

# **Modeling the effect of tumor size in early breast cancer**

## *Authors:*

Claire Verschraegen, Vincent Vinh-Hung, Gábor Cserni, Richard Gordon, Melanie E Royce, Georges Vlastos, Patricia Tai, Guy Storme.

## *Institutions:*

University of Texas M.D. Anderson Cancer Center, Houston, TX, USA, (MER);  
Department of Radiotherapy, Oncology Center, Vrije Universiteit Brussel, Jette, Belgium (VVH, GS);  
Surgical Pathology, Bács-Kiskun County Teaching Hospital, Kecskemét, Hungary (GC);  
Departments of Radiology and Biosystems Engineering, University of Manitoba, Winnipeg, Canada (RG);  
Gynecologic Oncology and Senology, Geneva University Hospitals, Geneva, Switzerland (GV);  
Department of Radiation Oncology, Saskatchewan Cancer Agency, Regina, Canada (PT);  
Division of Hematology Oncology, Cancer Research and Treatment Center, University of New Mexico, Albuquerque, NM, USA (CV).

## *Research support:*

Gábor Cserni was supported by a János Bolyai Research Fellowship from the Hungarian Academy of Sciences.

Richard Gordon was supported in part by a grant from Friends You Can Count On.

## *Scientific meetings:*

To be presented at the 4th European Breast Cancer Conference, Hamburg, March 16 - 20, 2004 (VVH, GS).

*Corresponding author:*

Claire Verschraegen,  
The University of New Mexico,  
Cancer Research and Treatment Center,  
900 Camino de Salud,  
Albuquerque, NM 87131  
Tel: 505-272 6760  
Fax: 505-272 2841  
email: cverschraegen@salud.unm.edu

*Key words:* functional form, generalized additive models, survival analysis, tumor growth, lymph node metastases.

*Notes:*

*Figure 5b:* With permission, Reprinted from The Journal of the American Statistical Association. Copyright 1992 by the American Statistical Association. All rights reserved.

*Figure 5c:* With permission, Reprinted from Breast Cancer Research and Treatment. Copyright 1998 by Kluwer Academic Publishers. All rights reserved.

## Abstract

*Background:* The purpose is to determine the type of relationship between tumor size and mortality in early breast carcinoma.

*Methods:* The data was abstracted from 83,686 cases registered in the Surveillance, Epidemiology, and End Results Program of women diagnosed with primary breast carcinoma between 1988 and 1997, presenting with a T1-T2 lesion and no metastasis, in whom axillary node dissection was performed: 58,070 women were node-negative (N0) and 25,616 node-positive (N+). Endpoint was death from any cause. Tumor size was modeled as a continuous variable by proportional hazards, using a generalized additive models procedure.

*Results:* Functionally, a Gompertzian expression  $\exp(-\exp(-(size-15)/10))$  provided a good fit to the effect of tumor size (in millimeters) on mortality, irrespective of nodal status. Quantitatively, for tumor size between 3 and 50 mm, the increase of crude cumulative death rate (number of observed deaths divided by number of patients at risk) increased with size from 10% to 25% for N0 and from 20% to 40% for N+.

*Discussion:* The functional relationship of tumor size with mortality is concordant with current knowledge of tumor growth. However, its qualitative and quantitative independence of nodal status is in contradiction with the prevailing concept of sequential disease progression from primary tumor to regional nodes. This argues against the perception that nodal metastases are caused by the primary tumor.

Abbreviations:

ANOVA	analysis of variance
BCS	breast conserving surgery
CMF	cyclophosphamide, methotrexate, 5-fluorouracil
DFS	disease free survival
ECOG	Eastern Cooperative Oncology Group
ER	estrogen receptor
GAM	generalized additive models
GBSG	German Breast Cancer Study Group
N+	node-positive
N0	node-negative
NSABP	National Surgical Adjuvant Breast and Bowel Project
OS	overall survival
PH	proportional hazards
PR	progesterone receptor
RT	Radiotherapy
SEER	Surveillance, Epidemiology, and End Results Program
ZEBRA	Zoladex Early Breast Cancer Research Association

## Introduction

As noted by the College of American Pathologists Consensus Statement in 1999 on prognostic factors in breast cancer, tumor size is one of the most powerful predictors of tumor behavior in breast cancer (1,2). It constitutes the basis of major staging systems (3,4), hence, tumor size is incorporated in almost all clinical reports.

Despite the clinical importance of tumor size, few studies have investigated the functional form linking size to outcome (*functional form* means a mathematical statement of the relationship between an explanatory variable and a dependent variable). As shown in Table 1, some studies suggested a linear effect of tumor size (5-14), whereas others found a nonlinear effect (15-26). Because of the diversity of endpoints used to evaluate outcome (local-regional or metastatic recurrence, cause-specific or overall survival, relative or absolute endpoints), and the complexity of the mathematical expressions, there is no apparent agreement. To add to the confusion, a "linear" relation, as implied by proportional hazards (PH) models, means that the risk is exponentially related to tumor size, but an "exponential-power" expression such as proposed by Michaelson et al (24,27) and other authors (18,21,22) describes a sigmoidal relationship (non-linear). The fact, that size (generally measured as the largest tumor dimension) is related to volume by a power expression, adds to the difficulty in interpreting the studies. The issue is further complicated by concepts of breast cancer disease –it has been argued that tumor size is of little consequence to the fate of the patient (15), or even inconsequential (28), when nodal status is taken into account– i.e. implying that tumor size has no intrinsic effect, that its

apparent effect as a harbinger of systemic disease disappears when a more potent indicator of systemic disease (lymph node involvement) is present.

The issue of identifying the functional form is more than an academic exercise. For example, if the functional form is sigmoidal, then its representation by a linear form would overestimate the prognostic role of large tumors, and underestimate the effect of small tumors, with the consequent risk of recommending inappropriate diagnostic procedures or treatments due to over/under-emphasis placed on tumor size.

In the present study, we investigate the functional form relating tumor size to mortality, taking into account nodal status, by using an extended survival modeling approach.

## Materials and methods

The SEER Program provides extensive cancer incidence data originating from several US registries (29). The "9-registries" database was used for data extraction. Selected patients were women without previous history of cancer presenting with a non-inflammatory invasive breast carcinoma, histologically confirmed and diagnosed between 1988 and 1997, with specified tumor size no larger than 50 mm (T1 and T2), strictly confined to breast (involvement of skin, hypodermis, pectoral muscles, and deep fixation were excluded), without distant metastasis, for whom the type of surgery was specified and axillary node dissection was performed. Post-operative radiotherapy was not a selection criterion, but pre-operative or intra-operative radiotherapy were excluded. Certain records were rejected because of data quality concerns (30): uncertain sequence of treatment, non-hospital based data records, month of diagnosis unknown, race unknown. Further screening of the data excluded cases with more than 50 nodes examined, cases with "0 month" of follow-up, and cases with age at diagnosis <25 years or >95 years. Event was defined as death from any cause. Follow-up cutoff date was December 31, 1999.

A Poisson regression approach, adapted from Therneau and Grambsch (31), was used to derive the effect of tumor size as a mathematical function. Briefly, the procedure starts by assuming that the effect of tumor size on the log hazard of mortality is linear. Tumor size (in millimeters) is incorporated as an untransformed continuous covariate in a proportional hazards model. A diagnostic of the PH model is performed using a Poisson

regression which implements the generalized additive models (GAM) procedure of Hastie and Tibshirani (32). The regression results in a non-parametric curve that represents the effect of tumor size. An ANOVA test of the linearity of the curve is performed. If the linearity test fails, a search is made for mathematical expressions of tumor size (i.e. a transform) which approximate the shape of the curve. The whole procedure is repeated, using each transform as a covariate replacing tumor size in the initial PH model until the condition of linearity is satisfied.

Other quantitative covariates included in the PH models were: age at diagnosis, number of nodes examined and year of diagnosis. Qualitative covariates were: registry area, race, marital status, tumor topography, histological type and grade, estrogen and progesterone receptor status, type of primary surgery, and administration of post-operative radiotherapy. All qualitative variables were converted or expanded as needed to dummy variables to allow coding using binary indicators, e.g. "married" versus "not married", "high grade" versus "not high grade", etc. Verification of the covariates found departures from the proportional hazard assumption as might be expected (33), but there was no notable impact on the results when alternative categorizations or subset analyses were performed. A first order interaction between type of surgery and post-operative radiotherapy (variable built by multiplying the two binary indicators) was included for consistency with a previous analysis (30).

A comparison with a more direct representation of mortality was performed using crude death rate, defined as the number of observed deaths at the follow-up cutoff date, divided by the number of patients at risk.

Node-positive cases were not considered until all the above analyses were performed in node-negative cases (34), providing the opportunity of a validation set. The mathematical expression found with node-negative cases was used in a PH analysis of the node-positive cases. The appropriateness of the transform was verified by the linearity test in this validation set.

All statistical analyses were performed using Splus (Insightful Corporation, Seattle, WA, USA) statistical software. The search for mathematical expressions representing the shapes of curves used a dictionary of curves provided by TableCurve 2D (Systat Software Inc, Richmond, CA, USA).

## Results

In the 2002 SEER release (29), 188,410 women were diagnosed with a breast tumor from 1988 to 1997, of whom 132,457 had a hospital based diagnosis of unilateral invasive carcinoma microscopically confirmed by a positive histology. A total of 83,686 cases matched the selection criteria, 58,070 were node-negative and 25,616 node-positive. The median follow-up was 73 months (range 1-143) for patients still alive at the follow-up cutoff date (December 31, 1999). Table 2 shows the characteristics of patients as a function of their nodal status. Figure 1 shows the distribution of tumor size measurements. Almost all qualitative and quantitative covariates included in the proportional hazards analyses were highly significant (supplemental material TableCharact.xls).

The PH analysis performed in node-negative cases showed a mortality hazard ratio of 1.027 [95% confidence interval 1.025-1.029] for a difference in tumor size of 1 mm. The GAM procedure showed a nonlinear effect of tumor size on the log hazard of mortality (Fig 2a). The test for non-linearity was statistically significant ( $\text{Chi}^2=72.57$ , 3 degrees of freedom,  $p<0.0001$ ). Several generic functions provided a good fit –fractional polynomial, logistic, lognormal, gompertzian. A good fit was also obtained with Michaelson's exponential-power (24,27). However, the gompertzian was found to be the most flexible without requiring excessive precision and a double exponential was selected to express the functional form (size in millimeters):

$$\exp(-\exp(-(size-15)/10)).$$

The test of linearity for PH using the transform was satisfied ( $\text{Chi}^2=3.82$ , 3 degrees of freedom,  $p=0.280$  i.e. non-significant non-linearity). The PH model was marginally improved (sum of squares of deviance residuals reduced from 43,541 to 43,488). The significance and hazard ratios of other covariates were unaffected.

The tumor size transform was scaled and drawn onto a plot of the crude death rates (Fig 2b). The graph shows a good concordance with the GAM plot (Fig 2a). The dispersion of the data points is concordant with the confidence intervals of the GAM plot.

Next, a PH analysis was performed in node-positive cases. The effect of tumor size appeared similar to that of node-negative cases (Fig 2c). The hazard ratio for tumor size was 1.021 [1.018-1.023]. The nonlinearity was statistically significant ( $\text{Chi}^2=85.13$ , 3 degrees of freedom,  $p<0.0001$ ). The transform derived from node-negative cases was used as a plug-in covariate replacing size in the PH analysis of node-positive cases. It satisfied the test of linearity ( $\text{Chi}^2=6.97$ , 3 degrees of freedom,  $p=0.070$  i.e. non-significant non-linearity).

As done for node-negative cases, the size transform was plotted onto the graph of crude death rates of patients with positive nodes (Fig 2d). It shows that the overall shapes are similar. A non-parametric spline smoothed curve of the death rates is superimposed in the same graph as a dashed curve. (The smoothed curve is similar to a moving average: it provides a visual indication of how much the size-transform differs from the average death rates.) The concordance was also good, except a small mismatch for tumor sizes

less than 10 mm. The slopes corresponding to the two curves (Fig 2b and 2d) are shown in Figure 3. Figure 3 shows the similarity of the global shape of the slopes while highlighting the differences. The peaks indicate where the maximal change of slope occurs. The plain curve –which is derived from modeling the node-negative cases– suggests that the transition occurs around 15 mm. The dashed curve –which is derived from smoothing the death rates of patients with positive nodes– suggests that the transition occurs around 20 mm. The irregularities of the dashed slope at larger tumor sizes appear attributable to the lack of precision of the data points at large size measurement. The small dashed inflexion at smaller tumor sizes suggests a secondary peak of increased death rate for very small node-positive tumors.

## Discussion

The results of the log hazard analysis show that patients with node-negative breast cancer have a linear increase of mortality with tumor size up to around 30 mm. The mortality risk plateaus beyond 30-50 mm (Fig 2). The relationship was modeled as a gompertzian double exponential function (Fig 4). The function was validated by finding an almost identical relationship in patients who had positive nodes (Fig 5).

There is no doubt that the selection of cases might have excluded high-risk patients. Patients with a more aggressive tumor would already have died or developed clinical metastases and therefore would have been excluded. However, because cases presenting at diagnosis with large tumors or with macroscopic metastases are much less frequent, it is unlikely that such a selection can be sufficient to explain the observed functional form of tumor size. Furthermore, the interpretation of a plateau in mortality risk as meaning "less aggressive tumor" is not a correct assumption. The data was also limited to registered cases for whom complete records were available. Results cannot be considered as representative of the whole population. Data on tumor size were used in this analysis without verification of the pathological examination. Errors might have resulted from difficulties of determining tumor extension, from inter-observer variations, or from transcription mistakes. The spikes in Figure 1 show that rounding occurred. Other limitations such as the lack of data on systemic treatment that might have influenced survival outcome have been mentioned in details elsewhere (35).

The GAM procedure allowed a multivariate analysis, which examined the deviation from the PH assumption, and searched for the amount of correction that was required to fulfill the PH requirement. Crude death rates were used to complement the GAM analysis. Although crude death rates do not take into account the length of follow-up and may be biased indicators of outcome, they appeared concordant with the results of the GAM modeling and remained immediately understandable.

We selected a gompertzian function to represent the functional form of tumor size, not because this model was more valid than any other, but for pragmatic reasons. One reason was the link to cytokinetics models (36,37). The second reason was the ease of interpretation. In the expression  $\exp(-\exp(-(size-15)/10))$ , representing a sigmoid, the value "15mm" corresponds to the transition from an "accelerating" to a "decelerating" growth (Fig 3). A third reason was the consideration that the precision of the parameters should not exceed the precision of size measurements (Fig 1).

A summary of the literature that modeled tumor size or reported at least 5 data points was presented in Table 1. The shape of the effect of tumor size on survival found in our study is comparable to the shape of nonlinear models that have been reported. For example, the expression (size in mm)

$$1-\exp(-0.00555*size^{1.714})$$

derived from the Michaelson's model (24), [also representative of a family of tumor control probability models (18,21,22)], gave an almost perfect fit in patients with negative nodes. Hence, there was no obvious "best" function. Figure 4 shows a selection of functions that appropriately described the shape within the data range of 3-50 mm.

Among the studies that reported a linear effect, four are of particular importance. One because it is the largest and also uses the SEER data (9), the three others because the data is collected from prospective clinical trials and analyzed with modern statistical tools (10-12). For the SEER study, node-negative cases were selected since this subset has been widely used by other authors (38-40) and has served as the basis of a linear model (14). Figure 5a displays the original paper's survival graph (9), converted to mortality log hazards by  $\log(-\log(\text{survival}))$  (41). Figures 5b and 5c display the functional form of tumor size respectively from the ECOG trials (node-positive) (10) and the NSABP B-14 trial (node-negative) (11), respectively. Figure 5d displays the output of GAM modeling using Sauerbrei and Royston data from a GBSG trial (node-positive) (12). Thus, even though we classified the studies as indicating a linear relationship, the shapes shown by comparable methods are similar to the one we describe in Figure 2. Despite different outcomes, various measurement scales, heterogeneous selection of patients, diverse mathematical expressions, all studies show a profound unity, which points to the same nonlinear functional form.

Cell kinetics studies have shown that, in general, tumor growth is constrained after an initial exponential phase, although some controversy exists over what type of function best represents that pattern, gompertzian (36,42) or non-gompertzian (43,44). Tumors have a tendency to exhibit a growth slowdown (45). Factors slowing down tumor growth after an initial exponential phase include epigenetic modifications, aberrant genetics (46-51), and interaction with the microenvironment (52-56) including hypoxia (57-61).

This short list of mechanisms suggests an extreme biological heterogeneity within tumors, within hosts, and between patients. Therefore, it is unlikely that an aggregate functional form such as derived from the SEER data can represent any single individual tumor growth pattern. Nevertheless, the observed relationship between tumor size and mortality, which ends in a plateau, is concordant with the biological mechanisms of tumor growth. However, a slow down in tumor growth does not mean that cancer cells become less aggressive – for example, cells that survive hypoxia are also the most resistant to treatments (58-60).

Axillary lymph node status has repeatedly been shown to be the single most important predictor of disease-free survival and overall survival (1,62). Nodal involvement may be an indicator of metastatic disease, the cause of death of breast cancer patients (63). Hence, the generally admitted conclusion that tumor size loses its prognostic role in cases of nodal involvement. The finding of an almost identical effect of

tumor size, both qualitatively (functional form) and quantitatively (by hazard ratios of untransformed size covariate, gompertzian transform, and change of crude death rate for a change of tumor size from 3 to 50 mm), was unexpected (Table 3). Nodal positivity caused a dramatic shift in the prognosis, but tumor size maintained its effect on mortality (Fig 2). In other words, although prognosis is worse for node positive patients, curves of tumor size versus mortality remain parallel.

These results suggest different processes through which tumor size and nodal status affect the risk of death. This is in contradiction with what should have been expected from a Halstedian paradigm of sequential progression from primary tumor to nodal metastasis, but fits the concept of breast cancer as a systemic disease (64), where the effect of tumor size represents the phenotypic manifestation of tumor growth, while the effect of nodal involvement represents the phenotypic manifestation of metastasis. The impact of local tumor growth is in keeping with a recent pooled-analysis that found a survival advantage when radiation was delivered to the breast after breast conserving surgery (65).

*In summary*, this study on a large population database suggests that the functional form of the effect of tumor size on all-cause mortality might be represented by a gompertzian type of function. It highlights that, on the average, each millimeter increment in size is associated with a considerable relative increase in mortality for smaller tumors than for larger tumors (Fig 3). Though the time scale for successive measurements of tumor size is missing in the present analyses, which thereby do not

allow extrapolation to growth rates, the consistency of our results with the literature argues for the importance of screening and early intervention. The study further finds that node-positive status does not abate the effect of tumor size. This finding emphasizes the need for a dual approach in the management of breast cancer that has to consider both processes of metastasis and tumor growth.

## **Acknowledgments**

We are are grateful to Dirk Van Den Berge, Donald Promish, Guy Soete, Hung Van Le, and Jan Van de Steene, for the encouraging comments during the design of the study.

Much of the present work owes to Tomasz Burzykowski who stimulated discussions and introduced us to different methods of extended survival analysis. Figure 5b was reproduced with kind permission from the American Statistical Association and Robert J. Gray. Figure 5c was reproduced with kind permission from Kluwer Academic/Plenum Publishers and John Bryant. We are grateful to Willi Sauerbrei for the availability of the GBSG-2 data.

**Table 1**

Literature: reports examining the type of relationship between tumor size and outcome.

Principal author	No. of patients analyzed	Period of diagnosis	Source of patients data; treatment	Nodal status	Outcome studied	Tumor size: data type (unit)	Functional form of the relationship between outcome and tumor size
Adair (5)	1,458	1940-1943	Memorial Hospital; radical mastectomy	N0, N+	relative survival (cause-specific)	categorized (cm)	linear?
Goldenberg (6)	381	?-1973	radical mastectomy	N0, N+	recurrence rate at 2 years	categorized (cm)	linear?
Haybittle (7)	387	1975-1980	Nottingham; simple mastectomy, triple node biopsy	N0, N+	log hazard for overall survival	continuous (cm)	linear
Carter (9)	24,740	1977-1982	SEER; at least 8 nodes examined	N0, N+	5-year relative survival	categorized (cm)	linear regardless of lymph node status
Gray (10)	2,404	1978-1987	Pooled ECOG trials of adjuvant therapy	N+	log-hazard ratio of tumor recurrence	continuous (mm?)	inconclusive (nonlinear but not statistically significant)
Bryant (11)	4,028	1982-1988	NSABP B-14 trial tamoxifen vs placebo	N0	log-hazard ratio of treatment failure	continuous (mm)	linear over the range of tumor size $\leq 5$ cm for both disease free and overall survival.
Sauerbrei (12)	686	1984-1989	GBSG-2 trial of adjuvant therapy	N+	log-hazard ratio of recurrence or death	continuous (mm)	linear
Langlands (14)	13,464	1977-1982	SEER data from Carter (9) reported by McGuire (40)	N0	% surviving at 5 years	categorized (cm)	linear
Fisher (15)	2,390	1957-?	Pooled NSABP trials, adjuvant therapy post mastectomy. Excludes randomized to radiotherapy and oophorectomy	N0, N+	recurrence and mortality rates	categorized (cm)	nonlinear
Koscielny (16)	2,648	1954-1972	Gustave Roussy	N0, N+	rate of distant relapse	categorized (cm)	nonlinear (function of volume)
Atkinson (17)	2,663	1955-1979	MD Anderson; primary surgery or radiation therapy	N0, N+	rate of distant metastasis	continuous (mm?)	nonlinear (function of volume)
Brenner (18,21,22)	463	1958-1974	Princess Margaret Hospital and Gustave Roussy data from Arriagada (20); primary treatment by radiotherapy alone	N0, N+	tumor control probability (freedom from local recurrence)	categorized (cm)	nonlinear (function of volume)
Coradini (23)	1,715	1981-1986	Milan; resectable tumor	N0	hazard ratio of disease recurrence	continuous (mm)	nonlinear
Michaelson (24)	1,352	1966-1990	Van Nuys Breast Center		15-year survival rate	categorized (mm)	nonlinear
Sauerbrei (25,26)	1,189	1990-1996	ZEBRA study of adjuvant goserelin vs. CMF in pre- and peri-menopausal patients; ER positive subgroup	N+	disease-free (DFS) and overall survival (OS)	continuous (mm)	nonlinear for DFS, linear for OS

**Table 2**

SEER 9-registries 1988-1997, women with T1-T2 breast carcinoma.

Characteristic (coding in proportional hazards)	Node-negative		Node-positive	
	n = 58,070		n = 25,616	
	n	% of total	n	% of total
SEER Central registries (vs. not-Central)	14,184	24.4	6,241	24.4
SEER Western registries (vs. not-West)	22,227	38.3	9,278	36.2
Race black (vs. not-black)	3,561	6.1	2,202	8.6
Married status (vs. not-married)	34,945	60.2	15,586	60.8
Tumor medial location (vs. not-medial)	9,714	16.7	2,808	11.0
Histology ductal (vs. not-ductal)	43,862	75.5	20,405	79.7
ER negative (vs. not-negative)	8,935	15.4	4,528	17.7
PR negative (vs. not-positive)	12,092	20.8	5,820	22.7
Grade 3-4 (vs. not-Grade 3-4)	15,239	26.2	9,868	38.5
Breast conserving surgery, BCS (vs. not-BCS)	25,332	43.6	7,594	29.6
Radiotherapy, RT (vs. no-RT)	23,412	40.3	9,045	35.3
BCS with RT (interaction BCS x RT)	22,391	38.6	6,104	23.8
T-stage (not included):				
T1	44,158	76.0	12,878	50.3
T2	13,912	24.0	12,738	49.7
Year of diagnosis (as continuous):				
1988-1992	28,019	48.3	12,567	49.1
1993-1997	30,051	51.7	13,049	50.9
Age (as continuous):				
<= 50	14,994	25.8	9,011	35.2
> 50	43,076	74.2	16,605	64.8
Nodes examined (as continuous):				
<10	10,514	18.1	3,699	14.4
>= 10 and < 15	19,196	33.1	7,786	30.4
>= 15	28,360	48.8	14,131	55.2
Nodes involved (as continuous):				
0	58,070	100.0	0	0.0
1	0	0.0	9,064	35.4
2-3	0	0.0	7,694	30.0
>= 4	0	0.0	8,858	34.6

**Table 3**

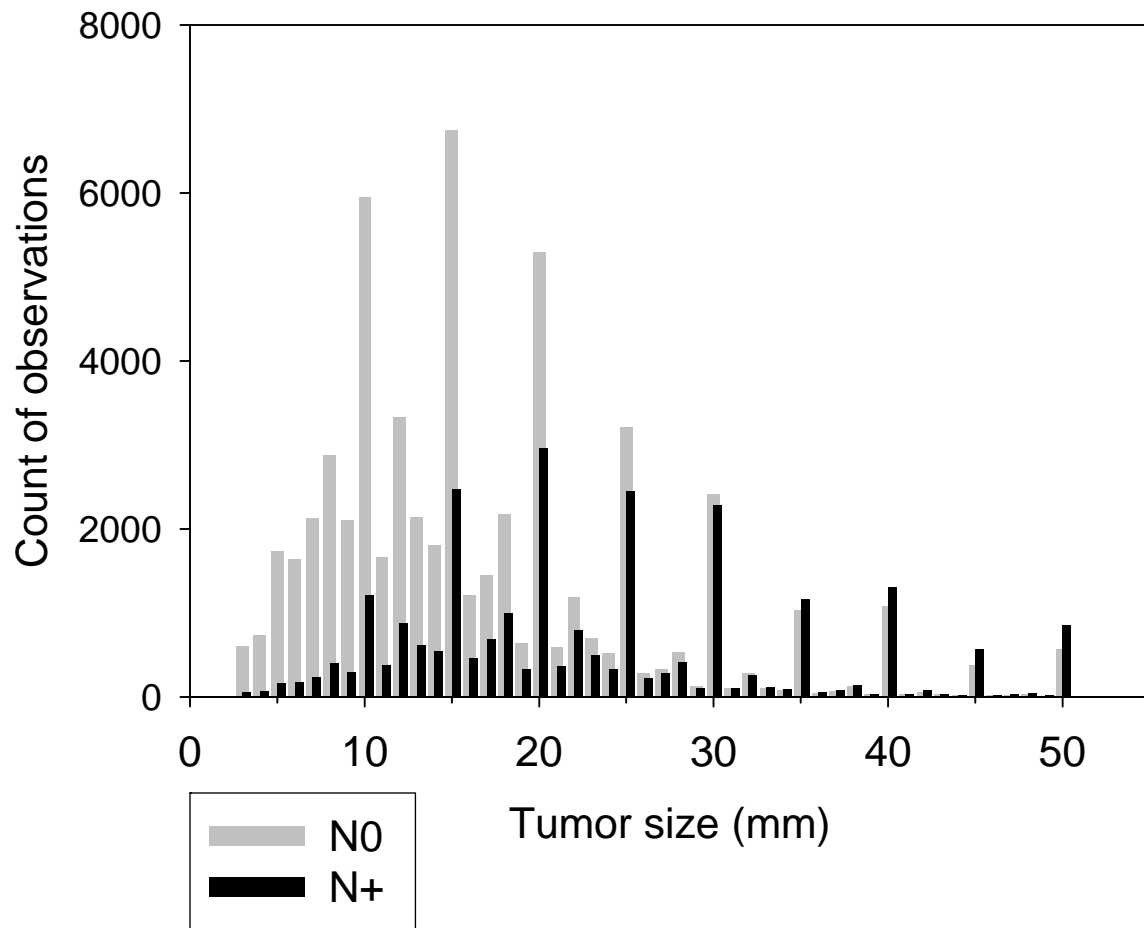
Effect of tumor size in relation to nodal status.

<b>Effect</b>	<b>Nodal status</b>	
	<b>N0</b>	<b>N+</b>
<b>Mortality hazard ratio of untransformed size covariate</b>	<b>1.03</b>	<b>1.02</b>
<b>Mortality hazard ratio of gompertzian transform size</b>	<b>2.98</b>	<b>2.71</b>
<b>Change of crude death rate for a change of tumor size from 3 to 50 mm</b>	<b>10% to 25% = 15%</b>	<b>20% to 40% = 20%</b>

## Figure 1

Frequency of tumor size measurements.

Gray: node-negative (N0). Black: node-positive (N+).



## Figure 2

Similarity of the effect of tumor size on mortality in node-negative (N0, left graphs) and in node-positive (N+, right graphs), shown by multivariate modeling (top graphs) and by crude death rates (bottom graphs).

(a) Functional form of the effect of tumor size on the log hazard of mortality based on GAM modeling in node-negative (N0) cases, with 95% confidence intervals (dotted curves).

(b) Crude death rates in function of tumor size, N0 cases.

Dot size represents a log10 function of the number of cases – small dots are <100 cases – larger dots represent up to 7,000 cases (see Figure 1).

Plain curve: plot of the scaled function  $\exp(-\exp(-(size-15)/10))$ .

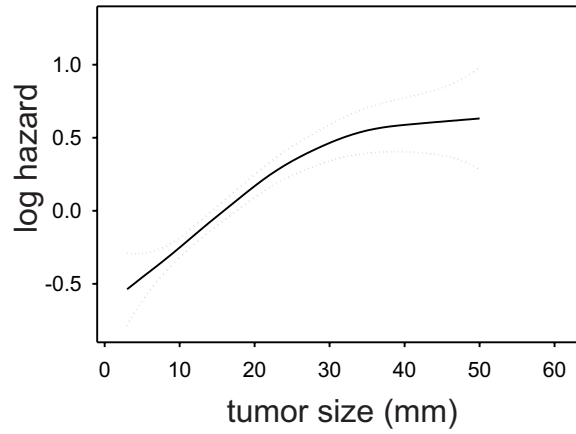
(c) Functional form based on GAM modeling in node-positive (N+) cases.

(d) Crude death rates in function of tumor size in N+. Dot size based on log10 function of the number of cases – small dots are <100 cases – larger dots represent up to 3,000 cases (see Figure 1).

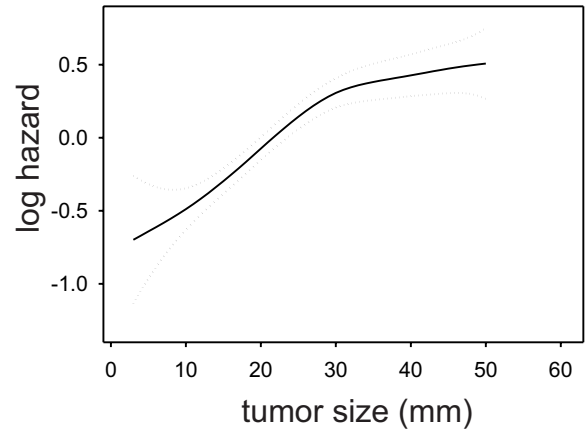
Plain curve: plot of the scaled function  $\exp(-\exp(-(size-15)/10))$ .

Dashed curve: spline smoothing of the crude death rate data points.

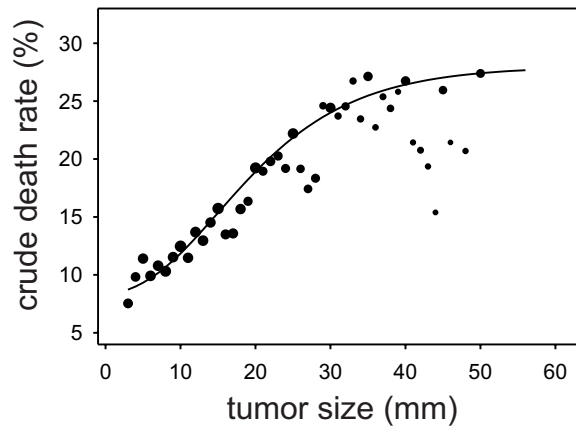
**(a) N0**



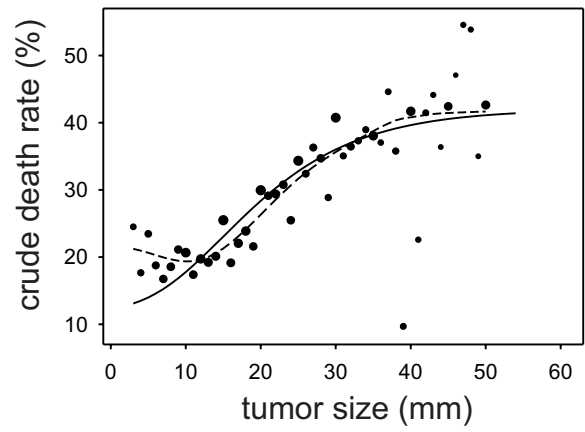
**(c) N+**



**(b) N0**



**(d) N+**

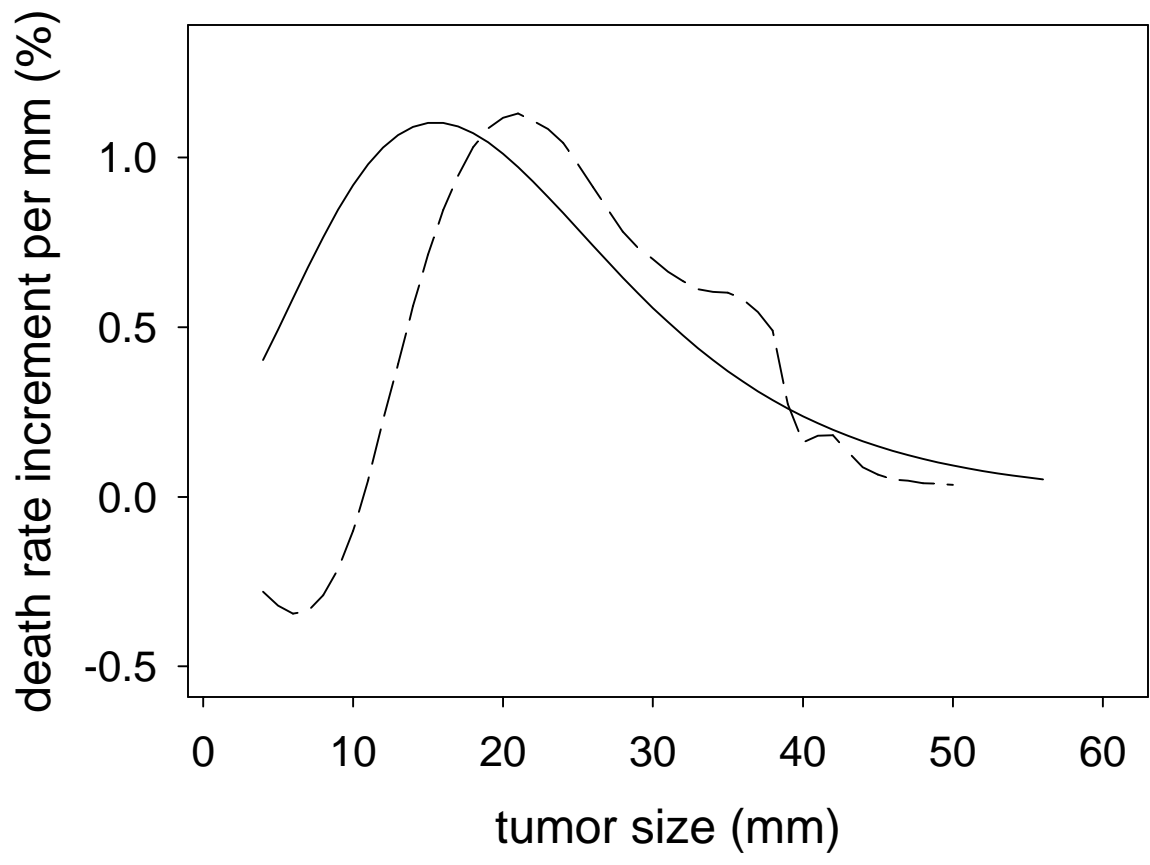


### Figure 3

Relative increase in risk of death for each 1 mm increment in tumor size.

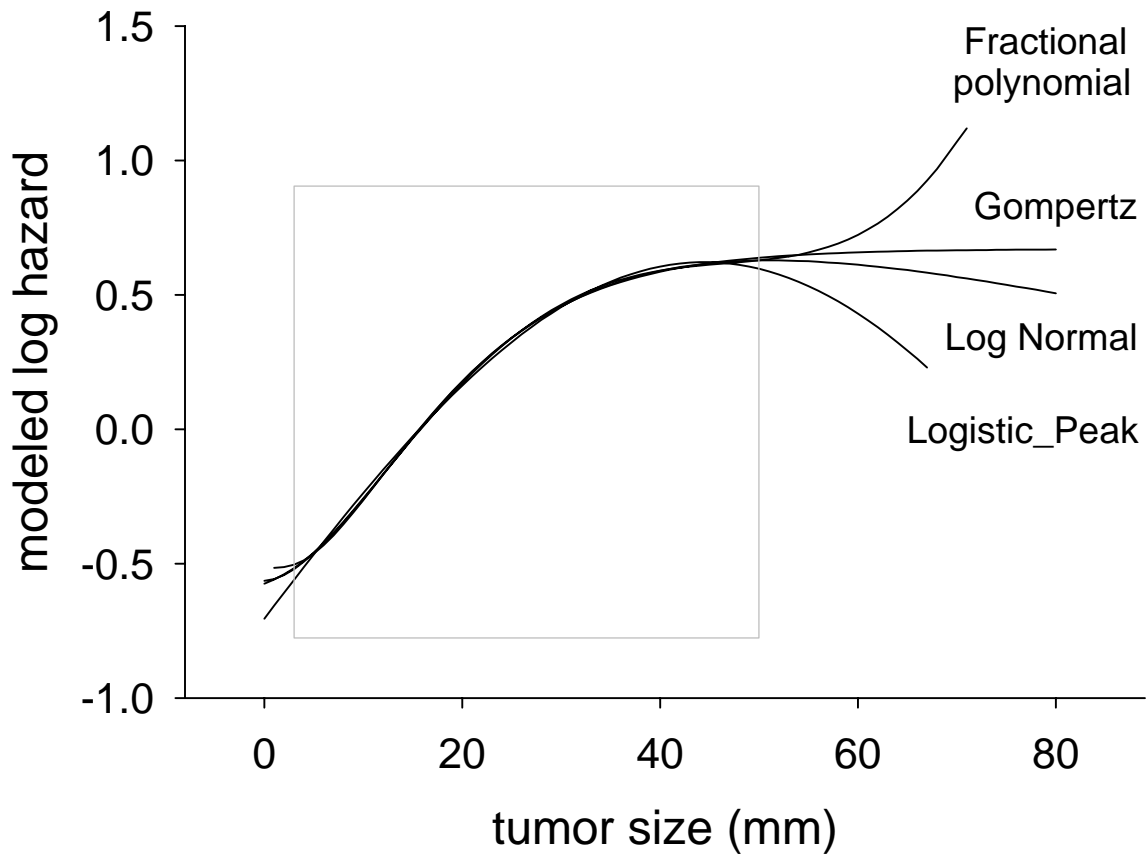
Plain curve: Plot of the slope of the scaled function  $\exp(-\exp(-(size-15)/10))$  based on modeling in node-negative cases.

Dashed curve: Plot of the slope of the smoothed crude death rates observed in node-positive cases.



## Figure 4

Selected nonlinear functions that provided an almost identical fit to the functional form of tumor size within the range of data 3-50 mm (gray square).



## Figure 5

Graphs from the literature data showing the similarity of the functional form of tumor size for different outcomes, in node-negative (N0) and in node-positive (N+).

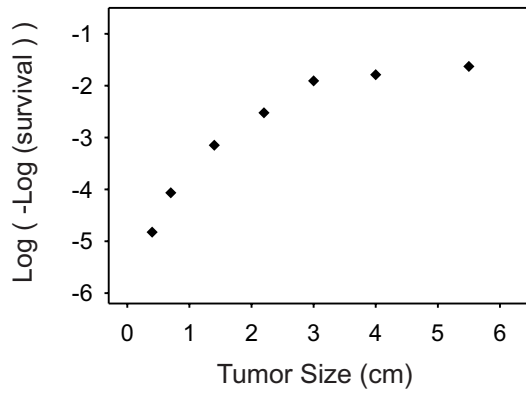
(a) Cumulative relative-mortality hazard in N0, adapted from Carter (9).

(b) Log hazard ratio of recurrence in N+, reprinted from Gray (10), with kind permission of the author and the American Statistical Association.

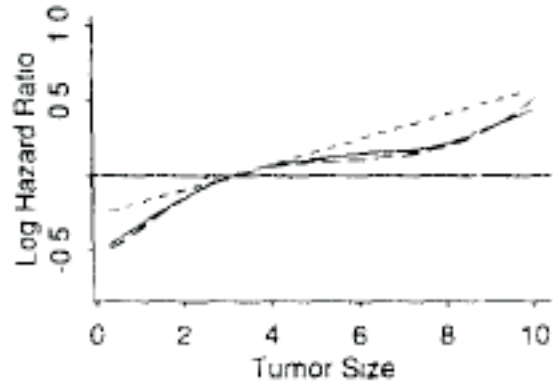
(c) Hazard ratio of treatment failure (local, regional or distant recurrence, or second primary tumor) in N0, reprinted from Bryant (11), with kind permission of the author and Kluwer Academic Publishers.

(d) Log hazard ratio of recurrence or death in N+. Functional form for tumor size based on the GBSG-2 data, from Willi Sauerbrei (13).

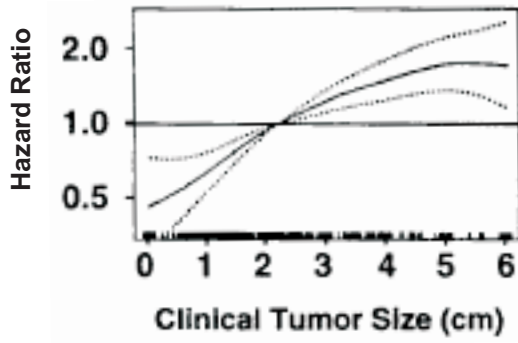
**(a) N0**



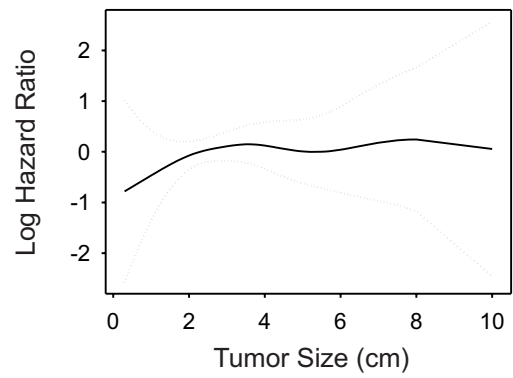
**(b) N+**



**(c) N0**



**(d) N+**



## REFERENCES

- (1) Fitzgibbons PL, Page DL, Weaver D, Thor AD, Allred DC, Clark GM, et al. Prognostic factors in breast cancer. College of American Pathologists Consensus Statement 1999. Arch Pathol Lab Med 2000;124:966-78.
- (2) Mirza AN, Mirza NQ, Vlastos G, Singletary SE. Prognostic factors in node-negative breast cancer: a review of studies with sample size more than 200 and follow-up more than 5 years. Ann Surg 2002;235:10-26.
- (3) Greene FL, Page DL, Fleming ID, Fritz AG, Balch CM, Haller DG, et al. AJCC Cancer Staging Handbook. TNM Classification of malignant tumors (6th Edition). New York: Springer Verlag; 2002.
- (4) Sobin (Editor) LH, Wittekind (Editor) Ch. TNM Classification of Malignant Tumours, 6th edition. New-York: John Wiley & Sons; 2002.
- (5) Adair F, Berg J, Joubert L, Robbins GF. Long-term followup of breast cancer patients: the 30-year report. Cancer 1974;33:1145-50.
- (6) Goldenberg IS, Carter AC, Friedell GH, Hankey BF. Identification of breast cancer patients with high risk of early recurrence after radical mastectomy. II. Clinical and pathological correlations. A report of the Primary Therapy of Breast Cancer Study Group. Cancer 1978;42:2809-26.
- (7) Haybittle JL, Blamey RW, Elston CW, Johnson J, Doyle PJ, Campbell FC, et al. A prognostic index in primary breast cancer. Br J Cancer 1982;45:361-6.

- (8) Todd JH, Dowle C, Williams MR, Elston CW, Ellis IO, Hinton CP, et al. Confirmation of a prognostic index in primary breast cancer. *Br J Cancer* 1987;56:489-92.
- (9) Carter CL, Allen C, Henson DE. Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. *Cancer* 1989;63:181-7.
- (10) Gray RJ. Flexible methods for analyzing survival data using splines, with application to breast cancer prognosis. *J Am Stat Assoc* 1992;87:942-51.
- (11) Bryant J, Fisher B, Gunduz N, Costantino JP, Emir B. S-phase fraction combined with other patient and tumor characteristics for the prognosis of node-negative, estrogen-receptor-positive breast cancer. *Breast Cancer Res Treat* 1998;51:239-53.
- (12) Sauerbrei W, Royston P, Bojar H, Schmoor C, Schumacher M. Modelling the effects of standard prognostic factors in node-positive breast cancer. German Breast Cancer Study Group (GBSG). *Br J Cancer* 1999;79:1752-60.
- (13) Sauerbrei W, Royston P. Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. *J R Stat Soc [Ser A]* 1999;162 Part 1:71-94.
- (14) Langlands AO, GebSKI V, Hirsch D, Tattersall MHN. Delay in the clinical diagnosis of breast cancer: estimating its effect on prognosis, with particular reference to medical litigation. *The Breast* 2002;11:386-93.
- (15) Fisher B, Slack NH, Bross ID. Cancer of the breast: size of neoplasm and prognosis. *Cancer* 1969;24:1071-80.

- (16) Koscielny S, Tubiana M, Le MG, Valleron AJ, Mouriesse H, Contesso G, et al. Breast cancer: relationship between the size of the primary tumour and the probability of metastatic dissemination. *Br J Cancer* 1984;49:709-15.
- (17) Atkinson EN, Brown BW, Montague ED. Tumor volume, nodal status, and metastasis in breast cancer in women. *J Natl Cancer Inst* 1986;76:171-8.
- (18) Brenner DJ. Dose, volume, and tumor-control predictions in radiotherapy. *Int J Radiat Oncol Biol Phys* 1993;26:171-9.
- (19) Sivaramakrishna R, Gordon R. Detection of breast cancer at a smaller size can reduce the likelihood of metastatic spread: a quantitative analysis. *Acad Radiol* 1997;4:8-12.
- (20) Arriagada R, Mouriesse H, Sarrazin D, Clark RM, Deboer G. Radiotherapy alone in breast cancer. I. Analysis of tumor parameters, tumor dose and local control: the experience of the Gustave-Roussy Institute and the Princess Margaret Hospital. *Int J Radiat Oncol Biol Phys* 1985;11:1751-7.
- (21) Dubben HH, Thames HD, Beck-Bornholdt HP. Tumor volume: a basic and specific response predictor in radiotherapy. *Radiother Oncol* 1998;47:167-74.
- (22) Buffa FM, Fenwick JD, Nahum AE. An analysis of the relationship between radiosensitivity and volume effects in tumor control probability modeling. *Med Phys* 2000;27:1258-65.
- (23) Coradini D, Daidone MG, Boracchi P, Biganzoli E, Oriana S, Bresciani G, et al. Time-dependent relevance of steroid receptors in breast cancer. *J Clin Oncol* 2000;18:2702-9.

- (24) Michaelson JS, Silverstein M, Wyatt J, Weber G, Moore R, Halpern E, et al. Predicting the survival of patients with breast carcinoma using tumor size. *Cancer* 2002;95:713-23.
- (25) Sauerbrei W, Schumacher M, Cuzick J, Jonat W, Kaufmann M. Prognostic factors determine groups with different prognoses in pre- and perimenopausal women with estrogen receptor-positive, node-positive breast cancer treated with goserelin (Zoladex<sup>R</sup>) or CMF - results from the ZEBRA study. Abstract 509. *Breast Cancer Res Treat* 2003;82 (Suppl 1):S121.
- (26) Jonat W, Kaufmann M, Sauerbrei W, Blamey R, Cuzick J, Namer M, et al. Goserelin versus cyclophosphamide, methotrexate, and fluorouracil as adjuvant therapy in premenopausal patients with node-positive breast cancer: The Zoladex Early Breast Cancer Research Association Study. *J Clin Oncol* 2002;20:4628-35.
- (27) Michaelson JS, Silverstein M, Sgroi D, Cheongsiatmoy JA, Taghian A, Powell S, et al. The effect of tumor size and lymph node status on breast carcinoma lethality. *Cancer* 2003;98:2133-43.
- (28) Beenken SW, Urist MM, Zhang Y, Desmond R, Krontiras H, Medina H, et al. Axillary lymph node status, but not tumor size, predicts locoregional recurrence and overall survival after mastectomy for breast cancer. *Ann Surg* 2003;237:732-8.
- (29) National Cancer Institute. Surveillance, Epidemiology, and End Results (SEER) Program Public-Use Data (1973-1999), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2002, based on the November 2001 submission. 2002. [[Http://seer.cancer.gov](http://seer.cancer.gov)].

- (30) 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.
- (31) Therneau TM, Grambsch PM. Modeling survival data: extending the Cox model. New York, NY: Springer-Verlag; 2000.
- (32) Hastie T, Tibshirani R. Generalized additive models. New York: Chapman and Hall; 1990.
- (33) 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 Research and Treatment* 1998;52:227-37.
- (34) Vinh-Hung V, Van Den Berge D, Van de Steene J, Soete G, Storme G. Functional form of tumor size and other quantitative variables in T1-T2 node negative breast cancer. Abstract No 3611. *Proc Am Soc Clin Oncol* 22, 898. 2003.

Ref Type: Conference Proceeding

- (35) Vinh-Hung V, Voordeckers M, Van de Steene J, Soete G, Lamote J, Storme G. Omission of radiotherapy after breast-conserving surgery: survival impact and time trends. *Radiother Oncol* 2003;67:147-58.
- (36) Gilewski TA, Dang C, Surbone A, Norton L. Section 13, Principles of chemotherapy: Chapter 38, Cytokinetics. In: Bast Jr RC, Kufe D W, Pollock R E, Weichselbaum R R, Holland J F, Frei III E, eds. *Holland-Frei Cancer Medicine*, 5e. Hamilton, Ontario: BC Decker Inc; 2000. p. 511-38.

- (37) Clare SE, Nakhlis F, Panetta JC. Molecular biology of breast cancer metastasis. The use of mathematical models to determine relapse and to predict response to chemotherapy in breast cancer. *Breast Cancer Res* 2000;2:430-5.
- (38) McGuire WL, Tandon AK, Allred DC, Chamness GC, Ravdin PM, Clark GM. Prognosis and treatment decisions in patients with breast cancer without axillary node involvement. *Cancer* 1992;70:1775-81.
- (39) McGuire WL, Clark GM. Prognostic factors and treatment decisions in axillary-node-negative breast cancer. *N Engl J Med* 1992;326:1756-61.
- (40) McGuire WL, Tandon AK, Allred DC, Chamness GC, Ravdin PM, Clark GM. Treatment decisions in axillary node-negative breast cancer patients. *J Natl Cancer Inst Monogr* 1992;11:173-80.
- (41) Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part I: basic concepts and first analyses. *Br J Cancer* 2003;89:232-8.
- (42) Norton L. A Gompertzian model of human breast cancer growth. *Cancer Res* 1988;48:7067-71.
- (43) Speer JF, Petrosky VE, Retsky MW, Wardwell RH. A stochastic numerical model of breast cancer growth that simulates clinical data. *Cancer Res* 1984;44:4124-30.
- (44) Retsky MW, Swartzendruber DE, Wardwell RH, Bame PD. Is Gompertzian or exponential kinetics a valid description of individual human cancer growth? *Med Hypotheses* 1990;33:95-106.

- (45) Hahnfeldt P, Panigrahy D, Folkman J, Hlatky L. Tumor development under angiogenic signaling: a dynamical theory of tumor growth, treatment response, and postvascular dormancy. *Cancer Res* 1999;59:4770-5.
- (46) Sato T, Akiyama F, Sakamoto G, Kasumi F, Nakamura Y. Accumulation of genetic alterations and progression of primary breast cancer. *Cancer Res* 1991;51:5794-9.
- (47) Dairkee SH, Smith HS. Genetic analysis of breast cancer progression. *J Mammary Gland Biol Neoplasia* 1996;1:139-51.
- (48) Hampl M, Hampl JA, Schwarz P, Frank S, Hahn M, Schackert G, et al. Accumulation of genetic alterations in brain metastases of sporadic breast carcinomas is associated with reduced survival after metastasis. *Invasion Metastasis* 1998;18:81-95.
- (49) Tomlinson IP. Mutations in normal breast tissue and breast tumours. *Breast Cancer Res* 2001;3:299-303.
- (50) Bertucci F, Viens P, Hingamp P, Nasser V, Houlgatte R, Birnbaum D. Breast cancer revisited using DNA array-based gene expression profiling. *Int J Cancer* 2003;103:565-71.
- (51) Frigyesi A, Gisselsson D, Mitelman F, Hoglund M. Power law distribution of chromosome aberrations in cancer. *Cancer Res* 2003;63:7094-7.
- (52) Ben Baruch A. Host microenvironment in breast cancer development: inflammatory cells, cytokines and chemokines in breast cancer progression: reciprocal tumor-microenvironment interactions. *Breast Cancer Res* 2003;5:31-6.

- (53) Neumark E, Sagi-Assif O, Shalmon B, Ben Baruch A, Witz IP. Progression of mouse mammary tumors: MCP-1-TNFalpha cross-regulatory pathway and clonal expression of promalignancy and antimalignancy factors. *Int J Cancer* 2003;106:879-86.
- (54) Schor SL, Schor AM. Phenotypic and genetic alterations in mammary stroma: implications for tumour progression. *Breast Cancer Res* 2001;3:373-9.
- (55) Shekhar MP, Werdell J, Santner SJ, Pauley RJ, Tait L. Breast stroma plays a dominant regulatory role in breast epithelial growth and differentiation: implications for tumor development and progression. *Cancer Res* 2001;61:1320-6.
- (56) Shekhar MP, Pauley R, Heppner G. Host microenvironment in breast cancer development: extracellular matrix-stromal cell contribution to neoplastic phenotype of epithelial cells in the breast. *Breast Cancer Res* 2003;5:130-5.
- (57) Hohenberger P, Felgner C, Haensch W, Schlag PM. Tumor oxygenation correlates with molecular growth determinants in breast cancer. *Breast Cancer Res Treat* 1998;48:97-106.
- (58) Hockel M, Vaupel P. Biological consequences of tumor hypoxia. *Semin Oncol* 2001;28:36-41.
- (59) Knowles HJ, Harris AL. Hypoxia and oxidative stress in breast cancer. Hypoxia and tumorigenesis. *Breast Cancer Res* 2001;3:318-22.
- (60) Boudreau N, Myers C. Breast cancer-induced angiogenesis: multiple mechanisms and the role of the microenvironment. *Breast Cancer Res* 2003;5:140-6.

- (61) Wouters BG, Koritzinsky M, Chiu RK, Theys J, Buijsen J, Lambin P. Modulation of cell death in the tumor microenvironment. *Semin Radiat Oncol* 2003;13:31-41.
- (62) Cserni G. Axillary staging of breast cancer and the sentinel node. *J Clin Pathol* 2000;53:733-41.
- (63) Nicolson GL. Breast cancer metastasis-associated genes: role in tumour progression to the metastatic state. *Biochem Soc Symp* 1998;63:231-43.
- (64) Fisher B. From Halsted to prevention and beyond: advances in the management of breast cancer during the twentieth century. *Eur J Cancer* 1999;35:1963-73.
- (65) 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:in press.

g34	N Rows	N(0)	N(1)	characteristic	Node-negative	p
0.0E+01	58579	42831	15748			
1	25107	15239	9868	SEER Central registries	14,184	24.4
				SEER Western registries	22,227	38.3
prneg	N Rows	N(0)	N(1)	Race black	3,561	6.1
0.0E+01	65774	45978	19796	Married status	34,945	60.2
1	17912	12092	5820	Tumor medial location	9,714	16.7
erneg	N Rows	N(0)	N(1)	Histology ductal	43,862	75.5
0.0E+01	70223	49135	21088	ER negative	8,935	15.4
1	13463	8935	4528	PR negative	12,092	20.8
				Grade 3-4	15,239	26.2
histduct	N Rows	N(0)	N(1)	Breast conserving surgery (BCS)	25,332	43.6
0.0E+01	19419	14208	5211	Radiotherapy (RT)	23,412	40.3
1	64267	43862	20405	BCS with RT	22,391	38.6
innerkwd	N Rows	N(0)	N(1)	T-stage:		0.0
0.0E+01	71164	48356	22808	T1	44,158	76.0
1	12522	9714	2808	T2	13,912	24.0
married	N Rows	N(0)	N(1)	Age:		0.0
0.0E+01	33155	23125	10030	<= 50	14,994	25.8
1	50531	34945	15586	> 50	43,076	74.2
rt	bc	N Rows	N(0)	N(1)	Nodes examined:	0.0
0.0E+01	0.0E+01	46798	31717	15081	<10	10,514
0.0E+01	1	4431	2941	1490	>= 10 and < 15	19,196
1	0.0E+01	3962	1021	2941	>= 15	28,360
1	1	28495	22391	6104	Nodes involved:	0.0
					0.0E+01	58,070
					1	0.0E+01
					2-3	0.0E+01
					>= 4	0.0E+01
						58070
rt	N Rows	N(0)	N(1)			
0.0E+01	51229	34658	16571			
1	32457	23412	9045			
bcs	N Rows	N(0)	N(1)			
0.0E+01	50760	32738	18022			
1	32926	25332	7594			

