Pergamon

European Journal of Cancer Vol. 32A, No. 11, pp. 1949–1956, 1996

Printed in Great Britain
0959-8049/96 $15.00 + 0.00

PII: S0959-8049(96)00232-8

Review

An Improved Statistical Approach: Can it Clarify the Role of New Prognostic Factors for Breast Cancer?

J.W. Chapman,[1,2] D. Murray,[1,3] D.R. McCready,[1,2] W. Hanna,[1,2] H.J. Kahn,[1,2] H.L.A. Lickley,[1,2] M.E. Trudeau,[1,2] B.G. Mobbs,[1,2] C.A. Sawka,[1,4] E.B. Fish[1,2] and K.I. Pritchard[1,4]

[1]The University of Toronto, Toronto, Ontario, Canada; [2]Henrietta Banting Breast Centre, Women's College Hospital, 60 Grosvenor Street, 5th floor, Toronto, Ontario, Canada M5S 1B6; [3]Pathology Department, St. Michael's Hospital, 30 Bond Street, Toronto, Ontario, Canada M5B 1W8; and [4]Toronto-Sunnybrook Regional Cancer Centre, The Ontario Cancer Treatment and Research Foundation, 2075 Bayview Avenue, Toronto, Ontario, Canada M4N 3M5

Recently, there has been a proliferation of new biomarkers, some of which may lead to an improved prognostic index or may influence treatment selection. However, there are methodological and statistical issues that require attention in assessing the role and use of these prognostic factors. Between 1977 and 1986, 1097 primary breast cancer patients were accrued for multidisciplinary research at the Henrietta Banting Breast Centre, Women's College Hospital; follow-up to 1990 is complete for 96% of the patients. Data for these patients are used here to illustrate strategies: (1) for the comparison of results from diverse assessments of biomarkers; (2) for the improved comparability of inter-laboratory results; (3) for the examination of the results from monoclonal or polyclonal antibody assays for possible clinically relevant bimodality; (4) for good statistical resolution of overlapping distributions; (5) that involve the use of quantitative values for prognostic factors whenever possible; and (6) for improved multivariate analyses. Good data handling and analyses may enable more accurate and rapid assessment of new prognostic factors, thereby expediting and improving their clinical application.

Key words: breast cancer, prognostic factors, cell cycle modelling, statistical analysis

Eur J Cancer, Vol. 32A, No. 11, pp. 1949–1956, 1996

INTRODUCTION

THERE IS, currently, an explosion of research concerning human genetics and its attendant implications for human cancer. Recent advances in molecular biology, molecular genetics and computer technology have led to work on the association between changes in DNA and disease prognosis [1, 2]. A host of new potential prognostic factors will require investigation in the near future. Clarification as to which of these many new factors carry substantive implications and are therefore functional, and which prove trivial or merely represent epiphenomena will be extremely important.

In vitro and animal models can often examine, in a preliminary way, the role of these potential prognostic markers and their biological significance. Biological assessment with blood samples, and tumour sampling using fine-needle aspiration material, or tumour tissue obtained by a variety of surgical procedures, allow us to examine many of these factors and indicate those patients most likely to relapse or conversely to respond to treatment. This may provide information which is more germane. For example, a variety of sampling and analytical methodologies permit assessment and rapid identification and/or quantification of the following variables which may have important implications for prognosis or treatment selection in humans: (a) the proportion or number of cells with an abnormal amount of DNA [3, 4]; (b) the proportion or number of tumour cells with different amounts of DNA [3, 4]; (c) the proportions

Correspondence to J.W. Chapman.
Received 25 Jul. 1995; revised 29 Apr. 1996; accepted 6 Jun. 1996.

or numbers of tumour cells in the various phases of the cell cycle (G_0/G_1, S, G_2 + M) [3, 4]; (d) the proportion or number of tumour cells undergoing proliferation [4, 5]; and (e) the proportion or number of cells displaying the presence of unusual gene segments [5].

A particular biomarker may be measured with a variety of methodologies, but most frequently with a single technique in each laboratory. Important questions arise about comparisons of one group's investigations with results from others. In clinical practice, patient values from different laboratories may or may not be comparable with each other or to reports in the literature.

In the past few years, there has been increasing recognition that fundamental changes may be required in the handling of research data and that new statistical methodologies may need to be developed [6–8]. Discussions and approaches have centred around the need for inference regarding the role of new prognostic factors, and the applications of decision theory methodology in establishing relevant categories for clinical use. A brief delineation of these dual approaches will be incorporated here. The breast cancer database from the multidisciplinary Henrietta Banting Breast Centre (HBBC) [9] provided the data concerning prognostic factors in breast cancer that are the focus of this paper.

DESCRIPTION OF DATA

Between 1 January 1977 and 31 December 1986, 1097 primary breast cancer patients were accrued at Women's College Hospital (WCH). Follow-up to 1990 is complete for 96% of patients while 85% had complete data for traditional prognostic variables (e.g. number of positive nodes, tumour size, oestrogen receptor (ER) and progesterone receptor (PgR)). Archival pathological material was available for nearly all patients, and has been or is being used to assess new markers such as ploidy and %S-phase measured using flow cytometry. The data used to illustrate certain statistical approaches was obtained from patient subgroups accrued after 1986.

ANALYSIS OF DATA

The HBBC database at WCH was used in the following ways.

Comparison of results from different methods for measuring biomarkers

(i) The association of ER values with disease-free survival was considered for patients with their first primary beast cancer diagnosed in 1985 and 1986 at the HBBC [10]. The receptors were assessed using two different methods, the immunocytochemical assay (ERICA) and the dextran-coated charcoal method (ER-DCC). The model improvements for adding the factors ERICA and ER-DCC in a stepwise regression procedure are assessed with the log likelihood ratio criterion ($-2\log R \sim \chi^2_{(1)}$). (ii) The frequency distribution of ERICA values ($n = 100$, see [11]) was compared with the distribution of ER values ($n = 717$, see [12]) measured by enzyme immunoassay (ER-EIA). These results were obtained from patients diagnosed in 1989 and 1990. The ER-EIA data were plotted on a logarithmic scale because biochemical ER assay values have consistently been reported to have a log normal distribution.

Standardisation of laboratory assay results

The ER-DCC and PgR-DCC assay values for 678 patients with primary invasive breast cancer accrued between 1977 and 1986 were used to demonstrate statistical standardisation of laboratory assay results. Frequency histograms were plotted here for the assay values in standardised log units [13, 14].

Clinical cut-off points and bimodality with new monoclonal/polyclonal antibodies

The low point in a frequency histogram of 154 ERICA values for patients accrued in 1985–86 is compared with the cut-off point suggested from a multivariate investigation of the effects of ERICA on distant recurrence [10].

Resolution of overlapping distributions

There were 415 node negative primary breast cancer patients accrued at the HBBC between 1977 and 1986. The %S-phase was estimated for 361 patients with Verity MODFIT, for 336 patients with POLY, and for 341 patients with RFIT [15]. Frequency histograms of S-phase in standardised log units were made for each method, for both diploid and aneuploid tumours.

Use of quantitative values

The frequency histograms constructed for ER-DCC and PgR-DCC were examined for evidence about the underlying distributional shape. The results of these investigations have been reported elsewhere [14, 16].

Univariate/multivariate analyses

The group of 378 node positive patients accrued between 1977 and 1986 were used [16] to examine the underlying assumption of proportional hazards for the Cox model. We compared the regression models obtained with stepwise and all subset variable selection using the Cox and accelerated failure time models (exponential, Weibull, log logistic, log normal). Model improvement was assessed with the log likelihood ratio criterion ($-2\log R \sim \chi^2$, with the number of degrees of freedom for the number of variables added to the model). The support for a model-type was assessed by comparing the maximum likelihood across model-type.

ANALYSIS OUTCOMES

Comparison of results from different methods for measuring biomarkers

There are three common ranges for sample size (number of cells assessed from tumours/tissue), each with interpretative implications: (a) with image analysis or immunocytochemical methods, several hundred cells are typically assessed. A pathologist may ensure that the results are obtained from tumour cells, but single tumour sections may not yield representative results for a heterogeneous tumour; (b) with flow cytometry, at least 20 000 cells are assessed. The results are more likely to be representative for heterogeneous tumours. A review of the corresponding section by the pathologist can provide some assurance that there was sufficient tumour included in the sample analysed, as well as an estimate of the percent of tumour in the specimen;

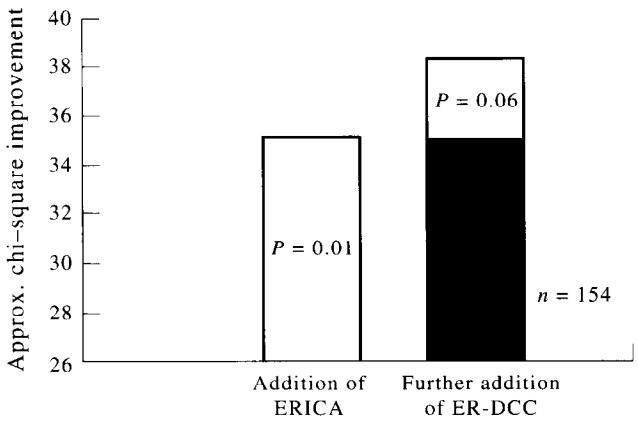

Figure 1. Stepwise chi-square improvements by adding ERICA, ER-DCC to the three variable model of T, Adjuvant therapy and Nodal status (modified from [10]).

and (c) with biochemical methods, millions of cells are usually assayed. Again a review of the contiguous section by a pathologist can confirm that the specimen contains tumour. Heterogeneous sections of tumour are likely to be included in the assay, but the actual amount of tumour in the specimen will be unknown.

Gundersen and associates [17, 18] have developed extensive stereological sampling techniques which provide an alternative approach of carrying out specific pathological assessments of tumour sections while assaying larger amounts of tumour. La Roye and Panzarella have demonstrated that these methods can yield clinically important improvements in the prediction of a patient's prognosis with breast cancer [19].

A particular prognostic factor may be measured by more than one methodology. Reports in the literature for a given prognostic factor should therefore be compared, first, within similar methodologies, and then across methodologies. Sometimes, a more direct comparison is possible for a single factor, assessed by different methodologies or using different sampling techniques, either within the same study or across studies with similar patients. We have used our ER data to illustrate this.

Figure 1 shows the stepwise model improvements for adding ER assessed by immunocytochemistry (ERICA) and by biochemistry (ER-DCC), to a three variable model of T, adjuvant therapy and nodal status. ERICA was consistently included in the best Cox stepwise regressions ($P < 0.01$), and there was weaker evidence of an association between DFS and the ER-DCC ($P = 0.06$) [10].

Figure 2a shows the ER values assessed by immunocytochemistry (ERICA) while Figure 2b shows the ER values obtained biochemically from enzyme immunoassay (ER-EIA). The histogram of ERICA values for the HBBC patients [11] indicated a separation of values analogous to the evidence of bimodality seen using the ER-EIA assay values [12].

In general, the persistence of an effect for a variable across methodologies could indicate the existence of an underlying biological phenomenon and imply validity for a prognostic factor, while inconsistent results may well be due solely to differences in tumour sampling.

Standardisation of laboratory assay results

In clinical practice, a physician will often have a single prognostic factor value available for a treatment decision, without knowledge about the particular laboratory procedures used to obtain that value or inter-laboratory comparability. An individual patient may be treated at several centres during the course of her disease, several laboratories may service a metropolitan region, and many laboratories may be involved in multicentre cooperative trials. It would be desirable to have a prognostic factor value which is comparable with values obtained elsewhere.

Historically, there have been problems with laboratory standards with regard to techniques, reagents, reaction temperatures, or processing times for various assays. These problems may make the inter-laboratory comparability of results poor, with shifts in assay values and differences in precision. A specific example of this for breast cancer would be ER and PgR. After decades of good quality control programmes [20, 21] for ER and PgR and the introduction of

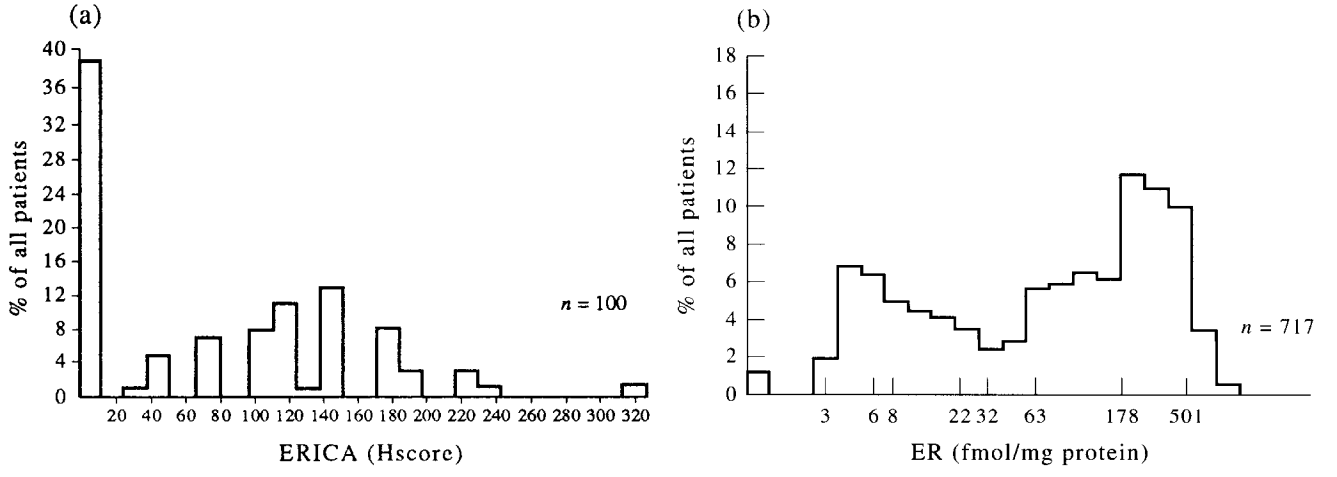

Figure 2. (a) The frequency distribution for ERICA Hscore = (the sum of the products of the number cells staining positive at intensities 0, 1 +, 2+, 3+ by intensity) (modified from [11]). (b) Frequency distribution of ER-EIA plotted on a logarithmic scale for tumours assessed by the laboratory serving the HBBC [12].

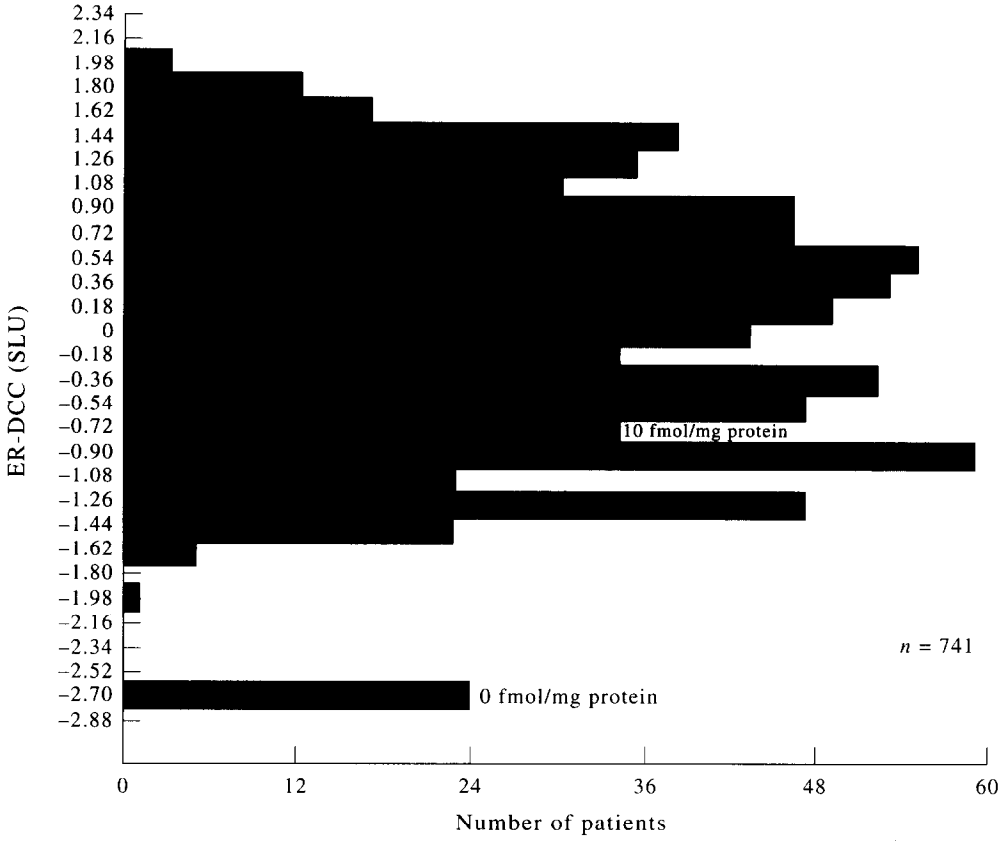

Figure 3. Frequency distribution of ER-DCC in standardised log units (SLU) obtained by statistical standardisation of log(ER + 0.5) to mean of 0, standard deviation of 1 (modified from [14]).

standard monoclonal antibody kits, there is still evidence that assay values in fmol/mg protein may not be well standardised across laboratories [12, 13, 22]. We have proposed statistical standardisation as an addition to the best current laboratory standardisation now achievable [13]. This was demonstrated for ER-DCC with Ontario Quality Control data from five laboratories and for the assay values for 184 primary breast cancer patients seen at the HBBC in 1985 and 1986 [13]. The standardisation process involved a logarithmic transformation to reduce asymmetry of the distribution of receptor assay values, stabilise the variance and obtain approximate normality. This was followed by statistical standardisation so that the distribution of receptor assay values for each laboratory or patient group had a mean of 0 and standard deviation of 1. Figures 3 and 4 show the ER-DCC and PgR-DCC assay values in standardised log units (SLU) for patients from the HBBC.

Statistically standardised units should provide greater comparability of assay values and thereby allow for more informed treatment decisions. The technique can be used for small laboratories or for small patient subgroups in clinical trials [13]. Standard normal tables could provide information about a patient's assay value relative to those for other breast cancer patients, in order to predict prognosis or response to therapy. As well, standardised units permit more comparable investigations of clinical cut-off point(s). The failure to standardise values before analysing the data for clinical investigations could lead to large differences in interpretation between studies, and when a number of laboratories are involved, could mask an effect.

An analogy might be drawn between this procedure and that used with INR* prothrombin (PT) times [23]. The rationale is similar in that standardisation procedures provide internationally comparable units for good patient management and research purposes.

The application of statistical standardisation would be relevant for any new biomarker, at least until the laboratory process and results are totally standardised in measurable laboratory units, which may never be completely possible. We used a logarithmic transformation for ER-DCC; with other variables, another Box-Cox transformation [24–26], such as a square root [25], might be appropriate to reduce the asymmetry, stabilise the variance and obtain approximate normality. As well, transformations have been found to improve the predictive effect of a prognostic factor [25]. In general, any distribution may be standardised to have a mean of 0 and standard deviation of 1 [27] even if approximate normality is not obtainable.

Clinical cut-off points and bimodality with new monoclonal/polyclonal antibodies

The use of prognostic information is simplified in clinical practice with the categorisation or dichotomisation of the values of a given factor according to patients' expected risk. The impetus to introduce new factors has led to the application of decision theory cut-off point analyses frequently

* INR = [patient PT/mean of reference PT]^{ISI} or log (INR) = ISI × log (PT patient/PT normal), where ISI is the International Sensitivity Index.

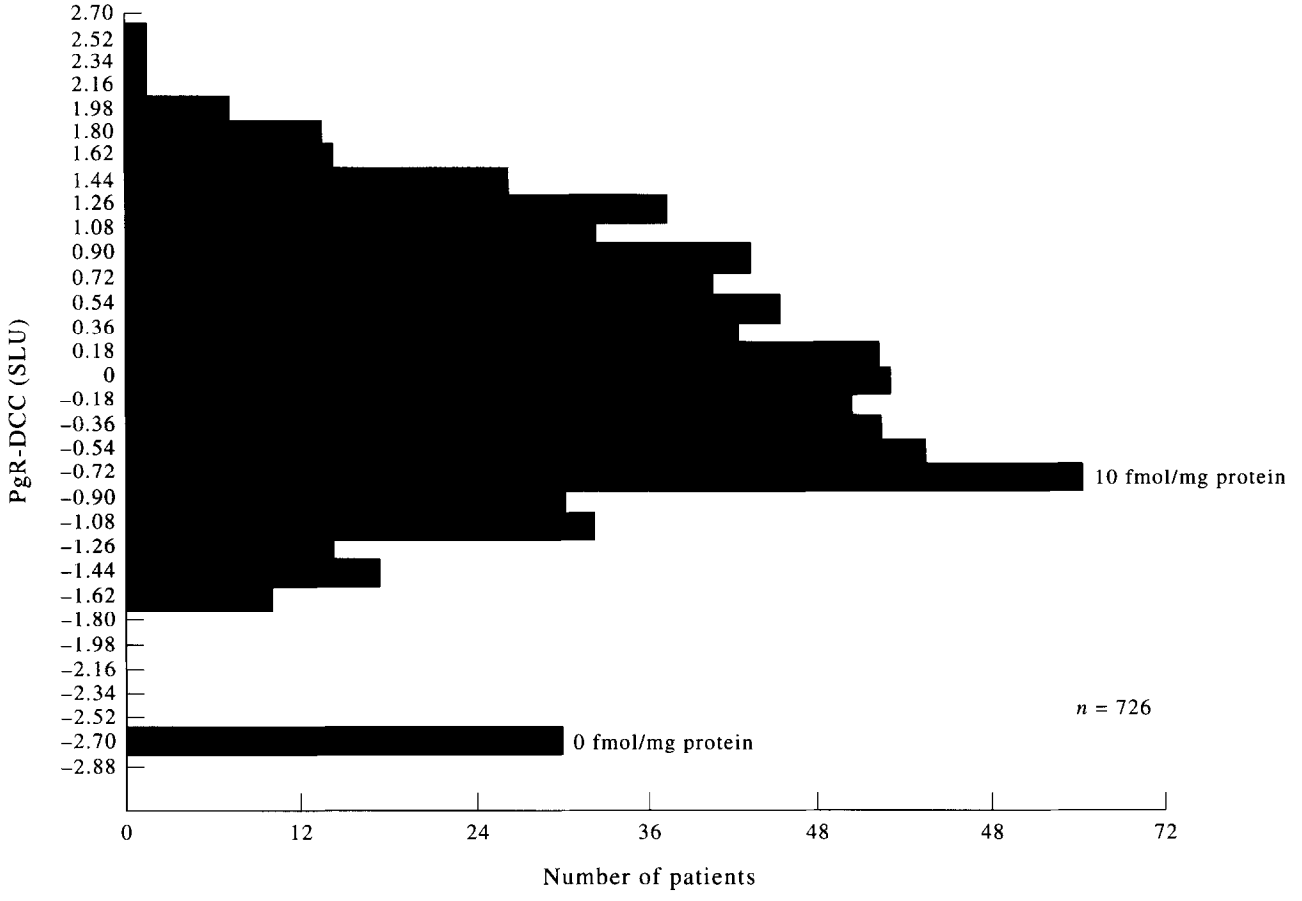

Figure 4. Frequency distribution of PgR-DCC in standardised log units (SLU) obtained by statistical standardisation of log(PgR + 0.5) to mean of 0, standard deviation of 1 (modified from [14]).

on continuous unimodal distributions. This often is carried out before there is consistent inferential evidence that the factor being investigated is associated with prognosis. This leads to an inaccurate assessment of the prognostic importance of the variable [7, 8].

We have found, however, that histograms of biomarkers measured with new monoclonal antibodies may show evidence of bimodal distributions; early follow-up with multivariate analysis has indicated that the central low point of the distribution may be biologically important. All of the Ontario laboratories performing biochemical assays for ER and PgR now use the Abbott enzyme immunoassay kits. With these, there has been evidence of bimodality for both ER-EIA and PgR-EIA, even after stratification by menopausal status [12].

There is similar evidence of bimodality for patient results using the immunocytochemical assay, ERICA, which shares one of the monoclonal antibodies used for ER-EIA [11]. The bimodal low point observed for ERICA at 5% cells staining positive, as shown in Figure 5, is at the same point as that observed in a multivariate investigation of the effects of ERICA cut-off point on disease-free survival [10], and this provides support for the concept that the bimodal low point observed for ERICA may represent a natural clinical cut-off point with respect to prognosis. The approach of investigating cut-off points after determining the important prognostic variables with multivariate analyses is supported by Altman and colleagues [7].

As mentioned above, any distribution may be statistically standardised [27]. The standardisation process will not eliminate the bimodal (or multimodal) low point(s), but the location of the low point(s) would be more comparable between laboratories or study groups when expressed in standardised units.

Resolution of overlapping distributions

The combination of new biological assessments with stains or fluorescence, along with sophisticated computer software systems, provides great potential for the current rapid quantitation of results for thousands of cells. The ability to scan simultaneously and measure several factors at once, or the same factor in several types of cells, may result in the need to achieve resolution of overlapping distributions in order to obtain meaningful estimates for a new biomarker [3–5]. An example of such an issue occurs in the estimate of the percent of cells in S-phase as measured by flow cytometry.

Frequency histograms for the amount of DNA in cells overlap for the cell cycle phases, G_0/G_1, S and G_2 + M. Gaussian distributions are usually fitted for G_0/G_1 and G_2 + M, and there are three common choices for S-phase [a broadened trapezoid (Verity MODFIT), rectangle, or polynomial]. A laboratory would generally use only one type of resolution. A consensus conference recommended the use of Verity MODFIT, and stratification of S-phase results by those for diploid and aneuploid tumours, followed by the

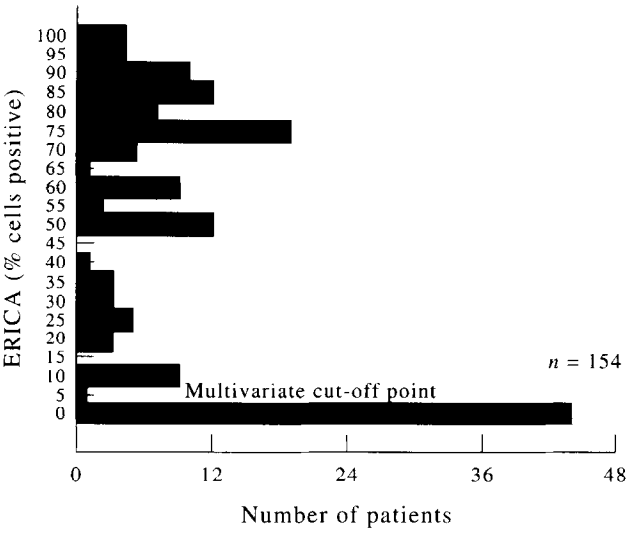

Figure 5. Frequency distribution of ERICA measured as % cells staining positive for ER. The multivariate cut-off point indicated was determined in an investigation of the effects of tumour size (T), number of lymph node metastases, ER-DCC, PgR-DCC, ERICA, and Adjuvant treatment on disease-free survival (modified from [10]).

use and reporting of tertile results (low, medium, high) [28].

We used statistical standardisation for S-phase so that the effects of S-phase cut-off points for the above three S-phase model choices could be investigated clinically [15]. Figure 6 shows frequency histograms of S-phase obtained by using Verity MODFIT, POLY and RFIT. We obtained similar clinical inferences in multivariate investigations for each of these S-phase model-types after statistical standardisation [15].

Use of quantitative values

ER and PgR are often reported as dichotomous (negative, positive) variables, but the frequency distributions of ER-DCC and PgR-DCC (Figures 3 and 4) are more suggestive that the receptor variables are continuous. We have recently found evidence supporting a clinical role for ER and PgR as continuous variables [14, 16], which confirms earlier results by others [29]. In a study of 378 node positive primary breast cancer patients [16], we found that better multivariate modelling was achieved when we used ER, PgR, tumour size and number of positive nodes as quantitative prognostic factors. The use of quantitative variables, rather than variables categorised in subgroups (e.g. tumour size in cms as opposed to ≤ 2, 2–5, > 5 cms), led to a significant overall model improvement. The magnitude of the improvement by using quantitative variables was similar to the effect of adding several new prognostic factors to a model in which the variables were divided into subgroups. The use of continuous variables whenever possible is supported by Altman and colleagues [7].

Univariate/multivariate analyses

A newly discovered biomarker may significantly improve our ability to predict prognosis or response to a therapeutic regimen, while another may simply be highly correlated with other important new factors. Furthermore, new factors

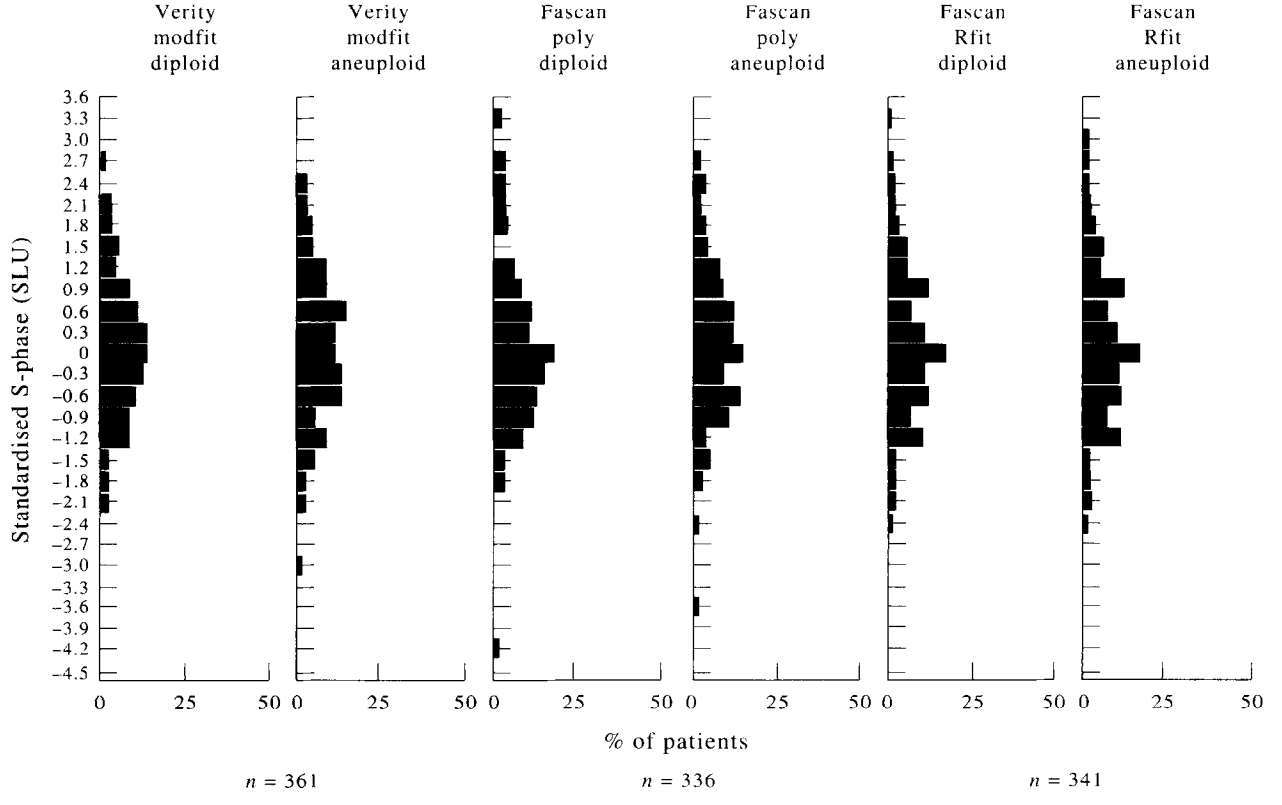

Figure 6. SLU is standardised log unit obtained by statistical standardisation of log(%S-phase + 0.5) to a mean of 0, standard deviation of 1 for diploid and aneuploid tumours, for Verity MODFIT, POLY and RFIT.

may be highly correlated with traditional factors such as nodal status, tumour size, ER and PgR, and age/menopausal status. As previously mentioned, the use of univariate analyses may bias results by inflating the relative importance of a factor [7, 8], and preventing the full recognition of the importance of other factors [7]. As well, univariate analyses usually involve the categorisation of a variable into subgroups rather than the use of quantitative values which leads to a loss of efficiency [7]. Multivariate investigations are recommended to determine the prognostic importance of multiple variables [7]. However, the standard use of a single Cox stepwise multivariate model to summarise a study may not be appropriate, especially if the traditional factors have been handled in a grouped manner, e.g. tumour size (T1, T2, T3, T4), grouped number of positive nodes (0, 1–3, ≥ 4), ER(+/–), PgR(+/–), age ($< 50, \geq 50$) rather than quantified as outlined above [16].

Several issues need to be considered in the selection of an appropriate type of multivariate method. The standard multivariate analysis in the last several decades has been a stepwise regression with a Cox proportional hazards model. The assumption of proportional hazards is not routinely checked; investigators rely on the general robustness of the non-parametric Cox model. Issues which should be considered for any multivariate analysis are: (a) the appropriateness of underlying model type assumptions; (b) requirements for categorisation of quantitative variables; (c) sample size; and (d) model building methodology (e.g. stepwise versus all subsets).

Alternative approaches to the Cox model include neural networks [26], recursive partitioning and amalgamation [30], general additive models [31], and accelerated failure time models (specifically, Exponential, Weibull, log logistic, log normal) [16]. (a) Any of these model types might be considered if there is evidence against the assumption of proportional hazards. (b) Recursive partitioning requires that quantitative variables be categorised into subgroups. While it is possible to investigate the effects of categorisation, it does not seem feasible to explore all possible combinations resulting from different variable categories. The neural networks, Cox, general additive models, and accelerated failure time models can all use quantitative variables directly. (c) Both neural networks and recursive partitioning will require larger sample sizes. For neural networks, the data are subdivided, before analyses, into two or three groups, one or two training sets and a validation set; sufficient sample sizes would be required for each subgroup. For recursive partitioning, the partitioning of the data into subgroups at each step may quickly reduce the power to detect variable effects with even moderate sample sizes. A neural network or recursive partitioning and amalgamation would seem appropriate only with large sample sizes in which there is sufficient power to detect differences in 2 or 3 data subgroups in the former instance, or after 4 or 5 nodes/branchings in the latter. (d) Stepwise model building does not recognise similarly important groups of variables, and this may result in inconsistencies as to which variables are reported to be important. Current commercial software facilitates 'all subset' investigations, which should be performed more routinely [16].

In a study of patients with node positive breast cancer carried out using the HBBC database [16], a check for the appropriateness of the Cox model and an 'all subset' variable selection strategy resulted in better multivariate modelling. The specific Pettit-type models [32] considered were the Cox, Exponential, Weibull, log logistic and log normal. The data strongly supported the log normal model over the examined alternatives. The model improvement obtained by using the log normal model, instead of the Cox was, again, of a magnitude such that several new prognostic factors might have been required in order to yield a similar improvement to that achieved by using the better model-type. It is not desirable to add new variables if they do not improve the overall modelling beyond that of the traditional variables assessed in a more optimal manner.

Our node positive study also provided a rationale for more routine use of 'all subset' regression analyses, since we obtained similarly important models with the inclusion of ER, the combined receptors (ER + / –, PgR + /–), and PgR. This parallels other reports on the importance of hormone receptors in different studies, even though we obtained this information from a single study. The reporting of a single multivariate model would seem appropriate only if it is substantially better than others. We have found as well that 'all subset' analyses were useful in identifying variables consistently associated with recurrence for node negative breast cancer [33].

CONCLUSION

Emerging technology may provide data for good modelling of disease progression, as the ability to detect early functional DNA changes becomes better established. This paper has outlined some issues which should be considered, in order to improve inferences with regard to these new biomarkers.

The increased diversification of methods for assessing a single factor, a frequent lack of inter-laboratory comparability for a single procedure, the use of data from a mixture of methods or different laboratories, and a premature dichotomisation of patient values for clinical use can all impede progress in patient care. Meanwhile, an approach, analogous to that adopted for prothombin times, of using statistically standardised units until there is good inter-laboratory comparability could shorten the assessment time for a new factor and hasten the adoption of those factors which are clinically beneficial. It is important in a research context that there be routinely more extensive examination of the data and the underlying assumptions being used for the analyses. Good data handling should facilitate rapid and accurate assessment of the role of new prognostic factors. While the examples and focus of this paper is on breast cancer, the concepts discussed have broad general applicability.

1. McGuire WL, Tandon AK, Allred DC, Chamness GC, Clark GM. How to use prognostic factors in axillary node-negative breast cancer patients. *J Natl Cancer Inst* 1990, **82**, 1006–1012.
2. Kallioniemi O-P, Isola J, Kallioniemi A, *et al.* Genetic events underlying breast cancer progression analyzed by comparative genomic hybridization. *Proc Am Assoc Cancer Res* 1994, **35**, 250 (Abstr. 1491).
3. Techniques in Diagnostic Pathology: Diagnostic Flow Cytometry. Edited by Coon JS, Weinstein RS. Williams and Wilkins, Baltimore, 1991, 115–146.

4. Viehl P. Guidelines to Clinical Aspiration Biology: Flow Cytometry. Igaku-Shoin, New York, 1991, 21–57.
5. Leslies KO, Howard P. Oncogenes and Antioncogenes in Human Breast Cancer. Pathology Annual 1992, Part 1. Edited by Rosen PP, Fechner RE. Appleton and Lange Norwalk, Connecticut, 1992, 321–342.
6. Clark GM. Integrating prognostic factors. *Breast Cancer Res Treat* 1992, **22**, 187–191.
7. Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using optimal cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst* 1994, **86**, 829–835.
8. Hilsenbeck SG, Clark GM, McGuire WL. Why do so many prognostic factors fail to pan out? *Breast Cancer Res Treat* 1992, **22**, 197–206.
9. Sawka CA, Pritchard KI, Lickley HLA, *et al.* The Henrietta Banting Breast Centre Database: a model for clinical research utilizing a hospital-based inception cohort. *J Clin Epidemiol* 1995, **48**, 779–786.
10. Hanna W, McCready DR, Chapman JW, *et al.* The predictive value of ERICA in breast cancer recurrence: a univariate and multivariate analysis. *Mod Pathol* 1993, **6**, 748–754.
11. Hanna WM, Kahn HJ, Chapman JW. The correlation of Ki67 growth factor and ERICA in breast cancer. *Mod Pathol* 1992, **5**, 220–223.
12. Mobbs BG, Chapman JW, Sutherland DJA, *et al.* Evidence for bimodal distribution of breast carcinoma ER and PgR values quantitated by enzyme immunoassay. *Eur J Cancer* 1993, **29A**, 1293–1297.
13. Chapman JW, Mobbs BG, Hanna WM, *et al.* The standardization of estrogen receptors. *J Steroid Biochem Mol Biol* 1993, **45**, 367–373.
14. Chapman JW, Mobbs BG, McCready DR, *et al.* Estrogen and progesterone receptor assay cut-points for primary breast cancer. *J Steroid Biochem Mol Biol*, in press.
15. Chapman JW, Murray D, Trudeau ME, *et al.* Optimal DNA index in a flow cytometry investigation of node negative breast cancer. *Proc Am Assoc Cancer Res* 1995, **36**, 210 (Abstr. 1249).
16. Chapman JW, Trudeau ME, Pritchard KI, *et al.* A comparison of all subset cox and accelerated failure time models with Cox step-wise regression for node positive breast cancer. *Breast Cancer Res Treat* 1992, **22**, 263–272.
17. Gundersen HJG, Bendtsen TF, Korbo L, *et al.* Some new, simple and efficient stereological methods and their use in pathological research and diagnosis. *APMIS* 1989, **96**, 379–394.
18. Gundersen HJG, Bagger P, Bendtsen TF, *et al.* The new stereological tools: disector, fractionator, nucleator and point sampled intercepts and their use in pathological research and diagnosis. *APMIS* 1988, **96**, 857–881.
19. LaRoye GJ, Panzarella T. Two techniques for measuring invasion in solid tumors. Evaluated in a retrospective study of 73 cases of breast carcinoma with 10-year follow-up. *APMIS* 1994, **102**, 103–111.
20. Cohen JL, Raam S, Gelman R. A blinded study of inter- and intra-laboratory variations in the performance of estrogen receptor (ER) assay. In Sarfaty GJ, Nash AR, Keightley DD. *Estrogen Receptor Assays in Breast Cancer, Laboratory Discrepancies and Quality Assurance.* New York, Masson Publishing, 1981, 43–56.
21. Ryan ED, Clark AF, Mobbs BG, Ooi TC, Sutherland DJA, Tustanoff ER. Inter-laboratory quality control of estrogen and progesterone receptor assays in breast cancer tissue using lyophilised cytosols. *Clin Biochem* 1985, **18**, 20–26.
22. Cren H, Lechevrel C, Roussel G, Goussard J. Evolution of immunoreactivity of monoclonal antibodies H222 and/or D547 used in the detection of breast cancer estrogen receptors. Varying reactivity of receptor isoforms. *J Steroid Biochem Mol Biol* 1991, **39**, 519–527.
23. International Normalized Ratio: Laboratory Monitoring of Oral Anticoagulants. Organon Teknika Product Information, 1994.
24. Box GEP, Cox DR. Analysis of Transformations. *J R Stat Soc, Series B* 1964, **26**, 211–252.
25. Muss HB, Thor AD, Berry DA, *et al.* c-erbB-2 expression and response to adjuvant therapy in women with node-positive early breast cancer. *N Engl J Med* 1994, **330**, 1260–1266.
26. Ravdin PM, Clark GM. A practical application of neural network analysis for predicting outcome of individual breast cancer patients. *Breast Cancer Res Treat* 1992, **22**, 285–293.
27. Kendall MG, Stuart A. *The Advanced Theory of Statistics.* Volume I, 3rd edition. Hafner Publishing Company, New York, 1969, 48.
28. Hedley DW, Clark GM, Cornelisse CJ, *et al.* Consensus review of the clinical utility of DNA cytometry in carcinoma of the breast. *Cytometry* 1993, **14**, 482–485.
29. Godolphin W, Elwood JM, Spinelli JJ. Estrogen receptor quantitation and staging as complementary prognostic indicators in breast cancer. A study of 583 patients. *Int J Cancer* 1981, **28**, 677–683.
30. Albain KS, Green S, LeBlanc M, Rivkin S, O'Sullivan J, Osborne CK. Proportional hazards and recursive partitioning and amalgamation analyses of the Southwest Oncology Group node-positive adjuvant CMFVP breast cancer database: a pilot study. *Breast Cancer Res Treat* 1992, **22**, 273–284.
31. Hastie T, Sleeper L, Tibshirani R. Flexible covariate effects in the proportional hazards model. *Breast Cancer Res Treat* 1992, **22**, 241–250.
32. Ciampi A, Chapman J, Hogg SA, Thiffault J. GENCOV: a Fortran program that generates randomly censored survival data with covariates. *Comp Meth Prog Biomed* 1989, **29**, 51–57.
33. Pritchard KI, Trudeau ME, Chapman JW, *et al.* Prognostic variables in node-negative breast cancer: an all subset analysis. *Proc Am Soc Clin Oncol* 1993, **12**, 68 (Abstr. 73).

Acknowledgements—This work was supported by the Women's College Hospital Research Fund and the Ontario Ministry of Health. We would like to thank Dr Denis MacDonald for advice regarding INR and for reviewing the manuscript.