Eur J Cancer, Vol. 27, No. 6, pp. 711–715, 1991.
Printed in Great Britain

0277–5379/91 $3.00 + 0.00

Specialist Interest Articles

A Multicentre Study into the Reliability of Steroid Receptor Immunocytochemical Assay Quantification

R.A. McClelland, D. Wilson, R. Leake, P. Finlay and R. I. Nicholson on behalf of the British Quality Control Group

Qualitative and semiquantitative assessments of oestrogen receptor and progesterone receptor positivity determined on previously immunocytochemically stained slides were performed by eight independent assessors. Concordance beween assessments of steroid receptor status was good (24/25, 96%). Interassessor variations in estimates of positive immunostaining levels were high, varying by between 10 and 75% for individual slides. In 2 cases estimates for the same section ranged between 15% nuclei positive and 90% nuclei positive. Wide variations were also recorded for slides stained for progesterone receptors. Results using an assessment procedure combining staining intensity and percentage positivity estimates were also subject to marked discordance. A computerised image analysis system, also used to assess slides gave results similar to the mean manually determined percentage positivity values. It is suggested that quality control of steroid receptor immunocytochemical quantification be considered and that automated image analysis may represent an accurate and valid means of achieving this.
Eur J Cancer, Vol. 27, No. 6, pp. 711–715, 1991

INTRODUCTION

THE VALUE of steroid receptor assays on breast cancer tissue, and in particular those for oestradiol receptors, in predicting prognosis and in selection of candidates for hormonal therapies is well established [1–5]. The presence of oestrogen receptors (ER), for example, is very closely associated with the successful outcome of tamoxifen [6, 7] or aminoglutethimide [8] therapies in advanced breast cancer. Furthermore, increased rates of response have been noted in association with higher quantities of receptor, as assessed by biochemical [9] and semiquantitative immunological methodologies [10]. The accurate quantification of these receptor levels may thus be deemed clinically more valuable than the qualitative status only determinations, but needs to be strictly comparable between laboratories. Immunocytochemical assay (ICA) procedures [11–13] can be applied to small tissue samples and enable the observation of immunolocalised receptor *in situ* at the cellular level. Problems, however, exist in the quantification of receptor levels within sections, and currently no accepted standardised procedure for this is in use.

This paper describes a multicentre collaborative study between eight independent assessors from seven British laboratories, all with experience in steroid receptor ICA staining and its interpretation. It reports the current level of accuracy obtained in the quantification of ER and progesterone receptors (PR) in stained breast sections. Selected immunostained specimens with a wide range of immunoreactivity were sent to each centre, in turn, for their assessment. Results were returned, collated and compared by a central laboratory (Breast Cancer Unit, Tenovus Institute).

Computer-assisted image analysis using software specifically created for the quantification of these assays through the comparative analysis of the optical properties of diaminobenzidene tetrachloride (DAB) positive stained nuclei and methyl green negative cell nuclei was available through the use of a Cell Analysis Systems (CAS) 200 analyser (CAS Lombard, Illinois) [14, 15]. Results from this were used for further comparison.

The study examines existing methods of immunocytochemical quantification and discusses their validity. It analyses the degree of interassessor variability by comparing results from a number of assessors using standardised specimens and procedures. The significance of the results obtained is discussed and recommendations made for the future study of the degree to which immunocytochemical quantification can be a reliable and valuable practice.

MATERIALS AND METHODS

Specimens

13 primary breast cancer specimens were selected, upon which oestrogen receptor content had previously been determined by ER-enzyme immunoassay using a monoclonal antibody based kit (ER-EIA, Abbott Diagnostics, Chicago). Of these 13, all demonstrated immunostaining using the ER-ICA monoclonal kit (Abbott Diagnostics, Chicago) [5, 10]. PR-ICA staining was performed on 12 of these samples by substituting anti-PR

Correspondence to R.A. McClelland.
R.A. McClelland, P. Finlay and R.I. Nicholson are at the Breast Cancer Laboratory and D. Wilson is at the Department of Biometry, Tenovus Institute for Cancer Research, Heath Park, Cardiff, Wales; and R. Leake is at the Department of Biochemistry, University of Glasgow, Glasgow, Scotland, U.K.
Revised 11 Mar. 1991; accepted 18 Mar. 1991.

antibody (68 kD) [12] at 10 μg/ml into the ER-ICA system. Positive staining indicative of the presence of PR was apparent in 9 of the 12 specimens. Methyl green was used as a counterstain in all the above to reveal receptor negative cells [16]. Haematoxylin, the recommended counterstain for these assays, was unsuitable for image analysis. In all cases tissue preservation and staining was good. Haematoxylin-eosin stained histological sections were available to accompany 7 of these.

Assessment procedures

Manual. Participants were requested to examine the slides and record their interpretations of the ER and PR staining of the tumour cell components of the slides by two separate means on standardised forms: (a) an estimate of the percentages of tumour cells failing to express immunoreactivity, i.e. the antibody negative cells and (b) an estimate of the percentage of tumour cells which were deemed weakly positive (i), moderately positive (ii) and strongly positive (iii). This enabled the calculation of an "*H* score" [17, 18] for each tumour by means of the formula (i × 1) + (ii × 2) + (iii × 3), with a maximum score of 300. Comments and criticisms of the study and/or of the specimens and staining were also requested. Results were returned upon completion to R.A.M. for collation and slides sent on to the next participating centre. In each case assessments from various centres were not revealed to other assessors and the collating centre assessed the slides first to avoid bias.

CAS analysis. Staining of the slides was further assessed by computer enhanced image-analysis using a CAS 200 image analyser equipped with specifically designed software for ER- and PR-ICA-stained breast section analysis, as previously described [7, 8]. In brief, thresholds for negativity are set using a negative control antibody–methyl green counterstained section. The system then assesses positive staining by a comparison of a red filtered digitised image of fields of the primary antibody stained specimen, in which both brown diaminobenzidene tetrahydrochloride stained receptor positive cells and methyl green counterstained receptor negative cells are visible, with a green filtered image. Here methyl green transmits rather than absorbs light rendering the negative cells invisible, whilst ER positive brown cells remain visible. Differences between images represent the degree of positivity within the specimen. The positive nuclear area is measured by a calculation of the number of standard sized screen pixels masked within the green image versus the red filtered image, and is recorded as "% positive area". A staining intensity based calculation recorded by the system, called "positive stain" is also obtained. This bears little comparability in its calculation to the manual *H* score and has not been used in this study.

Statistical calculations for bias. Levels of concordance between assessments received were tested using three different estimates of bias for percentage positive nuclei data. These were calculated against (a) the CAS estimates of "% positive area", (b) the group mean values for each slide and (c) the group median values for each slide.

Only two bias calculations could be made for estimates of the H scores because of the inapplicability of the CAS "positive stain" scores as target values. The definitions of bias are as follows.

(i) Intracentre bias relative to CAS estimates of % positive area: let X_{ij} be the value of a given variable for the ith slide and the jth centre, and let the number of slides be M and centres n which were 13 and 8, respectively, for ER results and 12 and 8 for PR. The mean intracentre bias relative to the CAS value is given by

$$b_j = \frac{1}{m}\sum_{i=1}^{n} b_{ij}, \text{ where } b_{ij} = \frac{(X_{ij} - \text{CAS}_i)}{\text{CAS}_i} \times 100.$$

The range of values are defined by the maximum and minimum values of b_{ij} for each value of j.

(ii) Intracentre bias relative to the group mean value for each slide: this estimate differs from the above in that whereas CAS_i values were independently obtained, the group mean values are calculated from data obtained from all centres. The group mean value for each slide was substituted into the above formulae in place of CAS_i. The range is as defined previously.

(iii) Intracentre bias relative to group median values for each slide: group median values replace those of the group mean for each slide in the above formulae. The range is as defined previously.

RESULTS

Assessments performed in accordance with the directives of the study were received from all centres.

Individual percentage positivity estimates ranged from 0% to 100% for PR slides and 15% to 100% for ER specimens. Similarly for *H* score calculations, PR slides were assessed between 0 and 300 and for ER from 20 to 270.

In all 13 ER cases, positive status was recorded from all assessors when presuming any level of staining as positive. CAS analysis also scored all ER cases positive. In 8 out of 12 cases in which it was measured PR status was recorded as positive by all assessors, and negative by all in 3 cases. In a single discrepant case PR status was assessed as positive by 4 assessors (range 1% to 5% positivity) and negative by 4 assessors. In this discrepant PR case CAS recorded a level of 4% positive area. Other CAS determined PR statuses were as manually reported.

Estimates of ER-ICA percentage positive cells are illustrated in Fig. 1a. Results for different centres are presented in a random order. Wide variations were recorded for some specimens. Sample D for example was reported as 15% positive by one assessor and 90% positive by another. Group mean values and the CAS positive area data are presented, and show a fair degree of correlation.

H score estimates for ER-ICA data showed similar degrees of discordance between assessors (Fig. 1b).

Figures 2a and 2b illustrate the results reported for PR-ICA stained slides and again show marked variance between assessments of the same slide.

In an attempt to further analyse the apparent lack of quantitative concordance between assessments and in particular the performance of individual laboratories and assessors, a rudimentary examination of variance and bias was undertaken. Data analysis, however, was limited by the small size of the study.

For each specimen 3 target means were calculated as described in the methods relating to the % of positive cells data; which were (i) the CAS value, (ii) the mean of all assessments for the specimen and (iii) the median value akin to this mean. The bias and relative bias recorded for each of the data received relative to these targets were calculated. Relative biases for assessment values equal to or in the region of zero were rejected in the analysis. Useful relative bias data were pooled for each centre and the range of these, a reflection of the centres performance,

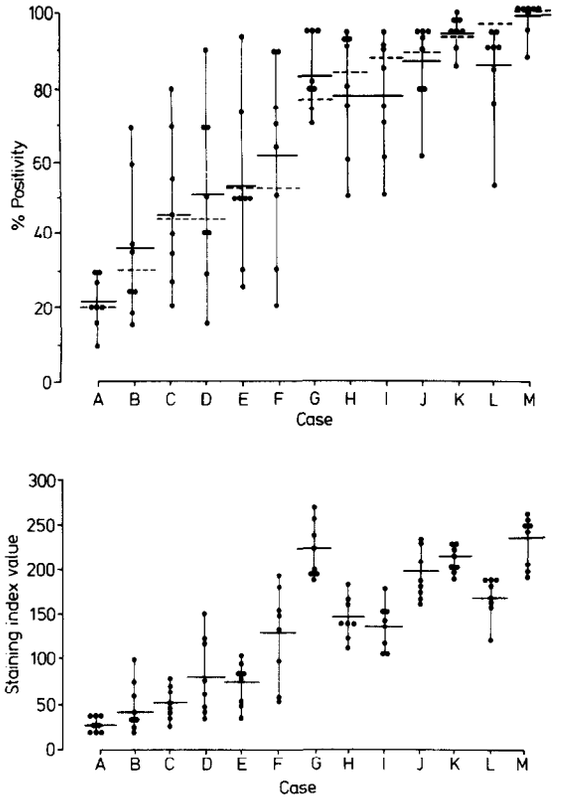

Fig. 1. Distribution of ER-ICA assessments received from each of 8 centres. Upper: % positivity scores (•), (○ represents all 8 results equivalent). (---) = CAS positive area and (——) = group mean values. Lower: staining index (*H* score) results and group mean.

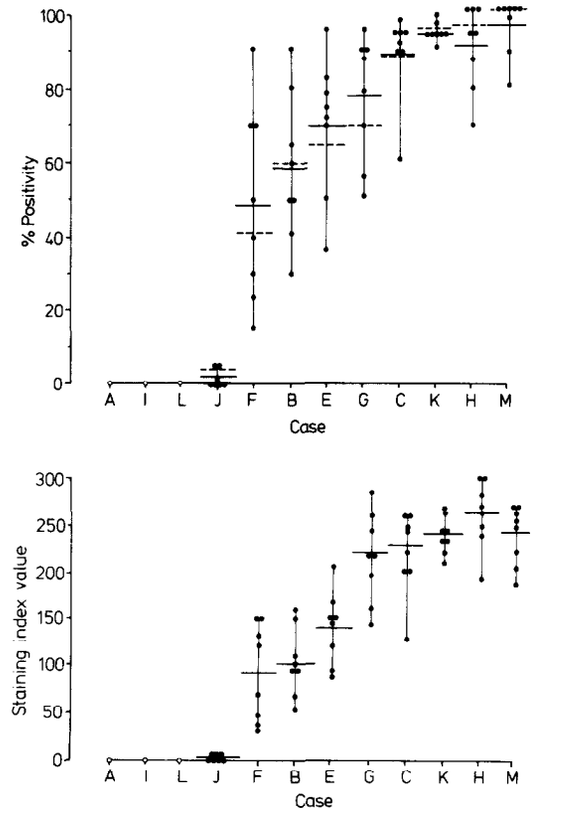

Fig. 2. PR-ICA assessments from 8 centres. Upper: % positivity scores (•), (○ represents all 8 results equivalent). (---) = CAS positive area and (——) = group mean. Lower: staining index (*H* score) results and group mean.

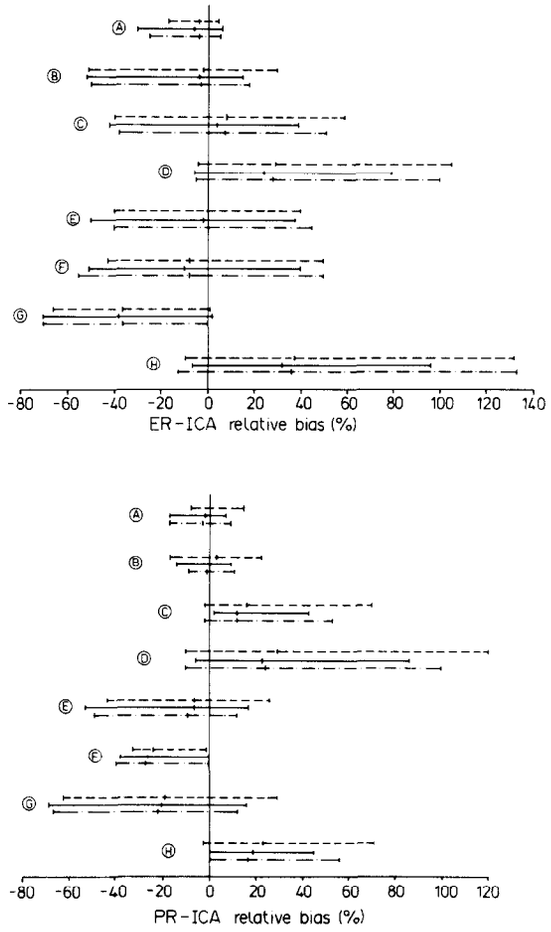

Fig. 3. Relative bias of grouped results shown by centre for estimates of % positivity. Bias is portrayed against three target values i (---), CAS positive area; ii (——), group mean value; iii (-•-), group median value. Upper = ER-ICA results and lower = PR-ICA results. Mean bias values are also indicated (-+-).

is portrayed in Figs 3a (for ER-ICA % positivity) and 3b (PR-ICA). Similar relative bias portrayals are presented for *H* score data. Here however, the CAS target mean is rejected as previously discussed since its calculation and that of the *H* score are not comparable [Figs 4a (ER-ICA) and 4b (PR-ICA)].

Wide ranging relative bias scores are apparent on all of these figures and demonstrate that some assessors consistently underestimated staining levels, others overestimated and others performed relatively well (as indicated by a centrally located, short bias range). In general however, data suggests centres arbitrarily labelled as (A) (B) (C) (E) and (F) produced reasonably good ER-ICA results with respect to the target values. Centres (D) and (H) frequently overestimated ER values both in terms of % positivity and *H* score, whilst (G) underestimated both. Bias for PR assessments was less than that for ER-ICA, yet only (A) (B) and (C) showed consistently good results relative to targets. Centres (D) and (H) overestimated PR staining levels and (E) (F) and (G) frequently underestimated.

DISCUSSION

The measurement of ER and PR levels in advanced breast cancer is of significant value in the selection of probable candidates for successful endocrine therapy [19]. Many reports indicate that the quantity of receptor levels correlate directly with response rates, with strongly positive specimens being more frequently hormone responsive [9, 10, 16]. In view of this, the

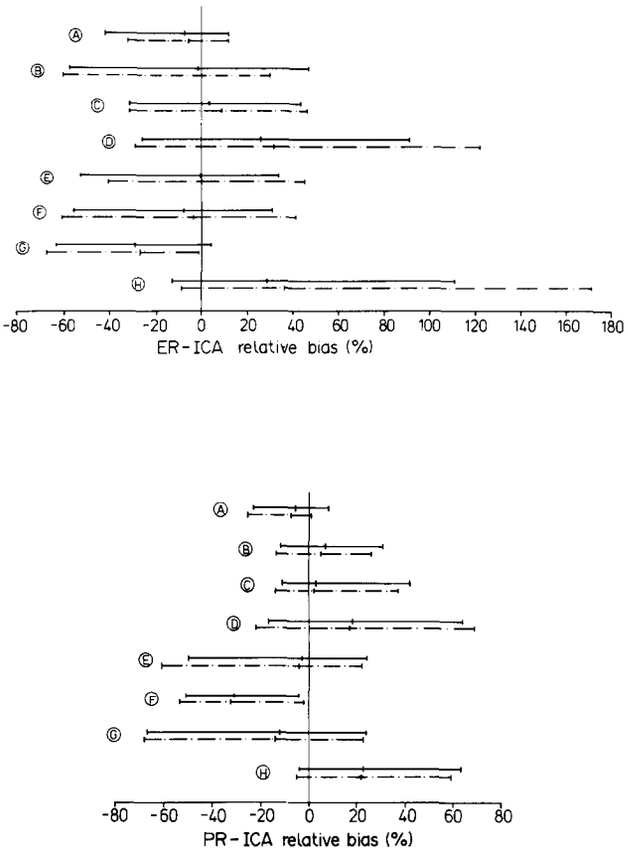

Fig. 4. Relative bias of grouped results shown by centre for *H* score calculations. Bias is portrayed against two target values i (——), group mean; ii (–•–), group median value. Upper = ER-ICA data and lower = PR-ICA data. Mean bias values are also indicated (–+–).

importance of accurate quantification of such assays is obvious. With the recent shift towards immunocytochemical methods for steroid receptor assay following the advent of reliable monoclonal antibody procedures the need for good quality control of the performance on quantification of assays is similarly crucial.

The centres involved in this study have many years of steroid receptor assay experience and developing expertise in the performance and interpretation of immunocytochemical assays for these receptors. Despite this, surprisingly large variations in the estimated levels of stained cells, as interpreted from the same slides were recorded. Assessors comprised biochemists and histopathologists but variant results were derived from all, when compared with mean values.

Explanations for the unexpectedly high levels of disparity observed are not immediately apparent. Some centres expressed concern over the use of methyl green as a counterstain for ER-ICA, having more experience of haematoxylin. As stated earlier, methyl green was used because it was the appropriate counterstain for assessment using the image analyser. It is however a less favoured nuclear stain than haematoxylin and may account in part for some overestimation of ER positivity if the less distinct green counterstained negative cells led to underestimations of their numbers. Such a question could be addressed with a repeat study using haematoxylin counterstained slides in association with the methyl green ones. It was further evident that a range of procedures for the calculation of semiquantitative indices were routinely being used among centres. The requirement of the study for a standardisation of these and the inexperience of some assessors in using this procedure may also have affected the poor correlation of *H* score results.

Some slides in particular caused greater difficulty between centres. These were slides of intermediate positivity, i.e. those expressing the greater degree of heterogeneity. Specimens of low or very high receptor content were given estimates of greater concordance.

Encouragingly, image analysis achieved results which were closely associated with the mean group performance and would therefore be acceptable as a standardised technique for such quality control assessment. CAS analysis has previously been shown to reduce variance in a similar study in the assessment of the stained sections [16].

Analysis of either percentage positivity, for which comparable CAS data is available, or of *H* score, a system which incorporates staining intensity into its calculation, and where comparable CAS scores are not possible, showed similar levels of disparity between manual assessments by centre. It was expected that *H* score calculations would give greater inconsistency than percentage positivity estimates since they require judgements of both positivity and intensity levels to be made. If such an assumption is unjustified then selection of the *H* score calculation as the method of choice might be preferable since it enables the incorporation of staining intensity and heterogeneity data into analyses.

Comparison of ER and PR stained slides showed similar levels of error when quantifying positive samples but good qualitative agreement between centres was noted throughout, for both assays.

In conclusion, this paper reports the findings of an initial study comparing the quantitative analyses of a number of assessors in the interpretation of ER and PR immunocytochemically stained breast tumour sections. Results suggest that unacceptably large variations in these subjective procedures are often observed and that some form of quality control system, possibly involving a computer-assisted image analysis system as a standard, be set up to overcome this problem. This study does not however attempt to address the potential problem of interlaboratory staining procedure variations, only prestained slides having been exchanged. Such variations must surely exist and are likely to reduce further the reproducibility of assessments. An investigation into this and other aspects of ICA quantification is being considered for future study.

1. Nicholson RI, Campbell FC, Blamey RW, Elston CW, George D, Griffiths K. Steroid receptors in early breast cancer: value in prognosis. *J Steroid Biochem* 1981, **15**, 193–199.
2. McGuire WL, Carbone PP, Vollmer EP, eds. *Estrogen Receptors in Human Breast Cancer*. New York, Raven Press, 1975.
3. Hawkins RA, White G, Bundred NJ, *et al.* Prognostic significance of oestrogen and progestogen receptor activities in breast cancer. *Br J Surg* 1987, **74**, 1009–1013.
4. Allegra JC, Lippman ME, Thompson EB, *et al.* Estrogen receptor status: an important variable in predicting response to endocrine therapy in metastatic breast cancer. *Eur J Cancer Clin Oncol* 1980, **16**, 323–331.
5. McClelland RA, Berger U, Miller LS, Powles TJ, Jensen EV, Coombes RC. Immunocytochemical assay for estrogen receptor: relationship to outcome of therapy in patients with advanced breast cancer. *Cancer Res* 1986, **46** (Suppl.) 4241s–4243s.
6. Nicholson RI, Griffiths K, Blamey RW, Bishop HM, Haybittle JL. Estrogen receptors and breast cancer. *Environ Health Perspect* 1981, **38**, 143–146.
7. Allegra JC, Lippman ME. Estrogen receptor determination predicts response to tamoxifen therapy. *Recent Results Cancer Res* 1980, **71**, 16–19.

8. Lawrence BV, Lipton A, Harvey HA, *et al.* Influence of estrogen receptor status on response of metastatic breast cancer to aminoglutethimide therapy. *Cancer* 1980, **45**, 786–791.
9. Vollenweider-Zerargui L, Barrelet L, Wong Y, Lemarchand-Beraud T, Gomez F. The predictive value of estrogen and progesterone receptors' concentrations on the clinical behavior of breast cancer in women. Clinical correlation on 547 patients. *Cancer* 1986, **57**, 1171–1180.
10. Walker KJ, Bouzubar N, Robertson J, *et al.* Immunocytochemical localization of estrogen receptor in human breast tissue. *Cancer Res* 1988, **48**, 6517–6522.
11. King WJ, Greene GL. Monoclonal antibodies localize oestrogen receptor in the nuclei of target cells. *Nature* 1984, **307**, 745–747.
12. Press MF, Greene GL. Localization of progesterone receptor with monoclonal antibodies to the human progestin receptor. *Endocrinology* 1988, **122**, 1165–1175.
13. Greene GL, Press MF. Immunochemical evaluation of estrogen receptor and progesterone receptor in breast cancer. In: Ceriani RL, ed. *Immunological Approaches to the Diagnosis and Therapy of Breast Cancer*. New York, Plenum Press, 1987.
14. Bacus JW, Grace LJ. Optical microscope system for standardized cell measurements and analyses. *Applied Optics* 1987, **26**, 3280–3293.
15. Bacus S, Flowers JL, Press MF, Bacus JW, McCarty KS. The evaluation of estrogen receptor in primary breast carcinoma by computer-assisted image analysis. *Am J Clin Pathol* 1988, **90**, 233–239.
16. McClelland RA, Finlay P, Walker KJ, *et al.* Automated quantitation of immunocytochemically localized estrogen receptors in human breast cancer. *Cancer Res* 1990, **50**, 3545–3550.
17. McCarty KS Jr, Miller LS, Cox EB, Konrath J, McCarty KS Sr. Estrogen receptor analyses: correlation of biochemical and immunohistochemical methods using monoclonal antireceptor antibodies. *Arch Pathol Lab Med* 1985, **109**, 716–721.
18. McClelland RA, Berger U, Miller LS, Powles TJ, Coombes RC. Immunocytochemical assay for estrogen receptor in patients with breast cancer: relationship to a biochemical assay and to outcome of therapy. *J Clin Oncol* 1986, **4**, 1171–1176.
19. Allred DC, Bustamante MA, Daniel CO, Gaskill HV, Cruz AB Jr. Immunocytochemical analysis of estrogen receptors in human breast carcinomas. *Arch Surg* 1990, **125**, 107–113.

Acknowledgements—We thank D. Brocklehurst, P. Cowen, I. Ellis, R. Hawkins, K. Healey, J. Teasdale and R.A. Walker of the British Quality Control Group for help in assessment of slides.

Eur J Cancer, Vol. 27, No. 6, pp. 715–720, 1991.
Printed in Great Britain

0277–5379/91 $3.00 + 0.00

Determination of Oestrogen Receptors: Application of the Passing–Bablock Linear Regression Technique for Comparison of Enzyme Immunoassay and Radioligand Binding Assay in 1841 Breast Cancer Tumours

S. Romain, C. Dussert and P.M. Martin

To test the qualities of two assays in the same laboratory on the same tumours, a single-point dextran-coated charcoal radioligand binding assay (RLA-DCC) and the Abbott enzyme immunoassay (EIA) were used to perform oestrogen receptor determinations on cytosols from 1841 breast cancers over a 2-year period. Statistical analysis of the data was performed by the Passing–Bablock linear regression technique. The final regression curve between EIA (y) and RLA-DCC (x) yielded $y = 1.187x$ fmol/mg of protein. However, a high variability in this correlation was observed from 1986 to 1988. This variability could be explained by calibration problems in the immunoassay kits and changes in our technical team. The binding assay appears to be more sensitive to the technicians' experience than the immunoassay. Other technical points are discussed, particularly cytosol preparation and KCl presence or absence in the homogenisation buffer. Finally, the Passing–Bablock and the least squares regression procedures are compared. The conditions allowing optimal correlation and routine determination reliability are defined and the correlation variability is discussed.
Eur J Cancer, Vol. 27, No. 6, pp. 715–720, 1991

INTRODUCTION

THE CLINICAL usefulness of oestradiol receptors (ER) assays in breast cancer specimens has been clearly established by correlations with patients' response to endocrine therapies and prognosis. For many years, ER determinations have been routinely performed in many breast cancer centres [1]. For optimal disease management, it seems important to accurately distinguish between ER negative and ER positive patients and to quantify the amount of ER present in the tumour tissue. Until recently, ER were quantified exclusively by radioligand binding assay. The Abbott enzyme immunoassay (EIA) using monoclonal antibodies now provides an alternative approach for ER assays in breast cancer [2–4]. Several groups have compared this enzyme immunoassay with conventional radioligand binding

Correspondence to S. Romain.
The authors are at the Laboratoire d'Oncologie Biologique-Biopathologie Tissulaire, Faculté de Médecine de Marseille-Secteur Nord, Boulevard Pierre Dramard, 13326 Marseille Cedex 15, France.
Revised 27 Feb. 1991; accepted 8 Mar. 1991.