PERGAMON

European Journal of Cancer 38 (2002) 1181–1188

European Journal of Cancer

www.ejconline.com

Prognostic role of serum CA15.3 in 362 node-negative breast cancers. An old player for a new game

M. Gion[a,*], P. Boracchi[b], R. Dittadi[a], E. Biganzoli[c], L. Peloso[a,b], R. Mione[a], C. Gatti[d], A. Paccagnella[d], E. Marubini[b,c]

[a]*Centro Regionale per lo Studio degli Indicatori Biochimici di Tumore, Ospedale Civile, ULSS12 Venice, Italy*
[b]*Istituto di Statistica Medica e Biometria, Università degli Studi di Milano, Milan, Italy*
[c]*Unità di Statistica e Biometria, Istituto Nazionale per lo Studio e la Cura dei Tumori, Milan, Italy*
[d]*Divisione di Oncologia Medica, Ospedale Civile, ULSS 12, Venice, Italy*

Received 6 March 2001; received in revised form 29 October 2001; accepted 13 November 2001

Abstract

The aims of the present investigation were to evaluate the association between serum CA15.3 levels and other biological and clinical variables and its prognostic role in patients with node-negative breast cancer. We evaluated 362 patients operated upon primary breast cancer from 1982 to 1992 (median follow-up 69 months). Serum CA15.3 was measured by an immunoradiometric assay. The association between variables was investigated by a Principal Component Analysis (PCA) and the prognostic role of CA15.3 on relapse-free survival (RFS) was investigated by Cox regression models adjusting for age, oestrogen receptor (ER), tumour stage, and ER×age interaction, with both the likelihood ratio test and Harrell's c statistic. The prognostic contribution of CA 15.3 was highly significant. Log relative hazard of relapse was constant until approximately 10 (U/ml) of CA15.3 and increased thereafter with increasing marker levels. CA15.3 showed a significant contribution using as a cut-off point a value of 31 U/ml. However, the contribution to the model of the marker as a continuous variable is much greater. From these findings, we can conclude that: (i) CA15.3 is a prognostic marker in node-negative breast cancer; (ii) its relationship with prognosis is continuous, with the risk of relapse increasing progressively from approximately 10 U/ml.

Keywords: CA15.3; MUC1; Serum markers; Breast cancer; Node-negative; Prognosis; Regression splines

1. Introduction

The assessment of the risk of relapse after curative treatment of primary breast cancer still represents a challenge. Although it is becoming less critical in patients without axillary metastasis, since clinicians tend to prescribe conventional adjuvant treatment in the majority of these patients [1], it remains pivotal for the choice of the type of therapy and for the selection of agents [2]. The number of tissue parameters potentially related to clinical outcome has sharply increased in the last decade and hundreds of prognostic factors are under evaluation [3]. Almost all these putative prognostic markers have been determined in the tumour tissue, while anecdotal data have been reported on serum markers [4,5]. This is not surprising, since tissue markers are a genuine and direct expression of the tumour phenotype. Conversely, the serum level of tumour-associated markers depends on several variables, including tumour burden, the production and release rate by normal tissues/organs, metabolism and clearance. However, in the last few years, the biological role of some mucin markers traditionally measured in blood has been supported by several findings. Of note, the products of the *MUC1* gene [6] have been extensively investigated, showing that they may elicit different effects favourable to tumour dissemination [7]. MUC1 protein may act as an anti-adhesive molecule since it reduces cell to cell aggregation and cell adherence to extracellular matrix components [8], thus facilitating tumour cell migration and metastatic spread. Circulating MUC1 elicits an immunosuppressive action either by switching T cells to a state of anergy [9] or inducing apoptosis of activated T cells [10]. Moreover, the overexpression of MUC1 protein

* Corresponding author. Tel.: +39-041-529-4535; fax: +39-041-529-4653.
E-mail address: cnabo@provincia.venezia.it (M. Gion).

0959-8049/02/$ - see front matter
PII: S0959-8049(01)00426-9

on the cell surface seems to protect cancer cells from immunorecognition and destruction [11]. Accordingly, high MUC1-associated marker levels in the blood are predictive of poor response in patients with metastatic breast cancer treated with immunotherapy [5]. These findings suggest that MUC1-associated markers may be considered as potential prognostic indicators. Several MUC-1 associated markers have been developed and used in clinical practice. The most widely evaluated is CA15.3, which measures the level of a mucin-like membrane glycoprotein shed from the tumour cell. The antigen is recognised through two monoclonal antibodies, 115D8 and DF3, raised against milk fat globule membranes and a membrane-enriched fraction of a metastatic breast cancer, respectively. So far, few papers have reported on the relationship between CA15.3 and prognosis in primary breast cancer [12–14], and have failed to find any prognostic role in node-negative cases [4].

The value of serum markers has been universally categorised as positive or negative with reference to cut-off points. Cut-off point based decision criteria have been extensively used in prognosis assessment. However, they have not led to the optimal clinical use of tissue biomarkers [15]. Recently, several studies have investigated the prognostic impact of variables in their original scale of measurement, showing continuous relationships and giving little support to the threshold relationships underlying the adoption of cut-off values [16–21]. This fact suggests that more accurate information may be obtained by processing data with statistical approaches for continuous variables [22]. This problem may affect also the use of serum markers for the purpose of prognosis assessment. The aim of the present investigation was to evaluate: (i) the association between serum levels of CA15.3 and other biological and clinical variables in patients with node-negative breast cancer not treated with adjuvant therapies; (ii) the prognostic role of CA15.3 accounting for the effect of other prognostic factors, using a statistical approach suitable for flexible modelling of continuous prognostic relationships.

2. Patients and methods

In the present investigation, we considered 599 patients operated upon primary breast cancer, without the involvement of the axillary lymph-nodes and without chemotherapy or hormonal therapy before and/or after the surgery, whose tissue were consecutively sent to the Venice Laboratory for steroid receptor and tumour markers determination from July 1982 to July 1992. The same cases were considered in a previous study [21]. Exclusion criteria were as follows: (a) no information on pT stage or pT_4 cases; (b) tumours other than infiltrating ductal, lobular and mixed carcinoma types; (c) pluricentric or bilateral tumours; and (d) concomitant or previous other neoplastic events leading finally to 473 patient records. For 376 patients, serum samples were available for CA15.3 determinations and for 373 of them information were available on oestrogen receptor (ER) and progesterone receptor (PgR). For 362 patients, follow-up information was collected.

Patients were treated with radical surgery (mastectomy or quadrantectomy) and axillary dissection. The type of surgery was not considered since it has not a significant prognostic role for the end-point considered in this study [23]. Pathological findings were assessed according to the World Health Organization criteria. Patients were followed-up with a clinical control every 4 months in the first year, every 6 months in the second year and yearly thereafter.

Relapse-free survival was considered as the time elapsed from the date of surgery to the date of the first neoplastic event (locoregional and distant relapse) or the last clinical examination. Times to other primary non-mammary tumours, contralateral breast tumours and deaths without the evidence of disease recurrence were considered as censored. With a median follow-up time of 69 months (first quartile: 53 months, third quartile: 90 months), 93 events were observed.

2.1. CA15.3 immunoassay

Serum samples were collected before the mastectomy and kept frozen (at -80° C) until the assay was performed. Serum CA15.3 levels were measured using a commercially available double-determinant immunoradiometric kit (ELSA-CA15.3, CIS Diagnostici, Santhià, Vercelli, Italy). The CA15.3 epitope is recognised by two monoclonal antibodies. The first monoclonal antibody (115D8) is raised against milk fat globule membranes, the second (DF3) against a membrane-enriched fraction of a metastatic breast cancer. The method precision, expressed as a coefficient of variation between 20 replicates of human serum pools with three different antigen concentrations, was lower than 8.4% intra-assay and lower than 9.6% between assays.

2.2. Steroid receptors assay

Tissue samples where obtained fresh from the operating room, handled on ice and stored in liquid nitrogen until assayed. Cytosol was prepared as previously described in Ref. [21]. ER and PgR were measured in the low-salt extract using the radioligand binding assay (RBA) method recommended by the European Organization for Research and Treatment of Cancer (EORTC) [24]. ER and PgR are expressed as fmol/mg of cytosol protein.

The total protein in the cytosol was measured by the Coomassie brilliant blue colorimetric assay (Bio-Rad, Anheim, CA, USA).

2.3. Statistical analysis

2.3.1. Correlation among the variables

The correlation between CA15.3 and each of the other continuous variables (age, ER and PgR receptor contents, tumour size) was measured by Spearman's correlation coefficient (r); due to the multiple testing procedure, the Bonferroni's correction was adopted to account for the inflation of the type I statistical error. Since pairwise associations do not take into account the concomitant relationships with the other variables, the study of the joint relationship among variables was investigated resorting to the Principal Component Analysis (PCA) [25]. This technique performs a rotation of the axes of the multivariate space of the original variables, along orthogonal directions of maximal variance. A new space is defined by principal components (PC), each of them being characterised by a percentage of the explained variance of the data. The biplot graph allows for the representation of both transformed observations and the variables; their projection as points and vectors, respectively, on the subspace defined by the first PC axes is useful to visualise correlation structures present in the data. The vectors corresponding to the variables are represented as arrows whose length indicate the proportion of the original variance explained, whereas their direction indicates the relative correlation with the considered PC axes. Two strongly correlated variables are projected close together or opposite with maximum alignment when the correlation was strongly positive or negative respectively, otherwise they tend to be projected with an angle of 90°. To perform a robust analysis with respect to distribution assumptions and possible non-linear correlation, a rank transformation was applied on each variable.

2.3.2. Prognostic role of the variables

The joint role of the prognostic variables (multivariate analysis) on RFS were investigated resorting to Cox regression models. Age, ER, PgR and CA15.3 were considered as continuous variables, whereas tumour size was considered as categorical (pT_1, $pT_{2–3}$). The natural logarithmic transformation (the constant term 1 was summed to the original variable values) was adopted for the biological variables because their distributions were positively skewed and proportional increments of their values were considered to be relevant for the assessment of their prognostic impact.

The proportional hazard assumption was tested as suggested by Grambsch and Therneau [26]. Since the prognostic role of age, ER, PgR, and tumour size were previously investigated in another case series [21], we adopt here as a baseline model previously identified for the above variables and we refer to the published paper for the details on the modelling strategy [21]. Briefly, age and ER effects were modelled by regression splines and tumour size by a dummy variable. On the grounds of previous findings, the interaction between ER and age was considered and PgR was not included into the model because it did not have a prognostic contribution [21]. Keeping the structure of the model previously published in Ref. [21], CA15.3 was then added by considering a restricted cubic spline with three knots.

The prognostic contribution of CA15.3 was evaluated (i) by a likelihood ratio test with a conventional significance level of 5%, and (ii) by comparing the predictive ability of the multivariate model with all variables with the predictive ability of the model without CA15.3. For this aim, the Harrell's c statistic [27] was used; c values were corrected for the optimism by an internal validation procedure. The relative decrement of optimism corrected c due to the exclusion of variables from the regression model was calculated as $(c_f - c_r)/(c_f - 0.5)*100$, where c_f and c_r refer to the full and reduced models (without the variable of interest), respectively. The relative decrement in c values due to pathological T stage was also calculated as a reference comparison, since it is a recognised prognostic factor. For descriptive purposes, model-estimated survival curves were plotted for selected combinations of covariate values.

Since CA15.3 was used as categorical variables in previous studies, the prognostic contribution of CA15.3 on the continuous scale was compared with those resulting from the categorisation with cut-off values of 10, 22 and 31 U/ml (the latter two among those most frequently reported in literature).

To graphically compare thresholds and continuous effects, a joint plot of the relative hazard estimated from multivariate models including CA15.3 as a dichotomous variable (using the above cut-off values) or the spline transformation, was provided as a function of CA15.3 values in the logarithmic scale. As a reference value for plotting continuous effects, the cut-off points were used to calculate the relative hazards. In such a way, a direct evidence of the differences between threshold and continuous estimates can be achieved according to the considered reference cut-off values.

Statistical analysis was performed by the S-Plus package.

3. Results

Table 1 summarises the main characteristics of the distributions of age, ER and pT; pT_2 and pT_3 cases were grouped together due to the low number of the latter ($n = 3$). Tumour grade was reported to be associated with prognosis in several investigations [28]. However, a large inter-observer variability was reported in the literature [29] and was confirmed also in the present investigation (data not shown). Therefore, it was not considered in the present study.

Table 1
Distributions of the main patient characteristics

Variable	Categories or units	Values
Age	Years	59 (49–68)[a]
ER	fmol/mg protein	56 (13–161)[a]
PgR	fmol/mg protein	41 (10–155)[a]
pT	T_1	220 (61%)[b]
	T_{2-3}	142 (39%)[b]
CA15.3	U/ml	17 (12–22)[a]

ER, oestrogen receptor; PgR, progesterone receptor.
[a] Median, first and third quartiles (Q1-Q3).
[b] Number of patients and percentage values.

The minimum and the maximum values obtained for CA15.3 were 5 and 194 U/ml, respectively. It is worth noting that for 361 patients values of CA15.3 are less or equal to 50.8, therefore 194 is considerably higher than all other values, but after a careful check no reasons were found to exclude the patient's record from the analysis.

3.1. Correlation among the variables

The correlation was evaluated on 340 subjects with available measures on all the considered variables; the limiting variable was tumour size (22 missing information); for these patients, only the information on pathological T stage was available, but the corresponding measure of tumour diameter was not reported. Spearman's rank correlation coefficients (r) showed values ranging between 0.162 and 0.006; a weak direct correlation between CA15.3 and tumour size was observed ($r=0.162$ $P=0.0029$). No correlations were found between CA15.3 and the other variables (age: $r=0.083$, $P=0.1126$; ER: $r=0.039$, $P=0.4597$; PgR: $r=0.006$, $P=0.9039$). Since pairwise associations do not take into account the joint relationships with the other variables, a further investigation of the association structure among prognostic factors was performed by a multivariate technique (PCA). The space identified by the first two PCs explains 55.4% of the total data variability (PC1 = 31.7%, PC2 = 23.7%). The projections of the original variables on the plane identified by PC1 and PC2 is shown in Fig. 1. The first axis (PC1) is mainly characterised by age, ER and PgR. The second axis (PC2) is mainly characterised by tumour size and CA15.3. The positive association between tumour size and CA15.3 was evident also from this analysis. The variables characterising the first axis (age, ER and PgR), as expected, appeared to be positively correlated. Overall, tumour size and CA 15.3 appeared to be poorly associated with age, ER and PgR.

3.2. Prognostic relationships

When the baseline model containing age, ER, tumour stage and the interaction term between ER and age was

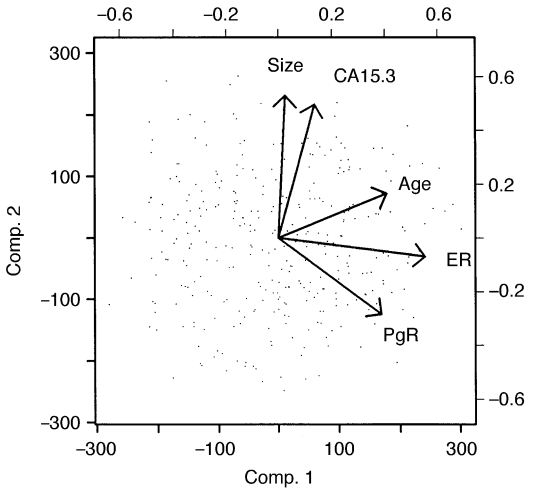

Fig. 1. Biplot graph of principal component analysis (PCA) on rank values of age, oestrogen receptor (ER) and progesterone receptor (PgR) contents, tumour size and CA15.3. The correlation between variables is shown on the plane defined by the first two principal components (Comp.1, Comp.2). Bottom and left axes are the coordinates of the observations, whereas top and right axes are the coordinates of the variables in the PC1, PC2 plane, respectively.

performed, the predictive ability of the model was $c=0.644$, and $c=0.615$ after the application of the correction for the optimism.

A model including CA15.3 was then performed and the predictive capability of the model increased to $c=0.659$, and $c=0.629$ after the application of the correction for the optimism. The relative decrement of predictive capability (optimism corrected c) due to the exclusion of CA15.3 from the full model was 11.2%, whereas for pT it was 37.5%. The prognostic contribution of CA 15.3 was highly significant (likelihood ratio test: $\chi^2=10.07$, degrees of freedom (d.f.) = 1, $P=0.0015$). The proportional hazard assumption on the model was tenable ($\chi^2=4.53$, $P=0.476$). CA15.3 was modelled by the non-linear term of a three-knots restricted cubic spline since the contribution of the linear term was not significant (Likelihood ratio test: $\chi^2=0.00031$, d.f. = 1 $P=0.986$). Results of the regression model are summarised in Table 2.

As sensitivity analysis, to evaluate the effect of the outlying value of CA15.3, we performed an additional model excluding the whole corresponding patient record. The prognostic contribution of CA 15.3 decreased, but was still significant (likelihood ratio test: $\chi^2=5.48$, d.f. = 1, $P=0.019$) and the estimate of regression coefficient for CA15.3 was lower than the corresponding value reported in Table 2 ($b=0.5644$, S.E. = 0.2319).

An additional analysis considering tumour size instead of tumour stage does not provide a substantial increment in the information already provided with pT. In fact, a linear effect was observed for tumour size since the contribution of the non-linear term was not

Table 2
(a) Results of the Cox model: estimated regression coefficients and their standard errors;[a] (b) Wald statistics for the contribution of variables in the regression model in (a)[b]
(a)

Variable		*b*	S.E.(*b*)
ER	Log(ER): L^1	−0.6214	0.2191
	Log(ER): NL^2	0.9899	0.2728
Age	Age: NL^3	−0.0778	0.1258
	Age: NL^4	−0.3453	1.5005
T stage	pT_{2-3} versus pT_1	0.7186	0.2128
Interaction			
ER×age	Log(ER): L^1×age: NL^3	0.0993	0.0569
	Log(ER): NL^2×age: NL^3	−0.1882	0.0677
	Log(ER): L^1×age: NL^4	−0.4784	0.5917
	Log(ER): NL^2×age: NL^4	1.2403	0.6025
CA15.3	Log(Ca15.3): NL^2	0.7385	0.2188

(b)

Factor	X^2	d.f.	*P* value
Log(ER): L + NL (factor + interaction terms)	15.89	6	0.0144
Age: NL (factor + interaction terms)	13.40	6	0.0371
pT stage	11.40	1	0.0007
Log(ER)×age:	11.70	4	0.0197
Log(CA15.3): NL	11.40	1	0.0007
Total	40.98	10	<0.0001

[a] *b*, estimated regression coefficient; S.E.(*b*) standard error of *b*; L^1, linear term; NL^2, non-linear term of a three-knots restricted cubic spline; NL^3, first non-linear term of a four-knots restricted cubic spline; NL^4, second non-linear term of a four-knots restricted cubic spline; ER, oestrogen receptor.
[b] X^2, Wald statistic; d.f., degrees of freedom; L, linear term; NL, non-linear terms.

significant ($\chi^2 = 2.01$, d.f. = 1, $P = 0.156$). As expected, the estimated regression coefficient and its sign ($b = 0.3664$, S.E. = 0.0586) indicates the increase of the relative hazard of relapse with the increase of 1 cm in tumour size. Moreover the interaction between tumour size and CA15.3 was not significant ($\chi^2 = 0.32$, d.f. = 1, $P = 0.57$), therefore a simple summation effect is expected between tumour size and CA15.3.

The relationship between CA15.3 and the logarithm of the relative hazard estimated by the multivariate regression model performed on the whole case series is shown in Fig. 2. Log relative hazard of relapse is constant until approximately 10 (U/ml) and increases thereafter with the increase of the marker values. A joint graphical representation of log relative hazard of relapse as a function of CA15.3 and other variables does not provide additional information because no interactions involving the above marker were included in the model.

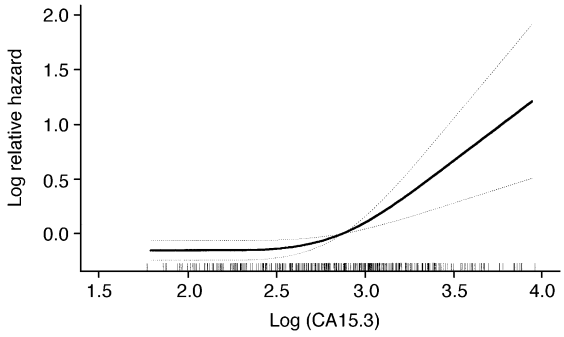

Fig. 2. Log relative hazard of relapse as a function of CA15.3 (continuous line). Dotted lines are 95% Confidence Limits.

We represented the RFS probability curves estimated by the model for the combinations of pT (pT_1,pT_{2-3}) and CA15.3 (10, 22 and 31 U/ml), fixing age and ER at their median values (56 fmol/mg protein and 59 years) (Fig. 3). These levels of CA15.3 have been arbitrarily chosen on the basis of the following criteria: 22 U/ml and 31 U/ml represent the lowest and the highest values among the cut-off points most frequently reported in the literature; 10 U/ml was the approximate value from which the risk or relapse began to increase with increasing CA15.3 concentrations (see Fig. 2).

A maximum difference of 24% was obtained by comparing 5 years relapse-free probability of patients with pT_1 and CA15.3 = 10 U/ml (89%) with that of patients with pT_{2-3} and CA15.3 = 31 U/ml (65%). Separately for the pT groups, the maximum difference of 17% was

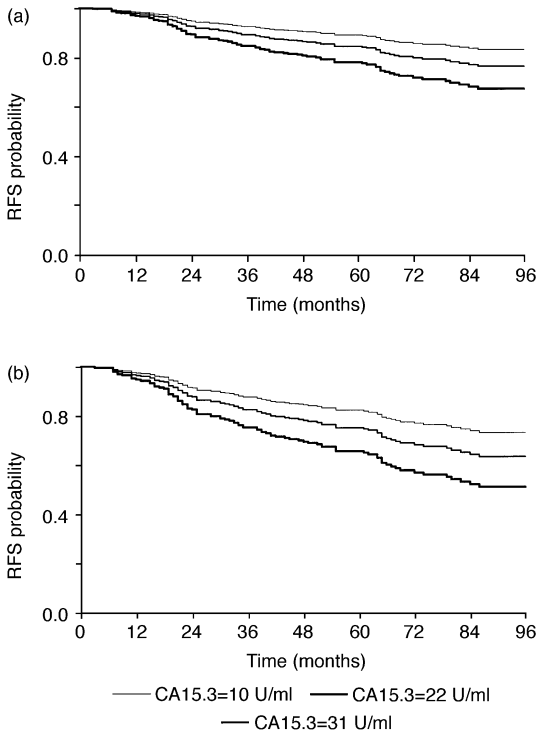

Fig. 3. Expected relapse-free survival (RFS) curves for different levels of CA15.3 in patients with, pT_1 (a) and pT_{2-3} tumours (b).

obtained for patients with pT_{2-3} tumours (Fig. 3b: 82% for CA15.3 = 10 U/ml versus 65% for CA15.3 = 31 U/ml), whereas, for patients with pT_1 tumours, a difference of 11% was observed (Fig. 3a: 89% for CA15.3 = 10 U/ml versus 78% for CA15.3 = 31 U/ml). However, a shrinkage of these differences could be expected in other case series.

We also evaluated the prognostic effect of CA15.3 dichotomised using the above-mentioned levels of 10, 22 and 31 U/ml as cut-off points. The contribution of the marker to the baseline model was not significant for the first two values: ($\chi^2 = 2.49$ d.f. = 1 $P = 0.114$) with $c = 0.651$ (corrected $c = 0.613$) using 10 U/ml and ($\chi^2 = 3.34$ d.f. = 1, $P = 0.068$) with $c = 0.653$ (corrected $c = 0.617$) using 22 U/ml. However, using as a cut-off point the value of 31 U/ml CA15.3 showed a significant contribution ($\chi^2 = 6.43$ d.f. = 1, $P = 0.011$) with $c = 0.654$ (corrected $c = 0.623$). However, the contribution to the model of CA15.3 as a continuous variable was greater. These findings supported the evidence that the relative hazard of relapse of patients within each class defined by cut-off values is heterogeneous. In Fig. 4 the estimated threshold effects are reported together with the continuous effects on CA15.3 adjusting to the median or modal values of continuous and categorical variables, respectively. Considering the range of values between 5 and 50.8 U/ml and adopting 22 U/ml as the cut-off reference value, an overestimate of the relative hazard for the lowest values and an underestimate for the highest is obtained when the threshold relationship is considered instead of the continuous one. In addition, adopting 31 U/ml as a cut-off and reference value, an overestimate of the relative hazard is obtained for all of the values in the considered range.

4. Discussion

The rapidly growing array of therapeutic agents addressed against molecular targets is sharply renewing the role of biomarkers for a comprehensive investigation of tumour phenotype and/or genotype [2]. However, the increasing awareness of women towards the importance of the early diagnosis of breast cancer is leading to a progressive reduction of the size of the detected tumour. This makes the choice of the panel of tumour markers to be measured more and more problematic, due to the decreasing amount of available tissue. New technologies, such as micro-array, are expected to overcome this problem in the near future. In the meantime, serum markers seem ideal candidates to tackle with the practical problem of reduced tissue availability. In addition, serum markers may provide information on the tumour phenotype at the relapse, when the collection of tissue specimen is usually not convenient. In spite of these putative roles, the association of preoperative marker levels with prognosis have been rarely evaluated. Horobin and colleagues [12] and Iaffaioli and colleagues [13] showed a weak prognostic role of CA15.3 levels in a small patient series without performing a multivariate analysis. Jaffaioli and colleagues [13] evaluated 60 patients with primary breast

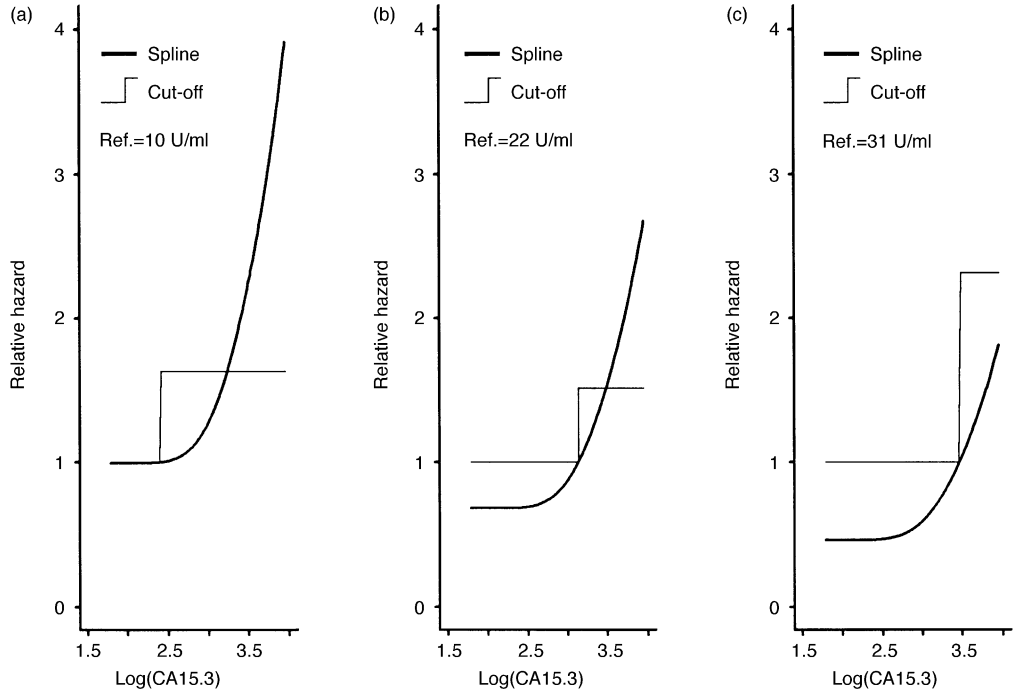

Fig. 4. Relative hazard of relapse as a function of CA15.3. Estimated continuous relationships are compared with those estimated from a threshold assumption. Reference values for the calculation of relative hazards are the cut-off points at 10, 22 and 31 U/ml.

cancer and dichotomised CA15.3 values using a cut-off point of 30 U/ml. They found a significant prognostic role of CA15.3 in node-positive patients, while the role of the antigen was not assessable in node-negative cases since no positive values of the antigen were found in this group.

More recently, Molina and colleagues [14] evaluated the relationship between CA15.3 and prognosis in a subset of 186 patients with primary breast cancer. They dichotomised CA15.3 values using a cut-off point of 35 U/ml; however, they did not specify how this cut-off was obtained. Using multivariate analysis, they did not find any significant association between CA15.3 positivity and prognosis. The authors did not evaluate node-positive and node-negative cases separately. Shering and colleagues [4] performed a study specifically aimed at evaluating the prognostic role of CA15.3 in primary breast cancer. They studied 368 patients with primary breast cancer, 184 with axillary metastasis and 184 without. They categorised the antigen values as positive or negative, according to a cut-off of 30.38 U/ml, obtained by the recursive partitioning technique. They found a significant association between CA15.3 positivity and prognosis in node-positive cases, but not in node-negative ones. The lack of prognostic significance of serum CA15.3 may be due, at least in part, to the categorisation of the antigen according to cut-off points. Moreover, their group of node-negative patients is not comparable with that evaluated in the present investigation. The majority of their patients (291/368) received adjuvant chemotherapy or hormoneal therapy. It should be assumed that adjuvant systemic treatments had been administered to approximately 58% of their node-negative patients. Therefore, their conclusions should be considered cautiously, since a predictive effect of CA15.3 on adjuvant therapies cannot be ruled out.

Practical reasons favoured in the past the adoption of cut-off based criteria for the evaluation of prognostic factors, namely, the chance to apply basic statistical techniques such as for example Kaplan and Meier curves and the use of indicator variables in regression models, giving a clear-cut presentation of the results. Problems related to the dichotomisation of biological variables have been usually underscored due to the consolidated habit of using cut-off points [30]. However, emerging evidences have shown that drawbacks are probably remarkable [15], the most relevant being the fact that prognostic relationships of biological variables are not expected to behave in a threshold manner. A continuous relationship between biological variables and prognosis has been proposed for p53 [16], ER and PgR [17], vascular endothelial growth factor (VEGF) [18], c-erbB-2 [19], pS2 and cathepsin D [20], and the cytosol levels of tissue polypeptide antigen (TPA) [21]. As far as serum CA15.3 is concerned, Shering and colleagues [4] reported that the antigen is significantly associated to both RFS and OS also when evaluated as a continuous variable. However, they did not report details on the statistical approach they used nor did they specify if the relationship occurred in either node-positive or node negative patients, or both. From the above considerations, it follows that variables should be initially analysed in their original scale of measurement, keeping the maximum information and avoiding gross biases in the estimates of the prognostic relationships. In a further phase, aiming to aid clinical decision-making, continuous variables could be divided into intervals. Suitable cut-off points should be provided on the ground of clinical and statistical considerations regarding the loss of information and bias.

In the present study, we studied the continuous relationship between CA15.3 and RFS, showing that prognosis tends to became increasingly unfavourable starting from a value of approximately 10 U/ml of CA15.3. If the antigen values are categorised as positive or negative according to a cut-off point, say 31 U/ml, the patient series is split into two groups that encompass 91% (below 31 U/ml) and 9% (over 31 U/ml) of the cases, respectively. However, cases with an estimated relative risk (reference value 5 U/ml minimum of CA15.3 concentrations) ranging from 1 to 2.16 are included in the CA15.3-negative group. Our data show that this subdivision is artifactual and may falsely reduce the real prognostic value of the marker, whereas the analysis of the variable in a continuous scale aims to preserve the prognostic information of CA15.3. The present study has been performed on samples taken at diagnosis and prior to any treatment from the widest series of patients with node-negative primary breast cancer not treated with adjuvant therapies reported so far. In this scenario, preoperative serum CA15.3 could provide relevant prognostic information, together with the most effective markers measured in tissues [3].

From these findings, we can draw the following conclusions:

1. CA15.3 is shown to be a prognostic marker in node-negative breast cancer;
2. the relationship of the marker with prognosis is continuous;
3. the risk of relapse increases progressively starting from a value of approximately 10 U/ml of CA15.3.
4. Dycotomic cut-off points aimed at an easier clinical application of the marker in decision-making could be obtained in an advanced phase of the statistical evaluation (i.e., after generation of the statistical model) being aware of the underlying continuous prognostic relationship.

On the basis of the present results, before considering CA15.3 for clinical decision-making, further studies

should be performed on other independent case series to confirm the observed prognostic relationships.

Acknowledgements

This work was partially supported by the Regione Veneto, the Association for the Application of Biotechnologies in Oncology (Gruppo ABO) and the Gruppo 3M Italia.

References

1. Goldhirsch A, Glick JH, Gelber RD, Senn HJ. Meeting highlights: international consensus panel on the treatment of primary breast cancer. *J Natl Cancer Inst* 1998, **90**, 1601–1608.
2. Gasparini G, Gion M. Molecular-targeted anticancer therapy. challenges related to study design and choice of proper endpoints. *Cancer J Sci Am* 2000, **6**, 117–131.
3. Gasparini G. Prognostic variables in node-negative and node positive breast cancer—editorial. *Breast Cancer Res Treat* 1998, **52**, 321–331.
4. Shering SG, Sherry F, Mcdermott EW, O'Higgins NJ, Duffy MJ. Preoperative CA 15-3 concentration predict outcome of patients with breast carcinoma. *Cancer* 1998, **83**, 2521–2527.
5. MacLean GD, Reddish MA, Longenecker BM. Prognostic significance of preimmunotherapy serum CA27.29 (MUC1) mucin level after active specific immunotherapy of metastatic adenocarcinoma patients. *J Immunother* 1997, **20**, 70–78.
6. Taylor-Papadimitriou J, Burchell J, Miles DW, Dalziel M. MUC1 and cancer. *Biochim Biophys Acta* 1999, **1455**, 301–313.
7. Agrawal B, Gendler SJ, Longenecker BM. The biological role of mucins in cellular interactions and immune regulation: prospects for cancer immunotherapy. *Mol Med Today* 1998, **4**, 397–403.
8. Wesseling J, van der Valk SW, Hilkens J. A mechanism for inhibition of E-cadherin-mediated cell-cell adhesion by the membrane-associated mucin episialin/MUC1. *Mol Biol Cell* 1996, **7**, 565–577.
9. Agrawal B, Krantz MJ, Reddish MA, Longenecker BM. Cancer-associated MUC1 mucin inhibits human T-cell proliferation, which is reversible by IL-2. *Nat Med* 1998, **4**, 43–49.
10. Gimmi CD, Morrison BW, Mainprice BA, *et al*. Breast cancer associated antigen, DF3/MUC1, induces apoptosis of activated human T cells. *Nat Med* 1996, **2**, 1369–1370.
11. Fung PYS, Longenecker BM. Specific immunosuppressive activity of epiglycanin, a mucin-like glycoprotein secreted by a murine mammary adenocarcinoma (TA3-HA). *Cancer Res* 1991, **51**, 1170–1176.
12. Horobin JM, Browing MC, McFarlane NP, *et al*. Potential use of tumour marker CA15–3 in the staging and prognosis of patients with breast cancer. *J R Coll Surg Edinburgh* 1991, **36**, 219–221.
13. Iaffaioli RV, Caponigro F, Esposito G, *et al*. Impact of preoperative CA 15–3 levels in operable breast cancer. Comparison with tissue polypeptide antigen (TPA) and carcinoembryonic antigen (CEA). *Int J Biol Markers* 1991, **6**, 21–24.
14. Molina R, Jo J, Filella X, *et al*. C-erb B-2 oncoprotein, CEA, and CA 15.3 in patients with breast cancer: prognostic value. *Breast Cancer Res Treat* 1998, **51**, 109–119.
15. Altman DG, Lausen B, Sauerbrei W, Schumacher M. Danger of using "optimal" cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst* 1994, **86**, 829–835.
16. Silvestrini R, Daidone MG, Benini E, *et al*. Validation of p53 accumulation as a predictor of distant metastasis at 10 years follow-up in 1400 node-negative breast cancer. *Clin Cancer Res* 1996, **2**, 2007–2013.
17. Hupperets PS, Volovics L, Schouten LJ, *et al*. The prognostic significance of steroid receptor activity in tumour tissue of patients with primary breast cancer. *Am J Oncol* 1997, **20**, 546–551.
18. Gasparini G, Toi M, Gion M, *et al*. Prognostic significance of vascular endothelial growth factor protein in node-negative breast carcinoma. *J Natl Cancer Inst* 1997, **89**, 139–147.
19. Dittadi R, Gion ?. More about prognostic importance of low c-erbB2 expression in breast cancer. *J Nat Cancer Inst* 2000, **92**, 1443–1444.
20. Dittadi R, Biganzoli E, Boracchi P, Salbe C, Mione R, Gatti C. Impact of steroid receptors, pS2 and cathepsin D on the outcome of N+ postmenopausal breast cancer patients treated with tamoxifen. *Int J Biol Markers* 1998, **13**, 30–41.
21. Gion M, Boracchi P, Dittadi R, *et al*. Quantitative measurement of soluble cytokeratin fragments in tissue cytosol of 599 node negative breast cancer patients: a prognostic marker possibly associated with apoptosis. *Breast Cancer Res Treat* 1999, **1602**, 1–11.
22. Biganzoli E, Boracchi P, Daidone MG, Gion M, Marubini E. Flexible modelling in survival analysis. Structuring biological complexity from the information provided by tumor markers. *Int J Biol Markers* 1998, **13**, 107–123.
23. Veronesi U, Salvadori B. Breast conservation is a safe method in patients with small cancer of the breast. Long-term results of three randomised trials on 1973 patients. *Eur J Cancer* 1995, **31**, 1574–1579.
24. EORTC Breast Cancer Cooperative Group. Revision of the standards for the assessment of hormone receptors in human breast cancer; report of the second EORTC Workshop. *Eur J Cancer* 1980, **16**, 1513–1515.
25. Lebart L, Morineau A, Piron M. *Statistique Exploratoire Multidimensionelle*. Paris, Dunod, 1995.
26. Grambsch P, Therneau T. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994, **81**, 515–526.
27. Harrel FE, Lee KL, Mark DB. Tutorial in biostatistics multivariable prognostic model: issues in developing models, evaluating assumption and adequacy, and measuring and reducing errors. *Stat Med* 1996, **15**, 361–387.
28. Neville AM, Bettelheim R, Gelber RD, *et al*. Factors predicting treatment responsiveness and prognosis in node-negative breast cancer. *J Clin Oncol* 1992, **10**, 696–705.
29. Davis BW, Gelber RD, Goldhirsch A, *et al*. Prognostic significance of tumor grade in clinical trials of adjuvant therapy for breast cancer with axillary lymph node metastasis. *Cancer* 1986, **58**, 2662–2670.
30. Altman DG, De Stavola BL, Love SB, Stepniewska KA. Review of survival analyses published in cancer journals. *Brit J Cancer* 1995, **72**, 518–551.