Biomarkers, 2009; 14(2): 130–136

informa
healthcare

RESEARCH ARTICLE

Serial determination of CEA and CA 15.3 in breast cancer follow-up: An assessment of their diagnostic accuracy for the detection of tumour recurrences

Luigi Mariani[1], Rosalba Miceli[1], Silvia Michilin[2], and Massimo Gion[3]

[1]*Unit of Medical Statistics and Biometry, Fondazione IRCCS Istituto Nazionale Tumori, Milan, Italy,* [2]*Associazione ABO per l'Applicazione delle Biotecnologie in Oncologia, Centro Regionale Indicatori Biochimici di Tumore, Consorzio Istituto Oncologico Veneto IRCCS, AULSS 12 – Venezia, Italy, and* [3]*Centro Regionale Indicatori Biochimici di Tumore, Consorzio Istituto Oncologico Veneto IRCCS, AULSS 12 – Venezia, Italy*

Abstract

We studied the diagnostic accuracy of carcinoembryonic antigen (CEA) and cancer antigen 15.3 (CA 15.3) in detecting breast cancer recurrence. Biomarker follow-up determinations, made over 900 patients, were related to local–regional or distant recurrence using statistical models for longitudinal data. The diagnostic accuracy was quantified in terms of sensitivity, specificity and Youden index. The biomarkers were poorly predictive of local–regional recurrence. As for distant recurrence, the best diagnostic accuracy was obtained considering the two biomarkers jointly and combining two positivity criteria: a value above the normal limit or a difference between two consecutive measurements greater than the critical difference for at least one biomarker. A third criterion, based on within-patient comparison between follow-up determinations and a baseline, failed to improve the above result. CEA and CA 15.3 might play a role in patient monitoring during follow-up for the search of distant recurrence.

Keywords: *Breast cancer; CEA; CA 15.3; sensitivity; specificity; decision criteria*

Introduction

Tumour biomarkers may have potentially useful applications in many clinical settings, such as tumour staging, treatment effectiveness monitoring and tumour recurrence detection during follow-up. The latest issue is currently of great interest, considering the availability of novel biologically targeted drugs the effectiveness of which might be optimal in the case of limited disease.

The correlation between increased circulating levels of carcinoembryonic antigen (CEA) or cancer antigen 15.3 (CA 15.3) and the presence of breast cancer is established (Molina et al. 1995, 2003, Nicolini et al. 1997a, 1997b, 2003a, Gion et al. 2002, Seregni et al. 2004, Park et al. 2008). However, the role of the above biomarkers in breast cancer follow-up has not been fully defined, and published guidelines do not yet recommend their routine assay (Molina et al. 2005, Harris et al. 2007). Moreover, CEA and CA15.3 have been evaluated using dichotomic criteria based on conventional normal limits in the majority of published studies, thus disregarding any disease-related dynamic pattern.

The aim of the present study was to evaluate the biomarker accuracy in detecting breast cancer recurrence and, in particular, if the introduction of 'dynamic criteria' may add information to dichotomic criteria.

Materials and methods

Samples

We used serial biomarker measurements taken during follow-up in a multicentre prospective trial involving

Address for Correspondence: Luigi Mariani, Unit of Medical Statistics and Biometry, Fondazione IRCCS Istituto Nazionale Tumori, Via G. Venezian 1, 20133 Milan, Italy. Tel: 0039-02-23903199. Fax: 0039-02-23902095; E-mail: luigi.mariani@istitutotumori.mi.it

(Received 10 October 2008; accepted 19 January 2009)

ISSN 1354-750X print/ISSN 1366-5804 online © 2009 Informa UK Ltd
DOI: 10.1080/13547500902770090

over 900 women with breast cancer, open to accrual between 1989 and 1994. The trial inclusion criteria were: age under 75 years; no residual disease after surgery for breast cancer; no previous or concomitant malignancy in other organ; postoperative radiotherapy and/or chemotherapy completed at least 6 months before entering the study; no clinical or laboratory evidence of benign disease of liver, kidney, pancreas or ovary (Gion et al. 1993, 1994, 1995, 2001).

Patient follow-up and therapeutic choices, either in the adjuvant setting or in the case of progressive disease, were performed according to the local clinical protocols.

Serum samples were collected before each clinical evaluation. The biomarkers were determined with the routine tools of each institution, using commercially available methods. A proper quality control programme, however, was carried out to monitor the accuracy and precision of the assays. Biomarker positivity, according to routine cut-off points, had to be verified in a second sample collected within a month; in case of confirmation, patients were assessed for local–regional or distant recurrence by means of clinical examination and imaging techniques. The study database was held by the coordinating centre (Centro Regionale Indicatori Biochimici di Tumore, Ospedale Civile – Azienda ULSS 12 Veneziana, Venezia, Italy).

Statistical methods

To assess the diagnostic accuracy of CEA and CA 15.3 we adopted the methodology proposed by DeLong et al. (1985), which allows investigation of the association between a test sequentially done over time and disease occurrence (in our study, local–regional or distant recurrence). The methodology is based on a discrete conditional logistic model which gives estimates of sensitivity, specificity, and positive and negative predictive value. Exact confidence interval calculations were not specifically addressed by DeLong et al. (1985); we thus applied normal approximation-based formulae for proportions to sensitivity and specificity.

Patients contributed biomarker data until disease onset or follow-up termination. DeLong et al. (1985) methodology application requires subdividing the observation period into intervals; we chose a 6-month window, that best reflected the temporal pattern of biomarker measurements and also allowed the investigation of a meaningful diagnostic anticipation. If repeated measurements occurred within intervals, we selected the one closest to the interval cut-off point. Follow-up was truncated at 10 years, beyond which data were sparse and no tumour recurrence was recorded.

We investigated three criteria for defining the biomarker positivity, two of which are dynamic as they take into account the biomarker variation over time within the same subject.

Cut-off

The standard laboratory upper normal limits of 5 ng ml^{-1} for CEA and 31 U ml^{-1} for CA 15.3 were used as cut-off points. Biomarker levels beyond such thresholds at a given time interval were considered positive. Although performed in different laboratories, biomarker measurements in the absence of disease recurrence were substantially stable over time and their distribution was comparable among different centres.

Critical difference

This dynamic approach was based on calculating the 'relative absolute difference' (RAD). Indicating with X_i the biomarker value in the ith interval and X_{i-1} that in the preceding interval, $RAD = [X_i - X_{i-1}]/[(X_i + X_{i-1})/2]$. The test is positive if RAD exceeds a 'critical difference' (CD), which is a function of the significance level α and the total intrasubject coefficient of variation, CV_T. The latter is usually calculated as $(CV_a^2 + CV_b^2)^{1/2}$ where CV_a is the analytical interassay coefficient of variation and CV_b the biological intrasubject variability (Fraser 2004). A more accurate approach for CV_T calculation proposed by Cohen et al. (2001) had a negligible effect on study results, and was therefore discarded in favour of the standard formula.

Three different levels of CD were investigated, henceforth named low, medium and high, respectively, equal to 30.8, 48.4 and 83.9% for CEA, and 20.3, 31.9 and 57.2% for CA 15.3. Values of CV_a and CV_b were obtained from quality control in the laboratory of the coordinating centre and the Westgard database (http://www.westgard.com/biodatabase1.htm), respectively, and plugged in for low and medium CD calculation; an estimate of CV_t obtained from the available data (Queraltó 2004) was used for high CD calculation. Significance levels were 5% one-sided, 1% two-sided and 5% one-sided for low, medium and high CD.

Change-point

A simple and general change-point identifier based on a Moving F statistic has been proposed by Riffenburgh and Cummins (2006). With this dynamic approach, a small number of data at the beginning of the series constitutes a baseline, and moving averages are calculated over groups of data taken from the remaining series. This generates moving F tests that, if significant, indicate that something has changed at some point in time with respect to the baseline. A point of change, once identified, is never reset. In our study, we used a baseline of four measurements for each biomarker and assumed that they were stable over time, as supported by exploratory analyses. Moving averages were calculated over

groups of three measurements and a two-sided 5% significance level was adopted for the F tests.

All criteria were assessed taking the biomarkers either individually or combined, namely by considering the test as positive at positivity of the single biomarker. We combined similarly the cut-off and the critical difference criteria, with which we also applied two different rules: 'ever positive' (the test may switch between positivity and negativity in each interval) and 'once positive, always positive' (once positive in a given interval, the test is considered positive in all the subsequent intervals).

Statistical analysis results were tabulated in terms of sensitivity, specificity, Youden index and two-sided Wald test *p*-value from the logistic model. The Youden index is a summary measure given by (sensitivity + specificity – 1); a value of the index equal to zero indicates the lack of diagnostic discrimination, a value of one perfect discrimination.

All the statistical analyses were carried out using the SAS™ software, version 6 (SAS Institute Inc., Cary, NC, USA). Further details on the above described application can be requested from the authors.

Results

From a total of 924 consecutive patients, 17 failed to provide sufficient information for the study. The remaining 907 patients, 509 (56%) node-negative and 398 (44%) node-positive, supplied a total of 10 844 CEA and CA 15.3 biomarker assays.

Twenty-seven local–regional recurrences and 132 distant recurrences were recorded; corresponding 10-year cumulative incidence estimates were 10% and 39%, respectively. These figures were remarkably affected by axillary nodes status. Among the 509 node-negative women, incidence estimates were 9% and 13% for local–regional and distant recurrence, while in the 398 node-positive women the figures were 22% and 55%, respectively.

Figure 1 shows the mean values of CEA and CA 15.3 according to recurrence status and follow-up year for patients without recurrence.

Table 1 shows the results obtained with the cut-off criterion according to the type of recurrence. For local–regional recurrence, statistical significance was never

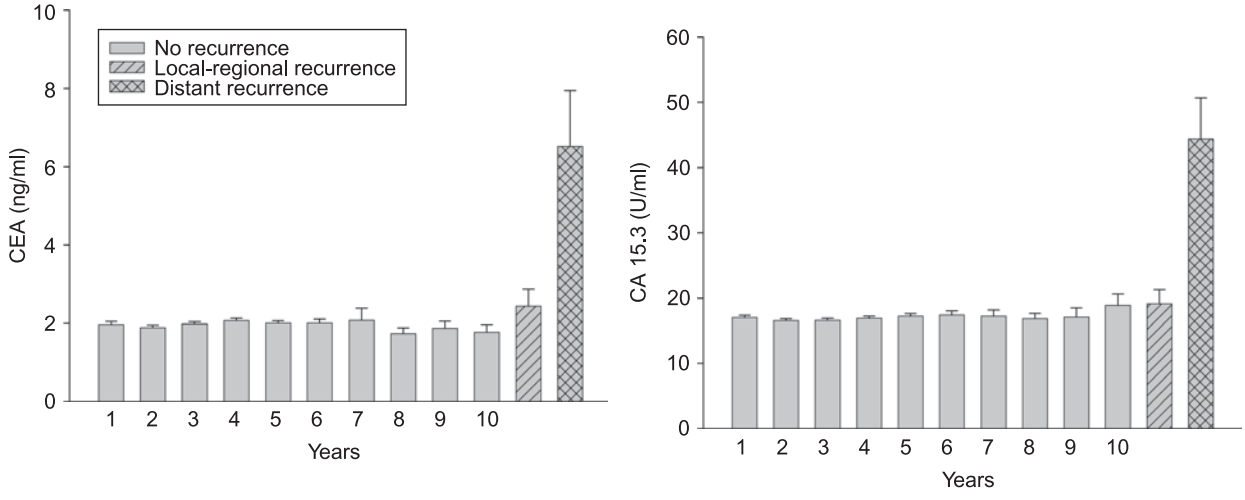

Figure 1. Bar charts showing the mean values of CEA (left panel) and CA 15.3 (right panel) according to recurrence status and follow-up time, for patients without recurrence. The error bars represent the standard error mean.

Table 1. Diagnostic accuracy of the cut-off criterion for detecting local–regional or distant tumour recurrence.

	Ever positive				Once positive, always positive			
	p-Value	Sensitivity (%)	Specificity (%)	Youden Index	*p*-Value	Sensitivity (%)	Specificity (%)	Youden Index
Local–regional recurrence								
CEA	0.060	12.5	95.8	0.08	0.280	12.5	93.2	0.06
CA15.3	0.106	12.5	95.1	0.08	0.104	20.8	89.7	0.10
CEA/CA15.3[a]	0.034	20.8	91.8	0.13	0.073	29.2	84.5	0.14
Distant recurrence								
CEA	<.001	30.8	96.3	0.27	<.001	34.6	93.7	0.28
CA15.3	<.001	40.2	95.8	0.36	<.001	44.9	90.2	0.35
CEA/CA15.3[a]	<.001	48.6	92.5	0.41	<.001	53.3	85.1	0.38

Cut-off values: 5 ng ml^{-1} for CEA and 31 U ml^{-1} for CA 15.3. *p*-Value at Wald's test in the DeLong et al. model (see Statistical methods section). [a]Joint test obtained by combining CEA and CA 15.3 information.

achieved and the Youden index figures were generally close to zero, denoting a poor diagnostic accuracy of both biomarkers. The findings did not improve when applying all the remaining criteria, and thus local–regional recurrence is henceforth disregarded in the presentation of study results.

Considering distant recurrence, *p*-values were significant with all the investigated criteria, indicating the existence of an association between the biomarkers and this kind of recurrence. With the cut-off criterion in particular (Table 1), the overall diagnostic accuracy, as reflected by the Youden index reaching a maximum of 0.41, was optimized by the 'ever positive' joint CEA/CA 15.3 test, for which sensitivity was 49% (95% confidence interval (CI) 40–58%) together with a still high 93% specificity (95% CI 92–94%). It is possible to observe that the results were generally better for CA 15.3 than CEA taken singly. The performance was generally slightly lower for the 'once-positive, always positive' test, in which the modest increase in sensitivity was traded off by a decline in specificity.

The performance of the critical difference criterion, possibly combined with the cut-off, is summarized in Table 2 considering the 'ever positive' rule. The low and medium critical differences generally yielded poor figures of specificity, while these figures were satisfactorily close to 95% for the high critical difference. In the latter case, the best overall performance was achieved again by joint CEA/CA 15.3 test and combining the cut-off and critical difference criteria (Youden index of 0.44). In this case, specificity was 86% (95% CI 85–87%) and sensitivity increased to an interesting 58% (95% CI 50–66%). The performance did not improve with the 'once-positive, always positive' rule (not shown).

Finally, Table 3 summarizes the results obtained with the change-point criterion, which failed to provide any substantial improvement over the cut-off and critical difference combined criteria. This analysis, however, was less powerful than those previously described, insofar a subset of 558 patients with sufficient data (four baseline and three follow-up measurements at least) had to be selected.

A search of whether or not axillary nodes status was able to affect sensitivity, specificity and the overall diagnostic accuracy failed to provide meaningful results with all criteria.

Discussion

Although established in other disease settings such as colon cancer (Renehan et al. 2002), the usefulness of tumour biomarkers in the follow-up of patients operated on for primary breast cancer is still debated. While some authors provided positive findings (Nicolini et al. 1997a, b, Sölétormos et al. 2004), the majority of clinical guidelines do not recommend their routine use, insofar neither decision criteria nor the optimal interval between serial determinations are adequately explored.

In the present study we compared the diagnostic accuracy of dynamic versus cut-off-based criteria for defining test positivity using data from a large prospective trial. Calculations of diagnostic accuracy measures must incorporate the time-varying amounts of information per individual, an issue that we addressed by adopting the model-based approach proposed by DeLong et al. (1985). A similar non-parametric methodology proposed by Emir et al. (1998) does not offer any intrinsic advantage. Among the three investigated criteria, the cut-off criterion is inherently cross-sectional, in that it is based on the distribution of biomarker values in the normal population, the other two criteria are dynamic

Table 2. Diagnostic accuracy of the critical difference or its combination with the cut-off criterion for detecting distant tumour recurrence ('ever positive' test).

	Critical difference				Cut-off/critical difference			
	p-Value	Sensitivity (%)	Specificity (%)	Youden Index	*p*-Value	Sensitivity (%)	Specificity (%)	Youden Index
Low critical difference								
CEA	<.001	38.3	79.7	0.18	<.001	50.5	77.3	0.28
CA15.3	<.001	62.6	74.6	0.37	<.001	68.2	72.8	0.41
CEA/CA15.3[a]	<.001	71.0	59.9	0.31	<.001	78.5	56.9	0.35
Medium critical difference								
CEA	<.001	29.0	88.4	0.17	<.001	43.9	85.7	0.30
CA15.3	<.001	53.3	85.0	0.38	<.001	61.7	82.5	0.44
CEA/CA15.3[a]	<.001	61.7	75.5	0.37	<.001	73.8	71.2	0.45
High critical difference								
CEA	<.001	17.8	96.4	0.14	<.001	35.5	93.1	0.29
CA15.3	<.001	30.8	95.8	0.27	<.001	47.7	92.3	0.40
CEA/CA15.3[a]	<.001	38.3	92.4	0.31	<.001	57.9	86.1	0.44

Cut-off values: 5 ng ml^{-1} for CEA and 31 U ml^{-1} for CA 15.3.Critical difference values (low, medium and high): 30.8, 48.4 and 83.9% for CEA, and 20.3, 31.9 and 57.2% for CA 15.3, respectively.*p*-Value at Wald's test in the DeLong et al model (see Statistical methods section).[a]Joint test obtained by combining CEA and CA 15.3 information.

Table 3. Diagnostic accuracy of the Change-point criterion for detecting distant tumour recurrence.

	p-Value	Sensitivity (%)	Specificity (%)	Youden Index
CEA	<.001	34.4	91.2	0.26
CA15.3	<.001	43.8	95.7	0.39
CEA/CA15.3[a]	<.001	51.6	88.3	0.40

p-Value at Wald's test in the DeLong et al. model (see Statistical methods section). [a]Joint test obtained by combining CEA and CA 15.3 information.

and take into account the within-subject variations in biomarker levels over time. Criteria of the latter kind are expected to be more effective when a biological quantity exhibits a high degree of individuality, and among them the change-point should behave better, being the analysis based on a multiplicity of data points.

By choosing different combinations of biomarkers, positivity criteria and related parameters (e.g. cut-off points, coefficients of variation, significance levels), huge numbers of decision rules are generated. One way to come over such a complexity is that of 'fine tuning' all possible rules to the available data, at the cost, however, of making the estimated diagnostic accuracy overly optimistic. We adopted a different strategy, namely a restricted search based on a predefined grid of parameter values and positivity criteria, coupled with an informal (descriptive) assessment of the diagnostic accuracy of each decision rule. While this strategy is not exhaustive, it nevertheless allowed the assessment of a wide spectrum of conditions in terms of sensitivity and specificity.

The results obtained with the above-described strategy can be summarized as follows.

First, CEA and CA 15.3 determination was a useful procedure for the diagnosis of distant recurrences, but not of local–regional recurrences, a finding that is in agreement with previous studies (Sölétormos et al. 1996, Nicolini et al. 2003a, b, Seregni et al. 2004, Laessig et al. 2007). While predictions were stronger with CA 15.3 than CEA, the best performance was achieved by combining information on the two biomarkers, namely by considering the test as positive whenever one of the two was above a given cut-off point. In particular, with the standard laboratory upper normal limits of 5 ng ml^{-1} for CEA and 31 U ml^{-1} for CA 15.3, the joint test was still able to provide high specificity (close to 95%) while increasing sensitivity to around 50%, starting from a 30–40% achieved by considering the two biomarkers individually.

This finding was coherent with the impressions emerging at the outset from exploratory data descriptions, as only some of the subjects with tumour recurrence showed an increase in both biomarkers and many others in only one of the two. Unfortunately, it was not possible to investigate whether or not such a behaviour was related to initial biomarker levels, insofar determinations prior to surgery were available in a small minority of patients (66 overall) and the values were always in the normal range.

Second, among the dynamic criteria, the change-point failed to substantially improve the above picture, but this result might be affected by the limitations in the available data, ending up in the choice of a short baseline series (only four measurements). A better picture emerged when combining the cut-off and critical difference criteria, using for the latter the highest value among those tested. In this case, specificity did not suffer (above or close to 90%), but sensitivity increased up to around 60% for the joint CEA/CA 15.3 test, which was indeed the best result achieved.

The above results were obtained by subdividing the follow-up time into 6-month intervals. Ideally, the choice of the interval width should be based on the estimation of the preclinical biomarker trajectory and the lead-time in particular (i.e. the mean time between the increase in biomarker levels and appearance of clinical signs of tumour recurrence). In our study this was hampered by the fact that the available measurements were too sparse in time for such a task. Nevertheless, our practical choice is justified by considering that the treatment of disseminated disease could not be considered a real advantage for a shorter interval and that performing regular semiannual measurements would easily fit into current follow-up programmes. Furthermore, to investigate biomarker effects extending beyond the interval, we also fitted models incorporating biomarker changes according to the 'once positive, always positive' rule, which is a way to allow for biomarker effects extending over time and is conceptually similar to the change-point criterion. These models, however, failed to provide interesting results, showing in particular a remarkable decline in specificity estimates, a finding suggesting that the actual lead-time tends to be relatively short.

Another possible concern is that the area under the ROC curve, as is generally believed, might be much more informative of the overall diagnostic accuracy than sensitivity and specificity. In fact, in a screening setting based on serial measurements, the target value for specificity needs to be very small, so as to avoid a too high risk of false-positives. As a consequence, only a small part of the area under the ROC curve would be of interest in the present context, thus losing much of its importance.

Likewise, while sensitivity and specificity are measures of intrinsic diagnostic accuracy, central to decision-making are the positive (PPV) and negative predictive values (NPV), whereby these measures reflect both the test performance and the prevalence of the event to be diagnosed or predicted. Actually, the DeLong et al. approach also allows the estimation of PPV and NPV,

based on a quantification of the cumulative risk of recurrence. Considering the rule based on the joint CEA/CA 15.3 test with cut-off and critical difference combination, yielding the best overall performance in our study, PPV and NPV were 73% and 76%, respectively. However, cumulative risk of recurrence in our series is probably higher than that expected in more recent years, owing to the fact that current patients tend to be diagnosed at an earlier stage and treated more effectively. Consequently, we did not systematically calculate predictive values, as their generalizability would be questionable.

Final comments consider the lack of biomarker assays standardization and of information on a possible confounder such as smoking. We note that a quality control program was actually run during the study, and that we verified *post hoc* that the biomarker distribution was comparable across centres, indicating that the heterogeneous laboratory methods were irrelevant. The role of smoking on biomarker levels is questionable (Fukuda et al. 1998), and stable smoking habits only affect interindividual variation, which does not influence the dynamic criteria.

The findings of the present study prompted us to design a two-arm randomized trial aimed at validating the present results and extend the investigation. In the 'intensified follow-up' (experimental) arm, CEA and CA 15.3 are to be measured every 3 months and additional radiological investigation, including positron-emission tomography scan, are performed in case of test positivity, the latter being defined in agreement with the best performing criteria in the present investigation. Among the many aspects to be considered, important extensions of the study will be the search for factors able to influence the biomarker diagnostic accuracy, and an assessment of the possible clinical benefit deriving from the anticipation of tumour recurrence diagnosis, which is possible only in a comparative trial.

Acknowledgements

The authors wish to acknowledge 'ABO Project 2010' and 'ASSOCIAZIONE ABO per l'applicazione delle biotecnologie in oncologia' for financial support.

***Declaration of interest*:** The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

Cohen A, Doveh E, Umansky T. 2001. How to assess the relative range of two consecutive laboratory measurements to monitor patients. *J Biopharm Stat* 11:177–191.

DeLong ER, Vernon WB, Bollinger RR. 1985. Sensitivity and specificity of a monitoring test. *Biometrics* 41:947–958.

Emir B, Wieand S, Su JQ, Cha S. 1998. Analysis of repeated markers used to predict progression of cancer. *Stat Med* 17:2563–2578.

Fraser CG. 2004. Inherent biological variation and reference values. *Clin Chem Lab Med* 42:758–764.

Fukuda I, Vamakado M, Kiyose H. 1998. Influence of smoking on serum carcinoembryonic antigen levels in subjects who underwent multiphasic health testing and services. *J Med Syst* 22:89–93.

Gion M, Cappelli G, Mione R, Vignati G, Fortunato A, Saracchini S, Biasioli R, Gulisano M. 1993. Variability of tumor markers in the follow-up of patients radically resected for breast cancer. *Tumour Biol* 14:325–333.

Gion M, Cappelli G, Mione R, Pistorello M, Meo S, Vignati G, Fortunato A, Saracchini S, Biasioli R, Giulisano M. 1994. Evaluation of critical differences of CEA and CA 15.3 levels in serial samples from patients operated for breast cancer. *Int J Biol Markers* 9:135–139.

Gion M, Barioli P, Mione R, Cappelli G, Vignati G, Fortunato A, Saracchini S, Biasioli R, Gulisano M. 1995. Tumor markers in breast cancer follow-up: a potentially useful parameter still awaiting definitive assessment. *Forza Operativa Nazionale sul Carcinoma Mammario (FONCaM). Ann Oncol* 6:31–35.

Gion M, Peloso L, Mione R, Vignati G, Fortunato A, Saracchini S, Biasioli R, Gulisano M, Cappelli G. 2001. Tumor markers in breast cancer monitoring should be scheduled according to initial stage and follow-up time: a prospective study on 859 patients. *Cancer J* 7:181–190.

Gion M, Boracchi P, Dittadi R, Biganzoli E, Peloso L, Mione R, Gatti C, Paccagnella A, Marubini E. 2002. Prognostic role of serum CA15.3 in 362 node-negative breast cancers. An old player for a new game. *Eur J Cancer* 38:1165–1166.

Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, Somerfield MR, Hayes DF, Bast RC. 2007. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol* 25:5287–5312.

Laessig D, Nagel D, Heinemann V, Untch M, Kahlert S, Bauerfeind I, Stieber P. 2007. Importance of CEA and CA 15-3 during disease progression in metastatic breast cancer patients. *Anticancer Res* 27:1963–1968.

Molina R, Zanón G, Filella X, Moreno F, Jo J, Daniels M, Latre ML, Gimenez N, Pahisa J, Velasco M, Ballesta AM. 1995. Use of serial carcinoembryonic antigen and CA 15.3 assays in detecting relapses in breast cancer patients. *Breast Cancer Res Treat* 36:41–48.

Molina R, Filella X, Alicarte J, Zanon G, Pahisa J, Munoz M, Farrus B, Ballesta AM. 2003. Prospective evaluation of CEA and CA 15.3 in patients with locoregional breast cancer. *Anticancer Res* 23:1035–1041.

Molina R, Barak V, Van Dalen A, Duffye MJ, Einarsson R, Gion M, Goike H, Lamerz R, Nap M, Sölétormos G, Stieber P. 2005. Tumor Markers in Breast Cancer – European Group on Tumor Markers Recommendations. *Tumor Biol* 26:281–293.

Nicolini A, Carpi A, Ferrari P, Anselmi L.1997a. Clinical use of tumor markers in the postoperative management of breast cancer patients: new concepts. *Int J Biol Markers* 12:181–186.

Nicolini A, Anselmi L, Michelassi C, Carpi A.1997b. Prolonged survival by 'early' salvage treatment of breast cancer patients: a retrospective 6-year study. *Br J Cancer* 76:1106–1111.

Nicolini A, Carpi A, Michelassi C, Spinelli C, Conte M, Miccoli P, Fini M, Giardino R.2003a. 'Tumour marker guided' salvage treatment prolongs survival of breast cancer patients: final report of a 7-year study. *Biomed Pharmacother* 57:452–459.

Nicolini A, Carpi A, Ferrari P, Pieri L.2003b. Utility of a serum tumour marker panel in the post-operative follow-up of breast cancer patients with equivocal conventional radiological examinations. *Tumour Biol* 24:275–280.

Park BW, Oh JW, Kim JH, Park SH, Kim KS, Kim JH, Lee KS. 2008. Preoperative CA 153 and CEA serum levels as predictor for breast cancer outcomes. *Ann Oncol* 19:675–681.

Queraltó JM. 2004. Intraindividual reference values. *Clin Chem Lab Med* 42:765–777.

Renehan AG, Egger M, Saunders MP, O'Dwyer ST. 2002. Impact on survival of intensive follow up after curative resection for colorectal cancer: systematic review and meta-analysis of randomised trials. *BMJ* 324:813–825.

Riffenburgh RH, Cummins KM. 2006. A simple and general change-point identifier. *Stat Med* 25:1067–1077.

Seregni E, Coli A, Mazzucca N. 2004. Circulating tumour markers in breast cancer. *Eur J Nucl Med Mol Imaging* 31:S15–S22.

Sölétormos G, Nielsen D, Schiøler V, Skovsgaard T, Dombernowsky P. 1996. Tumor markers cancer antigen 15.3, carcinoembryonic antigen, and tissue polypeptide antigen for monitoring metastatic breast cancer during first-line chemotherapy and follow-up. *Clin Chem* 42:564–575.

Sölétormos G, Nielsen D, Schiøler V, Mouridsen H, Dombernowsky P. 2004. Monitoring different stages of breast cancer using tumour markers CA 15-3, CEA and TPA. *Eur J Cancer* 40:481–486.