surgery, radiotherapy and nodal status in early breast cancer.

Patients and methods: Women selected from the SEER database, aged 40-69 with non metastasized T1-T2 breast cancer and in whom axillary node dissection was performed. The method uses proportional hazards regression models with interaction terms. Steps of the analysis are detailed.

Results: Significant statistical interactions were found between breast conserving surgery, radiotherapy, nodal involvement and T stage. In an integrated factorial presentation of results, patients undergoing breast conserving surgery without radiotherapy were not at survival disadvantage as compared with patients undergoing mastectomy, irrespective of T-stage and irrespective of nodal involvement. However, the addition of radiotherapy to breast conserving surgery was associated with a distinct survival advantage observed in all T stage and all levels of nodal involvement.

Conclusion: The analysis of interactions provides a comprehensive view of complex subgroup effects. The formal presentation of results might be useful as a framework to generate hypotheses for further investigations.

INTRODUCTION

Extreme heterogeneity of neoplastic cells from cell transformation to invasion and metastasis is well known $(_{1,2,3,4})$. Cells originating from the same tumor may express entirely different behavior $(_5)$. Gene expression levels from the same gene mutation are not identical $(_6)$. Circumstances of diagnosis and socio-demographic differences $(_{7,8})$ contribute to the heterogeneity. In view of this variability, both biological and non-biological, it cannot be expected that any given treatment would be identically efficient in all patients $(_9)$. Identifying subsets of patients that might or might not benefit from a treatment, termed "subgroup analyses", is a cornerstone for medical practice.

Typically, in a subgroup analysis treatment effects in multiple subsets of patients enrolled, e.g., in a clinical trial, will be systematically evaluated ($_{10,11}$). Such an analyses is subject to several problems, related to the risk of false-positive and false-negative findings ($_{12}$).

A false-positive finding is a 'statistically significant' result, which is due to chance rather than to a non-zero treatment effect. The probability of such a finding increases with the number of comparisons performed using a particular dataset ($_{13,14}$). For instance, even if treatment has no effect, when 100 subgroup tests is performed, each at 0.05 significance level, the probability of finding at least one statistically significant (and thus false-positive) result is 0.994. We are thus almost certain to get at least one 'positive' finding. In fact, the expected number of false-positive results will be $100 \times 0.05 = 5$ ($_{15}$).

A false-negative finding, on the other hand, is the 'statistically non-significant result', or failure to detect a treatment effect, when the effect in fact exists. The probability of such a finding increases with a decreasing sample size. Subgroup testing requires the data to be split and these smaller datasets will therefore have a reduced power to detect a treatment effect.

To address these issues, several recommendations have been

formulated. In essence, they suggest the use of a limited number of pre-planned comparisons within pre-specified groups; adjustment of significance levels for multiplicity of comparisons; adjustment of trial size; and inference based on formal tests of interaction in appropriate regression analyses (12>13>15>16>17>18).

While most of the recommendations are intuitively easily understood, the use of tests of interaction presents a practical problem. Most likely, this can be attributed to the somewhat more complex nature of the technique. From this point of view it is unfortunate that major textbooks sometimes do not list interaction as a keyword ($_{19}$), incorporate interaction terms in models without explanatory guidance ($_{20}$), or treat the more difficult issues with sentences like: "Interactions among more than two variables can be exceedingly complex" ($_{21}$). There is still a clear need for a careful explanation and exemplary applications of the technique, which could encourage its more extensive use.

The purpose of this paper is two fold, clinical and technical. Using conventional survival analysis models applied to a large population data, we found previously that radiotherapy was associated with a survival advantage in that population data (₂₂). In the present study we use the analysis of interaction to investigate whether the effect of post-operative radiotherapy may depend on type of surgery, stage of breast cancer, nodal status and extent of node examination. The clinical investigation is justified in view of the continuing debate on the role of radiotherapy in different subgroups of patients (₂₃). However, in the paper somewhat more attention will be given to the presentation of the analytical steps, to give more insight into the use of the tests of interaction.

PATIENTS AND METHODS

Patients were selected from the SEER (Surveillance, Epidemiology, and End Results) nine-registries database (₂₄) according to the following criteria: women diagnosed 1988-1998, aged 40-69, primary non-inflammatory histologically confirmed invasive carcinoma confined to breast, tumor largest diameter < 50 mm, no previous cancer, no known internal mammary node involvement or distant metastases, curative surgery performed with axillary dissection, and post-surgery radiation explicitly stated as delivered or not. Exclusion criteria were: non-hospital or non-clinic based record, unknown race, unknown month of diagnosis, discrepancy between nodal status and reported nodal involvement.

In the SEER data collection prior to 1998, receipt of

radiotherapy was defined as radiation delivered within 4 months of the initial treatment. Information on systemic treatment was not available for use as a variable. Hormone receptors status was available for 1990 and later.

The effect of treatments and patient characteristics on survival time was examined using the proportional hazards model (₂₅). The following variables were included in the model: patient registration area, age at diagnosis, race, marital status, multiple primaries indicator, histology, grade, estrogen receptor status, tumor size (largest diameter), number of nodes examined, number of positive nodes, type of surgery, and delivery of post-surgery radiation. Quantitative variables were used in the models as untransformed continuous variables. Qualitative data were re coded using binary variables.

Two variables are said to interact in their effect on the response if the effect of the first variable is different at the different levels of the second variable (26). Thus, the effect of the first variable (e.g., treatment) is different in the subgroups defined by the levels of the second variable (e.g., tumor size). In modelling, a simple way to describe the interaction is to include in the model the product of the two individual variables. The product is termed a "first order interaction", as it is the lowest level of interaction that can be encountered, which necessarily involves two variables since there can be no interaction with less than two variables. A test of statistical significance of the coefficient for the product is the test of interaction, i.e., the test of the assumption about the constant effect of one of the variables across the levels of the other one. Analogously, one can define, model, and test interactions of higher orders, i.e., involving three or more variables.

For the analyzed data, tests of interaction were performed by entering into the proportional hazard models selected multiplicative interaction terms between the following binary variables: the use of radiotherapy (RT) vs. no RT; the use of breast conserving surgery (BCS) vs. total mastectomy (TM); T2 vs. T1 stage; extensive (>10 nodes examined) vs. limited (≤ 10 nodes) node examination; extensive (≥ 4 nodes involved) vs. limited (< 4 nodes) nodal involvement (₂₇). In analogy with stepwise regression, we examined different orders of interaction, simplifying the model by excluding non significant terms.

All computations were performed using SAS v8.01 (SAS Institute Inc, Cary, NC, USA). Tests of significance and confidence intervals were computed using the Wald test $(_{20})$.

Two-sided 0.05 significance level was used to assess results of significance tests. A description of the SEER data regarding surgery and radiotherapy and of the verification of proportional hazards assumptions was provided in a previous analysis ($_{22}$).

RESULTS PATIENTS CHARACTERISTICS

There were 60,349 patients records matching the selection criteria. Follow-up cutoff date was December 31, 1998. Median follow-up of patients alive was 56 months (range 0-131 months). The total number of events (death from any cause) was 7,090.

The distribution of selected characteristics by tumor and by type of treatment is shown in Table 1. BCS patients tended to be 2-3 years younger and tumor size tended to be smaller than TM patients. BCS patients who did not receive RT tended to be 1 year younger than BCS patients who received RT. The extent of node examination was comparable among the different subgroups. The number of nodes involved was particularly large in N+ patients who had TM and post-surgery RT (Table 1). Other details have been presented elsewhere (22).

Note that all selected patients had an axillary dissection (ALND). When BCS or TM is mentioned, it means BCS with ALND or TM with ALND, respectively.

Figure 1

Table 1: Distribution of selected characteristics ±1 standard error, by stage and by type of surgery total mastectomy (TM) or breast conserving surgery (BCS) with or without radiation (RT). All patients had an axillary node dissection.

Stage	Treatment	n	n events	mean age (year)	mean tumor size (mm)	mean nb nodes examined	mean nb nodes involved	%Grade 3-4
	TM-noRT	14070	1211	57.1 ± 8.4	12.8 ± 4.7	16 ± 7	0 ± 0	23.2
	TM + RT	440	38	55.2 ± 8.6	13.4 ± 4.7	15±6	0 ± 0	26.8
T1N0	BCS-noRT	1826	106	54.9 ± 8.5	12.4 ± 4.6	15±6	0 ± 0	26.6
	BCS+ RT	15638	697	55.7 ± 8.2	12.0 ± 4.5	15±6	0 ± 0	22.1
	TM-noRT	5904	905	55.6 ± 8.7	30.5 ± 7.6	16±7	0 ± 0	40.2
-	TM + RT	350	68	54.3 ± 8.7	34.6 ± 9.4	16 ± 7	0 ± 0	42.3
12NU	BCS-noRT	494	39	52.8 ± 8.5	28.5 ± 6.8	15±7	0 ± 0	52.8
	BCS+ RT	2779	239	53.5 ± 8.4	27.5 ± 6.1	15±7	0 ± 0	44.4
	TM-noRT	4790	857	55.5 ± 8.6	15.0 ± 4.3	16±7	3±4	31.3
	TM + RT	764	164	54.4 ± 8.4	15.4 ± 4.5	17 ± 8	7±6	38.7
1114+	BCS-noRT	799	91	53.1 ± 8.1	14.5 ± 4.2	16 ± 7	4 ± 4	32.2
	BCS+ RT	3270	322	54.1 ± 8.2	14.4 ± 4.1	16 ± 7	3±4	31.2
	TM-noRT	5360	1554	54.7 ± 8.6	32.0 ± 8.2	17 ± 7	4±5	43.3
	TM + RT	1585	411	53.7 ± 8.3	34.2 ± 8.7	17 ± 7	8±7	51.7
12N+	BCS-noRT	502	91	52.3 ± 7.8	29.1 ± 6.6	16 ± 7	4±5	53.0
	BCS+ RT	1778	297	52.9 ± 8.3	28.6 ± 6.6	16±7	4 ± 4	48.1

MODELS EXAMINED

The result of the proportional hazards model without interactions is reported in Table 2. RT was associated with a statistically significant 10% reduction in mortality hazard (hazard ratio, HR=0.904) relative to no-RT. BCS was associated with a statistically significant 13% reduction of the hazard (HR=0.874), relative to TM. Note that, since the model does not include interactions, the effect of RT is assumed to be the same irrespectively of whether the patient was treated with BCS or TM. Thus, for both types of surgery, an important beneficial effect of RT might be inferred.

Figure 2

Table 2: Proportional hazards model without interaction. Hazard ratio >1, increased mortality hazard; <1, reduced hazard. Unit scale for continuous variables or reference level for categorical variables in square brackets. This is a "main effects" only model. Because interactions are significant (see latter Tables 3,4,6), the model is invalid.

			0.511 0.1	
variable	label	hazard ratio	95% confidence interval	р
SEER area [/east]				
Central (Iowa, New Mexico, Utah)	SEERCENT	0.948	0.891 - 1.010	0.0976
West (San Francisco, Hawaii, Seattle)	SEERWEST	0.903	0.854 - 0.954	0.0003
Age at diagnosis [year]	AGEDGC	1.023	1.020 - 1.026	<.0001
Race black [/non black]	RACEBLCK	1.519	1.408 - 1.640	<.0001
Married [/not married]	MARRIED	0.828	0.789 - 0.870	<.0001
Multiple Primaries [/single primary]	SEQNUM	1.511	1.419 - 1.608	<.0001
Histology ductal [/other]	HISTODUC	1.274	1.201 - 1.353	<.0001
Grade [/well-moderately differentiated]				
poor-undifferentiated	G34	1.596	1.501 - 1.696	<.0001
unknown	GX	1.143	1.073 - 1.217	<.0001
Estrogen receptor positive [/else]	ERPOS	0.752	0.714 - 0.792	<.0001
Tumor Size [mm]	EODSIZE	1.03	1.027 - 1.032	<.0001
Number of nodes examined [n]	EODNEXAM	0.978	0.974 - 0.982	<.0001
Number of nodes involved [n]	EODNPOS	1.092	1.087 - 1.097	<.0001
Radiotherapy [/no radiotherapy]	rt	0.904	0.840 - 0.973	0.0070
Breast conserving surgery [/total mastectomy]	bcs	0.874	0.807 - 0.946	0.0009

For illustrative purposes, we will consider another model, which includes the first order interaction (BCRT) between type of surgery and radiotherapy (Table 3). The interaction is expressed as the product of the binary variables BCS and RT. It assumes value 1 only if both BCS and RT were used, and 0 otherwise. Consequently, the term needs to be taken into account only in the computations for the subgroup of patients treated with both BCS and RT. The hazard ratio associated with the interaction term is estimated to be equal to 0.757 and is statistically significant. This result can be interpreted as follows: in the absence of RT, the mortality hazard for patients treated with BCS was equal to 1.071 of the hazard of patients not treated with BCS (TM patients). However, if RT was applied, the hazard for BCS treated patients was equal to $1.071 \times 0.757 = 0.811$ of the hazard of TM patients. Thus, the statistically significant interaction term implies a substantial modification of the effect of RT

by BCS: for BCS treated patients, RT is beneficial, but has no effect for patients treated with TM.

Figure 3

Table 3: A model with interactions between treatments and nodal status.

variable	label	hazard ratio	95% confidence interval	р
SEER area Central [/east]	SEERCENT	0.948	0.890-1.010	0.096
SEER area West [/east]	SEERWEST	0.910	0.862-0.962	0.001
Age at diagnosis [year]	AGEDGC	1.025	1.022-1.028	<.001
Race black [/non black]	RACEBLCK	1.515	1.404-1.635	<.001
Married [/not married]	MARRIED	0.831	0.791-0.873	<.001
Multiple Primaries [/single primary]	SEQNUM	1.512	1.421-1.610	<.001
Histology ductal [/other]	HISTODUC	1.228	1.157-1.303	<.001
Grade poor-undiff. [/well-mod.diff.]	G34	1.578	1.484-1.677	<.001
Grade unknown [/well-mod.diff.]	GX	1.142	1.073-1.217	<.001
Estrogen receptor positive [/else]	ERPOS	0.743	0.706-0.782	<.001
Tumor Size [mm]	EODSIZE	1.026	1.023-1.028	<.001
Number of nodes examined [n]	EODNEXAM	0.982	0.979-0.986	<.001
Number of nodes involved [n]	EODNPOS	1.067	1.061-1.073	<.001
Node positive N+ status [/hode negative]	n1	1.673	1.582-1.770	<.001
Radiotherapy [/ho radiotherapy]	rt	0.981	0.900-1.069	0.664
Breast conserving surgery [/Total mast.] Interactions:	bcs	1.071	0.956-1.200	0.237
BCS • RT	bort	0.757	0.653-0.878	0.002

It is worth noting here that, to investigate the possibility of the influence of BCS on the effect of RT using a subgroup analysis, two separate analyses of data would have to be performed: in BCS and TM patients. The two resulting estimates of the effect of RT would then have to be compared. An additional complication would be the need to adjust the estimates for the effect of other covariates. The model presented in Table 3 addresses all these issues simultaneously.

To investigate the possibility of a differential effect of RT in a more detail, a series of models containing interactions between RT, BCS, T-stage, number of positive nodes (0, 1 3, 4+ nodes), and extent of node examination (less than 10 or 10+ examined nodes) was considered. First, a model containing fourth-order interaction terms involving all the factors mentioned above, as well as all lower-order interactions, was fitted. Since it was found that the fourthorder terms were not significant, the model was simplified by deleting them and testing the joint significance of all third-order terms (for quadruples of factors). These terms were also found not to be significant. Hence, the model with only second-order (for triplets of factors) and lower-order interactions was considered. After deleting non-significant second-order terms (and, if possible, first-order interactions), the model presented in Table 4 was obtained. It was found that neither a small number of positive nodes (1 to 3) nor a small number of examined nodes (less than 10) changed the effect of treatments. The model suggests that T2 modifies the effect of both RT and BCS (as suggested by the presence of T2•RT•BCS interaction), while a large number of positive nodes (4+) influences mainly RT (T•(4+nodes)•RT

interaction).

Figure 4

Table 4: More comprehensive model with interactions: radiotherapy (RT vs. noRT), type of surgery (BCS vs. TM), T-stage (T2 vs. T1), limited node involvement (n1: 1-3 nodes involved), more extensive node involvement (n2: > 4 nodes involved), more extensive node examination (nexgt10: > 10 nodes examined).

		hazard	95% confidence	
variable	label	ratio	interval	р
SEER area Central [/east]	SEERCENT	0.946	0.889 - 1.008	0.085
SEER area West [/east]	SEERWEST	0.915	0.866 - 0.967	0.001
Age at diagnosis [year]	AGEDGC	1.025	1.022 - 1.028	<.001
Race black [/non black]	RACEBLCK	1.515	1.404 - 1.635	<.001
Married [/not married]	MARRIED	0.834	0.794 - 0.876	<.001
Multiple Primaries [/single primary]	SEGNUM	1.517	1.425 - 1.615	<.001
Histology ductal [/other]	HISTODUC	1.225	1.154 - 1.301	<.001
Grade poor-undiff. [/well-mod.diff.]	G34	1.552	1.460 - 1.650	<.001
Grade unknown [/well-mod.diff.]	GX	1.131	1.062 - 1.205	0.001
Estrogen receptor positive [/else]	ERPOS	0.744	0.707 - 0.783	<.001
Tumor Size [mm]	EODSIZE	1.017	1.014 - 1.021	<.001
Number of nodes examined [n]	EODNEXAM	0.989	0.984 - 0.993	<.001
Number of nodes involved [n]	EODNPOS	1.054	1.047 - 1.062	<.001
T2-stage [/T1]	t2	1.312	1.145 - 1.504	<.001
Limited node involvement [/0-3 nod.inv]	n1	1.547	1.457 - 1.642	<.001
Extensive node involvement [/0-3 nod.inv]	n2	2.454	1.942 - 3.102	<.001
> 10 nodes examined [/≤10]	nexgt10	0.891	0.809 - 0.980	0.018
Radiotherapy [/no radiotherapy]	rt	1.121	0.936 - 1.343	0.213
Breast conserving surgery [/Total mast.]	bcs	1.106	0.955 - 1.281	0.180
Interactions:				
T2 • >10 nodes examined	t2ne	1.034	0.909 - 1.176	0.613
T2 • RT	t2rt	0.941	0.748 - 1.184	0.606
T2 • BCS	t2bc	0.914	0.727 - 1.150	0.443
RT• BCS	rtbc	0.627	0.499 - 0.787	<.001
T2 • extensive node involvement	t2n2	1.058	0.789 - 1.417	0.708
extensive node inv • >10 nodes examined	n2ne	0.995	0.781 - 1.267	0.966
extensive node inv RT	n2rt	1.008	0.831 - 1.222	0.939
T2 • RT • BCS	t2rtbc	1.358	0.985 - 1.873	0.062
T2 • ext nod inv • >10 nodes exam.	t2n2ne	0.714	0.526 - 0.969	0.031
T2 • ext.nod.inv • RT	t2n2rt	0.771	0.598 - 0.995	0.045

To facilitate interpretation, a summary that presents the factorial layout derived from the model of Table 4 was computed (Table 5). The factorial table shows hazard ratios for RT and BCS for different combinations of T-stage and category of node positivity. The details of the computation procedure needed to derive the table from the model is shown in Table 6.

Figure 5

Table 5: Factorial summary of treatment effects: hazard ratios (95% confidence limits) representing estimated effects of T-stage, category of node involvement and interactions as modeled in Table 4. The values are based on tests of interaction and are adjusted for all listed variables of Table 4. All hazard ratios are relative to a single reference (stage T1, 0-3 nodes positive, total mastectomy TM, no RT), results for the different subgroups are directly comparable. Calculations detailed in Table 6.

No. of positive			Tl	Τ2		
nodes		No BCS (TM)	BCS	No BCS (TM)	BCS	
0-3	No RT	1	1.11 (0.95,1.28)	1.31 (1.14,1.50)	1.33 (1.07,1.64)	
	RT	1.12 (0.94,1.34)	0.78 (0.72,0.84)	1.38 (1.15,1.67)	1.19 (1.03,1.38)	
4+	No RT	2.45 (1.94,3.10)	2.71 (2.06,3.58)	3.41 (2.80,4.14)	3.44 (2.68,4.43)	
	RT	2.77 (2.13,3.62)	1.92 (1.47,2.51)	2.79 (2.26,3.45)	2.40 (1.92,3.02)	

Figure 6

Table 6: Calculations for the factorial summary Table 5. The labels represent the interaction effects given in Table 4.

No. of nositive			Tl	T2		
nodes		No BCS (TH)	BCS	No BCS (TM)	BCS	
	No RT	1	bcs	t2	t2*bcs*t2bc	
0-3	RT	rt	rt*bcs*rtbc	rt*t2*t2rt	rt*t2*t2rt* bcs*t2bc*rtbc* t2rtbc	
	No RT	n2	n2*bcs	n2*t2*t2n2	n2*t2*t2n2* bcs*t2bc	
4+	RT	n2*rt*n2rt	n2*rt*n2rt* bcs*rtbc	n2*rt*n2rt* t2*t2n2*t2rt* t2n2rt	n2*rt*n2rt* t2*t2n2*t2rt* t2n2rt*bcs*t2bc* rtbc*t2rtbc	

Example of calculations for RT and BCS in T1 with 4+ positive nodes: the relevant main effect parameters from Table 4 are bcs, rt and n2 (T1 is the reference category for T2); the relevant interaction effects are RT•BCS (rtbc) and 4+nodes•RT (n2rt). Consequently, the hazard ratio rt*n2*bcs*rtbc*n2rt is obtained. Table 5 shows the corresponding numerical value calculated from the estimates given in Table 4: $1.121 \times 2.454 \times 1.106 \times 0.627 \times 1.008 =$ 1.92.

Table 5 might be read in different ways depending on the center of interest. We will examine it from the perspective of treatment combinations: BCS vs. BCS+RT, TM vs. TM+RT, BCS vs. TM.

Among patients treated with breast conserving surgery, radiotherapy was associated with a reduced mortality hazard ratio, as compared to patients who did not receive radiotherapy. This reduction was consistent for all combinations of T stage and category of node positivity: the hazard ratios for RT vs. no RT (relative to T1 patients with 0 3 positive nodes treated with TM and no RT) in Table 5 were equal to 0.78 vs. 1.11 for T1 and 0 3 positive nodes, 1.92 vs. 2.71 for T1 and 4+ nodes, 1.19 vs. 1.33 for T2 and 0 3 nodes, and 2.40 vs. 3.44 for T2 and 4+ nodes. From Table 6 it follows that, using the estimated hazard ratios for the model presented in Table 4, the ratio 0.78 vs. 1.11 can be expressed as rt*rtbc, 1.92 vs. 2.71 as rt*rtbc*n2rt, 1.19 vs. 1.33 as rt*rtbc*t2rt*t2rtbc and 2.40 vs. 3.44 as rt*rtbc*t2rt*t2rtbc*n2rt*t2n2rt. Based on the model one can construct tests whether the products are equal to 1 (i.e., whether there is no difference in the mortality hazard ratio) [21]. The tests yield the following P-values: <0.001, 0.003, 0.285, 0.002 respectively. Thus, apart from T2-stage with 0-3 positive nodes, in the three other cases the reduction of the mortality hazard ratio associated with the use of radiotherapy in combination with breast conserving surgery was statistically significant.

Among patients treated with total mastectomy, radiotherapy – as compared to no use of radiotherapy – seemed to increase mortality hazard ratio, except for T2 stage with 4+ positive nodes (Table 5, columns TM, rows RT vs. no-RT: hazard ratios of 1.12 vs. 1 for T1 and 0 3 positive nodes, 2.77 vs. 2.45 for T1 and 4+ nodes, 1.38 vs. 1.31 for T2 and 0 3 nodes, and 2.79 vs. 3.41 for T2 and 4+ nodes). Significance of these differences can be achieved by testing whether the following ratios are equal to 1: rt, rt*n2rt, rt* t2rt and rt*t2rt*n2rt*t2n2rt (see Table 6). The tests yield the following P-values: 0.21, 0.18, 0.46, 0.001, respectively. Thus, apart from T2-stage with 4+ positive nodes, in the three other cases the reduction of the mortality hazard ratio associated with the use of radiotherapy was not statistically significant.

Breast conserving surgery in the absence of radiotherapy did not seem to have any effect as compared to total mastectomy (Table 5, rows No RT, columns BCS vs. TM: hazard ratios of 1.11 vs. 1 for T1 and 0 3 positive nodes, 2.71 vs. 2.45 for T1 and 4+ nodes, 1.33 vs. 1.31 for T2 and 0 3 nodes, and 3.44 vs. 3.41 for T2 and 4+ nodes). In this case the significance of the differences can be verified by testing whether bcs=1 (P=0.18) and bcs*t2bc=1 (P=0.91). The results are not significant but, in view of the wide confidence intervals presented in Table 5, no firm conclusions can be drawn.

DISCUSSION

As had been commented by Harrell $(_{21})$, the complexity and the difficulty of interpretation of interactions increase steeply when it involves more than two variables. Raw results such as those shown in Table 4 may be difficult to understand at first sight and are prone to erroneous interpretation. Consider for example the non-significant variables labeled rt and bcs in Table 4. Does it mean that neither the type of surgery nor radiation have any effect?

When answering the above question one needs to consider the form of the model presented in Table 4. In particular, the presence of interaction terms involving rt and bcs should be taken into account. Note that when the interaction terms were not included in the proportional hazards model (Table 2), both rt and bcs were statistically significant and indicated a reduced mortality hazard. In Table 4, one can see that they become non-significant, but there are statistically significant interaction terms involving one or both of them. This indicates that radiotherapy and BCS did influence the risk of death, but their effect depended on other factors. For instance, the interaction terms t2rt, t2bc and t2rtbc suggest that, the effects of radiotherapy and BCS differed for different T stages. This can be clearly seen from Table 5: the mortality hazard ratio of T1 patients with 0-3 positive nodes treated with radiotherapy and BCS was 0.78 as compared to the reference group, while for the otherwise comparable T2 patients the ratio was 1.19. One might thus infer that the use of radiotherapy and BCS might be justified for T1 patients with a limited or no node involvement, but not for comparable T2 patients. It is this type of conclusion that is sought in subgroup analyses.

The above example shows that the estimated effects of variables depends on the structure of the model, and in particular, on the presence of interactions involving them. It also indicates that interpretability is not a trivial issue. To that aim, we find that the construction of a factorial summary table, like Table 5, is useful, if not almost necessary. It enhances the presentation of results and might facilitate the comprehension of how statistical interaction may relate to subgroup analyses.

Significant interactions detected in Table 4 indicate that the simpler models presented in Tables 2 and 3 are inappropriate. The results of interaction tests might be summarized as follow:

• The use of radiotherapy was in general beneficial

for patients treated with breast conserving surgery with axillary dissection, but the magnitude of the positive effect depended on T stage and the number of positive nodes.

- For patients treated with total mastectomy the application of radiotherapy in general did not seem to have a favorable effect, except for T2 patients with 4+ positive nodes.
- Breast conserving surgery without radiotherapy appeared equivalent to mastectomy without radiotherapy, although one cannot rule out the possibility of a false-negative finding.

A few remarks on the validity of the aforementioned are worth mentioning. First, they were obtained as a result of an analysis that was exploratory in nature. That is, it did not start with a pre specified set of hypotheses to be confirmed, as it might be the case in, e.g., a clinical trial. For this reason, no adjustment for multiple testing was applied. Admittedly, while this increases the statistical power of the analysis, it also increases the risk of false positive findings. Therefore, the conclusions based on the statistically significant results of performed tests should be treated as statements that would require an independent verification.

Secondly, since the conclusions are based on an analysis of observational data, one needs to consider a possibility of bias. The imbalance in the distribution of different patients characteristics between the treatment groups (Table 1) indicates that the data may be subject to some selection processes. The imbalances observed in Table 1 are not of the main concern, since their effect can be removed by using the models like those presented in Tables 2 4. However, a hidden bias might still exist, if treatment was systematically allocated to some categories of patients for undocumented reasons related to patient's prognosis. Because the presence of a hidden bias can never be excluded, results based on observational studies like the present one cannot be accepted without consideration of randomized trials or other collateral evidence.

Regarding the clinical plausibility, the present results for breast conserving surgery confirm previous findings ($_{22}$), and are concordant with a recent meta-analysis of breastconserving surgery and radiation that found a survival advantage with radiation ($_{28}$). For patients treated with mastectomy, the results suggest a favorable effect of radiotherapy in higher risk patients (T2 stage with more nodes involved), but no significant effect in lower risk patients. This seems in discrepancy with clinical trials of post-mastectomy radiotherapy that found no significant interaction with tumor size or with the number of positive lymph nodes ($_{29}$), or a significant interaction between radiotherapy and number of nodes removed only in patients who survived less than 4 years ($_{30}$). We have currently no explanation. A tentative hypothesis is the confounding by the variability of lymph node dissection that would require more complex modeling ($_{31}$), but this is an investigation that is beyond the purpose of the present analysis.

CONCLUSION

Statistical guidelines recommend the use of formal tests of interactions instead of subgroup analyses. However, examples of application of the procedure are rare. We applied systematic stepwise tests of interactions within proportional hazards models to the study of a large registry database of breast cancer patients, in order to investigate the changes in the effect of radiotherapy related to the use of breast conserving surgery and other clinical factors. The major finding is that in the absence of radiotherapy, breast conserving surgery patients presented no survival disadvantage as compared with mastectomy patients, while with the addition of radiotherapy, a substantial survival advantage was observed in all subgroups of patients.

ACKNOWLEDGMENT

Vincent Vinh-Hung (Oncology Center, Academic Hospital, Vrije Universiteit Brussel, Belgium) and Tomasz Burzykowski (Limburgs Universitair Centrum, Center for Statistics, Diepenbeek, Belgium) were the original authors. The paper underwent several revisions. Reviewers were not sure whether it addressed methodological or clinical audience. For statisticians, the topic is not new. For clinicians, it is too difficult. The original authors could not pursue this work further, nevertheless they gave full permission for the present submission.

References

1. Poste G, Fidler IJ. Pathogenesis of cancer metastasis. Nature 1980;283:139-46.

2. Schirrmacher V. Cancer metastasis - Experimental approaches, theoretical concepts, and impacts for treatment strategies. Adv Cancer Res 1985;43:1-73.

3. Price JT, Bonovich MT, Kohn EC. The biochemistry of cancer dissemination. Crit Rev Biochem Mol Biol 1997;32:175-253.

4. Wyke JA. Overview - burgeoning promise in metastasis research. Eur J Cancer 2000;36:1589-94.

5. Verschraegen CF, Giovanella BC, Mendoza JT, Kozielski AJ, Stehlin JS, Jr. Specific organ metastases of human melanoma cells injected into the arterial circulation of nude

mice. Anticancer Res 1991;11:529-35. 6. Hedenfalk I, Duggan D, Chen Y, Radmacher M, Bittner M, Simon R, et al. Gene-expression profiles in hereditary breast cancer. N Engl J Med 2001;344:539-48. 7. Norredam M, Groenvold M, Petersen JH, Krasnik A. Effect of social class on tumour size at diagnosis and surgical treatment in Danish women with breast cancer. Soc Sci Med 1998;47:1659-63. 8. Bradley CJ, Given CW, Roberts C. Race, socioeconomic status, and breast cancer treatment and survival. J Natl Cancer Inst 2002;94:490-6. 9. Rothwell PM. Can overall results of clinical trials be applied to all patients? Lancet 1995;345:1616-9. 10. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. N Engl J Med 1987;317:426-32. 11. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet 2000;355:1064-9. 12. Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey SG. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and falsenegatives. Health Technol Assess 2001;5:1-56. 13. Buyse ME. Analysis of clinical trial outcomes: some comments on subgroup analyses. Control Clin Trials 1989;10:187S-94Š. 14. Schmoor C. On problems related to subgroup analyses in clinical trials. Onkologie 2000;23:346-51. 15. Bender R, Lange S. Adjusting for multiple testing - when and how? J Clin Epidemiol 2001;54:343-9. 16. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. JAMA 1991;266:93-8. 17. Oxman AD, Guyatt GH. A consumers guide to subgroup analyses. Ann Intern Med 1992;116:78-84. 18. Scott PE, Campbell G. Interpretation of subgroup analyses in medical device clinical trials. Drug Inf J 1998;32:213-20 19. Draper NR, Smith H. Applied regression analysis. Third edition. New York, NY: John Wiley & Sons, Inc.; 1998. (Page 488-500). 20. Therneau TM, Grambsch PM. Modeling survival data: extending the Cox model. New York, NY: Springer-Verlag; 2000. (Page 39-185). 21. Harrell FE, Jr. Regression modeling strategies with applications to survival analysis and logistic regression. Charlottesville, VA: University of Virginia; 2000. (Page 15). 22. Vinh-Hung V, Burzykowski T, Van de Steene J, Storme G, Soete G. Post-surgery radiation in early breast cancer: survival analysis of registry data. Radiother Oncol 2002;64:281-90. 23. Early Breast Cancer Trialists' Collaborative Group. Favourable and unfavourable effects on long-term survival of radiotherapy for early breast cancer: an overview of the randomised trials. Lancet 2000;355:1757-70. 24. National Cancer Institute. Surveillance, Epidemiology, and End Results (SEER) Program Public-Use Data (1973-1998), National Cancer Institute, DCCPS, Cancer Surveillance Research Program, Cancer Statistics Branch, released April 2001, based on the August 2000 submission. 2001. Available at: Http://seer.cancer.gov. 25. Cox DR. Regression models and life-tables. J R Stat Soc B 1972;34:187-220. 26. Box GEP, Draper NR. Empirical model-building and response surfaces. New York, NY: John Wiley & Sons;

1987. (Page 109). 27. Sauer R, Schulz KD, Hellriegel KP. [Radiation therapy after mastectomy--interdisciplinary consensus puts and end to a controversy. German Society of Senology]. Strahlenther Onkol 2001;177:1-9.

28. Vinh-Hung V, Verschraegen C. Breast-conserving surgery with or without radiotherapy: pooled-analysis for risks of ipsilateral breast tumor recurrence and mortality. J Natl Cancer Inst 2004;96:115-21.

29. Overgaard M, Hansen PS, Overgaard J, Rose C, Andersson M, Bach F, et al. Postoperative radiotherapy in high-risk premenopausal women with breast cancer who receive adjuvant chemotherapy. Danish Breast Cancer Cooperative Group 82b Trial. N Engl J Med 1997;337:949-55.

30. Overgaard M, Jensen MB, Overgaard J, Hansen PS, Rose C, Andersson M, et al. Postoperative radiotherapy in high-risk postmenopausal breast-cancer patients given adjuvant tamoxifen: Danish Breast Cancer Cooperative Group DBCG 82c randomised trial. Lancet 1999;353:1641-8.

31. Vinh-Hung V, Burzykowski T, Cserni G, Voordeckers M, Van de Steene J, Storme G. Functional form of the effect of the numbers of axillary nodes on survival in early breast cancer. Int J Oncol 2003;22:697-704.

Author Information

Anne-Marie Danzel, M.D. Algemene Praktijk Kantoor