Eastern HEALTH

24 October 2005

Philip Hassen, Chief Executive Officer
Canadian Patient Safety
Suite 1414, 10235 – 101 Street
Edmonton, Alberta
T5J 3G1

Dear Phil:

Further to our recent discussions and e-mails, we continue to work towards resolving the concerns which were identified with the ER/PR testing in St. John's. We have also gained an appreciation for the potential that this is an issue for more than just Newfoundland. To that end, we continue to seek the advise of people throughout the country as to how this can be appropriately followed-up. The lack of a recognized national body in this area may be a challenge.

In the meantime, I am forwarding some of the literature that was provided to me for your review.

Sincerely,

George Tilley
President and Chief Executive Officer

GT:jmp

Encls.

Assessment of Tissue Estrogen and Progesterone Receptor Levels: A Survey of Current Practice, Techniques, and Quantitation Methods

Lester J. Layfield, MD,* Dilip Gupta, MD,* and Eoghan E. Mooney, MB, MRCPath†

Department of Pathology, University of Utah, Salt Lake City Utah, and †Department of Pathology and Laboratory Medicine, National Maternity Hospital and St. Vincent's University Hospital, Dublin, Ireland

Abstract: The assessment of steroid hormone receptors in resected breast carcinoma tissue is currently the standard of practice. The traditional method for assessment of receptor status is the ligand binding assay. More recently, immunohistochemistry (IHC) has become a popular method for such testing. Despite the widespread use of IHC and the availability of many antibodies, standardization of quantitative IHC for assessment of estrogen and progesterone receptors has not been achieved. While the College of American Pathologists (CAP) offers a Quality Assurance (QA) program for IHC quantitation of estrogen receptor (ER) and progesterone receptor (PgR), no universal standard is currently recognized in assessment of ER and PgR by IHC. We surveyed 300 laboratories within the United States for their current practices regarding the assessment of ER and PgR status in breast cancer tissue specimens. Eighty usable responses were received. Forty-nine (61%) laboratories performed the assay in-house, while the remainder sent the material out for assay. All responding laboratories performing their steroid receptor analysis in-house used the IHC technique. Forty-three (80%) laboratories answering the question on material accepted for analysis performed the assay only on paraffin-embedded material, three (6%) used either paraffin block or frozen material, and two (4%) used only frozen material. Eighty-eight percent of laboratories performing steroid receptor analysis in-house used a manual quantitation technique. Four (8%) used computer-assisted image analysis, and a single laboratory used laser scanning cytometry. Eight different antibodies were used among the 44 laboratories documenting the antibody supplier, and for any given commercially prepared antibody a wide variety of dilutions were used, with the exception of the standard solution used with the Ventana antibody. Of the laboratories using manual estimation techniques, 61% simply estimated the percentage of positive cells, 29% evaluated both the intensity of staining and percentage of nuclei staining, 6% used formal H-score analysis, 2% evaluated only intensity of nuclear staining, and 2% mainly counted the percentage of nuclei staining for ER but used a formal H score in the assessment of PgR. Cutoff points for the separation of positive and negative results varied widely, with some laboratories assessing any demonstrable positivity as a positive result, while others required as many as 19% of the nuclei to stain before a specimen was declared positive. Standardization techniques differed considerably among laboratories. Eighty-six percent used the CAP program for QA. While all laboratories utilized some form of intralaboratory control for assessment of ER and PgR, the nature of that control varied from laboratory to laboratory. Our survey indicates that a majority of laboratories perform their steroid hormone receptor

Address correspondence and reprint requests to: Lester J. Layfield, MD, Department of Pathology, University Of Utah, Salt Lake City, UT.

© 2000 Blackwell Science Inc., 1075-122X/00/$15.00/0
The Breast Journal, Volume 6, Number 3, 2000 189–196

analysis in-house using IHC. There is considerable variability in the antibodies utilized, the dilutions applied, and the quantitation method and level of expression used to dichotomize specimens into positive and negative groups. Finally, no universal control for interlaboratory standardization appears to exist. ■

Key Words: breast carcinoma, estrogen receptor, progesterone receptor, steroid hormone assay

Estrogen receptor (ER) and progesterone receptor (PgR), by their interaction with their respective steroid hormones play important roles in regulating the proliferation and differentiation of normal breast epithelium (1). The level of steroid hormone receptor expression in breast carcinoma cells is believed to be associated with the responsiveness of the neoplastic cells to circulating estrogen and progesterone. During the past quarter century, many studies have measured tissue levels of ER and PgR by biochemical methods and correlated them with both prognosis and response to hormone therapy (2–5). Within the past decade, a variety of antibodies against both the ERs and PgRs have become available on a commercial basis. Many studies have compared the results of immunohistochemically determined steroid receptor values with those obtained by ligand binding analysis. In general, the correlation has been good (6–11). Despite agreement of results derived from individual antibodies used in immunohistochemistry (IHC) determinations and the ligand binding technique, significant variability has been documented in the results obtained by IHC using different commercially available antibodies (12,13). While variability between the results achieved by different antibodies may exist, the overall value of IHC-determined ER and PgR levels for the prediction of response to hormonal therapy and overall prognosis appears high (9,14–16). Some studies have documented IHC determination of ER to be superior to the ligand binding assay for the prediction of response to adjuvant endocrine therapy in breast cancer (17).

Despite the documented value and accuracy of ER and PgR assayed by IHC, a wide variety of antibodies and quantitation techniques are currently in use. The extent to which these variabilities in technique and material affect the predictive value and standardization of IHC determination of ER and PgR is unknown. Of significance, there is no universally accepted control for standardization of the assays of ER and PgR by IHC. Hence interlaboratory comparisons of steroid receptors as determined by IHC may not be entirely valid. The authors are aware of only a single quality assurance/quality control (QA/QC) program within the United States for IHC determination of steroid receptors, again bringing into question the overall consistency of results obtainable between laboratories. Despite such problems, the overall robustness of the method appears to compensate for the lack of precision associated with the technique. Several laboratories are working to refine protocols and standardization methods and Riera et al. (18) have recently proposed tissue cell culture lines as a universal control.

In order to clarify the current status of ER and PgR level determinations, we surveyed 300 laboratories to determine their methods and interpretative approach for the estimation of steroid hormone receptors in breast tissue and whether they take part in a regional or national QA/QC program in this field. Herein we report the results of that survey.

MATERIALS AND METHODS

Questionnaires were mailed with return addressed and stamped envelopes to 300 hospital pathology departments. The first 200 questionnaire recipients were drawn from the College of American Pathologists (CAP) directory by randomly selecting four pathologists from each state. An additional 100 pathologists with interest and expertise in the area of breast pathology were selected on the basis of a literature search for publications concerning estrogen and progesterone receptors.

The questionnaire contains questions relating to size and type of hospital practice. The questionnaire asked if they routinely ordered ER and PgR assays on newly diagnosed breast carcinomas and whether this analysis was done in-house or if it was sent out. If they performed in-house analysis, questions about the method used [IHC, dextran-coated charcoal (DCC) assay, polymerase chain reaction (PCR), or flow cytometry], type of material accepted [paraffin embedded, frozen tissue, or fine needle aspiration (FNA)], type of antibody, and dilution employed were included. The pathologists were also asked how quantitation was performed (image analysis, manually calculating the nuclei staining percentage, estimating the intensity of staining, H score), what the cutoff point was for differentiating positive and negative results, what protocol was used for standardization, and what controls were used. Lastly, the laboratories were asked whether they participated in any of the interinstitutional QA programs and, if so, which one.

In addition, 150 questionnaires were mailed with return addressed and stamped envelopes to hematology/

oncology departments in all 50 states. One hundred and ten oncologists were program directors at teaching hospitals. The remaining were randomly selected oncologists based at community hospitals. The questionnaire contains questions relating to the method used by the laboratory for the analysis of ER and PgR, the quantitation method, and the cutoff point used to separate positive and negative results. The questionnaire asked whether their treatment approach changed following a switch from DCC to IHC, whether they equated immunohistochemical expression of ER and PgR to specific femtomol values, and whether they required quantitation or merely positive and negative results. The oncologists were also asked whether they were influenced in choice of therapy by PgR status, or whether they gave tamoxifen regardless of steroid hormone receptor status to all postmenopausal patients. They were also asked if ER and PgR status had ever changed during treatment. Finally, their opinion regarding current recommendations (19,20) on chemotherapy and endocrine therapy and the duration of such treatment (21) (2 years versus 5 years) was sought.

Following mailing of the survey questionnaire, 3 months were allowed to pass before closure of the data collection period, allowing for adequate response time. The responses were entered on a spreadsheet program (Excel 7.0, Microsoft, Redmond, WA) and analyzed.

RESULTS

Responses were received from 80 of the 300 questionnaire recipients, geographically representing 35 states. Thirty (38%) were obtained from institutions describing themselves as community/general hospitals, seven from nonacademic tertiary care centers, 41 from academic tertiary care centers, 1 from a reference laboratory, and 1 did not indicate the type of institution. Respondents were almost equally distributed between hospitals with fewer than 250 beds (22), hospitals between 250 and 500 beds (26), and hospitals with more than 500 beds (29). Three respondents did not state the size of their institution.

Seventy-seven of the 80 (96%) responding practices routinely performed ER and PgR analysis on both in situ and invasive carcinoma of the breast. Two performed it only on invasive carcinoma, and a single respondent did not indicate his practice pattern. Both institutions performing ER and PgR analysis only on invasive carcinomas were community hospitals. Forty-nine of the 80 (61%) respondents performed steroid receptor hormone analysis in their own laboratories, 30 routinely sent tissue out for such analysis, and 1 laboratory performed in-house IHC but sent out tissue to have DCC ligand binding assay performed on some specimens. Table 1 shows the distribution of institutions performing steroid receptor analysis by hospital type. Academic tertiary care hospitals were most likely to perform the assays within their institution. Similarly, hospitals with more than 500 beds were more likely to perform steroid receptor analysis in their own laboratories. Community/general hospitals were most likely to send out ER and PgR analyses (23; 77%). A wide variety of academic tertiary care medical centers and specialty commercial laboratories were used for referral of tissue for steroid hormone analysis.

All institutions performing ER and PgR analysis in-house used immunohistochemistry. A single institution performed in-house IHC but sent out material for ligand binding assay in selected cases. Forty-three (80%) of the institutions answering the question on tissue acceptable for analysis performed ER/PgR analysis only on paraffin tissue, 6 (11%) performed the analysis on paraffin-embedded and FNA material, 2 (4%) performed the assay on paraffin-embedded, frozen, and FNA material, while 2 (4%) performed the assay only on frozen material. One (2%) respondent performed the analysis on paraffin-embedded and frozen material.

Methods of quantitation varied among the laboratories responding to our survey. Forty-two (88%) used a manual counting method for quantitation, 4 (8%) used computer-aided image analysis, and 1 (2%) used laser scanning cytometry. One additional laboratory (2%) used computer-assisted image analysis between 1988 and 1998 but recently changed to a manual technique. Of the laboratories utilizing a manual counting method for quantitation of ER and PgR levels, 30 (61%) manually counted the number of tumor cell nuclei staining positively and calculated a staining percentage. Fourteen respondents (29%) used both the percentage of positive nuclei and the intensity of staining. Three laboratories (6%) used formal H-score analysis. One laboratory (2%) measured only the intensity of staining. A single

Table 1. Hospital Type

Type	Number	Percentage
Community/general	30	37.5
Nonacademic tertiary	7	8.75
Academic tertiary	41	51.25
Reference laboratory	1	1.25
No response	1	1.25
Total	80	100

Table 2. Suppliers of Antibodies Used for ER Analysis by Responding Laboratories

Type	Number	Percentage
Dako	17	39
Ventana	15	34
Novacastra	3	7
AMAC	2	5
Immunotech	4	9
Abbott	1	2
Biogenex	1	2
Zymed	1	2
Total	44	100

Table 3. Range of Dilutions Used

ER	
Dako	5–2,000
Ventana	Prediluted
AMAC	50–500
Novacastra	40–200
Immunotech	50–400
PR	
Dako	20–160
Biogenex	25–350
Novacastra	50–100

laboratory manually counted the nuclei staining percentage for ER but employed the H-score technique for PgR analysis.

Table 2 shows the different commercial suppliers of antibodies used for IHC analysis of ER and PgR. Antibodies were supplied by eight different companies, and antibody dilutions varied considerably, as indicated in Table 3. Cutoff points for separation of positive and negative results varied widely among laboratories. Even when the H-score system was used, the cutoff point was not uniform. Table 4 shows the cutoff points reported by the respondents. Some laboratories accepted any visually detectable staining as indicative of positive ER while others required the nuclei staining percentage to be as high as 20% before a tissue specimen was considered positive. Two laboratories (4%) did not interpret the results as either positive or negative, but simply estimated the nuclei staining percentage and intensity of staining present, leaving interpretation to the clinicians.

Protocols for standardization are listed in Table 5. There was considerable variation, with reliance on manufacturer protocols, DCC validation, or CAP survey. The controls used for standardization were also inconsistent among laboratories, with the majority of institutions (42; 74%) using known positive and negative cases as their controls. Table 6 shows the types of controls used by the responding institutions. Sixty-nine of the 80 institutions responding (86%) took part in the CAP program. The remaining 11 institutions did not take part in any QA/QC program.

Only 26 (17%) useful responses were received from the 150 questionnaires sent out to directors of hematology/oncology programs (representing 17 states). All 26 respondents routinely ordered steroid hormone receptor analysis on all newly diagnosed breast carcinomas. One of these did not routinely request such an analysis on intraductal carcinoma. Twenty-five respondents believed the technique used for the analysis was IHC, while one received information from the ligand binding (DCC)

Table 4. Criteria Used to Separate Positive and Negative Results

	Number	Percentage
10% positive cells	15	34
5% positive cells	12	28
Any positivity	4	9
Combination formula (number of cells and intensity)	3	7
<9% = negative, 9–19% = borderline, >19% = positive	2	4
Different criteria for ER and PgR	2	4
20% ER, 5% PgR		
20% ER, 10% PgR		
Number and intensity supplied for clinician interpretation	2	4
Combination of number and intensity		
10% positive with ≥2 (scale of 1–4)	1	2
H score of 50	1	2
H score of 10	1	2
San Antonio score		
1–2 = negative, 3 = borderline, 4–8 = positive	1	2
Intensity (SI 0–3), PP = percentage 0–4, IS = SI × PP		
0–1 = negative, 2 or more = positive	1	2
Total	45	100

Table 5. Protocols Used for Standardization

Protocol	Number
Positive and negative controls	3
DCC validation	3
Ventana automated stainer	2
Dako protocol	2
Do not know (DNK)	1
In-house standardized protocol (Techmate instrumentation) and standardized commercial reagents	1
Parallel testing for new kits and new antibody lot	1
CAP survey	1
Modified Techmate	1
All slides reviewed by the director	1
Ventana/microwave antigen retrieval	1
No protocol	1
DCC and reference laboratories	1
HIER target antigen retrieval (Dako)	1
CAS 2000	1
Manufacturers' guidelines/journals/textbooks	1

analysis. Of the 25 individuals experiencing a shift in analytic technique from DCC to IHC, only one changed their treatment approach because of the modification in technique. Thirteen of 25 (52%) hematologists/oncologists equated negative or low IHC values of ER and PgR with specific femtomol values. The other 12 did not directly correlate IHC results with femtomol levels.

In agreement with the results of the survey of pathologists, there was considerable variation in the cutoff points used by hematologists and oncologists to separate positive from negative ER results. These values ranged from 1 to 30%. Four hematologists/oncologists (18%) did not know the value used for stratifying ER results into positive and negative. Table 7 shows the distribution of cutoff points used by the responding hematologists and oncologists. Eleven respondents required quantitation in their practice, while 16 required only a statement of positive or negative. Thirteen of 25 respondents agreed with the recommendations for treatment of early breast cancer as stated in the *British Journal of Cancer* (19) and in the *Review of Seminars in Oncology* (20).

Table 7. Distribution of Thresholds for Establishing ER Positivity Used by Hematologists/Oncologists

Threshold	Number	Percentage
No response	7	26
10%	5	19
Do not know	4	15
10 fmol	4	15
5%	2	7
30%	2	7
1%	1	4
20% = negative, <30% = borderline, >30% = positive	1	4
Laboratory does it	1	3
Total	27	100

Sixteen of 25 (64%) responding hematologists/oncologists would not treat an ER-negative carcinoma in a postmenopausal patient with tamoxifen. Twenty-three of 27 responding hematologists/oncologists stated that they were influenced in their treatment decisions by the presence or absence of PgR positivity in the neoplasm. Twenty-five of 26 hematologists/oncologists stated that they had modified their treatment plans based on the ER or PgR status of the patient. Finally, 20 of 27 responding hematologists/oncologists stated that they had not changed the duration of treatment after publication of the Swedish Breast Cancer Cooperative Group results in *Journal of the National Cancer Institute* (21).

DISCUSSION

Since the recognition of the relationship between ER and PgR levels in breast carcinoma and patient prognosis and response to hormonal therapy, the assessment of steroid hormone receptors has become a widely accepted component in the examination of breast carcinomas (23). Initially assessment of ER and PgR levels was performed using a ligand binding technology (DCC) (2–5). More recently, ER and PgR assessment by IHC has become popular, if not the predominant technique. Many

Table 6. Type of Controls Used for ER and PgR Analysis

Control	Number	Percentage
Known positive and negative cases	42	71
Internal and external breast tissue control	9	17
EIA assayed breast cancers	1	2
Endometrium	1	2
ER, breast; PgR, endometrium	1	2
Abbott ER-ICA, PgR-ICA control slides for frozen and cytology	1	2
Control cases with $H < 10$, 10–100, and >100	1	2
Multitumor blocks	1	2
Total	57	100

studies have correlated the results of IHC with those obtained by the ligand binding method (6–12), and have confirmed the relationship of ER and PgR with patient prognosis and response to hormonal therapy (1,14–17). Despite the near uniformity in finding a high correlation between IHC and ligand binding assay results and good predictive value for IHC, these studies have used a variety of antibodies, antibody dilutions, staining and quantitation techniques, and cutoff points (23). The variation within the literature is reflected within community practice. Many of the issues relating to IHC determination of steroid hormone receptor levels reflect basic issues in quantitative IHC as discussed by a panel of experts during a recent meeting of the European Section of the International Academy of Pathology (Nice, France, October 1998).

In order to assess the current status of ER and PgR assay methods in the United States, we undertook a mail survey investigating the methodologies, controls, quantitation techniques, and cutoff points utilized by a variety of academic and nonacademic laboratories. Our study found a wide variation in the antibodies, dilutions, quantitation techniques, and cutoff points used by the respondents. While these variations do not in and of themselves negate the clinical significance of steroid hormone receptor analysis, the existence of such variations raises the potential for clinically significant discordance in reported steroid hormone receptor values between laboratories.

Commercial and large academic medical center laboratories perform approximately 38% of all ER and PgR assays, but the majority of such assays are performed in-house by local laboratories. Nearly all of these laboratories report using IHC for the assessment of ER and PgR. A number of observations can be made on the basis of this study. First, a confounding variable for interlaboratory comparisons of steroid hormone receptor results is the variety of materials accepted by various laboratories. Eighty-one percent of laboratories accept only paraffin-embedded material for analysis, but 11% used only frozen tissue or both frozen and paraffin-embedded tissue for analysis. In addition, 4% also accepted FNA specimens. Second, various antibodies were used. The laboratories in our study employed antibodies supplied by eight different manufacturers. Seventy-three percent of institutions used an antibody supplied by either Dako (Carpinteria, CA) or Ventana (Tucson, AZ). As shown in Table 3, even when using the same antibody, various institutions employed widely different antibody dilutions for their assays. These differences in antibodies and dilutions may have significant impact on the quantitative assessment of ER and PgR by IHC.

Third, methods for the quantitation of IHC results varied considerably between respondents. The majority (88%) of laboratories completing the survey questionnaire used various manual techniques for quantitation in which estimates of nuclei staining percentage were made. Multiple manual quantitation methods exist in addition to simply estimating the nuclei staining percentage. Twenty-nine percent of laboratories using a manual quantitation technique employed a technique where both the number of positive cells and the intensity with which the cell nuclei stained were estimated. Six percent used formal H-score analysis (6). We did not obtain information on the threshold of staining intensity used to accept a nucleus as positive for quantitation purposes. Neither did we obtain information on cell selection techniques. Clearly differences in counting technique can affect whether a neoplasm is designated positive or negative for ER and PgR. Recommendations for cell counting have been published (11).

Fourth, the cutoff points used for the assignment of breast cancer cell populations as positive for ER or PgR differ between laboratories. Our survey documented at least a fourfold variation in the nuclei staining percentage used by laboratories for the assignment of positivity. Twelve laboratories (27%) used a 5% cutoff point to designate a specimen as positive and four laboratories (9%) used a value of 20% as their cutoff point. Fifteen laboratories (33%) used a 10% nuclear positivity rate as an indicator of a positive ER level. Variability in cutoff point results in discordance of results even when methodologic aspects of the assay are identical. Such variability in threshold for positivity highlights the need for laboratories to include cutoff points in their reports. Thresholds used with other quantitation methods (H score) also varied. A few laboratories simply assessed the number of positively staining cells and the intensity of staining present, allowing the clinicians to interpret the data. Thus wide variability in practice exists and such variability may have a significant impact on the documentation of the presence or absence of clinically significant levels of ER and PgR.

Standardization protocols varied widely among the respondents to our survey. Only 22 respondents (28%) answered our query concerning their standardization protocol. Of these, three used DCC validation as their standardization technique, while a majority relied on manufacturer's guidelines, previously tested positive and negative intralaboratory controls, or simply review

of all slides by the laboratory director. Controls for the standardization of ER and PgR assays varied among the laboratories. The majority (52%) used previously assayed tissue blocks known to be positive and negative. This offered a degree of intralaboratory consistency but did not allow external verification of laboratory assay levels. Nine laboratories utilized internal breast tissue controls for the assessment of steroid hormone receptors in associated neoplastic tissue. Occasionally laboratories utilized endometrium as a positive control. As stated by Riera et al. (18), no universal control exists for external validation of steroid hormone receptor assays by IHC. The lack of such a control complicates comparisons between laboratories using different antibodies, dilutions of antibodies, and modifications of the IHC technique. Recently, cultured cells have been suggested as a control for quantitative immunocytochemical analysis of ER levels (18). Widespread utilization of such a standard control should increase the comparability of ER results performed at different laboratories.

Only 26 responses (17%) from our survey of 150 directors of hematology/oncology programs were received, meaning conclusions based on this small dataset should be made with caution. However, certain trends were noted. First, the respondents routinely ordered ER and PgR analysis on all newly diagnosed breast carcinomas, implying general acceptance by oncologists for measurement of ER and PgR. The majority of responding oncologists did not alter their treatment approach when their laboratory switched from the DCC to the IHC methodology. Variability existed among clinicians in how they equated IHC expression to femtomol values. Approximately half of the respondents did not equate negative or low IHC values with specific femtomol values.

In agreement with our laboratory survey findings, the threshold for calling a result positive varied widely among oncologists. Values associated with a positive result by IHC varied from 1 to 30%. Of equal importance, 59% of the responding hematologists/oncologists treating breast cancer patients did not require quantitative data but merely desired a statement by the laboratory as to whether the assay was positive or negative. This finding is of particular interest in light of a recent study showing that very high levels of ER are associated with an unfavorable prognosis (22). Simply dividing ER values into positive and negative may yield incomplete and misleading information. The reporting of femtomol equivalents, nuclei staining percentage, or the stratification of results into negative, borderline, intermediate, and high levels may be more clinically useful.

There does not appear to be uniformity in approach to the interpretation and utilization of ER and PgR data by oncologists. Only 13 respondents (48%) agreed with recently published recommendations concerning the use of endocrinology and chemotherapy in patients with breast cancer (19,20). The majority of oncologists responding to our questionnaire acknowledged that ER or PgR status had modified their treatment of patient's with breast cancer (92%). Despite this reliance on steroid hormone receptor assay results, fully one-third of responding oncologists would treat a postmenopausal patient whose carcinoma was ER negative with tamoxifen. The results of the Swedish BCCG study (21) appeared to change the treatment approach of only a minority of oncologists responding to our survey (26%).

The oncologists varied significantly in the threshold they used to classify specimens as positive or negative. Reported cutoff points for positivity varied from any staining to a cutoff point of at least 30% of nuclei staining. When specific percentages were given, there was a sixfold variation in the cutoff point (5–30%). Such variability in interpretative thresholds renders interlaboratory comparisons of ER and PgR results difficult if only positive and negative assessments are reported. Interlaboratory comparisons are more easily achieved if laboratories record the nuclei staining percentage, cutoff point used, and interpretation of the results rather than simply reporting the specimen as positive or negative.

The CAP QA program is commonly used by laboratories assessing ER and PgR in breast tissue, but its level of success in ensuring interlaboratory uniformity was not assessed by this survey. Further studies into the effectiveness of this program would be of value both to pathology laboratories performing steroid hormone receptor analysis as well as to oncologists interpreting the results.

Steroid hormone receptor assay by IHC appears robust enough to maintain the correlation with prognosis established by DCC, despite the many variations discussed. This interpretation is supported by the observation that the majority of studies in the literature using variable techniques, antibodies, titers, and cutoff points still report good correlation of the IHC results with DCC assays and demonstrate good predictive and prognostic value for the test (1,6–17). As in many areas of surgical pathology, reliability may exceed reproducibility (24). This may lead some clinicians and pathologists to adopt a nihilistic approach to standardization. However, it may be that important prognostic information associated with steroid hormone receptor levels is being

concealed by the imprecision of current IHC methods. The development of a universal control and improved standardization methods should improve the validity of interlaboratory comparison of the results of ER and PgR measurement by the IHC technique.

REFERENCES

1. Allred DC, Harvey JM, Berardo M, Clark GM. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol* 1998;11:155–68.

2. Allegra JC, Lippman MC, Thompson EB, *et al.* Estrogen receptor status: an important variable in predicting response to endocrine therapy in metastatic breast cancer. *Eur J Cancer* 1980;16:323–31.

3. Clark GM. Prognostic and predictive factors. In: Harris JR, Lippman ME, Marrow M, Heliman S, eds. *Diseases of the breast.* Philadelphia: Lippincott-Raven, 1996:461–85.

4. Henson JC, Longeval E, Mattheiem WH, Debocl MC, Sylvester RJ, Leclercq G. Significance of quantitative assessment of estrogen receptors for endocrine therapy in advanced breast cancer. *Cancer* 1977;39:1971–77.

5. Campbell FC, Blamey RW, Elston CW, *et al.* Quantitative oestradiol receptor values in primary breast cancer and response of metastases to endocrine therapy. *Lancet* 1981;2: 1317–19.

6. McCarty KS Jr, Miller LS, Cox EB, Konrath J, McCarty KSS. Estrogen receptor analysis. Correlation of biochemical and immunohistochemical methods using monoclonal antireceptor antibodies. *Arch Pathol Lab Med* 1985; 109:716–21.

7. Sondergaard G, Pedersen KO, Paulsen SM. Estrogen receptor analyses in breast cancer: comparison of monoclonal immunohistochemical and biochemical methods. *Eur J Cancer Clin Oncol* 1989;25:1425–29.

8. Kinsel LB, Szabo E, Greene GL, Konrath J, Leight GS, McCarty KS Jr. Immunocytochemical analysis of estrogen receptors as a predictor of prognosis in breast cancer patients: comparison with quantitative biochemical methods. *Cancer Res* 1989;49:1052–56.

9. Pertschuk LP, Kim DS, Nayer K, *et al.* Immunocytochemical estrogen and progestin receptor assays in breast cancer with monoclonal antibodies: histopathologic, demographic, and biochemical correlations and relationship to endocrine response and survival. *Cancer* 1990;66:1663–70.

10. Kell DL, Kamel OW, Rouse RV. Immunohistochemical analysis of breast carcinoma estrogen and progesterone receptors in paraffin-embedded tissue. Correlation of clones ERID5 and IA6 with a cytosol-based hormone receptor assay. *Appl Immunohistochem* 1993;1:275–81.

11. Esteban JM, Kandalaft PL, Mehta P, Odom-Maryon TL, Bacus S, Battifora H. Improvement on the quantification of estrogen and progesterone receptors on paraffin-embedded tumors by image analysis. *Am J Clin Pathol* 1993;99:32–38.

12. Layfield LJ, Conlon DH, Dodge R, Saria E, Kerns BJ. Immunohistochemically determined estrogen and progesterone receptor levels: a comparison of three commercially available antibodies with the ligand-binding assay. *Breast J* 1996;2:362–71.

13. Mauri FA, Veronese S, Frigo B, *et al.* ERID5 and H222 (ER-ICA) antibodies to human estrogen receptor protein in breast carcinoma: results of a multicentric comparative study. *Appl Immunohistochem.* 1994;2:157–63.

14. Esteban JM, Ahn C, Mehta P, Battifora H. Biological significance of quantitative estrogen receptor immunohistochemical assay by image analysis in breast cancer patients. *Am J Clin Pathol* 1994;102:158–62.

15. Andersen J, Poulson HS. Immunohistochemical estrogen receptor determination in paraffin-embedded tissue. Prediction of response to hormonal treatment in advanced breast cancer. *Cancer* 1989;64:1901–8.

16. Pertschuk LP, Feldman JG, Eisenberg KB, *et al.* Immunocytochemical detection of progesterone receptor in breast cancer with monoclonal antibody. Relation to biochemical assay, disease free survival and clinical endocrine response. *Cancer* 1988;62:342–49.

17. Harvey JM, Clark GM, Osborne K, Allred DC. Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. *J Clin Oncol* 1999;17:1474–81.

18. Riera J, Simpson JF, Tamayo R, Battifora H. Use of cultured cells as a control for quantitative immunocytochemical analysis of estrogen receptor in breast cancer. The Quiegel method. *Am J Clin Pathol* 1999;111:329–35.

19. Buzdar A. The place of chemotherapy in the treatment of early breast cancer. *Br J Cancer* 1998;78:16–20.

20. Hudis CA, Norton L. Adjuvant drug therapy for operable breast cancer. *Semin Oncol* 1996;23:475–93.

21. Rutqvist LE, Hatschek T, Ryden S, *et al.* Randomized trial of two versus five years of adjuvant tamoxifen for postmenopausal early stage breast cancer. *J Natl Cancer Inst* 1996; 88:1543–49.

22. Thorpe S, Christensen IJ, Rasmussen BB, Rose C. Short recurrence-free survival associated with high oestrogen receptor levels in the natural history of post menopausal, primary breast cancer. *Eur J Cancer* 1993;29A:971–77.

23. EORTC Breast Cancer Cooperative Group. Revision of the standards for the assessment of hormone receptors in human breast cancer. *Eur J Cancer* 1980;16:1513–15.

24. Cramer SF. Interobserver variability in surgical pathology. *Adv Pathol and Lab Med* 1996;6:3–82.

Rhodes, A, Jasani, B, Balaton, A.J., Miller, K. D.
Immunohistochemical Demonstration Of Oestrogen And Progesterone Receptors: Correlation Of Standards Achieved On In House Tumors With That Achieved On External Quality Assessment Material In Over 150 Laboratories From 26 Countries.
J Clin Pthol 2000;53:292-301

The aim of this study was to investigate the sensitivity of immunohistochemical ER/PR testing on tissue samples processed in their own labs as compared to the sensitivity on samples that were part of an external QA program.
Results showed a significant positive correlation between the results of the QA samples and the in house samples. This was felt to show that the QA slides could be used as an indicator of the lab's performance.
They also found that most labs low or medium expression results were higher when done by the external lab.
Of note:

- The article identified that six labs published results validating their technique, but because of anonymity reasons they were not identified
- Although the article refers to labs as having "many similarities....in the proportion of tumors confirmed to be receptor positive" it does not state what that value is.

Rhodes, A, et al
Reliability of Immunohistochemical Demonstration Of Oestrogen in Routine Practice: Interlaboratory Variance in the Sensitivity of detection and evaluation of scoring systems.
J Clin Pthol 2000;53:125-130

This study, conducted by the same authors as noted above, investigates interlaboratory variance to determine rate of false negatives.
Results showed there is a high rate of intervariability, the false negative rates are between 30% and 60%, caused by methodology.
Of note:

- In view of the increasing use of ER/PR, it is vital a good QA procedures are in place
- While the majority of labs had little difficulty in obtaining the strongly positive results, 62.7% demonstrated a wide variety of results when reporting low ER results
- This variance was not caused by interpretation, but by the variations in the sensitivity of the IHC method
- Again, the authors say that threshold values should be determined by gauging against clinical outcomes

57-78% ER positive

Zafrani, B. et al.
High Sensitivity And Specificity Of Immunohistochemistry For The Detection Of Hormone Receptors In Breast Carcinoma: Comparison With Biochemical Determination In A Prospective Study Of 793 Cases. *Histopathology* 2000, 37, 536-545

This prospective study compared the ER/PR results from immunohistochemistry (IHC) method of testing to results from enzyme immunoassay (EIA).
Of the 793 cases, 77% were infiltrating ductal breast cancer; with ER positivity of 78% for EIA and 81% for IHC. PR rates were 69% and 65% respectively.
Discrepancies in results were found in the 18% infiltrating lobular cases: ER/PR positivity was 84% with IHC testing and 45% with EIA.
The study found that IHC testing was more sensitive, specific and economical and should be the new standard "provided that good quality assurance procedures are respected".

The authors also concluded that a consensus was needed to determine the best cut-off value, validated by correlation with clinical outcomes.

Of note:

- There are a lot of positivity rates in this article
- ER false negative rate for EIA was 6%; false positive rate was 2%

Wells, C.A. et al
Consistency Of Staining And Reporting Of Oestrgen Receptor Immunocytochemistry Within The European Union – An Inter-Laboratory Study
Virchows Arch (2004) 445:119-128

This study wished to assess the variability of ER testing using IHC. Controls slides were distributed and the results analyzed.
The authors found that there were variations in the technical and interpretative parts of the test, but this variation was less for the strongly positive and negative tumors. More variation came from the weakly positive. They concluded that external quality assessment is needed.

Of note:

- It was noted that older publications found that biochemical assays found a 60% ER positivity rate and 8% true ER negative rate
- They also state that the QA processes for IHC have not been developed and there is a concern of high false negative results. ✓
- The article cites that the side effects of Tamoxifan are severe
- Controls should be a known positive, known negative and a very weak positive for each run ✓

Regitnig, P. et al
Quality Assurrance for Detection of estrogen and Progesterone Receptors by Immunohistocemistry in Austrian pathology Laboratories.
Virchows Arch (2002) 441:328-334

This study aimed to assess the quality of IHC testing and the variability of results by different Austrian labs. The labs used a variety of methods; manual systems, Ventana autostainer, Dako autostainer and a LabVision autostainer. They found the automated staining gave better results, but internal and external QA should be done. Internal controls should be strongly positive, weakly positive and negative.

Of note:

- The authors found that false positive and false negative results were very low
- The authors state that the American Society of Clinical Oncology recommends that the IHC technique is not validated enough (2000)
- This project was undertaken by the Austrian Society of Pathology

J Clin Pathol 2000;53:292–301

Immunohistochemical demonstration of oestrogen and progesterone receptors: correlation of standards achieved on in house tumours with that achieved on external quality assessment material in over 150 laboratories from 26 countries

A Rhodes, B Jasani, A J Balaton, K D Miller

Abstract

Aims—To investigate the sensitivity of immunohistochemical (IHC) assays for oestrogen receptors (ER) and progesterone receptors (PR) achieved by laboratories on breast tumours fixed and processed in their own department, and to compare this with the degree of sensitivity they achieve on tumours circulated as part of an external quality assessment (EQA) programme.

Methods—On 10 occasions between April 1994 and June 1998, histological sections from breast cancers showing various degrees of expression of ER and PR were circulated for IHC staining to laboratories participating in the UK national external quality assessment scheme for immunocytochemistry (UK NEQAS-ICC). The staining of these tumours, in addition to that of tumours fixed and processed in the participants own laboratories (in house tumours), was assessed by a panel of four assessors, using the established UK NEQAS-ICC scoring system. For a selected assessment run, the degree of expression of participants in house tumours was evaluated by means of the semiquantitative quick score method.

Results—Although the scores awarded for the staining of in house tumours were generally higher than those awarded for the staining of UK NEQAS tumours, there was also a significant positive correlation between the two sets of scores. Using the quick score method of evaluation for one of the assessment runs, 47% of in house tumours were classified as having a high degree of ER expression. Of the remaining cases, a significant proportion initially classified as having only low or medium expression of ER were found to have higher expression when stained by the organising laboratory. The UK NEQAS-ICC centre's routine assay for hormonal receptors was found to be 90–100% efficient in achieving optimal demonstration of breast tumours from over 150 different laboratories.

Conclusions—The significant positive correlation between the results obtained on the UK NEQAS tumours and the in house tumours provides evidence for the view that results achieved on EQA material are accurate indicators of in house laboratory performance. Although most laboratories adequately detected tumours with high receptor expression, a large proportion of in house tumours classified initially by participants' staining as being of low or medium ER expression had a higher degree of expression when stained by the UK NEQAS-ICC centre. The efficiency of the organising centre's routine IHC method for ER and PR in optimally demonstrating participants in house breast tumours shows that variations in fixation and tissue preparation are not limiting factors preventing a different laboratory achieving optimal demonstration.

(*J Clin Pathol* 2000;53:292–301)

Keywords: immunohistochemistry; oestrogen receptors; progesterone receptors; external quality assessment

Department of Histopathology, UCL Medical School, University Street, London WC1E 6JJ, UK
A Rhodes

Department of Histopathology, UCL Medical School
K D Miller

Department of Pathology, University of Wales College of Medicine, Cardiff CF4 4XN, UK
B Jasani

Centre de Pathologie, 20 Avenue de la Gare, 91570 Bievres, France
A J Balaton

Correspondence to:
Mr Rhodes
email: rmkdhcn@ucl.ac.uk

Accepted for publication
20 October 1999

Recent leading articles have emphasised the importance of establishing the oestrogen receptor (ER) status of women with breast cancer.[1 2] Other articles have reported on the degree of variability that exists between laboratories when demonstrating ER by immunohistochemistry (IHC) on the same cases.[3 4] The largest of these studies looked at the results obtained by 200 different participants of an external quality assessment (EQA) programme on slides circulated to these laboratories by the EQA scheme and containing tumours with differing degrees of ER expression.[3]

Although many view the results of EQA as a useful gauge of a laboratory's ability to perform adequate staining for ER and progesterone receptors (PR) on paraffin wax embedded sections,[5 6] it could be argued that the system suffers from a number of drawbacks. Distributed material is limited and consists of tissue that has been fixed, processed, and prepared under different conditions, no matter how slight, to those used by the participating laboratory. This is thought to be important because there is a view that the IHC assay optimised for use on in house material cannot be expected to produce results of the same quality on tissues fixed and processed in a different laboratory. Consequently, it is thought by some that the quality of IHC achieved on material distributed by an EQA scheme does not reflect

Table 1 Details of tumours circulated by UK NEQAS-CC for assessments between April 1994 and June 1998

Run no.	Date	Number of labs	Tumour type(s)	Receptor	Degree of expression	
					IHC	Cytosol
26	4/94	37	IDC	ER	100%, +++	Not performed
29	2/95	66	IDC (mets)	ER	100%, +++	Not performed
32*	12/95	98	Inf./DCIS	ER	100%, +++	Not performed
			ILC		80–90% ++	Not performed
34*	8/96	105	IDC	ER	90–100%, +++	ER = 248
			ILC		75–80% , ++	ER = 29
			IDC		0%	ER = 0
36	1/97	118	IDC	PR/ER†	90%, +++	ER = 66, PR = 307
38*	4/97	175	IDC	ER	90–100%, +++	ER = 122
			IDC		90%, ++	ER = 41
			IDC		0%	ER = 7
39*	8/97	176	IDC	PR/ER†	99–100%, +++	ER = 309, PR = 1582
			ILC		75%, ++	ER = 11, PR = 21
			IDC		1%, +	ER = 7, PR = 6
40*	12/97	192	IDC	ER	99%, +++	ER=15
			IDC		99%, +++	ER=11
			IDC		90%, ++	ER = 12
41	4/98	178	IDC	ER and PR‡	ER = 75–80%, ++	ER = 10
					PR = 0%	PR = 2
42*	6/98	205	IDC	ER	100%, +++	ER = 70
			IDC		90%, ++	ER = 41
			IDC		15%, +	ER = 65

All tumours were fixed for 24 hours in 10% neutral buffered formalin.
ER/PR expression of tumours by IHC is described in terms of the proportion of invasive nuclei staining (%) and the average staining intensity (+, ++, +++).
Cytosol values of expression are in fmol/mg protein.
*Composite blocks containing more than one tumour were used for these assesments.
†Participants not stocking an antibody to PR were asked to demonstrate ER instead. On these runs tumours were chosen that have very similar expression of ER and PR, as determined by IHC at the organising laboratory.
‡On this assessment, participants were asked to demonstrate both ER and PR.
ER, oestrogen receptor; IDC, infiltrating ductal carcinoma; IHC, immunohistochemistry; Inf., infiltrating; mets, metastatic; PR, progesterone receptor.

the standards achieved by a laboratory on tumours fixed and processed in house, which might account for a large percentage of its work load. Problems of this type are encountered with most EQA programmes that circulate material for analysis, and have yet to be resolved completely.[7–11]

To investigate these limitations, our study evaluates the immunostaining of the UK NEQAS-ICC organising laboratory and tests its validity as a reference standard for ER and PR. It then investigates and compares the performance achieved by laboratories on in house tumours to that achieved by the same laboratories on the tumours circulated by the scheme.

Table 2 Details, as given on the participants returning questionnaire, of the 152 in house breast carcinomas stained and submitted for assessment (run 41)

Type of tumour	n	%
Adenocarcinoma	5	3.3%
Breast carcinoma (unspecified)	66	43.4%
Ductal carcinoma (grade unspecified)	47	30.9%
Infiltrating ductal carcinoma (grade I)	4	2.6%
Infiltrating ductal carcinoma (grade II)	13	8.6%
Infiltrating ductal carcinoma (grade III)	5	3.3%
Colloid breast carcinoma	1	0.7%
Cribiform	1	0.7%
Infiltrating lobular carcinoma	4	2.6%
Intracystic papillary	1	0.7%
Metastatic disease	4	2.6%
Mucinous	1	0.7%
Tubular carcinoma	2	1.3%
Total	152	100%

Materials and methods

TUMOURS CIRCULATED BY THE EQA SCHEME

Laboratories participating in the UK NEQAS-ICC programme for steroid hormonal receptors were sent, at each assessment, two unstained slides containing histological tissue sections of formalin fixed and paraffin wax processed breast tumours showing different degrees of hormonal expression. Each participant was asked to demonstrate ER and/or PR and to return the stained slide(s) to the UK NEQAS coordinating centre for assessment of staining quality. Table 1 shows details of the tumours circulated at each assessment from April 1994 to June 1998. Although tested in the organisers laboratory, most of these tumours were fixed and processed in the laboratories of participants, from where they were kindly donated. Whenever possible, tumours that also had their receptor status determined biochemically by the ligand binding assay (LBA) were used.

IN HOUSE TUMOURS

In parallel with the immunostaining of slides circulated by the scheme, participants were asked to demonstrate the same receptor on their own in house tumour, and to submit this stained slide for assessment of staining quality, along with two additional unstained tissue sections from the same tissue block. Tables 2 and 3 give examples of the types of tumours submitted by participants and the various fixatives used. These unstained slides were then stained by the UK NEQAS organising laboratory along with its routine workload, utilising its routine methodology and reagents, for the demonstration of hormonal receptors (table 4). They were then coded and filed alongside the participants own immunostaining before routine assessment. The UK NEQAS-ICC organising centre also stained the participants' in house tumours so that a comparison could be

Table 3 Details of fixation for the 152 in house breast carcinomas submitted for run 41

Type of fixative	Duration	Temperature	n	%
10% Formal saline	18–48 hours	Ambient	80	52.6%
10% Formal saline	72 hours	Ambient	2	1.3%
10% Formal saline	24 hours	37°C	2	1.3%
10% Neutral buffered formalin	24–48 hours	Ambient	44	28.9%
10% Neutral buffered formalin	72 hours	Ambient	1	0.7%
10% Neutral buffered formalin	5 hours	Ambient	1	0.7%
Acidic formal alcohol	24–48 hours	Ambient	3	2.0%
Acidic formal alcohol	10 hours	25°C	1	0.7%
Acidic formal alcohol	48 hours	Ambient	1	0.7%
Bouin's fluid	12–24 hours	Ambient	2	1.3%
Bouin's Hollande	Not given	Ambient	1	0.7%
Bouin Duboscq-Brasil	6–24 hours	Ambient	3	2.0%
Carson's fluid	24 hours	Ambient	1	0.7%
Formal calcium	24 hours	Ambient	1	0.7%
7.5% Formalin	24 hours	Ambient	1	0.7%
Information not given			8	5.3%
Total			152	100%

Unless stated otherwise, it was assumed fixation was performed at room temperature (ambient).

made between this staining and that produced by the participant on the same case.

ASSESSMENT OF SLIDES

An expert panel of four, comprising consultant pathologists and biomedical and clinical scientists, assessed the quality of the IHC independently on a single blind basis, with each assessor awarding marks out of 5 for each of the coded slides. The four individual marks were then added together to give a total mark out of 20. Marks were awarded by comparing the proportion and intensity of tumour nuclei staining in the participant's slide, to that achieved on duplicate sections of the same cases by the UK NEQAS-ICC organising centre. A total mark > 12 out of 20 indicates acceptable immunostaining and a pass at assessment, a mark of 10–12 out of 20 is considered to be suboptimal and borderline, whereas a total mark < 10 out of 20 is given for staining that is of unacceptable quality and represents a failure at assessment. One of the main criteria by which staining is deemed unacceptable is when < 10% of receptor positive tumour nuclei are clearly demonstrated in a tumour that has been shown by the UK NEQAS-ICC organising centre to express > 10% ER or PR positive nuclei. To ensure assessor concordance, within ± 1 mark, the slides were marked in batches of 20. On the marking of the 20th slide, all scores were read out, when there was a difference of greater than 1 mark between any of the assessors' individual marks, the respective slide was reviewed until a consensus was reached—that is, all the assessors gave the same mark, within ± 1 mark.

For a selected assessment (run 41), the participants in house staining and that achieved by the organising centre on sections from the same tumours was assessed using the semi-quantitative quick score method of evaluation.[13–15] With this method, the intensity of the immunohistochemical reaction as viewed under the light microscope was recorded as either: 0, negative (no staining of any nuclei even at high magnification); 1, weak (only visible at high magnification); 2, moderate (readily visible at low magnification); or 3, strong (strikingly positive even at low power magnification). The proportion of tumour nuclei showing positive staining was also recorded as either: zero (0), approximately 1–25% (1), 26–50% (2), 51–75% (3), or 76–100% (4). The score for intensity was then added to the score for proportion, giving the quick score with a range of 0–7. In the case of composite blocks (n = 17), the tumour showing the lowest amount of expression was assessed using this method, and where participants had submitted normal breast tissue, these were not submitted for quick score evaluation. Before the assessors' (AR, BJ) evaluation of these slides, all were randomised using numbers generated by a Microsoft Excel program. Determination of the degree of concordance between the quick scores of the participants' immunostaining of their own in house tumour and the quick scores of the UK NEQAS laboratory's immunostaining of the same tumour was established using Cohen's κ coefficient. The proportion of cases that showed the same or a higher degree of expression when stained by the organising laboratory was analysed by means of the χ^2 test, as was the proportion of cases that showed the same or a higher degree of expression when stained by the participant's laboratory. The degree of assessor concordance when using the quick score method to evaluate stained slides was measured using Goodman and Kruskal's γ statistic.

VALIDATION OF ASSESSMENT THRESHOLDS AND OPTIMAL SENSITIVITY

Of the participating laboratories, six were identified as having published studies clinically

Table 4 The main technical parameters used by participants and the organising laboratory of UK NEQAS-ICC in the immunohistochemical demonstration of oestrogen receptors (ER) and progesterone receptors (PR)

	Oestrogen receptor: run 41			Progesterone receptor: run 39		
	Ag retrieval	1° Ab clone	Detection	Ag retrieval	1° Ab clone	Detection
Clinically validated labs (6)	MW (4) PC (2)	1D5 (5) 6F11 (1)	LSAB (1) St ABC (3) ABC (2)	MW (3) PC (2) None (1)	1A6 (3) KD68 (3)	LSAB (1) St ABC (4) ABC (1)
All UK NEQAS-ICC participating labs	MW (96) PC (60) PC in MW (18) Other (4)	1D5 (131) 6F11 (41) Other (6)	St ABC (61) ABC(50) LSAB (23) Other (44)	MW (106) PC (55) PC in MW (5) Other (10)	1A6 (76) 1D5 (52) 6F11 (10) Other (38)	St ABC (64) ABC (43) LSAB (41) Other (28)
UK NEQAS organising lab	PC	1D5	St ABC	PC	1A6	St ABC

Numbers shown in parenthesis relate to the number of laboratories using each antibody or system for run 41 (ER – 178 returns) and run 39 (PR and ER, – 176 returns)
In the runs for which PR was requested, laboratories not stocking an antibody to PR were requested to demonstrate ER.
1°, primary; ABC, avidin biotin complex; Ab, antibody; Ag, antigen; MW, microwave oven pretreatment; LSAB, labelled St ABC; "Other", antibodies or systems each used by less than four participants, or incomplete data; PC, pressure cooker pretreatment; St ABC, streptomyces ABC.

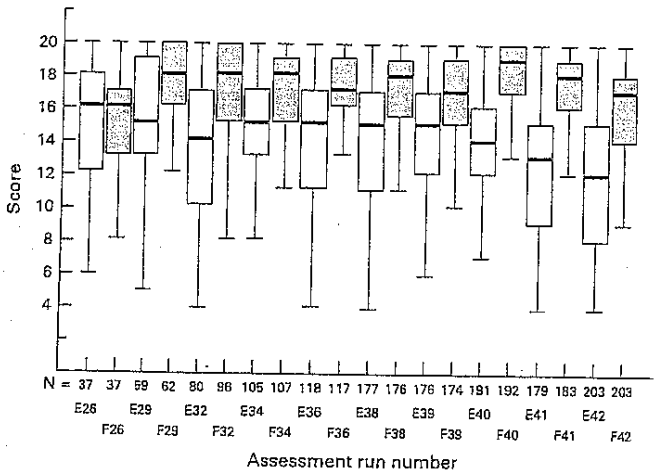

Figure 1 Box plot to show the relation between the UK NEQAS scores achieved by participants on the slides circulated by UK NEQAS (unshaded boxes labelled with run number and the letter "E") and the scores achieved by the same participants on their own in house control slides (shaded boxes labelled with the run number and the letter "F") between April 1994 and June 1998. The maximum score attainable was 20 and the minimum score attainable was 4. The bold line across each box indicates the median score. N, number of laboratories participating in each assessment run.

validating their technique. These studies are not referred to here because this would identify the laboratories concerned, and in so doing transgress the UK NEQAS code of practice, which confers anonymity to all participants.[15] The proportion of these laboratories achieving staining comparable with that of the UK NEQAS organising centre on the tumours circulated at assessment by the scheme and on their own in house tumours was determined for all the assessments between April 1994 and June 1998. For the in house tumours, the Wilcoxon two sample matched pairs signed ranks test was also used to test for differences in distributions between the marks awarded to the expert centres and those awarded to the UK NEQAS organising laboratory for staining of the same tumours.

COMPARISON OF THE SCORES AWARDED TO PARTICIPANTS FOR STAINING OF IN HOUSE TUMOURS AND THE STAINING OF UK NEQAS-ICC TUMOURS

The UK NEQAS scores for all the assessments conducted for ER and/or PR between April 1994 and June 1998 were compared by the use of box plots and by establishing the proportion of participants achieving acceptable staining at each assessment on the two sets of slides. The Wilcoxon two sample matched pairs test was used to test for differences in distributions. Spearman's rank coefficient was used to test for correlations between the results achieved by participants on the UK NEQAS tumours circulated at assessment and on their own in house tumours.

Results

VALIDATION OF ASSESSMENT THRESHOLDS AND OPTIMAL SENSITIVITY

Comparison of the scores of expert centres and those of the UK NEQAS organising centre on UK NEQAS tumours

For all the assessments (but one) conducted between April 1994—June 1998, all of the six expert centres demonstrated > 10% nuclei in tumours deemed by the organising centre to be hormonal receptor positive. The exception was assessment run 42, where two of the expert laboratories stained < 10% of invasive nuclei in a low expressing infiltrating ductal carcinoma, considered by the UK NEQAS organising centre to be ER positive.

Correlation with biochemical values of UK NEQAS tumours circulated at assessment

Of the tumours used for assessments by the scheme between April 1994 and June 1998, 17 of 21 had been initially tested before assessment using both the LBA and IHC. Of these, all were similarly receptor positive or negative with either assay, using an arbitrary threshold value of 10% or greater of invasive tumour nuclei stained by IHC or 10 fmol/mg protein or greater with the LBA, as designating receptor positive status.

Comparison of the scores of expert centres and those of the UK NEQAS organising centre on the same in house tumours

The Wilcoxon two sample matched pairs signed ranks test was used to test for differences in distributions between the scores awarded to the expert centres for the quality of immunostaining of their own in house tumour and the marks awarded to the UK NEQAS organising laboratory for staining of the same tumours. This test revealed no significant difference, either in the routine scores

Table 5 Assessment runs April 1994 to June 1998; differences in the distribution of marks for UK NEQAS and in house slides, measures of correlation between the two sets of scores, and details of the pass rates at each assessment

	Distribution differences (E) and (F)		Spearman's correlation (r')			Pass rate	
Run no.	Wilcoxon (Z)	p Value	r'	SE	p Value	NEQ (E)	In house (F)
26	−0.189	0.850	0.436	0.152	0.007	73%	81%
29	−3.612	< 0.0001	0.314	0.126	0.018	74%	94%
32	−5.705	< 0.0001	0.462	0.099	< 0.0001	60%	89%
34	−4.962	< 0.0001	0.485	0.092	< 0.0001	75%	90%
36	−7.333	< 0.0001	0.336	0.090	< 0.0001	70%	97%
38	−7.990	< 0.0001	0.280	0.070	< 0.0001	68%	91%
39	−6.464	< 0.0001	0.356	0.069	< 0.0001	73%	94%
40	−10.705	< 0.0001	0.307	0.068	< 0.0001	64%	97%
41	−10.345	< 0.0001	0.304	0.068	< 0.0001	56%	98%
42	−10.769	< 0.0001	0.327	0.064	< 0.0001	43%	87%

Pass rate refers to the proportion of participants achieving a score > 12 out of 20 (acceptable staining).
NEQ (E) refers to UK NEQAS slides circulated to each participating laboratory.
In house (F) refers to the participants own in house tumours.

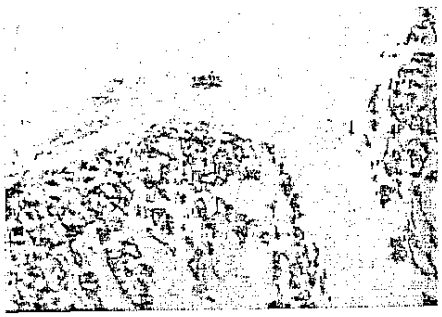

Figure 2 Results of immunohistochemistry for oestrogen receptors (ER) performed by the UK NEQAS organising laboratory on the low expressing (cytosol assay ER, 10 fmol/mg protein), ER positive, infiltrating ductal carcinoma circulated by UK NEQAS-ICC for assessment run 41.

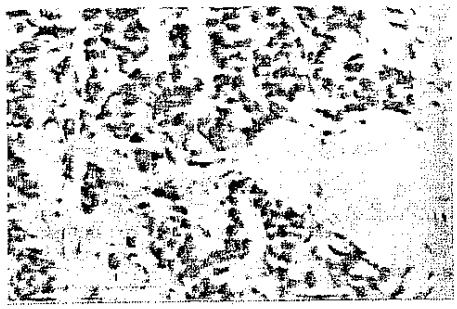

Figure 3 High power detail of the same section shown in fig 2.

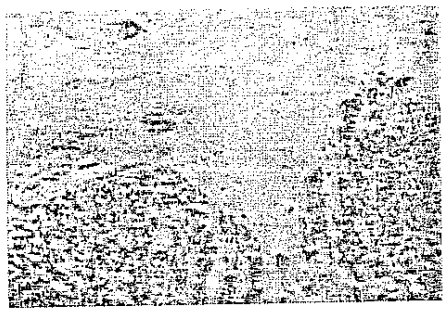

Figure 4 Results of immunohistochemistry for oestrogen receptors (ER) performed by laboratory "X" on the low expressing infiltrating ductal carcinoma shown in figs 2 and 3. The UK NEQAS score awarded to laboratory "X" for this staining was 8 out of 20. Laboratory "X" considered this tumour to be ER negative.

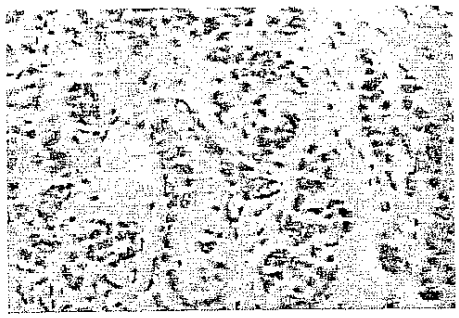

Figure 5 High power detail of the same section shown in fig 4.

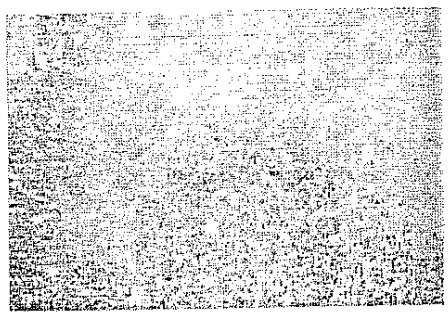

Figure 6 Results of immunohistochemistry for oestrogen receptors (ER) performed by laboratory "X" on the high expressing in house tumour submitted by laboratory "X" for run 41 (UK NEQAS score, 16 out of 20, quick score, 4).

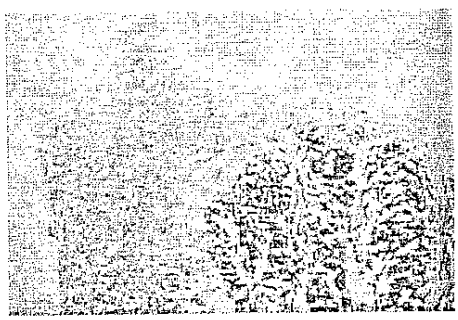

Figure 7 Results of immunohistochemistry for oestrogen receptors (ER) performed by the UK NEQAS organising laboratory on the tumour of laboratory "X" shown in fig 6 (UK NEQAS score, 20 out of 20, quick score, 7).

($Z = -0.170$; $p = 0.865$) or the quick scores generated for run 41 ($Z = -0.647$; $p = 0.518$).

COMPARISON OF THE UK NEQAS SCORES AWARDED AT ASSESSMENT FOR THE QUALITY OF IMMUNOSTAINING ON THE SLIDES CIRCULATED AND THOSE ACHIEVED BY THE SAME PARTICIPANTS ON IN HOUSE SLIDES FROM APRIL 1994 TO JUNE 1998

For nine out of the 10 assessments analysed, the median for the scores that participants achieved on in house tumours was higher than the median for the scores achieved on UK NEQAS tumours (fig 1). The Wilcoxon signed ranks test showed a highly significant difference in the distribution of marks for these nine runs ($p < 0.0001$; two tailed; table 5). The interquartile range was also frequently smaller for the scores achieved on in house sections, indicating less spread in the results. The proportion of participants achieving acceptable staining was always greater on the in house tissues than on the tumours circulated at assessment. These differences ranged from just 8% for run 26 to 44% for run 42 (table 5). However, Spearman's test showed a significant positive correlation between the routine scores awarded for the staining of UK NEQAS tumours and the staining of in house tumours. This relation is seen for all the assessments for ER/PR conducted between April 1994 and June 1998 (table 5, figs 2–9).

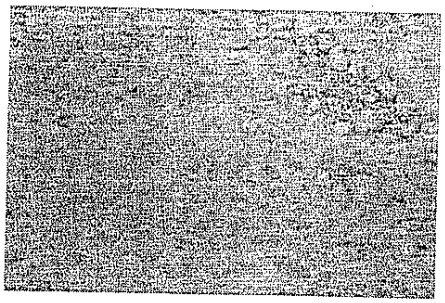

Figure 8 Results of immunohistochemistry for oestrogen receptors (ER) performed by laboratory "Y" on the high expressing in house tumour submitted by laboratory "Y" (UK NEQAS score, 13 out of 20, quick score, 3). Laboratory "Y" scored 6 out of 20 on the low expressing infiltrating ductal carcinoma shown in figs 2–5 and considered this tumour to be ER negative.

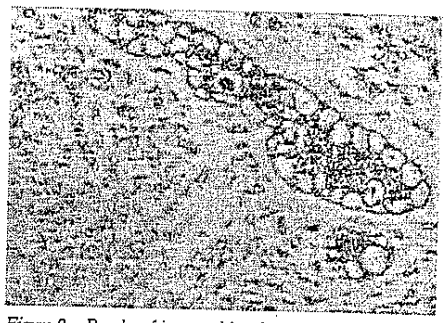

Figure 9 Results of immunohistochemistry (ICH) for oestrogen receptors (ER) performed by the UK NEQAS organising laboratory on the tumour of laboratory "Y" shown in fig 8 (UK NEQAS score, 20 out of 20, quick score, 7).

COMPARISON OF THE QUICK SCORES (RUN 41) ON PARTICIPANTS' STAINING OF THEIR OWN IN HOUSE TUMOURS AND THE QUICK SCORES OF THE SAME TUMOURS WHEN STAINED BY THE UK NEQAS-ICC ORGANISING CENTRE

The number of participants who submitted two unstained slides containing sections of breast tumour, along with their own laboratories immunostaining of that tumour, for run 41 was 152 (85% of the total returns). The remaining 26 participants (15%) did not provide unstained slides, or only ones of normal breast tissue. Table 2 details the types of tumours submitted, as described in the returned questionnaires. The initial Wilcoxon test indicated a highly significant difference between the quick scores for the in house tumours when stained by the participant and when stained by the organising laboratory ($Z = -6.814$; $p < 0.0001$; two tailed). Table 6 shows the degree of expression of the 152 tumours as evaluated using the quick score method on both the slides stained by the participants and duplicate slides stained by the organising laboratory. The proportion of cases designated as high expressers was 51.3% (n = 78) by the participants' staining and 80.9% (n = 123) by the UK NEQAS organising laboratory, with concordance on 72 cases (55.8%; κ coefficient, −0.091; p = 0.043). The proportion of cases designated as medium expressers was 27.6% (n = 42) by the participants' staining and 12.5% (n = 19) by the organising laboratory, with concordance on seven cases (13.0%; κ coefficient, −0.495; p < 0.0001). The proportion of cases designated as low expressers was 15.8% (n = 24) by the participants' staining and 4.6% (n = 7) by the organising laboratory, with concordance on just three cases (10.7%; κ coefficient, −0.316; p < 0.0001). Lastly, the proportion of cases designated as negative was 5.3% (n = 8) by the participants' staining and 2.0% (n = 3) by the organising laboratory, with concordance on three cases (37.5%; κ coefficient, not applicable). Overall, there was agreement on the degree of expression, as defined by the participants' staining and that of the organising laboratory, in 96 of the 152 cases (63.2%; κ coefficient, −0.026; p = 0.291).

Table 7 details the analysis of the tumours showing less than high ER expression by the participants' IHC. Of the 42 cases classified as medium ER expressers and obtaining quick scores of 4 and 5, 69% (p = 0.014) were shown to have higher expression when stained by the UK NEQAS organising laboratory and achieved quick scores that were higher by 2 marks or more. Of the 24 cases initially classified as low expressers and obtaining quick scores of 3 or 2, 83% (p = 0.001) were shown to have higher expression when stained by the UK NEQAS organising laboratory, with quick scores that were higher by 2 marks or more (table 7). Lastly, five of the eight cases classified by the participants' IHC as being ER negative and having quick scores of zero were shown to be ER positive when stained by the UK NEQAS reference laboratory, with one having a quick score of 2, two quick scores of 3, and two quick scores of 6.

EVALUATION OF THE EFFICIENCY OF THE UK NEQAS ORGANISING LABORATORY'S ROUTINE METHOD IN STAINING TUMOURS SUBMITTED FROM PARTICIPATING LABORATORIES

Using the standard UK NEQAS scoring system, for six of the seven assessment runs at

Table 6 The degree of oestrogen receptor (ER) expression of 152 tumours from 152 laboratories participating in assessment run 41, as defined by the participants' IHC assays and the UK NEQAS organising laboratory's IHC assay

Degree of ER expression	*Participant laboratory*		*Organising laboratory*		*Level of concordance*				
	n	*%*	*n*	*%*	*n*	*%*	*% of total*	*κ Value*	*p Value*
High expressing tumours (quick scores of 7 and 6)	78	51.3	123	80.9	72	55.8	47.4	−0.091	0.043
Medium expressing tumours (quick scores of 5 and 4)	42	27.6	19	12.5	7	13.0	4.6	−0.495	< 0.0001
Low expressing tumours (quick scores of 3 and 2)	24	15.8	7	4.6	3	10.7	2.0	−0.316	< 0.0001
Negative tumours (quick scores of 0)	8	5.3	3	2.0	3	37.5	2.0	NA	NA
All tumours (quick scores the same, within ± 1 mark)					96	63.2	63.2	−0.026	0.291

ICH, immunohistochemistry; NA, not applicable.

Table 7 The proportion of the 74 participating laboratories in house tumours submitted for run 41 that showed a higher degree of oestrogen receptor (ER) expression when tested by the UK NEQAS organising laboratory*

ER expression (participant)	n	ER expression (organising lab)	n	Quick score (participant)	Quick score (organising lab)	%	χ^2 Value	p Value
Medium	42	High	29	5 4	7 6 or 7	69%	6.095	0.014
Low	24	High/Medium	20	3 2	> 4 > 3	83%	10.667	0.001
Negative	8	High Low	5	0	6, 6 3, 3 2	63%	0.500	0.480

*Achieved a quick score that was greater by 2 marks or more.

which the UK NEQAS organising laboratory stained participants' in house slides, the median of the scores awarded to the organising laboratory was either equal or greater to the median of the scores awarded to the participants on these same in house slides (fig 10). The interquartile range for the scores achieved by the organising laboratory was smaller than the interquartile range of the participants scores on all seven occasions, indicating less spread in the results. The Wilcoxon signed rank test showed that the distribution of scores awarded to the UK NEQAS organising laboratory was significantly higher overall ($Z = -6.190$; $p < 0.0001$; two tailed), and individually in four of the seven assessments. For the remaining three runs, there was no significant difference between the two sets of scores (table 8).

Using the quick score method of evaluation, the technique used by the UK NEQAS organising laboratory was 99% efficient ($p < 0.0001$) in demonstrating the 152 different tumours submitted by participants for run 41, at either the first or second attempt. The overall efficiency achieved by participants using various different methods was 65% ($p < 0.0001$; table 9).

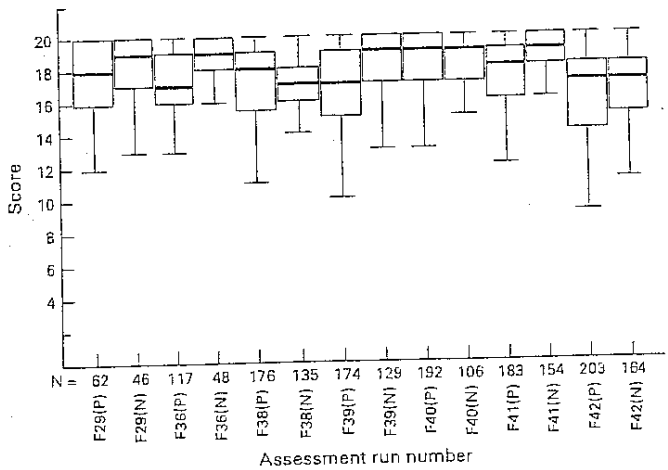

Figure 10 Box plot to compare the scores achieved by participants on their own in house breast tumours and the scores awarded for the UK NEQAS organising laboratory's immunostaining of duplicate sections of the same tumours. The plot shows seven assessment runs between February 1995 and April 1998 for which the UK NEQAS-ICC organising laboratory stained participants' in house slides. Run numbers are labelled F29–F42. The boxes labelled "P" refer to the participants' scores, whereas the boxes labelled "N" refer to the scores awarded to the UK NEQAS organising laboratory.

MEASURES OF ASSESSOR CONCORDANCE WHEN EVALUATING IHC SENSITIVITY BY THE QUICK SCORE METHOD FOR RUN 41

Goodman and Kruskal's γ statistic showed highly significant observer concordance between the assessors (AR, BJ) when using the quick score method to evaluate slides, with values of 0.949 ($p < 0.0001$) and 0.960 ($p < 0.0001$) for the staining of in house tumours by the participants and by the UK NEQAS organising laboratory, respectively.

Discussion

For the accurate assessment of the results achieved by different laboratories participating in EQA it is essential to validate the standards against which optimal sensitivity is defined. In our study, we have sought to validate these standards in various ways. Comparison of the results deemed to be optimal by the organising centre with those achieved by participants of the scheme who are known to have clinically validated their results has revealed many similarities, both in the proportion of the UK NEQAS tumours confirmed to be receptor positive and in the quick scores generated on the in house tumours for run 41. Of the tumours used for assessment by the scheme between April 1994 and June 1998, 81% had been initially tested using both the LBA and IHC. Of these, all were similarly receptor positive or negative with either assay, using a threshold value of 10% or greater of invasive tumour nuclei stained by IHC and 10 fmol/mg protein or greater with the LBA, as designating receptor positive status. Although the use of any threshold value is arbitrary, we have used this cut off point because of its use in several studies that correlate IHC receptor assay results with clinical and biochemical values.[16–20] We have also shown previously that this is the threshold most commonly used by the laboratories participating in UK NEQAS-ICC.[3] Also imperative to our study is the reproducibility of the methods of evaluation used to assess the quality of IHC. The reproducibility of the routine UK NEQAS scoring system was ensured at assessment by the checking of assessor concordance after every 20 slides. A highly significant degree of concordance with the quick score evaluations was confirmed by Goodman and Kruskal's γ statistic.

Between April 1994 and June 1998, UK NEQAS-ICC conducted 10 assessment runs for ER or PR. During this period, the pass rate on in house tumours remained high (81–97%), whereas that on the distributed UK NEQAS

Table 8 The differences in distributions of the routine marks awarded to participants for the staining of their own in house tumour and those awarded to the UK NEQAS-ICC organising laboratory for staining of the same tumour (February 1995 to June 1998)

Run no.	Distribution differences		Pass rate (scores > 12/20)	
	Wilcoxon (Z)	p Value	Participant	NEQAS organising lab*
29	−2.079	0.038	93%	100%
36	−2.752	0.006	97%	98%
38	−0.105	0.916	93%	98%
39	−4.416	< 0.0001	94%	97%
40	−0.606	0.544	99%	97%
41	−5.982	< 0.0001	97%	99%
42	−0.026	0.979	87%	90%

*The scores awarded to the UK NEQAS-ICC organising laboratory for immunochemical staining of participants in house tumours was performed for research purposes only.

slides fell, particularly for the later runs (runs 40–42). The reasons for these differences are twofold. First, many in house slides submitted for assessment contain just one tumour with high ER/PR expression, and thus are easier to stain than the UK NEQAS tumours. For example, for the one assessment (run 41) subjected to quick score evaluation, at least 47% of the in house tumours submitted were judged to be high expressing tumours by both the participant and the UK NEQAS organising laboratory (table 6). In contrast, the scheme has circulated slides from composite blocks comprising ER positive/PR positive tumours with progressively lower amounts of expression, particularly for runs 40, 41, and 42 (table 1).

The second reason for the difference does not appear to be possible biological differences in ER expression of the tumours used, but more probably differences in the way the tissues have been prepared. In cellular pathology, a multitude of variables affect a specimen the moment it is removed from the patient—for example, delay in fixation, type of fixation, duration of fixation, fixation temperature, paraffin wax processing schedule, and so on.[21–25] Individually, or in combination, these variables might have an effect on the efficiency and reliability of the immunohistochemical assay to demonstrate various antigens.[23] To minimise their effect, the technologist will have adjusted the methodology (over a period of time) to achieve consistently optimum results, according to his or her laboratory's expectations of the desired standard. When presented with tissue subjected to a different set of fixation and processing variables, as is the case with EQA and referred material, the efficiency of the in house method might fail to achieve the optimal result to varying degrees. This probably is the most likely reason for the differences seen between the scores that participants achieved for in house tumours and for the UK NEQAS slides, over the range of antigen expression included in the tumours examined.

Although the results achieved by participants on the UK NEQAS tumours were significantly different to those obtained on the in house tumours, Spearman's coefficient revealed that there was a significant positive correlation between the two sets of scores (table 5). To understand why there is a significant difference, and yet still a significant correlation, it is necessary to consider the relation between IHC assay sensitivity and the degree of receptor expression by the tumours under investigation. Suboptimal IHC assay sensitivity when used to stain ER or PR in a low expressing receptor positive tumour (for example, a UK NEQAS tumour) usually results in < 10% of invasive nuclei being stained and a failure at assessment. When applied to a high expressing in house tumour, this same degree of IHC sensitivity is also suboptimal because some invasive receptor positive nuclei that should be demonstrated are not. However, this is unlikely to result in < 10% of the tumour nuclei being stained, purely on the basis of the large number of receptor epitopes available. Consequently, participants who fail on the low expressing receptor positive UK NEQAS tumour tend to achieve lower scores than they should on their high expressing in house tumour. However, they do not usually fail (a score < 10 out of 20) if the proportion of nuclei demonstrated is equal to, or greater than, the designated 10% threshold. Conversely participants with high IHC assay sensitivity, who achieve a relatively high score on the low expressing receptor positive UK NEQAS tumour, tend to score very high marks on their in house tumour. This relation between assay sensitivity and the proportion of nuclei demonstrated in low expressing ER positive tumours and high expressing ER positive tumours is illustrated in figs 2–9.

The main implication of this correlation is that the IHC sensitivity achieved by laboratories on tumours circulated by UK NEQAS-ICC at assessment is a reflection of the sensitivity that the same laboratories achieve on tumours fixed and processed in their own laboratory (in house tumours). This is the first time evidence has been obtained in support of the view that the IHC results achieved on EQA material are accurate indicators of in house laboratory performance.

It has been shown previously that there is a significant positive correlation between the sensitivity achieved by the same laboratories on tumours of differing expression when these tumours are stained as a composite block.[3] In

Table 9 Relative efficiency of the immunohistochemical (IHC) assay of the UK NEQAS organising laboratory in achieving optimal demonstration of oestrogen receptors in 152 breast carcinomas, fixed and processed in 152 different laboratories

Efficiency of IHC assays	n	%	χ^2 Value	p Value
Proportion of 152 cases showing the same* or a greater level of expression** (UK NEQAS organising laboratory's assay)	150	99%	144.105	> 0.0001
Proportion of 152 cases showing the same† or a greater level of expression‡ (participating laboratories' assays)	99	65%	13.921	> 0.0001

*IHC assay results awarded the same quick score as participant (±1 mark)
**IHC assay results awarded a quick score that was greater by 2 marks or more than that awarded for participant's IHC.
†IHC assay results awarded the same quick score as UK NEQAS organising laboratory (±1 mark).
‡IHC assay results awarded a quick score that was greater by 2 marks or more than that awarded for UK NEQAS organising laboratory's IHC.

our study, we show that there is a similar correlation between suboptimal demonstration of in house tumours and suboptimal demonstration (< 10% nuclei staining) of relatively low expressing ER positive tumours circulated by an EQA scheme.

At present, there is a tendency to overlook the implications of suboptimal staining of tumours with relatively high amounts of ER or PR expression. Reiner and colleagues[12] and MacGrogan and colleagues[24] showed that patients whose carcinomas contained high numbers of hormone receptor positive cells (> 30%, > 50%, > 70%) had a better overall survival than those patients whose tumours had fewer receptor expressing cells. This was in general agreement with the results of Barnes and colleagues[13 14] and Walker *et al*, who showed a high rate of recurrence occurring in patients whose tumours contained high proportions of ER negative cells.[25] Hawkins has suggested that it is possible for ER IHC results to be divided into a minimum of four categories (negative, low, medium, and high) and still provide prognostic/predictive information similar to that provided, as a continuum, by a sensitive and quantitative biochemical assay.[26] Our present study found that 69% (p = 0.014) of tumours initially classified as medium expressers by participants' staining were subsequently shown to be high expressers when stained by the organising centre, and that 83% (p = 0.001) of those classified as low expressers were shown to be medium or high expressers (table 7). In addition, five of eight completely ER negative tumours were found to be ER positive. Two of these tumours were subsequently classified as high expressers, with quick scores of 6, and three low expressers, with quick scores of 3 or 2, although all with 10% or greater of the tumour nuclei staining. Obviously, the clinical importance of producing false negative staining is greater than that of suboptimally staining relatively high expressing ER positive tumours.[1] However, all these tumours were from patients who, according to the criteria of Reiner *et al*,[12] Barnes *et al*,[13 14] Walker *et al*,[25] and Hawkins,[26] would have a better overall survival than that predicted by the initial in house IHC assays.

This raises the important question as to whether anything short of optimal immunostaining is acceptable for ER testing, the results of which are likely to influence overall clinical management. A number of other immunocytochemical markers also appear to fall into this category, c-erb-2 being one of the best examples.[27–30]

Evaluation of the efficiency of the UK NEQAS reference laboratory's routine technique in analysing the in house tumours submitted between February 1995 and June 1998 using the standard UK NEQAS scoring system gives an efficiency ranging from 90% to 100% (table 8). The in depth analysis of run 41, using the quick score method to evaluate staining of 152 in house tumours, shows the organising laboratory's routine method to be 99% efficient in achieving an equivalent or greater sensitivity to that of the participating laboratory from where the tumours were submitted (table 9). This degree of sensitivity was achieved on the first, or second, attempt.

These results clearly indicate that the variations in fixation and paraffin wax processing that have been used by participating laboratories on the tumours submitted for assessment to date are not limiting factors preventing a different laboratory achieving a similar or greater degree of sensitivity for hormonal receptors. This is supported by a detailed analysis conducted by Williams *et al*, which showed that variations in immunostaining as a result of variations in fixation and processing regimens could be overcome by heat mediated antigen retrieval.[23]

The organising laboratory's standard technique uses routine commercial antibodies and reagents—that is, the same antibodies or reagents used by numerous participants at assessment and by some of the laboratories that have clinically validated their results (table 4). Although some participants use different clones to the ones used by the UK NEQAS organising laboratory, the technical comparisons performed after recent assessments do not show one clone to ER or PR to be significantly superior to another.[31] The same applies for the differing secondary detection systems. This leaves the efficiency of the heat mediated antigen retrieval step as the most likely factor preventing some participants from achieving optimal demonstration of hormonal receptors. Because all the clones to ER currently used at assessment, and most of those used to PR, necessitate heat mediated antigen retrieval for use on routinely processed tissues,[32–36] the degree of sensitivity ultimately achieved with these clones is directly dependant on how well the heat mediated antigen retrieval step has been performed. A multicentre study, involving 15 French laboratories, found the duration of the antigen retrieval step to be the crucial factor preventing some of the participating laboratories producing adequate results for ER on tissues fixed in a different laboratory.[37] Subsequent extension of the heat mediated antigen retrieval time allowed these laboratories to achieve optimal results. This supports the findings of our study, which suggest inefficiencies in the heat mediated antigen retrieval step might be the most important factor responsible for poor IHC demonstration of hormonal receptors. It is beyond the remit of our present investigation to provide an in depth technical analysis of the different variables that affect the efficiency of heat mediated antigen retrieval. To date, the few publications on this subject have mainly restricted their investigations to the efficiencies of the buffers used.[38–41] Other papers have compared the efficiency of the various heating methods available—for example, microwave ovens versus pressure cookers.[42 43] However, a comprehensive study is clearly required to compare the relative merits of all the different systems, particularly with respect to their efficiencies in the demonstration of low ER/PR positive breast tumours that have been fixed and processed under differing conditions. This would provide valuable information in the

formulation of recommended technical guidelines for optimal IHC demonstration of ER and PR. In turn, this should help in the standardisation of the technique, increase the sensitivity of detection for some laboratories, and ultimately help ensure that hormonal receptor positive cases are not erroneously reported as hormonal receptor negative.

We thank E Anderson, D Barnes, R Baumann, L Bobrow, V LeDoussal, and R Golouh for providing us with invaluable assistance, and all the participants of UK NEQAS-ICC, without whom this study would not have been possible.

1 Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet* 1998;351:1451–67.
2 Elledge RM, Osborne CK. Oestrogen receptors and breast cancer: it is time for individualised treatment based on oestrogen receptor status. *BMJ* 1997;314:1843–44.
3 Rhodes A, Jasani B, Barnes D, *et al.* The reliability of immunohistochemical demonstration of oestrogen receptors in routine practice: interlaboratory variance in the sensitivity of detection and evaluation of scoring systems. *J Clin Pathol* 2000;53:125–30.
4 Lambkin HA, Dunne P, McCarthy PM. Standardisation of estrogen-receptor analysis by immunohistochemistry: an assessment of interlaboratory performance in Ireland. *Applied Immunohistochemistry* 1998;6:103–7.
5 Barnes DM, Millis RR, Beex LVA, *et al.* Increased use of immunohistochemistry for oestrogen receptor measurement in mammary carcinoma: the need for quality assurance. *Eur J Cancer* 1998;34:1677–82.
6 Groupe d'Etude des Recepteurs Hormonaux par Immunohistochemie. FNCLCC/AFAQAP. Recommandations pour l'évaluation immunohistochimique des récepteurs hormonaux sur coupes en paraffine dans les carcinomes mammaires. *Ann Pathol* 1996:16:144–8.
7 Shahangian S. Proficiency testing in laboratory medicine: uses and limitations. *Arch Pathol Lab Med* 1998;122:15–30.
8 Naito HK, Kwak Y-S, Hartfiel JL, *et al.* Matrix effects on proficiency testing materials: impact on accuracy of cholesterol measurements in laboratories in the nation's largest hospital system. *Arch Pathol Lab Med* 1993;117:345–51.
9 Ross JW, Myers GL, Gilmore BF, *et al.* Matrix effects and the accuracy of cholesterol analysis. *Arch Pathol Lab Med* 1993;117:393–400.
10 Lasky FD. Achieving accuracy for routine clinical chemistry methods by using patient specimen correlations to assign calibrator values: a means of managing matrix effects. *Arch Pathol Lab Med* 1993;117:412–19.
11 Long T. Statistical power in the detection of matrix effects. *Arch Pathol Lab Med* 1993;117:387–92.
12 Reiner A, Nuemeister B, Spona J, *et al.* Immunocytochemical localisation of oestrogen and progesterone receptors and prognosis in human primary breast cancer. *Cancer Res* 1990;50:7057–61.
13 Barnes DM, Harris WH, Smith P, *et al.* Immunohistochemical determination of oestrogen receptors: comparison of different methods of assessment of staining and correlation with clinical outcome of breast cancer patients. *Br J Cancer* 1996;74:1445–51.
14 Barnes DM, Millis RR. Oestrogen receptors: the history, the relevance and the methods of evaluation. In: Kirkham N, Lemoine NR, eds. *Progress in pathology 2*. Edinburgh: Churchill Livingston, 1995:89–114.
15 Participation in UK NEQASs. In: *UK NEQAS report and directory*, 3rd ed. Sheffield: UK NEQAS Office, 1998:15.
16 Pellicer EM, Sundblad A. Evaluation of antibodies to oestrogen receptors. *Applied Immunohistochemistry* 1994;2:141.
17 De Mascarel I, Soubeyran I, MacGrogan G, *et al.* Immunohistochemical analysis of estrogen receptors in 938 breast carcinomas. Concordance with biochemical assay and prognostic significance. *Applied Immunohistochemistry* 1995;3:222–31.
18 Pertschuk LP, Feldman JG, Kim Y-D, *et al.* Estrogen receptor immunocytochemistry in paraffin embedded tissues with ER1D5 predicts breast cancer endocrine response more accurately than H222Spg in frozen sections or cytosol-based ligand-binding assays. *Cancer* 1996;77:2514–19.
19 Ferno M, Andersson C, Gilta F, *et al.* Oestrogen receptor analysis of paraffin sections and cytosol samples of primary breast cancer in relation to outcome to adjuvant tamoxifen therapy. *Acta Oncol* 1996;35:17–22.
20 Soubeyran I, Quenel N, Coindre J-M, *et al.* PS2: a marker improving prediction of response to neoadjuvant tamoxifen in post-menopausal breast cancer patients. *Br J Cancer* 1996;74:1120–5.
21 Larsson LI. Tissue preparation methods for light microscopic immunohistochemistry. *Applied Immunohistochemistry* 1993;1:2–16.
22 van Diest PJ, van Dam P, Henzen-Longmans SC, *et al.* A scoring system for immunohistochemical staining: consensus report of the task force for basic research EORTC-GCCG. *J Clin Pathol* 1997;50:801–4.
23 Williams JH, Mepham BL, Wright DH. Tissue preparation for immunocytochemistry. *J Clin Pathol* 1997;50:422–8.
24 MacGrogan G, Soubeyran, De Mascarel I, *et al.* Immunohistochemical detection of progesterone receptors in breast invasive ductal carcinomas: a correlative study of 942 cases. *Applied Immunohistochemistry* 1996;4:219–27.
25 Walker KJ, Bouzubar N, Robertson J, *et al.* Immunocytochemical localization of estrogen receptor in human breast cancer. *Cancer Res* 1998;48:6517–22.
26 Hawkins RA. How best to express oestrogen receptor activity. *Br J Cancer* 1996;74:1329–30.
27 Allred DC, Harvey JM, Berardo M, *et al.* Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol* 1998;11:155–68.
28 Press MF, Hung G, Godolphin W, *et al.* Sensitivity of HER-2/neu antibodies in archival tissue samples: potential source of error in immunohistochemial studies of oncogene expression. *Cancer Res* 1994;54:2771–7.
29 Pegram MD, Lipton A, Hayes DF, *et al.* Phase II study of receptor-enhanced chemosensitivity using recombinant humanized anti-p185$^{HER2/neu}$ monoclonal antibody plus cisplastin in patients with HER2/neu-overexpressing metastatic breast cancer refactory to chemotherapy treatment. *J Clin Oncol* 1998;16:2659–71.
30 Seidman AD, Beselga J, Yao T-J, *et al.* HER2/neu-overexpression and clinical taxane sensitivity: a multivariate analysis in patients with metastatic breast cancer (MBC). *Proceedings of the American Society of Clinical Oncology* 1996;15:104.
31 Rhodes A. UK NEQAS for immunocytochemistry: review of run 44. *J Cell Pathol* 1999;4:114–18.
32 Saati TAL, Clamens S, Cohen-Knafo E, *et al.* Production of monoclonal antibodies to human oestrogen receptor (ER) protein using recombinant ER (RER). *Int J Cancer* 1993;55:651–4.
33 Hendricks JB, Wilkinson EJ. Comparison of two antibodies for evaluation of estrogen receptors in paraffin-embedded tumours. *Mod Pathol* 1993;6:765–70.
34 Sannino P, Shousha S. Demonstration of oestrogen receptors in paraffin sections of breast carcinoma using microwave oven processing. *J Pathol* 1993;170(suppl):201.
35 Kell DL, Kamel OW, Rouse RV. Immunohistochemical analysis of breast carcinoma estrogen and progesterone receptors in paraffin-embedded tissue, correlation of clones ER1D5 and 1A6 with a cytosol-based hormone receptor assay. *Applied Immunohistochemistry* 1993;1:275–81.
36 Leong ASY, Milios J. Comparison of antibodies to oestrogen and progesterone receptors and the influence of microwave antigen retrieval. *Applied Immunohistochemistry* 1993;1:282–8.
37 Balaton AJ, Mathieu MC, Le Doussal V (on behalf of the Groupe d'Etude des Recepteurs Hormonaux par Immunohistochimie/FNCLCC). Optimization of heat-induced epitope retrieval for estrogen receptor determination by immunohistochemistry on paraffin sections: results of a multi-centric comparative study. *Applied Immunohistochemistry* 1996;4:259–63.
38 Morgan JM, Navabi H, Schmid KW, *et al.* Possible role of tissue bound calcium ions in citrate-mediated high temperature antigen retrieval. *J Pathol* 1994;174:301–7.
39 Pileri SA, Roncador G, Ceccarelli C, *et al.* Antigen retrieval techniques in immunohistochemistry: comparison of different methods. *J Pathol* 1997;183:116–23.
40 Cattoretti G, Suurmeijer AJH. Antigen unmasking on formalin-fixed paraffin embedded tissues using microwaves: a review. *Adv Anat Pathol* 1995;2:2–9.
41 Leong ASY, Milios J, Leong FJ. Epitope retrieval with microwaves: a comparison of citrate buffer and EDTA with three commercial solutions. *Applied Immunohistochemistry* 1996;4:201–7.
42 Norton AJ, Jordan S, Yeomans P. Brief, high temperature heat denaturation (pressure cooking): a simple and effective method of antigen retrieval for routinely processed tissues. *J Pathol* 1994;173:371–9.
43 Miller K, Auld J, Jessup E, *et al.* Antigen unmasking in formalin fixed routinely processed paraffin wax-embedded sections by pressure cooking: a comparison with microwave oven heating and traditional methods. *Adv Anat Pathol* 1995;2:60–4.

J Clin Pathol 2000;53:125–130

Reliability of immunohistochemical demonstration of oestrogen receptors in routine practice: interlaboratory variance in the sensitivity of detection and evaluation of scoring systems

A Rhodes, B Jasani, D M Barnes, L G Bobrow, K D Miller

Abstract

Aims—To investigate interlaboratory variance in the immunohistochemical (IHC) detection of oestrogen receptors so as to determine the rate of false negatives, which could adversely influence the decision to give adjuvant tamoxifen treatment.

Methods—To ensure that similar results are obtained by different institutions, 200 laboratories from 26 countries have joined the UK national external quality assessment scheme for immunocytochemistry (NEQAS-ICC). Histological sections from breast cancers having low, medium, and high levels of oestrogen receptor expression were sent to each of the laboratories for immunohistochemical staining. The results obtained were evaluated for the sensitivity of detection, first by estimating threshold values of 1% and 10% of stained tumour cells, and second by the Quick score method, by a panel of four assessors judging individual sections independently on a single blind basis. The results were also evaluated using participants' own threshold values.

Results—Over 80% of laboratories were able to demonstrate oestrogen receptor positivity on the medium and high expressing tumours, but only 37% of laboratories scored adequately on the low expressing tumour. Approximately one third of laboratories failed to register any positive staining in this tumour, while one third showed only minimal positivity.

Conclusions—There is considerable interlaboratory variability, especially in relation to the detection of breast cancers with low oestrogen receptor positivity, with a false negative rate of between 30% and 60%. This variability appears to be caused by minor differences in methodology that may be rectified by fine adjustment of overall technique.

(*J Clin Pathol* 2000;53:125–130)

Keywords: immunohistochemistry; oestrogen receptors; interlaboratory variation

Department of Histopathology, University College London Medical School, Rockefeller Building, University Street, London WC1E 6JJ, UK
A Rhodes
K D Miller

Department of Pathology, University of Wales College of Medicine, Cardiff, UK
B Jasani

Hedley Atkins ICRF Breast Pathology Laboratory, Guy's Hospital, London SE1, UK
D M Barnes

Department of Histopathology, Addenbrookes NHS Trust, Cambridge, UK
L B Bobrow

Correspondence to:
Mr Rhodes
email: rmkdhcr@ucl.ac.uk

Accepted for publication
6 August 1999

The importance of establishing the oestrogen receptor status of tumours for the treatment of women with breast cancer has recently been emphasised.[1] The authors concluded that the fundamental question to be asked when predicting the likely outcome for a particular woman receiving adjuvant tamoxifen treatment is not whether she is young or old, with or without nodal involvement, or receiving chemotherapy—but whether or not her tumour is completely oestrogen receptor negative. Oestrogen receptor status is now often established by an immunohistochemical (IHC) test employing monoclonal antibodies.[2–4] This assay has been shown to be at least as sensitive as the biochemical ligand binding assay[5 6] and has the advantages of being applicable to small tumours and Tru-Cut biopsy samples, and of allowing only tumour cells to be assessed for oestrogen receptor status. The IHC assay can be conducted inexpensively[7 8] on routinely processed tissue sections, with no need for specialised equipment. Consequently in many countries IHC analysis has become the chosen technique for establishing oestrogen receptor status in a routine pathology setting.[9 10]

In view of the increasing use of the oestrogen receptor IHC assay, it is vital that good quality assurance procedures are in place to assess the quality of the assays carried out by different laboratories.[10] The United Kingdom national external quality assessment scheme for immunocytochemistry[11] (UK NEQAS-ICC) currently assesses the quality of many immunohistochemical techniques carried out in the majority of UK clinical laboratories and in various laboratories based outside the United Kingdom. Since April 1994 the scheme has provided an external quality assessment (EQA) programme for the demonstration of oestrogen and progesterone receptors on routinely processed breast tumours.

In this paper we report on the degree of variability between 200 laboratories in demonstrating oestrogen receptors by immunohistochemistry on the same cases. The main aim of the study was to establish the proportion of laboratories able to demonstrate oestrogen receptors reliably in a weakly positive tumour, as there is a danger that these tumours could be erroneously reported as negative if the IHC assay is not of adequately high sensitivity.

Methods

Laboratories participating in the UK NEQAS-ICC programme for steroid hormone receptors (table 1) were sent two unstained slides containing histological tissue sections of formalin fixed and paraffin processed breast tumours showing different levels of receptor expression. Included in the composite tumour block, comprising three different oestrogen receptor positive infiltrating ductal carcinomas

Table 1 Countries with laboratories participating in the UK NEQAS-ICC programme for steroid hormone receptors

Country	No of Labs
Australia	6
Austria	1
Belgium	3
Canada	1
Denmark	3
Finland	1
France	25
Germany	4
Greece	1
Hong Kong	1
Hungary	1
Ireland	14
Malaysia	1
Malta	1
New Zealand	1
Norway	1
Portugal	1
Saudi Arabia	1
Singapore	2
South Africa	1
Slovenia	1
Sultanate of Oman	1
Spain	1
Sweden	9
Switzerland	9
United Kingdom	138
USA	2

Table 2 Evaluation of the staining achieved by the organising centre and participants who are known to have validated their immunohistochemical assay by published clinical studies

	Low expressor (X)			Medium expressor (Y)			High expressor (Z)			
Lab‡	Quick score	% nuclei positive	Result*	Quick score	% nuclei positive	Result*	Quick score	% nuclei positive	Result*	Cut off used*
a	3	>10	+	3	>10	+	6	>10	+	20%
b	2	>10	+	5	>10	+	6	>10	+	H score†
c	2	>10	−	7	>10	+	6	>10	+	15%
d	0	0	−	3	>10	+	6	>10	+	10%
e	2	>10	+	5	>10	+	6	>10	+	2%
f	2	>1	+	5	>10	+	6	>10	+	5%
g¶	3	>10	+	5	>10	+	6	>10	+	10%

*The result using participants' own threshold value.
†H score – cut off value of 50.
‡In order to preserve anonymity, the laboratories have been coded by letters.
¶Results of the initial testing conducted by the UK NEQAS-ICC organising laboratory.

(X, Y, and Z), was some normal glandular breast tissue which acted as an internal control. In order to ensure that all sections contained a similar proportion of oestrogen receptor positive cells, every 100th section was immunostained for oestrogen receptors by the organising laboratory. Each participant was asked to demonstrate oestrogen receptors and to return the best stained slide, along with their own in-house control slide and a completed questionnaire giving methodological details (including details of the threshold value used by the laboratory), to the UK NEQAS-ICC coordinating centre for assessment. An expert panel of four, comprising pathologists (BJ, LB) and biomedical and clinical scientists (AR, DB), examined the slides and assessed the quality of the IHC assay performed by each laboratory.

METHODS OF EVALUATION

For the purposes of the present study, the "Quick" score method of assessment[12 13] was used to assess the range of immunostaining performed by the participating laboratories. With this method the intensity of the immunohistochemical reaction as viewed under the light microscope was recorded as follows: 0, negative (no staining of any nuclei even at high magnification); 1, weak (only visible at high magnification); 2, moderate (readily visible at low magnification); 3, strong (strikingly positive even at low power magnification). The proportion of tumour nuclei showing positive staining was also recorded as: 0 (none); 1 (approximately 1–25%); 2 (26–50%); 3 (51–75%); or 4 (76–100%). The score for intensity was added to the score for proportion, giving the Quick score, with a range of 0–7 for each individual tumour.

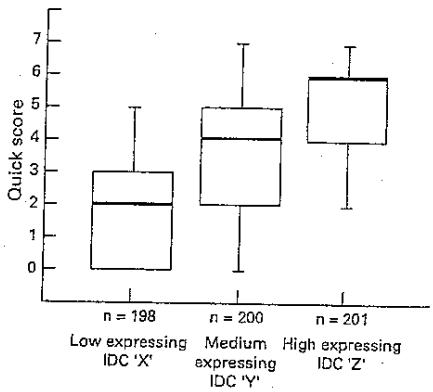

Figure 1 Distribution of the results of the "Quick" score evaluation conducted on the three infiltrating ductal carcinomas (IDC), X, Y, Z, used at assessment. The bold line represents the median score, the bottom and top of the boxes, the 1st and 3rd quartiles, respectively, and the range bars, the lowest and highest scores, respectively. The slightly different numbers for the three tumours reflect loss of tissue from the microscope slides; n, number of participating laboratories.

The proportion of cells stained in each tumour in the composite block was also recorded as either 0, ⩾ 1% but < 10%, or ⩾ 10%. The absence or presence of staining of the nuclei of non-neoplastic ducts in adjacent tissue was also recorded. This served as an internal control. Slides which failed to show any staining in the normal internal control or which showed excessive non-specific immunostaining in the stromal component were deemed unsatisfactory and were excluded from statistical analysis.

OESTROGEN RECEPTOR STATUS OF THE REFERENCE TUMOURS X, Y, AND Z

From the UK NEQAS participants, six were identified as having published clinical studies relating oestrogen receptor positivity to tamoxifen treatment. These studies are not referred to in this paper as this would identify the laboratories concerned and in so doing transgress the UK NEQAS code of practice which confers anonymity to all participants.[14] The assessment results from these laboratories and the initial testing performed by the organising centre were used to establish the oestrogen receptor status of the tumours X, Y, and Z, and are recorded in table 2. Additional confirmation of the oestrogen receptor positive status was provided in the form of the results of previous biochemical assays conducted on these cases.

STATISTICAL ANALYSIS

Median values were established for the Quick scores achieved by participating laboratories on the infiltrating ductal carcinomas (IDC) labelled X, Y, and Z. Spearman's rank coefficient was used to test for correlation between the level of sensitivity achieved on the three different tumours and differences in the proportion of laboratories showing oestrogen receptor positivity at various threshold values was tested by means of the χ^2 test. Kendall's coefficient of

Table 3 Correlation of the Quick scores achieved by 190 laboratories on the infiltrating ductal carcinoma (IDC) with low oestrogen receptor expression with the scores achieved by the same laboratories on the IDC with medium oestrogen receptor expression

Quick scores, low expressor (X)	*Quick scores, medium expressor (Y)*							
	0.00	*2.00*	*3.00*	*4.00*	*5.00*	*6.00*	*7.00*	*Total*
0.00	19	18	4	12	5	2	1	61
2.00	2	20	9	21	16	3	3	74
3.00		3	3	6	14	3	8	37
4.00			1	3	6	1	4	15
5.00						2	1	3
6.00								
7.00								
Total	21	41	17	42	41	11	17	190

The Quick scores for the 4% of laboratories where staining was recorded as "uninterpretable" have been removed from the analysis.
Spearman correlation = 0.557, standard error = 0.053; significance = $p < 0.0001$.

Table 4 Correlation of the Quick scores achieved by 190 laboratories on the infiltrating ductal carcinoma (IDC) with low oestrogen receptor expression with the scores achieved by the same laboratories on the IDC with high oestrogen receptor expression

Quick scores, low expressor (X)	*Quick scores, high expressor (Z)*							
	0.00	*2.00*	*3.00*	*4.00*	*5.00*	*6.00*	*7.00*	*Total*
0.00		16	5	17	7	14	2	61
2.00	1	1	2	13	19	33	5	74
3.00				3	5	25	4	37
4.00					1	9	5	15
5.00						2	1	3
6.00								
7.00								
Total	1	17	7	33	32	83	17	190

The Quick scores for the 4% of laboratories where staining was recorded as "uninterpretable" have been removed from the analysis.
Spearman correlation = 0.528, standard error = 0.055; significance = $p<0.0001$.

Table 5 Correlation of the Quick scores achieved by 190 laboratories on the infiltrating ductal carcinoma (IDC) with high oestrogen receptor expression with the scores achieved by the same laboratories on the IDC with medium oestrogen receptor expression

Quick scores, medium expressor (Y)	*Quick scores, high expressor (Z)*							
	0.00	*2.00*	*3.00*	*4.00*	*5.00*	*6.00*	*7.00*	*Total*
0.00	1	12	3	4		1		21
2.00		4	3	14	9	11		41
3.00		1	1	3	4	8		17
4.00				8	10	23	1	42
5.00				4	7	24	6	41
6.00					1	6	4	11
7.00					1	10	6	17
Total	1	17	7	33	32	83	17	190

The Quick scores for the 4% of laboratories where staining was recorded as "uninterpretable" have been removed from the analysis.
Spearman correlation = 0.661, standard error = 0.044; significance = $p<0.0001$.

concordance (Kendall's W) was used to determine the level of agreement between assessors.

Results

When the staining results were analysed by the Quick score (fig 1) the median scores were 2 for tumour X (low oestrogen receptor expressor), 4 for tumour Y (medium oestrogen receptor expressor), and 6 for tumour Z (high oestrogen receptor expressor).

Spearman's rank coefficient showed a highly significant positive correlation between the level of sensitivity achieved by individual laboratories on the tumours of differing oestrogen receptor expression (tables 3–6).

When only the proportion of nuclei stained in the tumours was evaluated, 99.0% of participants demonstrated 10% or more of the nuclei of the high expressor, while 99.5% demonstrated 1% or more. For the medium expressor, 84.5% demonstrated 10% or more of nuclei, while 88.0% demonstrated 1% or more. For the low expressor, 37.3% demonstrated 10% or more of tumour nuclei, with 66.3% demonstrating 1% or more (fig 2). When the threshold values used by participants to designate a tumour as either oestrogen receptor positive or oestrogen receptor negative were used, the proportion of assays which would have recorded the high, medium, and low expressing tumours as oestrogen receptor positive fell to 98.0%, 80.0%, and 32.8%, respectively (for all evaluations, $p < 0.0001$, two tailed). Approximately one third of participants failed to demonstrate any tumour nuclei at all in the low expressor (fig 3).

Kendall's coefficient of concordance revealed a significant level of concordance between assessors in the evaluation of slides (Kendall's $W = 0.014$, $p = 0.040$).

Discussion

With immunocytochemistry for oestrogen receptors, it is a commonly observed phenomenon that the first sign of a fall in sensitivity of the IHC technique is a diminution in staining intensity, and this is followed by a reduction in the proportion of tumour nuclei demonstrated. For this reason, three methods of evaluation were used to assess one or both of these criteria.

The Quick score method was included on the basis that it was a previously validated system for evaluating oestrogen receptor status of each of the tumours,[12 13] in conjunction with a simple but clinically validated 10% oestrogen receptor positive threshold.[15–19] This threshold is commonly used by many laboratories to differentiate between breast tumours which are likely to respond to tamoxifen treatment and those which are not (table 7). We also included the recently recommended 1% threshold value, considered to be clinically relevant by some

Table 6 Comparison of the Quick scores achieved on the infiltrating ductal carcinoma (IDC) with medium oestrogen receptor expression with the proportion of nuclei stained in the IDC with low oestrogen receptor expression

% Nuclei stained, low expressor (X)	*Quick scores, medium expressor (Y)*							
	0.00	*2.00*	*3.00*	*4.00 (median)*	*5.00*	*6.00*	*7.00*	*Total*
No nuclei stained	18	17	3	12	5	2	1	58
Some nuclei stained but less than 10%	2	16	7	20	11		2	58
10% or greater		7	7	10	25	9	14	72
Total	20	40	17	42	41	11	17	188

Spearman correlation = 0.539, standard error = 0.057; $p < 0.0001$.

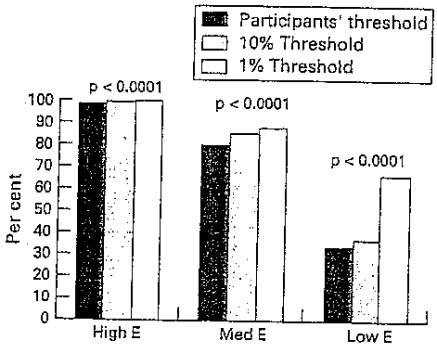

Figure 2 The proportion of laboratories from which immunohistochemistry reliably demonstrated the intraductal carcinomas X, Y, Z as being oestrogen receptor positive. χ^2 values were as follows:

High oestrogen receptor expressing tumour:
Proportion demonstrating oestrogen receptor positivity using own threshold value: 98.0% (n = 176), χ^2=167.201.*
Proportion demonstrating 10% or more nuclei: 99.0% (n = 198), χ^2 = 192.080.
Proportion demonstrating 1% or more nuclei: 99.5% (n = 199), χ^2 = 196.020.

Medium oestrogen receptor expressing tumour:
Proportion demonstrating oestrogen receptor positivity using own threshold value: 80.0% (n = 178), χ^2 = 65.528.*
Proportion demonstrating 10% or more nuclei: 84.5% (n = 143), χ^2 = 95.220.
Proportion demonstrating 1% or more nuclei: 88.0% (n = 176), χ^2 = 115.520.

Low oestrogen receptor expressing tumour:
Proportion demonstrating oestrogen receptor positivity using own threshold value: 32.8% (n = 58), χ^2 = 21.023. Proportion demonstrating 10% or more nuclei: 37.3.0% (n = 73), χ^2 = 12.755.*
Proportion demonstrating 1% or more nuclei: 66.3% (n = 130), χ^2 = 20.898.

**Where no threshold value was given it was assumed that, regardless of the value used, (a) no nuclear staining will always represent an oestrogen receptor negative status; (b) staining awarded a Quick score of >4 will always represent an oestrogen receptor positive status. p Values refer to all three threshold values and are two tailed.*

workers.[6 8 21] Positive IHC assays using this cut off value has been associated with a large improvement in disease-free survival in patients receiving adjuvant tamoxifen (~30% at five years), with nearly one tenth of all oestrogen receptor positive patients investigated having only 1–10% of oestrogen receptor positive nuclei in their tumours.[21] Lastly, the oestrogen receptor status of the tumours was evaluated using the threshold values employed in the participants' own laboratories.

The overall analysis showed that while the majority of laboratories had little difficulty in demonstrating the tumours with high oestrogen receptor expression, a significant proportion (62.7%, p < 0.0001) failed to demonstrate 10% or more of the nuclei of the low expressor (fig 2). Interestingly there was a three way split in these results, with approximately one third of the assays staining no nuclei at all, one third staining some nuclei but less than 10%, and one third staining 10% or more (fig 3). Clearly with such wide interlaboratory variation in the assay sensitivity, a 10% threshold value used in one laboratory is unlikely to be applicable in another. The same would apply to the Quick score, with relatively large interquartile ranges of 0–3 for the low expressing carcinoma and 2–5 for the medium expressing carcinoma (fig 1). This interlaboratory variance is not caused by inconsistencies at the time of evaluation, as the level of agreement between individual assessors was good, as it was in a previous study,[13] but instead it was caused by variations in the sensitivity of the IHC method. Consequently the oestrogen receptor status (positive or negative) of these tumours and the predicted response to adjuvant tamoxifen treatment are considerably influenced by which laboratory has performed the assay.

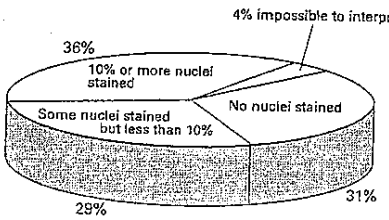

Figure 3 The proportions of 200 laboratories from which immunohistochemistry demonstrated either no nuclei, some nuclei but less than 10%, and 10% of more, in the "low" oestrogen receptor expressing infiltrating ductal carcinoma.

The choice of threshold value could compensate for the slightly differing levels of IHC sensitivity observed between laboratories. It has been recommended that threshold values should always be gauged against clinical outcome.[13] Consequently laboratories with different assay sensitivities could theoretically obtain the same result on the same tumour, as long as individual threshold values have been carefully adjusted to clinical outcome (assuming a similar proportion of patients respond to adjuvant tamoxifen treatment in different populations). In order to make allowance for this, the oestrogen receptor status of the tumours used in the present study was also established, using the participants' own threshold values. The fact that the interlaboratory variance persisted and if anything increased when the laboratories' chosen threshold values were used (fig 2) indicates that these would not compensate entirely for the differences in sensitivity observed between laboratories.

The positive oestrogen receptor status of the three tumours used in this study, as determined by the organising centre, is ratified by the results of the biochemical analyses. Furthermore the results of all six of the expert laboratories known to use clinically validated oestrogen receptor assays indicated that the high and medium expressing tumours were oestrogen receptor positive, and four of the six agreed that the low expressing tumour was positive, using either their own threshold value or a 10% cut off. Yet further support for the view that all the tumours were oestrogen receptor positive was obtained indirectly from the significant correlation between the Quick scores achieved on the medium expressing tumour and the proportion of nuclei stained on the low expressing tumour (table 6). Approximately 70% of laboratories who achieved higher than the median Quick score of 4 on the medium expressing tumour demonstrated ≥ 10% of nuclei in the low expressing tumour. In contrast only 18% of those scoring less than 4 on the medium expresser demonstrated ≥ 10% of nuclei in the

Table 7 The methods of evaluation for oestrogen receptors used by UK NEQAS-ICC participants

Threshold value	*Frequency (No of labs)*	*%*
10% or more of tumour nuclei demonstrated[13–19]	106	50.0
Histo (H) score[20 22 23]	17	8.1
20% or 25% and more of tumour nuclei demonstrated	13	6.1
5% or more of tumour nuclei demonstrated	10	4.7
Quick score[15 17 22]	6	2.8
1% or more of tumour nuclei demonstrated[6 8 21]	3	1.4
Category score[13 22]	2	0.9
50% or more of tumour nuclei demonstrated	2	0.9
Values known but each account for less than 0.9% of total	8	3.8
Unknown (information not provided by participant)	45	21.2
Total	212	100

low expresser. Consequently a Quick score of less than the median value on a relatively high oestrogen receptor expressing tumour correlates with < 10% of nuclei staining on the low expresser, while a Quick score greater than the median correlates with ≥ 10% of nuclei staining on the low expresser.

The significant positive correlation between the level of sensitivity achieved by the same laboratories on the different tumours (tables 3–6) indicates that less than optimum sensitivity on relatively high expressing tumours equates to poor and sometimes inadequate demonstration of very low expressers. This is because in the low expressing tumours the amount of oestrogen receptor present is much closer to the designated threshold value, and a slight fall in sensitivity can result in the number of nuclei demonstrated being below this value.

Interestingly, of all the threshold values investigated, the recently recommended 1% threshold value[6 8 21] would result in a significant number of laboratories recording all three categories of tumour used in the present study, including the low expressing intraductal carcinoma, as oestrogen receptor positive (fig 2). The reason for this is that the 1% threshold alone would make sufficient allowance for the observed interlaboratory variation in IHC sensitivity. However, it must be emphasised that a 1% threshold could result in detection of a higher proportion of oestrogen receptor positive unresponsive tumours from laboratories using a more sensitive method of detection. Hence, as emphasised by Barnes *et al*, a reasonable balance must be achieved between sensitivity and specificity in order to more accurately predict the proportion of patients likely to benefit from hormone treatment.[10 13]

Once improvement in interlaboratory consistency in carrying out the IHC assay has been achieved, it will be possible to address two outstanding questions: first the "accuracy" of the assay, and second the choice of cut off point. In the past, when the cytosol assay was used, there was always a small number of oestrogen receptor "negative" cases that responded to endocrine treatment. It is not clear whether these were genuinely negative or whether there was insufficient tumour in the sample used to prepare the cytosol. The advantage of IHC is that the presence of tumour can be confirmed by eye. Conversely there are also unresponsive oestrogen receptor positive cases. This may happen because the tumour burden is so great that treatment is ineffective or it could reflect the presence of oestrogen receptor in normal epithelial cells; again negative staining of tumour cells can now be checked visually.

The question of the cut off values remains a topic of much discussion. These may well differ according to whether the assay is to provide prognostic or predictive information. Much experience has been gained from the treatment of metastatic disease but less is available from the adjuvant setting. The increased use and improvements in quality of IHC will enable critical examination of relations between different cut off points and response. This in turn will lead to a consensus as to the "correct" values and make comparisons between studies easier.

CONCLUSIONS

In this study, we have investigated the ability of laboratories participating in the United Kingdom NEQAS-ICC for hormonal receptors to demonstrate positive staining in mammary carcinomas shown by experienced laboratories to have an oestrogen receptor positive status. The difficulties experienced by some laboratories in achieving this goal are highlighted and have since been communicated to the participants, with special emphasis on the false negative results. The reasons for the underachievement by some laboratories may lie in variations in the sensitivity of the overall staining technique. The sensitivity of the IHC assay is determined by several variables, which include the quality and concentration of the primary antibody used, the power of the antigen retrieval, and the secondary detection systems and quality of the fixation of the tissue. A superficial comparison of these variables among the assay systems used by different laboratories has failed to reveal any that are predominantly responsible for the differences observed. However, quality assurance is a continual process and the ongoing cycle of assessment runs, currently in progress for the oestrogen receptor IHC assay, may show that a combination of these factors is responsible for the observed interlaboratory variance. Better optimisation of such factors is needed to ensure that the results produced in one laboratory are comparable with those produced in another. This in turn may allow the chosen set of prognostic/therapeutic threshold values for selecting treatment for both primary and metastatic breast cancers to be safely applicable in the majority of laboratories offering the specialist oestrogen receptor IHC assay service.

We thank Elizabeth Anderson, Andre Balaton, Rudolf Baumann, and Rastko Golouh for providing us with invaluable assistance, and all the participants of UK NEQAS-ICC without whom this study would not have been possible.

1 Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet* 1998;351:1451–67.
2 Saati TAL, Clamens S, Cohen-Knafo E, *et al*. Production of monoclonal antibodies to human oestrogen receptor (ER) protein using recombinant ER (RER). *Int J Cancer* 1993;55:651–4.
3 Hendricks JB, Wilkinson EJ. Comparison of two antibodies for evaluation of estrogen receptors in paraffin-embedded tumours. *Mod Pathol* 1993;6:765–70.

4 Sannino P, Shousha S. Demonstration of oestrogen receptors in paraffin sections of breast carcinoma using microwave oven processing. *J Pathol* 1993;170(suppl):201.
5 Keil DL, Kamel OW, Rouse RV. Immunohistochemical analysis of breast carcinoma estrogen and progesterone receptors in paraffin-embedded tissue, correlation of clones ER1D5 and 1A6 with a cytosol-based hormone receptor assay. *Appl Immunohistochem* 1993;1:275–81.
6 Clark GM, Harvey JM, Osborne CK, *et al.* Estrogen receptor status determined by immunohistochemistry is superior to biochemical ligand binding assay for evaluating breast cancer patients. *Proc Am Soc Clin Oncol* 1997;16:129.
7 Leong ASY, Milios J. Comparison of antibodies to oestrogen and progesterone receptors and the influence of microwave antigen retrieval. *Appl Immunohistochem* 1993;1:282–8.
8 Elledge RM, Osborne CK. Oestrogen receptors and breast cancer: it is time for individualised treatment based on oestrogen receptor status. *BMJ* 1997;314:1843–44.
9 Elias JM. A phoenix arisen—estrogen receptor immunohistochemistry. *J Histotechnol* 1997;20:7–10.
10 Barnes DM, Millis RR, Beex LVA, *et al.* Increased use of immunohistochemistry for oestrogen receptor measurement in mammary carcinoma: the need for quality assurance. *Eur J Cancer* 1998;34:1677–82.
11 UK NEQAS for Immunocytochemistry. Department of Histopathology, UCL Medical School, Rockefeller Building, University St, London WC1E 6JJ, UK.
12 Reiner A, Nuemeister B, Spona J, *et al.* Immunocytochemical localisation of oestrogen and progesterone receptors and prognosis in human primary breast cancer. *Cancer Res* 1990;50:7057–61.
13 Barnes DM, Harris WH, Smith P, *et al.* Immunohistochemical determination of oestrogen receptors:comparison of different methods of assessment of staining and correlation with clinical outcome of breast cancer patients. *Br J Cancer* 1996;74:1445–51.
14 Participation in UK NEQASs. In: *UK NEQAS report and directory 1998*, 3rd ed. Sheffield: UK NEQAS office, 1998: 15.
15 Pellicer EM, Sundblad A. Evaluation of antibodies to oestrogen receptors. *Appl Immunohistochem* 1994;2:141.
16 De Mascarel I, Soubeyran I, MacGrogan G, *et al.* Immunohistochemical analysis of estrogen receptors in 938 breast carcinomas. Concordance with biochemical assay and prognostic significance. *Appl Immunohistochem* 1995;3:222–31.
17 Pertschuk LP, Feldman JG, Kim Y-D, *et al.* Estrogen receptor immunocytochemistry in paraffin embedded tissues with ER1D5 predicts breast cancer endocrine response more accurately than H222Spg in frozen sections or cytosol-based ligand-binding assays. *Cancer* 1996;77:2514–19.
18 Ferno M, Andersson C, Giita F, *et al.* Oestrogen receptor analysis of paraffin sections and cytosol samples of primary breast cancer in relation to outcome to adjuvant tamoxifen therapy. *Acta Oncol* 1996;35:17–22.
19 Soubeyran I. Quenel N, Coindre J-M, *et al.* PS2: a marker improving prediction of response to neoadjuvant tamoxifen in post-menopausal breast cancer patients. *Br J Cancer* 1996;74:1120–5.
20 Kinsel LB, Szabo E, Greene GL, *et al.* Immunocytochemical analysis of estrogen receptors as a predictor of prognosis in breast cancer patients: comparison with quantitative biochemical methods. *Cancer Res* 1989;49:1052–6.
21 Allred DC, Harvey JM, Berado M, *et al.* Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol* 1998;11:155–68.
22 Barnes DM, Millis RR. Oestrogen receptors:the history, the relevance and the methods of evaluation. In: Kirkham N, Lemoine NR, eds. *Progress in pathology 2*. Edinburgh: Churchill Livingstone, 1995:89–114.
23 McCarty KS, Miller LS, Cox EB, *et al.* Estrogen receptor analyses: correlation of biochemical and immunohistochemical methods using monoclonal anti-receptor antibodies. *Arch Pathol Lab Med* 1985;109:716–21.

Histopathology 2000, 37, 536–545

High sensitivity and specificity of immunohistochemistry for the detection of hormone receptors in breast carcinoma: comparison with biochemical determination in a prospective study of 793 cases

B Zafrani, M-H Aubriot, E Mouret,[1] P de Crémoux,[2] Y de Rycke,[1] A Nicolas, E Boudou,[2] A Vincent-Salomon, H Magdelénat[2] & X Sastre-Garau

Departments of Pathology, [1]Biostatistics, [2]Pathophysiology and Pharmacology, Institut Curie, Paris, France

Date of submission 7 February 2000
Accepted for publication 6 April 2000

Zafrani B, Aubriot M-H, Mouret E, de Crémoux P, de Rycke Y, Nicolas A, Boudou E, Vincent-Salomon A, Magdelénat H & Sastre-Garau X

(2000) *Histopathology* 37, 536–545

High sensitivity and specificity of immunohistochemistry for the detection of hormone receptors in breast carcinoma: comparison with biochemical determination in a prospective study of 793 cases

Aims: The hormone receptor (HR) status of breast cancer is an important prognostic factor and predictive parameter of the response to hormone therapy. Enzyme immunoassay (EIA) is currently the standard for determination of HR, but immunohistochemistry (IHC) represents a potentially useful alternative. We used IHC to determine HR status in a large prospective study and compared the results to those obtained by EIA. This study was designed to determine which technique should be used in daily practice in our institution which manages a large number of patients. *Methods and results*: Oestrogen (ER) and progesterone (PgR) receptor status was evaluated in a prospective series of 793 infiltrating breast cancers by IHC in paraffin-embedded tissue sections, using antibodies 6F11 and 1A6, with a rigorous quality control of the methodology. ER were found to be significantly expressed in 81% of cases after IHC analysis and in 78% of cases by EIA. For PgR, the respective rates of positivity were 65% and 69%. The tumour HR level detected by either technique was significantly correlated with the value of tumour size, histological grade and S-phase fraction. A significant link was observed between the percentage of labelled cells after IHC analysis and the amount of protein detected by EIA. Critical analysis of discordance found that, in the group of invasive lobular carcinomas, the rate of HR positivity was higher with IHC (84%) than with EIA (45%) and that, in the overall population, IHC was more specific than EIA, since cases with nonrelevant positivity related to intraductal normal or neoplastic cells expressing HR could be discarded. The cost of IHC analysis was found to be about one-third of that of EIA. *Conclusions*: IHC is more sensitive, specific and economical than EIA. It should constitute the new standard technique provided that good quality assurance procedures are respected.

Keywords: biochemical assay, breast carcinoma, hormone receptors, immunohistochemistry

Introduction

The value of biochemical hormone receptor assay to predict breast cancer response to therapy and overall survival has been established for 20 years by an extensive literature on this subject. Most of this literature has been based on the dextran-coated

Address for correspondence: B Zafrani, Department of Pathology, Institut Curie, 26, rue d'Ulm, 75248 Paris Cedex 05, France.
e-mail: brigitte.zafrani@curie.net

charcoal ligand-binding assay and, more recently, on the enzyme immunoassay. However, it requires a minimal amount of representative tissue which must be immediately frozen. Delays in tissue processing, sparsely cellular tumours, contamination with non-neoplastic receptor-rich normal tissues, are well recognized causes of inaccuracy of such cytosol-based assays.[1–3] Nonetheless, the cytosol-based assay is generally regarded to be the standard against which new methods are measured.

Numerous reports in the literature show convincing evidence that immunohistochemical methods, particularly when performed on frozen tissue, provide as good or better predictors of response to hormone therapy or overall survival, than biochemical methods.[4,5] Due to recent improvements in immunohistochemical methodology as well as the use of more specific monoclonal antibodies the same can now be said about immunohistochemistry performed on paraffin-embedded tissue.[6–9] However, reliable quantitative immunohistochemistry is mandatory for clinical application of hormone receptor determination.[10] Although immunohistochemistry is easier to perform, less expensive, applicable to all tumours irrespective of their size and is more specific than biochemical methods, it has been criticized because of the lack of quality control and standardization of staining assessment.

This study was designed to validate the use of immunohistochemistry for the detection of hormone receptors in our Institution. Quality control procedures and participation to an external assessment scheme were set up, and validation was performed by comparing immunohistochemical expression (IHC) with enzyme immunoassay (EIA), considered to be the reference, and by correlating the results of the two techniques with prognostic factors. Possible causes for discrepancies were analysed in order to understand the specific contribution of immunohistochemistry.

Patients and methods

Between July 1997 and July 1998, 793 patients treated at Institut Curie for primary or recurrent infiltrating breast carcinoma underwent evaluation of their tumour hormone receptor status by biochemical assay and immunohistochemistry.

Five hundred and fourteen patients with primary breast cancer were treated by surgery: limited excision or mastectomy and axillary dissection. Fifty-four patients with a breast recurrence underwent mastectomy without axillary dissection. The remaining 225 patients were treated by neoadjuvant chemotherapy and sampling of their tumour was obtained by a needle core procedure before treatment.

BIOCHEMICAL ASSAY (EIA)

Cytosol oestrogen (ER) and progesterone (PgR) receptors were measured by enzyme immunoassay. A representative sample of tumour tissue was obtained from the surgical specimen, immediately frozen in liquid nitrogen and stored at −80 °C until the assay. Assays were always run within a week of collection. Tissue homogenization was performed in TK buffer (10 mM Tris, 10% glycol, 10 mM sodium molybdate, 0.5 mM dithiothreitol, 0.4 M KCl) and centrifuged at 105 000 *g* for 1 h. ER and PgR were assayed in the supernatant (cytosol fraction), according to the recommendations of the EORTC Receptor Study Group (EORTC Breast Cooperative Group, 1980). ER and PgR were assayed using an enzyme immunoassay (ER-EIA and PR-EIA, Abbott Laboratories, IL, USA) as previously described.[11] ER and PgR levels were expressed relative to tissue weight (fmol/gT), DNA content (fmol/μg DNA), and protein content of the sample (fmol/mg protein). The DNA content was measured on the 105 000 *g* pellet by a fluorimetric assay.[12] Cytosol proteins were quantified with the Pierce protein assay (Pierce, Rockford, IL, USA). The hormone receptor status (+ or −) was defined in terms of the hormone receptor content per gram of tissue, DNA content and protein content of the sample. The cut-off value for ER and PgR positivity was determined as 500 fmol/g tissue, which approximately corresponds to 15 fmol/mg protein and 250 fmol/μg DNA. This cut-off is higher than the sensitivity of the assay (5 fmol/mg protein), but corresponds to a clinically relevant threshold below which the response to endocrine treatment is low.

IMMUNOHISTOCHEMICAL PROCEDURE

All steps of the procedure were carefully controlled in order to achieve reproducible quantificative immunostaining.

Tissue fixation

Surgical specimens were delivered to the surgical pathology laboratory immediately after removal. Tumours were then cut into 4–5 mm slices to allow uniform penetration of the fixative and placed in AFA (75% absolute alcohol−20% formalin−5% acetic acid) for 24 h. Needle-core biopsies were fixed immediately after the needling procedure under the same conditions.

Sample selection
The size of the sample used for immunohistochemistry was determined in order to avoid detachment of the tissue on the glass slide during microwave heating. A 15 mm^2 tissue sample including the tumour and its interface with normal breast tissue was prepared.

Slide preparation
Tissue sections were placed on Capillary Gap Plus slides (Dako, Trappes, France), dried for 60 min at 58 °C and left overnight at 37 °C.

Heat-induced epitope retrieval
Tissue sections were immersed in 0.01 M citrate buffer, pH 6.1 (400 ml) and exposed to 650 W microwave for 5 min to reach 97–98 °C. This temperature was maintained for 15 min by controlling the power. This heating time was optimized after several trials and adjusted to obtain optimal sensitivity. Keeping the temperature below 100 °C allowed a constant volume of buffer to be maintained and avoided detachment of the sections. Sections in buffer were allowed to cool at room temperature for 20 min.

Antibodies and Reagents
ER 6F11 monoclonal antibody (Novocastra, Le Perray en Yeulines, France) diluted to 1 : 50, PgR 1A6 monoclonal antibody (Novocastra) diluted to 1 : 40, and biotinylated horse anti-mouse immunoglobulin, diluted to 1 : 200, were used as primary and secondary antibodies, respectively. Detection was performed by the avidin–biotin–peroxidase complex (ABC, Vector Coger Company, Paris, France). 3,3′-Diaminobenzidine (1 mg/ml) was used as chromogen and light haematoxylin as counterstain.

Procedure
The Biotek Horizon machine using the capillary action method was used. After the epitope retrieval procedure, sections were rinsed in distilled water, and left in 0.01 M phosphate-buffered saline (PBS), pH 7.6. Non-specific binding was blocked by washing buffer, containing carrier protein (Dako). Sections were then washed in PBS Tween 0.05% and incubated with primary antibodies for 60 min, at room temperature. After rinsing in PBS Tween, they were incubated with the secondary antibody for 25min. and were rinsed again before blocking of endogenous peroxidase activity. After further washing, the ABC complex was applied for 25 min. Sections were rinsed, reacted with DAB, rinsed and counterstained.

CONTROLS

A specifically prepared block was used as external control. Three tissue specimens of tumours (with known levels of receptor activity) with strong staining, weak staining or no staining for ER were included in a paraffin block. Histological sections were taken from this block and included in every run of ER detection. The staining of this control section was checked and compared with that of the previous analysis to detect minor variations in the staining intensity. The same procedure was used for PR analysis. When a decreased intensity was observed, the procedure was repeated. Normal ductal/lobular breast epithelium located at the periphery of the tumour was used as an internal control when available.

MICROSCOPIC INTERPRETATION AND SCORING

Tumours were classified and graded according to the WHO classification,[13] and the Elston and Ellis grading system.[14] Immunohistochemical interpretation took into account nuclear staining of invasive cancer cells only. Stained cells of the in-situ component or normal glands were assessed separately but not considered in the final report. Results were expressed in terms of the percentage of positive cells (0–100%) and intensity of staining (weak, moderate, strong). A threshold of 10% positive cells at any intensity was considered positive.

The slides were scored by all pathologists after six months' training on the multihead microscope. Immunohistochemical results were reported without knowledge of the biochemical results.

STATISTICAL ANALYSIS

Comparisons between hormone receptor status assessed by the two methods and prognostic factors were performed using the Pearson chi-squared test. The correlation between immunohistochemistry (IHC) and enzyme immunoassay (EIA) was described by sensitivity and specificity values, using dichotomous variables (positive/negative). The EIA method was used as the reference method. We performed a linear regression to assess the quantitative correlation between IHC and EIA methods. We used logit transformation for the percentage of ER and PgR positive cells and logarithm transformation for ER and PgR levels assessed by biochemistry. We performed linear regression analysis to explain the variation of the percentage of positive cells by ER and PgR levels expressed in terms of protein content. *P*-values greater than 0.05 were reported as non-significant (n.s).

Table 1 Clinicopathological characteristics of the 793 cases

Patient and tumour characteristics	Number (%)
Age	
< 35 years	19 (2)
≥ 35 years	774 (98)
Material	
Needle-core biopsy	225 (28)
Excisional biopsy	568 (72)
Histological type	
Infiltrating ductal	613 (77)
Infiltrating lobular	145 (18)
Others	35 (5)
Histological grade	
I	210 (27)
II	310 (40)
III	263 (33)
Not evaluable	10
Histological size (pT)*†	
pT1 < 5 mm	6 (1)
pT1 < 20 mm	364 (65)
pT2 < 50 mm	175 (31)
pT3 > 50 mm	15 (3)
*Vascular invasion**	
Not seen	378 (64)
Present	190 (36)
Axillary nodal status†*	
Negative nodes	317 (63)
Positive nodes < 4	119 (24)
Positive nodes ≥ 4	65 (13)
No axillary dissection	54
Ploidy	
Diploid	166 (33)
Aneuploid	338 (67)
Not done	289
S-phase fraction	
< 4%	299 (68)
≥ 4%	142 (32)
Not done	352

*Data evaluated only for the 568 tumours obtained at the time of surgery. †Missing data for some cases.

In order to assess any interobserver variation, a considerable number of cases were reviewed at the end of the study by two pathologists (BZ, MHA). Sixty cases, without knowledge of the biochemical and immunohistochemical results, were selected by drawing lots and reviewed: interobserver variation in the scoring of results was evaluated by the kappa test. In addition, all cases showing discordant results between the two methods (i.e. immunohistochemistry negative/biochemistry positive and conversely, immunohistochemistry positive/biochemistry negative) were re-analysed.

Results

The clinicopathological features of the 793 cases are shown in Table 1. The great majority of patients were older than 35 years. Seventy-seven per cent of the tumours were infiltrating ductal carcinomas. Twenty-seven per cent of cases were classified as grade I, 40% as grade II and the remaining 33% as grade III. Some pathological data (i.e. histological size, vascular invasion, axillary nodal status, ploidy and S-phase fraction) could not be assessed in every case because of treatment procedures (needle core biopsy, no axillary dissection) or ancillary techniques not performed (flow cytometry).

IMMUNOHISTOCHEMICAL RESULTS

By immunohistochemistry, 644 cases (81.2%) were oestrogen receptor positive and 518 cases (65.3%) were progesterone receptor positive. According to the EIA method, 617 cases (77.8%) were oestrogen positive and 549 cases (69.2%) were progesterone receptor positive.

Table 2 shows the relation between staining intensity and percentage of positive cells. As the staining intensity appeared stronger, the percentage of labelled cells increased. The mean percentage of ER-positive cells was 38% when the intensity was weak, 69% when the intensity was moderate and increased to 90% when the intensity was strong. The mean percentage of PgR-positive cells was 32% when the intensity was weak, 50% when the intensity was moderate and increased to 70% when the intensity was strong.

CORRELATION WITH PROGNOSTIC FACTORS

Table 3 shows the correlation between immunohistochemical (IHC) and biochemical (EIA) determination of hormone receptors, and prognostic factors. Hormone receptor status determined by immunohistochemistry or biochemistry was not correlated with age assessed as a dichotomized variable. ER status evaluated by the two

Table 2. Relationship between staining intensity and mean percentage of positive cells in the 644 ER positive and 518 PgR positive cases

	Oestrogen receptor		Progesterone receptor	
Staining intensity	Number of cases	Mean positive cells (%)	Number of cases	Mean positive cells (%)
Weak	57	38	123	32
Moderate	192	69	220	50
High	395	90	175	70

assays was correlated with histological type ($P < 10^{-4}$, and $P = 0.01$ for IHC and EIA, respectively), as well as PgR evaluated by immunohistochemistry ($P = 0.02$), whereas PgR evaluated by EIA was not. A strong correlation was found between both receptors, regardless of the method used, and histological grade ($P < 10^{-4}$). ER and PgR determination by IHC was correlated with vascular invasion ($P < 0.01$ and $P = 0.08$, respectively), whereas no correlation was found with axillary nodal status for either method. ER status determined by both assays was strongly correlated with ploidy ($P < 10^{-4}$, and $P = 0.0003$ for IHC and EIA, respectively), as well as PgR status determined by biochemistry ($P = 0.03$), whereas PgR evaluated by immunohistochemistry was not. Both receptors were highly correlated with S-phase fraction on both assays.

COMPARISON OF RESULTS OBTAINED BY IMMUNOHISTOCHEMISTRY AND BIOCHEMISTRY (QUALITATIVE ASSESSMENT)

The relation between ER and PgR status, evaluated by immunohistochemistry and biochemistry is listed in Table 4. Concordance of the two assays for ER determination was equal to 92%. Discordant results were observed in 63/793 cases. Considering the biochemical assay as the reference, sensitivity and specificity of immunohistochemistry were 97.1% (95% CI) and 74.4% (95% CI), respectively. Concordance of the two assays for PgR determination was equal to 82.9%. In 135/793 cases, results were discordant. Sensitivity and specificity of immunohistochemistry were 84.9% (95% CI) and 78.7% (95% CI), respectively.

Comparison between the two methods was performed according to histological type (Tables 5 and 6). For the 613 ductal carcinomas, concordance of the two assays was equal to 93% and 83% for ER and PgR determination, respectively. Sensitivity and specificity were almost identical to those observed for the whole group. For the 145 lobular carcinomas, concordance of the two methods was equal to 85.5% and 79% for ER and PgR determination, respectively. By considering the biochemical method as the reference, the specificity of immunohistochemistry was 16% for ER determination and 55% for PgR determination.

COMPARISON OF RESULTS OBTAINED BY IMMUNOHISTOCHEMISTRY AND BIOCHEMISTRY (QUANTITATIVE ASSESSMENT)

Linear regression analysis showed a very strong correlation between the percentage of ER and PgR positive cells and the ER and PgR levels assessed by biochemistry (ER: $r = 0.73$, $P < 10^{-4}$ and PgR: $r = 0.69$, $P < 10^{-4}$). The relation between the percentage of ER positive cells at any intensity and the ER levels by protein content is shown, as an example, on Figure 1.

ANALYSIS OF DISCORDANCES

Discordant results between the two methods were observed in 63/793 cases for ER (8%) and in 135/793 cases for PgR (17%). Among the 63 discordant cases for ER, 45 were IHC positive/EIA negative, whereas 18 were IHC negative/EIA positive. Among the 135 discordant cases for PgR, 52 were IHC positive/EIA negative, whereas 83 cases were IHC negative/EIA positive.

ANALYSIS OF IHC POSITIVE/EIA NEGATIVE CASES

Figure 2 shows the distribution of the percentage of stained cells of 45 ER cases. The percentage of positive cells was clearly above the cut-off value with more than 50% stained cells in the very great majority of cases. Similar data were obtained for 52 PgR cases, although the percentage of stained cells was in a lower range than that observed for ER (data not shown).

Table 3 Correlation between prognostic factors and ER/PR status determined by immunohistochemistry and biochemistry

	Immunohistochemistry				Biochemistry			
Prognostic factors	ER+ (%)	*P-value*	PgR+ (%)	*P-value*	ER+ (%)	*P-value*	PgR+ (%)	*P-value*
Age								
< 35 years	79		79		79		79	
≥ 35 years	81.2	*n.s*	65	*n.s*	77.7	*n.s*	69	*n.s*
Histological type								
Infiltrating ductal	78.8		63.2		76.1		69.2	
Infiltrating lobular	94.5		75.2		86.9		72.4	
Others	69.6	$< 10^{-4}$	63.6	*0.02*	72.7	*0.01*	57.5	*n.s*
Histological grade								
I	97.6		77.6		90		79.5	
II	88.4		72.3		86.8		75.8	
III	59.3	$< 10^{-4}$	47.5	$< 10^{-4}$	57.4	$< 10^{-4}$	54	$< 10^{-4}$
Histological size								
≤ 20 mm	86.2		69.7		81.4		71.9	
> 20 mm	78.4	*0.01*	62.6	*0.09*	72.6	*0.01*	60.5	*< 0.01*
Vascular invasion								
Not seen	87.3		69.9		80.4		69.6	
Present	76.8	*< 0.01*	62.6	*0.08*	75.3	*n.s*	65.8	*n.s*
Axillary nodal status								
Negative nodes	87.1		70.6		81.3		71.6	
Positive nodes < 4	84.1		66.4		78.2		69.7	
Positive nodes ≥ 4	81.5	*n.s*	60	*n.s*	83.1	*n.s*	66.2	*n.s*
Ploidy								
Diploid	88.6		67.5		86.7		75.3	
Aneuploid	78.9	$< 10^{-4}$	59.8	*n.s*	71.1	*0.0003*	63.6	*0.03*
S-phase fraction								
< 4%	85.6		67.8		84.2		73.5	
≥ 4%	61.2	$< 10^{-4}$	50	*0.0003*	59.8	$< 10^{-4}$	54.9	*0.0001*

As the 'false negative' EIA results could have been due to intratumour heterogeneity, the EIA assay was repeated in 22 cases.

ANALYSIS OF IHC NEGATIVE/EIA POSITIVE CASES

Sixteen of eighteen cases assessed for ER had no stained cells, whereas 2/18 cases showed 5% stained cells. Forty-two of eighty-three cases assessed for PgR had no stained cells, whereas the remaining 41/83 cases showed 1–5% stained cells.

As scoring of immunostaining cases can be a source of error, particularly when the percentage of stained cells is close to the threshold, all cases with stained cells near the threshold were reviewed independently, at the end of the study, by two observers (BZ, MHA). In all, except one case, the first scoring was confirmed. In addition, evaluation of interobserver variation was assessed on a series of 60 specimens. Disagreement between the first result (assessed by any pathologist participating in the study) and the second result (assessed by MHA without knowledge of the first scoring) was evaluated by the kappa test. Kappa values were 0.91 and 0.89 for ER and PgR scoring, respectively.

In addition to validating the scoring of the infiltrative cells of these tumours, analysis of normal glands

Table 4 Correlation between immunohistochemical and biochemical determination of ER and PgR status

Immunohistochemical detection	Biochemical assay		
	ER+	*ER−*	*Total*
ER+	599	45	644
ER−	18	131	149
Total	617	176	793
	PgR+	*PgR−*	*Total*
PgR+	466	52	518
PgR−	83	192	275
Total	549	244	793

entrapped within the tumour and in-situ component was performed. Strongly positive stained cells were observed in the in-situ component and/or normal glandular epithelium in 30% of cases (i.e. 8/18 cases and 24/83 cases assayed for ER and PgR, respectively).

Discussion

Immunohistochemical assessment of hormone receptors is regarded as a good alternative for biochemical assessment.[15–17] It has the advantage of a morphological technique allowing selective analysis of tumour cells, excluding stromal cells and benign epithelium (Figure 3). It is also the only method available when tumour size is too small to allow partition for biochemical assay. However, immunohistochemistry also presents certain weaknesses related to variation in the staining technique and assessment of staining.[10] This study was therefore conducted to validate the use of immunohistochemistry in the determination of steroid receptors in our Institution. The aim was to provide accurate, reliable and reproducible quantification of oestrogen and progesterone receptors. Validation was performed by comparing

Table 5 Ductal carcinomas: correlation between immunohistochemical and biochemical determination of ER and PgR status (613 cases)

Immunohistochemical detection	Biochemical assay		
	ER+	*ER−*	*Total*
ER+	454	29	483
ER−	12	118	130
Total	466	147	613
	PgR+	*PgR−*	*Total*
PgR+	355	31	386
PgR−	69	158	227
Total	424	189	613

Table 6 Lobular carcinomas: correlation between immunohistochemical and biochemical determination of ER and PgR (145 cases)

Immunohistochemical detection	Biochemical assay		
	ER+	*ER−*	*Total*
ER+	121	16	137
ER−	5	3	8
Total	126	19	145
	PgR+	*PgR−*	*Total*
PgR+	92	18	110
PgR−	13	22	35
Total	105	40	145

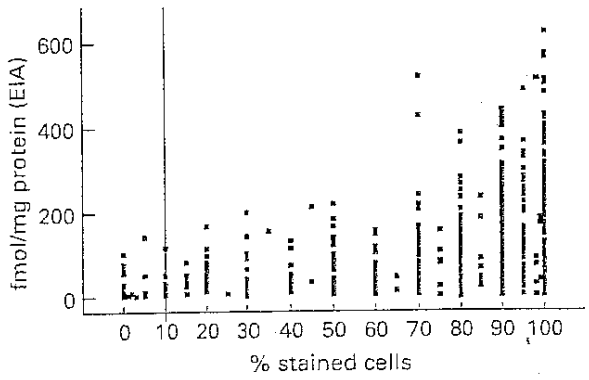

Figure 1 Distribution of percentage of ER cells stained by immunohistochemistry as a function of ER level determined by EIA assay.

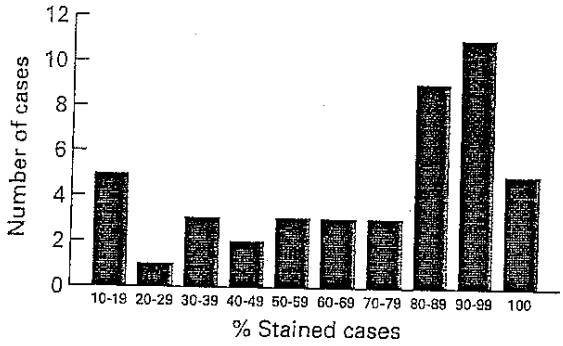

Figure 2 Percentage of ER positive cells in 45 immunohistochemistry positive/EIA negative cases.

immunohistochemical results with those obtained with enzyme immunoassay and correlating both results with prognostic factors.

Our results showed that immunohistochemistry is as accurate as biochemistry to detect hormone receptor positive tumours. The immunohistochemical method is more sensitive than the biochemical method in the identification of rare positive tumour cells, as demonstrated in our series for infiltrating lobular carcinomas, as 16/19 cases considered to be ER negative by EIA were IHC positive and 18/40 cases considered to be PgR negative by EIA were IHC positive. Immunohistochemistry is also more specific as it is able to differentiate, under microscopic visualization, infiltrative tumour cells from in-situ tumour cells or benign cells that might be stained. In our series, 30% of IHC negative/EIA positive cases were related to the presence

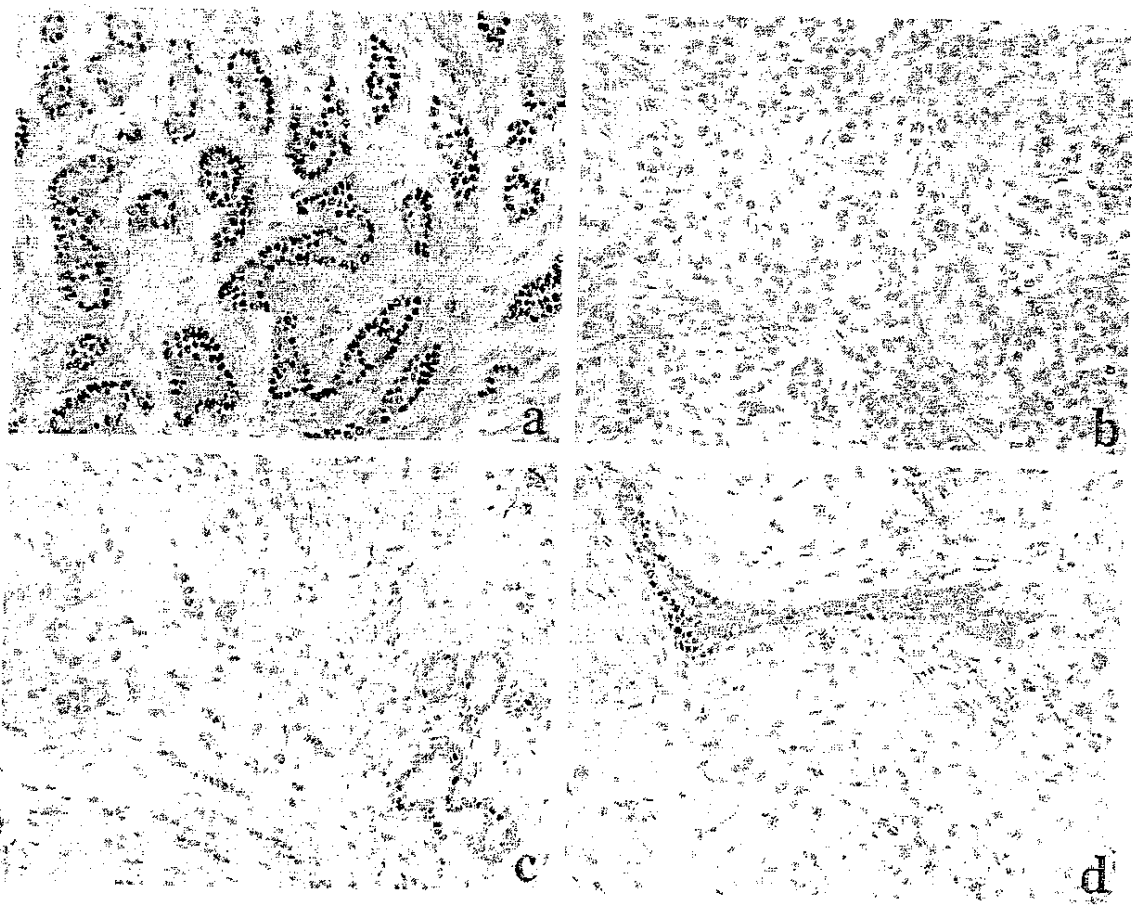

Figure 3 a, Ductal carcinoma, oestrogen receptor positive (90% stained cells, strong staining intensity); b, ductal carcinoma, progesterone receptor positive (20% stained cells, weak to moderate staining intensity); c, paucicellular lobular carcinoma, oestrogen receptor positive (40% stained cells, moderate staining intensity), normal glandular epithelium in the bottom right corner; d, ductal carcinoma, oestrogen negative (normal stained glandular epithelium entrapped within tumour).

of stained cells observed in normal glands or in the in-situ component within a negative infiltrative tumour.

In our experience, quantification of oestrogen and progesterone receptors was highly reproducible. Variations in the staining technique were excluded by a rigorous quality control procedure. Furthermore, our study showed a very high level of interobserver agreement in the staining assessment, as demonstrated by kappa values of 0.91 for ER and 0.89 for PgR.

Oestrogen and progesterone receptor expression was detected in 81.2% and 65.3% of our cases, respectively. These figures are similar to those already reported in the literature for ER (59–78%)[17–22] and PgR (52–71%),[23–25] although it is difficult to compare the results of different series because of the various types of material (frozen/paraffin-embedded sections), epitope retrieval technique and primary antibody used. There is also no consistent scoring method in the literature. Some authors consider only the invasive tumour component[22,24,26] while others consider both invasive and in-situ components,[20] and in some studies this is not specified.[18,21,27] There is no general agreement on the threshold, which is generally chosen arbitrarily. However, Barnes *et al.*,[28] in a series of 170 women receiving first-line tamoxifen treatment for metastatic disease, showed that regardless of the scoring assessment, immunohistochemical results were closely related to patient outcome. In a recent study by Harvey,[9] the definition of ER positivity was calibrated by the clinical outcome. In this series, as few as 1–10% weakly stained cells were shown to be positive cases.

Correlation between immunohistochemistry and enzyme immunoassay, in our series, was 92% and 83% for ER and PgR, respectively. These figures are also in agreement with reported results, although, to our knowledge, this is the first large series comparing enzyme immunoassay and immunohistochemistry performed with these new antibodies.[29] Some studies have investigated the correlation with prognostic factors.[19,22–24,27] Our results, determined either by immunohistochemistry or immunoassay, confirmed those obtained by others. ER positivity was correlated with other indicators of good prognosis, such as low histological grade, small tumour size, absence of vascular invasion, diploidy or low S-phase fraction. PgR positivity was also correlated with histological grade, tumour size, vascular invasion and S-phase fraction. No correlation was found between hormone receptor and axillary nodal status. The age of the patients was not correlated, in our experience, with the receptor status, although, we classified patients according to young age (< 35 years) rather than menopausal status or middle age (50 years) as in other studies.[18,19,22–24,30]

In conclusion, immunohistochemical determination of hormone receptor status in breast carcinoma is, in our experience, as sensitive but more specific than biochemical determination. It has the advantage of being easier to perform, is applicable to all tumours irrespective of their size and is less expensive. However, a consensus is certainly needed to define the best cut-off value which will be ultimately validated by clinical studies evaluating the correlation with prognosis and response to endocrine treatment.

Acknowledgements

The authors thank Christine Lainé-Bidron, Sylviane Bouthinon and Sonia Rivault for their assistance in collecting the biochemical and immunohistochemical data.

References

1. Borjesson B, Cauchi M, Compton P *et al.* Identification of the major sources of error in estrogen receptor measurements for individual laboratories using both tissue and cytosol samples. *Eur. J. Cancer Clin. Oncol.* 1989; 25; 1079–1086.
2. King W, DeSombre E, Jensen E. Greene G. Comparison of immunocytochemical and steroid-binding assays for estrogen receptor in human breast tumors. *Cancer Res.* 1985; 45; 293–304.
3. Parl F, Posey Y. Discrepancies of the biochemical and immunohistochemical estrogen receptor assays in breast cancer. *Hum. Pathol.* 1988; 19; 960–966.
4. Allred D, Bustamante M, Daniel C, Gaskill H, Cruz AJ. Immunocytochemical analysis of estrogen receptors in human breast carcinomas. Evaluation of 130 cases and review of the literature regarding concordance with biochemical assay and clinical relevance. *Arch. Surg.* 1990; 125; 107–113.
5. Reiner A, Neumeister B, Spona J, Reiner G, Schemper M, Jakesz R. Immunocytochemical localization of estrogen and progesterone receptor and prognosis in human primary breast cancer. *Cancer Res.* 1990; 50; 7057–7061.
6. Andersen J, Thorpe S, Rose C, Christensen I, Rasmussen B, Poulsen H. Estrogen receptor in primary breast cancer estimated in paraffin-embedded tissue. A study if its usefulness compared to dextran-coated charcoal assay. *Acta Oncol.* 1991; 30; 685–690.
7. Andersen J. Determination of estrogen receptors in paraffin-embeded tissue. Techniques and the value in breast cancer treatment. *Acta Oncol.* 1992; 31; 611–627.
8. Raymond W, Leong A. Oestrogen receptor staining of paraffin-embedded breast carcinomas following short fixation in formalin: a comparison with cytosolic and frozen section receptor analyses. *J. Pathol.* 1990; 160; 295–303.
9. Harvey J, Clark G, Osborne C, Allred D. Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. *J. Clin. Oncol.* 1999; 17; 1474–1481.
10. Barnes D, Millis R, Beex L, Thorpe S, Leake R. Increased use of immunohistochemistry for oestrogen receptor measurement in

mammary carcinoma: the need for quality assurance. *Eur. J. Cancer* 1998; 34; 1677–1682.
11. Goussard J, Magdelenat H, Martin P. Monoclonal antibodies for progesterone receptor assays and polymorphism studies in breast cancer. Comparison with radioligand assays. *Bull. Cancer* 1988; 75; 771–782.
12. Fiszer-Szafars B, Szafars D, Guevazade de Murillo A. A general, fast and sensitive micromethod for DNA determination. *Anal. Biochem.* 1981; 110; 165–170.
13. Scarff R, Torloni H. *Histological Typing of Breast Tumors.* Geneva: World Health Organization, 1968.
14. Elston C, Ellis I. Pathological prognostic factors in breast cancer. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* 1991; 19; 403–410.
15. Bosman F, de Goeij A, Rousch M. Quality control in immunocytochemistry: experiences with the oestrogen receptor assay. *J. Clin. Pathol.* 1992; 45; 120–124.
16. Taylor CR. Paraffin section immunohistochemistry for estrogen receptor. *Cancer* 1996; 77; 2419–2422.
17. Battifora H, Mehta P, Ahn C, Esteban JM. Estrogen receptor immunohistochemical assay in paraffin-embedded tissue. A better gold standard? *Appl. Immunohistochem.* 1993; 1; 39–45.
18. Molino A, Micciolo R, Turazza M *et al.* Prognostic significance of estrogen receptors in 405 primary breast cancers: a comparison of immunohistochemical and biochemical methods. *Breast Cancer Res. Treat.* 1997; 45; 241–249.
19. Alberts S, Ingle J, Roche P *et al.* Comparison of estrogen receptor determinations by a biochemical ligand-binding assay and immunohistochemical staining with monoclonal antibody ER1D5 in females with lymph node positive breast carcinoma entered on two prospective clinical trials. *Cancer* 1996; 78; 764–772.
20. Saccani Jotti G, Johnston S, Salter J, Detre S, Dowsett M. Comparison of new immunohistochemical assay for oestrogen receptor in paraffin wax embedded breast carcinoma tissue with quantitative enzyme immunoassay. *J. Clin. Pathol.* 1994; 47; 900–905.
21. De Negri F, Campani D, Sarnelli R *et al.* Comparison of monoclonal immunocytochemical and immunoenzymatic methods for steroid receptor evaluation in breast cancer. *Am. J. Clin. Pathol.* 1991; 96; 53–58.
22. De Mascarel I, Soubeyran I, MacGrogan G *et al.* Immunohistochemical analysis of estrogen receptors in 938 breast carcinomas. Concordance with biochemical assay and prognostic significance. *Appl. Immunohistochem.* 1995; 3; 222–231.
23. Gasparini G, Pozza F, Dittadi R, Meli S, Cazzavillan S, Bevilacqua P. Progesterone receptor determined by immunocytochemical and biochemical methods in human breast cancer. *J. Cancer Res. Clin. Oncol.* 1992; 118; 557–563.
24. MacGrogan G, Soubeyran I, De Mascarel I *et al.* Immunohistochemical detection of progesterone receptors in breast invasive ductal carcinomas. A correlative study of 942 cases. *Appl. Immunohistochem.* 1996; 4; 219–227.
25. Soomro S, Shousha S. Demonstration of progesterone receptors in paraffin wax sections of breast carcinoma. *J. Clin. Pathol.* 1990; 43; 671–674.
26. Biesterfeld S, Schröder W, Steinhagen G *et al.* Simultaneous immunohistochemical and biochemical hormone receptor assessment in breast cancer provides complementary information. *Anticancer Res.* 1997; 17; 4723–4730.
27. Golouh R, Vrhovec I, Bracko M, Frkovic-Grazio S. Comparison of standardized immunohistochemical and biochemical assays for estrogen and progesterone receptors in breast carcinoma. *Pathol. Res. Pract.* 1997; 193; 543–549.
28. Barnes DM, Harris WH, Smith P, Millis RR, Rubens RD. Immunohistochemical determination of oestrogen receptor: comparison of different methods of assessment of staining and correlation with clinical outcome of breast cancer patients. *Br. J. Cancer* 1996; 74; 716–721.
29. Bevitt D, Milton I, Piggot N *et al.* New monoclonal antibodies to oestrogen and progesterone receptors effective for paraffin section immunohistochemistry. *J. Pathol* 1997; 183; 228–232.
30. Molino A, Micciolo R, Turazza M *et al.* Estrogen receptors in 699 primary breast cancers: a comparison of immunohistochemical and biochemical methods. *Breast Cancer Res. Treat.* 1995; 34; 221–228.

Virchows Arch (2004) 445:119–128
DOI 10.1007/s00428-004-1063-8

ORIGINAL ARTICLE

C. A. Wells · J. P. Sloane · D. Coleman · C. Munt · I. Amendoeira · N. Apostolikas · J. P. Bellocq · S. Bianchi · W. Boecker · G. Bussolati · C. E. Connolly · P. Dervan · M. Drijkoningen · I. O. Ellis · C. W. Elston · V. Eusebi · D. Faverly · P. Heikkila · R. Holland · J. Jacquemier · M. Lacerda · J. Martinez-Penuela · C. De Miguel · J. L. Peterse · F. Rank · A. Reiner · E. Saksela · B. Sigal-Zafrani · M. Sylvan · B. Borisch · G. Cserni · T. Decker · H. Kerner · J. Kulka · P. Regitnig · A. Sapino · A. M. Tanous · S. Thorstenson · E. Zozaya
The European Working Group for Breast Screening Pathology

Consistency of staining and reporting of oestrogen receptor immunocytochemistry within the European Union—an inter-laboratory study

Received: 7 April 2004 / Accepted: 28 May 2004 / Published online: 24 June 2004

The European Working Group for Breast Screening Pathology is supported by the European Commission through the European Breast Cancer Network

C. A. Wells (✉)
Department of Histopathology,
St. Bartholomew's Hospital Medical School,
West Smithfield, London, EC1A 7BE, United Kingdom
e-mail: c.a.wells@qmul.ac.uk
Tel.: +44-20-76018451
Fax: +44-20-76018543

J. P. Sloane
Department of Pathology,
Liverpool, United Kingdom

D. Coleman · C. Munt
Cancer Screening Evaluation Unit,
Institute of Cancer Research,
Sutton Surrey, United Kingdom

I. Amendoeira
Department of Pathology,
Instituto de Patologia & Imunologia Molecular da Universidade do Porto,
Porto, Portugal

N. Apostolikas
Department of Pathology, Saint Savvas Hospital,
Athens, Greece

J. P. Bellocq
Service d'Anatomie Pathologique,
Hopital de Hautepierre, Strasbourg, France

S. Bianchi
Department of Human Pathology and Oncology,
University of Florence, Italy

W. Boecker
Gerhard-Domagk Institut fur Pathologie,
Universitat et Munster,
Munster, Germany

G. Bussolati · A. Sapino
Istituto di Anatomia e Istologia Patologica,
Torino, Italy

C. E. Connolly
Department of Pathology,
Clinical Sciences Institute, University College Hospital,
Galway, Ireland

P. Dervan
Pathology Department,
Mater Hospital, University College,
Dublin, Ireland

M. Drijkoningen
Pathologische Ontleedkunde, University Hospital,
Leuven, Belgium

I. O. Ellis · C. W. Elston
Department of Histopathology, City Hospital,
Nottingham, United Kingdom

V. Eusebi
Sezione Anatomia Patologica M. Malpigni,
Universita di Bologna,
Ospedale Bellaria, Bologna, Italy

D. Faverly
CMP Laboratory, Bruxelles, Belgium

P. Heikkila · E. Saksela
Haartman Institute, University of Helsinki,
Helsinki, Finland

R. Holland
National Expert & Training Centre for Breast Cancer Screening,
Academisch Ziekenhuis Nijmegen,
Nijmegen, The Netherlands

J. Jacquemier
Institut Paoli Calmettes, Marseille, France

M. Lacerda
Laboratorio De Histopatologica,
Centro Regional De Oncologia De Coimbra,
Coimbra, Portugal

J. Martinez-Penuela · E. Zozaya
Department of Pathology, Hospital de Navarra,
Pamplona, Spain

C. De Miguel
Department of Pathology, Hospital Virgen del Camino,
Pamplona, Spain

J. L. Peterse
Department of Pathology,
The Netherlands Cancer Institute,
Amsterdam, The Netherlands

F. Rank
Center of Laboratory Medicine and Pathology,
Department of Pathology,
Rigshospitalet, Copenhagen University Hospital,
Copenhagen, Denmark

A. Reiner
Institute of Pathology, Donauspital,
Wien, Austria

B. Sigal-Zafrani
Institut Curie, Section Medicale et Hospitaliere,
Paris, France

M. Sylvan
Department of Clinical Pathology and Cytology,
Huddinge University Hospital,
Stockholm, Sweden

B. Borisch
Centre Médicale Universitaire,
Genève, Switzerland

G. Cserni
Department of Pathology,
Bács-Kiskun County Teaching Hospital,
Kecskemét, Hungary

T. Decker
Department of Pathology,
The Breast Unit, HELIOS Medical Centre,
Berlin, Germany

H. Kerner
Department of Pathology,
Rambam Hospital,
Haifa, Israel

J. Kulka
2nd Institute of Pathology, Semmelweis University,
Budapest, Hungary

P. Regitnig
Institut fur Pathologie der Karl-Franzens Universitat Graz,
Graz, Austria

A. M. Tanous
Division d'Anatomie Pathologique,
Laboratoire National de Santé,
Luxembourg

S. Thorstenson
Kalmar County Hospital,
Kalmar, Sweden

Abstract To assess the variability of oestrogen receptor (ER) testing using immunocytochemistry, centrally stained and unstained slides from breast cancers were circulated to the members of the European Working Group for Breast Screening Pathology, who were asked to report on both slides. The results showed that there was almost complete concordance among readers (kappa=0.95) in ER-negative tumours on the stained slide and excellent concordance among readers (kappa=0.82) on the slides stained in each individual laboratory. Tumours showing strong positivity were reasonably well assessed (kappa=0.57 and 0.4, respectively), but there was less concordance in tumours with moderate and low levels of ER, especially when these were heterogeneous in their staining. Because of the variation, the Working Group recommends that laboratories performing these stains should take part in a external quality assurance scheme for immunocytochemistry, should include a tumour with low ER levels as a weak positive control and should audit the percentage positive tumours in their laboratory against the accepted norms annually. The Quick score method of receptor assessment may also have too many categories for good concordance, and grouping of these into fewer categories may remove some of the variation among laboratories.

Keywords Oestrogen receptor · Immunohistochemistry · Quality assessment · Variability

Introduction

Since the advent of breast screening and detection of breast cancers using mammography, tumours in general are being detected at a smaller size. The necessity to obtain oestrogen receptor (ER) status on these very small tumours has led to the measurement of receptor status using immunocytochemistry. The older biochemical methods of dextran-coated charcoal and enzyme-linked immunoassay are unable to provide results on these small tumours which often require the entire tumour to be sampled for accurate diagnosis. Even in the sampling of larger tumours for biochemical assay, the problems of

sampling error, with benign areas having a low level of oestrogen positivity being included with a negative tumour in the homogenate, led to certain ER-negative tumours being falsely classified as low expressers of ER. Conversely, areas of tumour from high ER-expressing tumours composed predominantly of the fibro-elastotic core could be falsely classified as ER poor. Older publications based on the biochemical methods put the level of response to anti-oestrogens at approximately 60% of those tumours that were ER rich, but also acknowledged a response rate of about 8% in totally ER-negative tumours. For these reasons, coupled with the availability of high-quality antibodies, the assessment of receptors using the biochemical methods has largely been superseded by immunocytochemistry on tissue sections [1, 4, 17]. This has the following advantage: direct visualisation of the cells expressing the receptor enables an assessment of whether these are tumour cells or are associated benign tissue. More recent publications seem to agree that totally ER-negative tumours identified using immunocytochemistry do not respond to hormonal therapy [6] and that these should be preferentially treated with other therapeutic modalities, although some debate persists about women with tumours expressing very low levels of receptor.

Most workers are now agreed that immunocytochemistry for ER is the method of choice [1, 4, 5, 17], and a number of scoring systems have been developed in an attempt to semi-quantify the test. The most complex of these is the H-score, which is still used by a number of laboratories [7]. More recently, the recommended method by the European Organization for the Research and Treatment of Cancer (EORTC) working group has been the modified Quick score [6, 9], which has been adopted by some national guidelines [12]. This has the advantage that it is somewhat quicker to perform than the H-score and still correlates well with clinical response [1]. Some laboratories have used 10% ER positivity of any type as the threshold for considering breast cancers as ER positive. However, with the Quick score, a score above two (e.g., above 1% of weakly positive cells) has been reported to predict responsiveness to endocrine treatment [6]. One of the strengths of the old biochemical methods, however, was that the technical aspects of the methodology were rigorously quality controlled through the EORTC and biochemical networks using a vial of known oestrodiol content. Quality control of the immunocytochemical test is not so well developed, and there is some concern that tumours having low levels of ER may be scored as negative in some laboratories, leading to the possibility that these women will not benefit from anti-oestrogen therapy. Previous publications have concentrated on the variation in technical methods among laboratories [11, 14] but have not investigated the inter-observer variation in scoring.

Materials and methods

Slides from 30 cases of breast carcinoma were cut from the paraffin block. One set of slides was stained with haematoxylin and eosin (H&E) for grading, another set of slides from the same tissue blocks was stained using immunocytochemistry for ER in the University of Liverpool laboratory and a final set of unstained slides was also prepared.

One slide stained with H&E, one stained for ER and one unstained slide were sent to each laboratory participating in the European Working Group. For each tumour, members of the group were asked the following.

a. To score the Liverpool-stained slides of the tumours for the percentage of positive cells in 10% bands (0, 1–10, 11–20, 21–30, 31–40, 41–50, 51–60, 61–70, 71–80, 81–90 and 91–100) and to assess the intensity of staining on a four-point scale of negative, weakly positive, moderately positive and strongly positive, assessing also whether the staining is homogeneous or heterogeneous. Only nuclear staining was assessed.
b. To stain the unstained sections in their own laboratory using their standard routine method and to score these slides in the same way.
c. To grade the tumour on the H&E section.

Although the Liverpool-stained slides were there at the same time, the members of the Working Group were asked to score them independently. They were blinded to the original Liverpool reports of the cases.

Readers, all members of the Working Group because of their recognised expertise in breast pathology, submitted their assessment of the staining on standard forms designed for the purpose to the Cancer Screening Evaluation Unit (CSEU) in Sutton, Surrey. The results of the submitted forms were analysed anonymously. The CSEU then calculated the variation in assessment of the staining both on the slides stained in Liverpool and on the slides stained in the individual reader's laboratory using kappa statistics. Due to the large number of classes involved in the assessment using 10% bands, the bands were amalgamated into the groups used in the recommendations for the modified Quick score [1]. The reason for the analysis using this method was that the Quick score method was adopted as the recommended method after the circulations were performed. Kappa scores were calculated for the variation in percentage positivity for each case and for positive versus negative using a 10% positive cell staining cut-off for staining intensity and for the presence or absence of tumour heterogeneity. The scores were calculated according to the recommended method using the modified Quick score [6], and kappa statistics were performed on the scores 0–8.

Results

The results for the percentage of positive cells are given in Table 1; not all cases were reported by all pathologists; cases with less than 10 assessments were excluded.

While the results for the Liverpool-stained slide (i.e. interpretation only—line B) showed excellent correlation on kappa statistics for receptor negativity (0.95), the slides stained in the participant's own laboratories (line A) were a little more variable but still extremely good (kappa=0.82). When the variation in assessment of percentages was taken into account, there was reasonable correlation on assessment with a kappa score of 0.57 on the Liverpool-stained slide (Fig. 1 and Fig. 2), while this drops to only moderate correlation (kappa=0.4) when the inter-laboratory variation is taken into account. This in-

Table 1 Oestrogen receptor—percentage of positive cells. *A* own laboratory slide, *B* Liverpool slide

Slide		0	10–30	40–60	70–90	100	Majority opinion	% Of readers in agreement with majority
11595/97	A	11	0	0	0	0	0	100
	B	10	1	0	0	0	0	91
13955/97	A	1	0	1	9	0	70–90	82
	B	0	0	1	11	0	70–90	92
22338/97	A	1	0	1	9	0	70–90	82
	B	0	0	0	9	3	70–90	75
26342/97	A	2	2	4	4	0	Tie 40–60/70–90	-
	B	0	4	5	4	0	40–60	56
1/98	A	1	1	4	11	2	70–90	58
	B	0	0	0	15	5	70–90	75
4/98	A	0	0	3	6	12	100	57
	B	0	0	0	3	17	100	85
5/98	A	0	0	2	12	7	70–90	57
	B	0	0	0	4	16	100	80
7/98	A	0	1	0	10	9	70–90	50
	B	0	0	0	6	14	100	70
8/98	A	0	0	3	6	12	100	57
	B	0	0	0	5	15	100	75
9/98	A	0	0	1	4	15	100	75
	B	0	0	0	3	17	100	85
10/98	A	0	1	4	14	1	70–90	70
	B	0	0	2	15	3	70–90	75
11/98	A	0	1	4	13	3	70–90	62
	B	0	0	1	18	1	70–90	90
12/98	A	15	5	0	0	0	0	75
	B	17	2	1	0	0	0	85
13/98	A	0	1	5	13	2	70–90	62
	B	0	0	1	16	3	70–90	80
15/98	A	1	4	8	6	1	40–60	40
	B	0	2	10	8	0	40–60	50
16/98	A	0	0	2	6	13	100	62
	B	0	0	0	7	13	100	65
17/98	A	21	0	0	0	0	0	100
	B	20	0	0	0	0	0	100
18/98	A	0	0	2	12	7	70–90	57
	B	0	0	0	15	5	70–90	75
19/98	A	0	2	2	11	4	70–90	58
	B	0	0	0	16	4	70–90	80
20/98	A	0	2	5	9	4	70–90	45
	B	0	0	3	16	1	70–90	80
21/98	A	21	0	0	0	0	0	100
	B	20	0	0	0	0	0	100
22/98	A	19	2	0	0	0	0	90
	B	20	0	0	0	0	0	100
23/98	A	21	0	0	0	0	0	100
	B	20	0	0	0	0	0	100
24/98	A	19	0	0	0	0	0	100
	B	22	0	0	0	0	0	100
25/98	A	3	12	4	1	0	10–30	60
	B	0	11	10	1	0	10–30	50
26/98	A	4	6	3	7	0	70–90	35
	B	1	7	7	7	0	Tie10–30/40–60/70–90	-
Total	A	140	40	58	163	92	Grand total	493
	B	130	27	41	179	117		494
Kappa	A	0.82	0.20	0.06	0.27	0.33	Overall kappa	0.40
	B	0.95	0.27	0.26	0.45	0.52		0.57
Total ≤10%	A	163		Kappa	0.80			
	B	142		Kappa	0.91			
Total >10%	A	330		Kappa	0.80			
	B	352		Kappa	0.91			

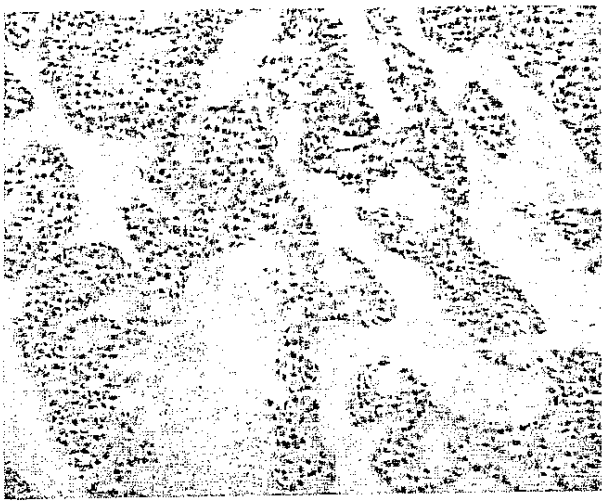

Fig. 1 Medium-power view of a Liverpool-stained oestrogen-receptor-positive slide (case 11/98), which had one of the best scores for inter-observer concordance in reporting (×200)

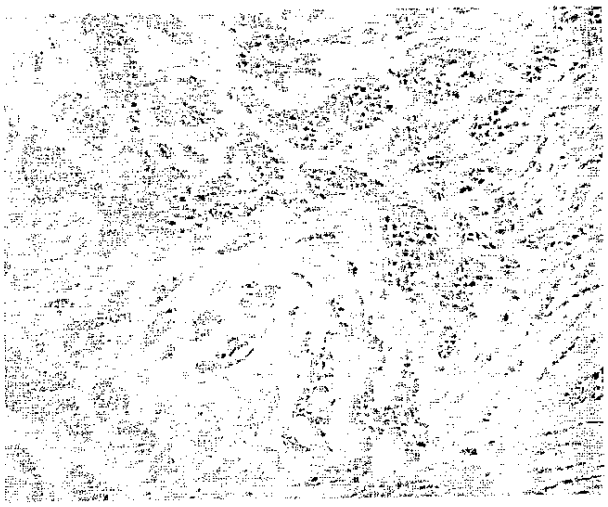

Fig. 2 Medium-power view of a Liverpool-stained oestrogen-receptor-positive slide (case 26342/97), which had the worst inter-observer concordance in reporting (×200)

creased variability must be interpreted as due to technical variability among methods used in each laboratory.

When the cut-off point for ER-positive tumours was based on the >10% value used in some laboratories, the correlation between laboratories was excellent (kappa=0.80) and even better for interpretation only (kappa=0.91).

The results for intensity are given in Table 2. Variation of the intensity reading by the pathologists is only assessed on tumours with positive staining, and, hence, the kappa statistics are lower than if negative staining was included. As in any three-way classification, the middle grade (i.e. intermediate in this case) shows poorer correlation than the two other grades. A similar result is found for intensity as with percentage positive cells in that the overall kappa value is 0.35 for interpretation only, which falls to 0.23 when inter-laboratory technical variation is also taken into account. These figures are low, as they exclude those cases with no staining, but could clearly be improved by closer quality assurance.

The results of the assessment of homogeneity or heterogeneity are given in Table 3. Again, the results show moderate agreement for the Liverpool-stained slide with a kappa value of 0.48, but this also excludes those cases with no staining. The level of agreement falls significantly when inter-laboratory variation is taken into account and is poor at 0.28.

The results of the analysis recalculating the results to give a Quick score are given in Table 4 for slides stained in each laboratory and Table 5 for the Liverpool-stained slide. For interpretation alone (Table 5), the results are excellent for lack of staining (0.95) and good for the highest Quick score of 8 (0.62). Kappas in the intermediate groups QS 3–7 are, however, not so good, with poor correlation within the groups. The results with inter-laboratory variation included show increased variability as before, still with an excellent score for negative tumours (0.82) but a slightly lower score of 0.45 for a Quick score of 8. Looking at the original results for each tumour, one can see that a small number of cases appear to be responsible for the majority of the variation. The cases with the most variation (23642/97, 15/98, 25/98 and 26/98) all scored highly for heterogeneity of staining, and this may well be an area where counting systems may be less consistent, as the area chosen for counting will vary from observer to observer.

Discussion

ER and progesterone-receptor analyses are now mandatory tests on breast cancer to determine appropriate therapy. There is now good evidence that tumours negative for ER are highly unlikely to respond to endocrine manipulation unless they happen to fall into the small proportion of tumours (approximately 2–5%) that are ER negative, progesterone-receptor positive. The older biochemical methods of assessment of tumour positivity have largely been superseded by the immunocytochemical method. This has gained favour due to the ready availability of good specific antibodies to the receptor proteins and due to the decreasing size of tumours detected using mammography. This latter problem has made it impossible to persist with the older biochemical methods, as there is insufficient tumour for analysis without compromising diagnosis.

ER status using immunocytochemistry is now crucial to therapy, especially in pre-menopausal women, where the side-effects of unnecessary Tamoxifen may be severe. Even in post-menopausal women, the side-effects are not negligible, and the rationale for giving irrelevant therapy in receptor-negative tumours is, at best, dubious.

The variation in general immunocytochemical staining is assessed by external quality assessment schemes run by

Table 2 Oestrogen receptor—intensity of stain. *A* own laboratory slide, *B* Liverpool slide

Slide		Weak	Moderate	Strong	Majority opinion	% Of readers in agreement with majority
11595/97	A	0	0	0	Negative	100
	B	1	0	0	Negative	95
13955/97	A	1	6	3	Moderate	60
	B	1	7	4	Moderate	58
22338/97	A	1	3	6	Strong	60
	B	0	3	9	Strong	75
26342/97	A	0	7	3	Moderate	70
	B	1	8	4	Moderate	62
1/98	A	6	11	2	Moderate	58
	B	1	13	6	Moderate	65
4/98	A	0	4	16	Strong	80
	B	0	0	20	Strong	100
5/98	A	1	3	17	Strong	81
	B	0	2	18	Strong	90
7/98	A	1	3	17	Strong	81
	B	0	2	18	Strong	90
8/98	A	1	1	19	Strong	90
	B	0	0	19	Strong	100
9/98	A	1	3	16	Strong	80
	B	0	2	18	Strong	90
10/98	A	3	12	5	Moderate	60
	B	1	9	10	Strong	50
11/98	A	2	6	13	Strong	62
	B	0	8	12	Strong	60
12/98	A	5	0	0	Negative	75
	B	3	0	0	Negative	85
13/98	A	3	8	9	Strong	45
	B	0	9	11	Strong	55
15/98	A	10	7	2	Weak	53
	B	5	14	1	Moderate	70
16/98	A	2	2	17	Strong	81
	B	0	0	20	Strong	100
17/98	A	0	0	0	Negative	100
	B	0	0	0	Negative	100
18/98	A	2	5	14	Strong	67
	B	0	6	14	Strong	70
19/98	A	3	5	11	Strong	58
	B	0	6	14	Strong	70
20/98	A	5	10	5	Moderate	50
	B	4	12	4	Moderate	60
21/98	A	0	0	0	Negative	100
	B	0	0	0	Negative	100
22/98	A	2	0	0	Negative	90
	B	0	0	0	Negative	100
23/98	A	0	0	0	Negative	100
	B	0	0	0	Negative	100
24/98	A	0	0	0	Negative	100
	B	0	0	0	Negative	100
25/98	A	11	5	1	Weak	65
	B	9	13	0	Moderate	59
26/98	A	10	6	0	Weak	62
	B	14	7	0	Weak	67
Total	A	70	107	176	Grand total	353
	B	40	121	202		363
Kappa	A	0.26	0.10	0.33	Overall kappa	0.23
	B	0.38	0.21	0.47		0.35

the National External Quality Assessment Scheme for Immunocytochemistry (NEQAS) scheme in the UK [14]. In other European countries, similar initiatives have been successfully established [11]. Although ER staining has been included in the UK scheme recently [16], the scheme does not assess the variation in reading the slides and in determining the variation in positivity and negativity of the test. Many laboratories are still using the simple method of counting the percentage of positive cells, although the intensity of staining appears also to be important. In this regard, there is a necessity to standardise the quality of the staining technique and the assessment to ensure that the level of the receptors is consistently estimated. Indeed, it is entirely possible that the response of some tumours with low Quick scores reported in the lit-

Table 3 Oestrogen receptor—distribution of positive cells. *A* own laboratory slide, *B* Liverpool slide

Slide		Homogeneous	Heterogeneous		% Of readers in agreement with majority opinion
11595/97	A	0	0		-
	B	0	1		100
13955/97	A	4	4		Tie
	B	6	4		60
22338/97	A	4	2		67
	B	6	2		75
26342/97	A	0	8		100
	B	0	9		100
1/98	A	12	6		67
	B	18	1		95
2/98	A	0	0		-
	B	0	0		-
3/98	A	0	0		-
	B	0	0		-
4/98	A	17	4		81
	B	19	1		95
5/98	A	12	8		60
	B	16	3		84
6/98	A	0	0		-
	B	0	0		-
7/98	A	18	3		86
	B	18	0		100
8/98	A	18	3		86
	B	20	0		100
9/98	A	17	3		85
	B	19	1		95
10/98	A	13	7		65
	B	10	10		Tie
11/98	A	8	12		60
	B	6	13		68
12/98	A	1	5		83
	B	2	3		60
13/98	A	13	7		65
	B	15	4		79
14/98	A	0	0		-
	B	0	0		-
15/98	A	5	14		74
	B	6	14		70
16/98	A	21	0		100
	B	19	1		95
17/98	A	1	0		100
	B	1	0		100
18/98	A	18	3		86
	B	18	2		90
19/98	A	12	7		63
	B	17	3		85
20/98	A	10	10		Tie
	B	7	12		63
21/98	A	1	0		100
	B	1	0		100
22/98	A	1	2		67
	B	1	0		100
23/98	A	1	0		100
	B	1	0		100
24/98	A	1	0		100
	B	1	0		100
25/98	A	0	17		100
	B	0	22		100
26/98	A	4	12		75
	B	4	17		81
Total	A	212	137	Grand total	349
	B	231	123		354
Kappa	A	0.28	0.28	Overall kappa	0.28
	B	0.48	0.48		0.48

Table 4 Oestrogen receptor—Quick score calculated from original submitted results of intensity and percentage positivity (own laboratory slide)

Case	Score of own laboratory slide						
	0	3	4	5	6	7	8
11595/97	11	0	0	0	0	0	0
13955/97	1	0	0	1	0	6	3
22338/97	1	0	0	0	2	2	6
26342/97	2	0	0	2	3	3	2
1/98	1	1	0	2	4	9	2
4/98	0	0	0	0	2	2	16
5/98	0	0	0	0	1	5	15
7/98	0	0	0	1	1	2	16
8/98	0	0	0	0	2	2	17
9/98	0	0	0	0	2	2	16
10/98	0	0	1	2	2	10	5
11/98	0	0	0	3	2	3	13
12/98	15	5	0	0	0	0	0
13/98	0	1	0	1	5	4	9
15/98	1	2	2	5	4	4	2
16/98	0	0	0	1	2	1	17
17/98	21	0	0	0	0	0	0
18/98	0	0	0	0	3	5	13
19/98	0	2	0	0	2	5	10
20/98	0	1	1	1	6	6	5
21/98	21	0	0	0	0	0	0
22/98	19	2	0	0	0	0	0
23/98	21	0	0	0	0	0	0
24/98	19	0	0	0	0	0	0
25/98	3	6	4	2	5	0	0
26/98	4	2	4	1	5	4	0
Kappa	0.82	0.10	0.09	0.04	0.04	0.13	0.45
					Overall kappa		0.39

Table 5 Oestrogen receptor—Quick score calculated from original submitted results of intensity and percentage positivity (Liverpool slide)

Case	Score of Liverpool slide						
	0	3	4	5	6	7	8
11595/97	10	1	0	0	0	0	0
13955/97	0	0	0	1	0	7	4
22338/97	0	0	0	0	0	3	9
26342/97	0	0	0	3	6	3	1
1/98	0	0	0	0	1	13	6
4/98	0	0	0	0	0	0	20
5/98	0	0	0	0	0	2	18
7/98	0	0	0	0	0	2	18
8/98	0	0	0	0	0	0	19
9/98	0	0	0	0	0	2	18
10/98	0	0	0	0	2	9	9
11/98	0	0	0	0	1	7	12
12/98	17	2	0	1	0	0	0
13/98	0	0	0	0	0	10	10
15/98	0	0	2	1	11	5	1
16/98	0	0	0	0	0	0	20
17/98	20	0	0	0	0	0	0
18/98	0	0	0	0	0	6	14
19/98	0	0	0	0	0	6	14
20/98	0	0	0	0	7	9	4
21/98	20	0	0	0	0	0	0
22/98	20	0	0	0	0	0	0
23/98	20	0	0	0	0	0	0
24/98	22	0	0	0	0	0	0
25/98	0	2	5	7	8	0	0
26/98	1	4	2	7	3	5	0
Kappa	0.95	0.07	0.11	0.19	0.29	0.23	0.62
					Overall kappa		0.55

erature is due to the technical differences in the quality of the immunocytochemical staining [11, 15].

The results of this circulation demonstrate that there is some variation in reporting of the level of ER positivity, but encouragingly the variation, on the whole, is very small for completely negative tumours, even when inter-laboratory variation is taken into account. This means that a negative result may be generally relied upon for therapy, unlike some of the older biochemical results. Similar reliability for ER-negative and also for progesterone-receptor-negative results have also been demonstrated in a study performed in a number of pathology departments in Austria [11]. Tumours with a high degree of tumour heterogeneity may cause problems in arriving at a consistent level of positivity, but this heterogeneity would also give rise to difficulties with biochemical methods, depending on which bit of the tumour is submitted for biochemical analysis.

The results do, however, indicate the need for some quality control of the technical aspects of immunocytochemical staining as well as guidance in the aspects of reporting of the stains. As Table 1 shows, there were a few "false-negative" cases reported, coming from 11 different laboratories, but there was one laboratory responsible for five of these false negatives. To reduce and avoid such discrepancies, the Working Group recommends that laboratories performing ER and progesterone-receptor tests for therapeutic reasons take part in one of the circulations of quality assurance of immunocytochemistry, such as the NEQAS scheme in the UK [2, 11, 13, 14]. It is also helpful to include a composite block with a known positive, a known negative and a very weakly positive tumour as a control with each run. The assessment of the strength of staining of internal controls in each slide can also help to remove variations. Comparison of the strength of staining with the strongest cell in the normal tissue allows a rather more objective assessment of the staining intensity. It is also recommended that the percentage of tumours positive and negative per year be calculated and that variation from the normally accepted distribution be investigated [8].

Variation in the assessment of the Quick score is interesting in that there is poor concordance between observers in allocating a specific Quick score to tumours except in very strongly positive or completely negative tumours, and this suggests that the eight-point score may have too many categories for good inter-laboratory concordance. It may be more consistent to stratify tumours into those that are completely negative (Quick score=0), negative with low receptor levels (Quick score=1 or 2), weakly positive (Quick score=3 or 4), moderately positive (Quick score=5 or 6), or strongly positive (Quick score=7 or 8). This is consistent with the known response rates to therapy and may remove some insignificant variation among laboratories in assessment of receptor levels [3]. When this was done in the present study (Table 6), the kappa statistics were clearly shown to be better and excellent for strongly positive (Quick score 7&8) cases and for negative cases (Quick score=0). Where assessment was made on the previously stained slide, the concordance of the moderately positive cases was also reasonable (kappa=0.43). This concordance disappears when the same cases are stained in individual laboratories, suggesting a technical variation, which needs to be addressed [10, 13].

Table 6 Oestrogen receptor—kappa scores calculated by broad bands from the above table

Score	Own slide	Liverpool slide
0	0.82	0.95
3.4	0.19	0.17
5.6	0.10	0.43
7.8	0.55	0.82
Overall kappa	0.51	0.76

In summary, the variation in assessment of receptor status among laboratories has two components, a technical variation and an interpretative variation. These variations are less pronounced in negative tumours and in strongly positive tumours, but significant variation is seen in tumours that express low or moderate levels of receptor. Technical external quality assessment schemes and inclusion of a weakly positive control may help reduce technical variation, and guidelines on assessment of receptors may also help to reduce the interpretative variation.

References

1. Barnes DM, Hanby AM (2001) Oestrogen and progesterone receptors in breast cancer: past, present and future. Histopathology 38:271–274
2. Barnes DM, Millis RR, Beex LVAAM, Thorpe SM, Leake RE (1998) Increased use of immunohistochemistry for oestrogen receptor measurement in mammary carcinoma: the need for quality assurance. Eur J Cancer 34:1677–1682
3. Early Breast Cancer Trialists Collaborative Group (1998) Tamoxifen for early breast cancer: an overview of randomised trials. Lancet 351:1451–1467
4. Elledge RM, Green S, Pugh R, Allred DC, Clark GM, Hill J, Ravdin P, Martino S, Osborne CK (2000) Estrogen receptor (ER) and progesterone receptor (PgR), by ligand-binding assay compared with ER, PgR and pS2, by immuno-histochemistry in predicting response to tamoxifen in metastatic breast cancer: a Southwest Oncology Group Study. Int J Cancer 89:111–117
5. Ferrero-Pous M, Trassard M, Le Doussal V, Hacene K, Tubiana-Hulin M, Spyratos F (2001) Comparison of enzyme immunoassay and immunohistochemical measurements of estrogen and progesterone receptors in breast cancer patients. Appl Immunohistochem Mol Morphol 9:267–275
6. Harvey JM, Clark GM, Osborne CK, Allred DC (1999) Oestrogen receptor status by immunohistochemistry is superior to the ligand-binding assays for predicting response to adjuvant therapy in breast cancer. J Clin Oncol 17:1474–1485
7. Lacroix M, Querton G, Hennebert P, Larsimont D, Leclercq G (2001) Estrogen receptor analysis in primary breast tumors by ligand-binding assay, immunocytochemical assay, and northern blot: a comparison. Breast Cancer Res Treat 67:263–271
8. Leake R (2000) Detection of the oestrogen receptor (ER). Immunohistochemical versus cytosol measurements. Eur J Cancer 36[Suppl 4]:S18–S19
9. Leake R, Barnes D, Pinder S, Ellis I, Anderson L, Anderson T, Adamson R, Rhodes T, Miller K, Walker R (2000) Immuno-

histochemical detection of steroid receptors in breast cancer: a working protocol. UK Receptor Group, UK NEQAS, The Scottish Breast Cancer Pathology Group, and The Receptor and Biomarker Study Group of the EORTC. J Clin Pathol 53:634–635
10. Lee H, Douglas-Jones AG, Morgan JM, Jasani B (2002) The effect of fixation and processing on the sensitivity of oestrogen receptor assay by immunohistochemistry in breast carcinoma. J Clin Pathol 55:236–238
11. Regitnig P, Reiner A, Dinges HP, Hoefler G, Mueller-Holzner E, Lax SF, Obrist P, Rudas M, Quehenberger F (2002) Quality assurance for estrogen and progesterone receptors by immunohistochemistry in Austrian pathology laboratories. Virchows Arch 441:328–334
12. Rhodes A, Jasani B, Balaton AJ, Barnes DM, Miller KD (2000) Frequency of oestrogen and progesterone receptor positivity by immunohistochemical analysis in 7016 breast carcinomas: correlation with patient age, assay sensitivity, threshold value, and mammographic screening. J Clin Pathol 53:688–696
13. Rhodes A, Jasani B, Balaton AJ, Miller KD (2000) Immunohistochemical demonstration of oestrogen and progesterone receptors: correlation of standards achieved on in house tumours with that achieved on external quality assessment material in over 150 laboratories from 26 countries. J Clin Pathol 53:292–301
14. Rhodes A, Jasani B, Barnes DM, Bobrow LG, Miller KD (2000) Reliability of immunohistochemical demonstration of oestrogen receptors in routine practice: interlaboratory variance in the sensitivity of detection and evaluation of scoring systems. J Clin Pathol 53:125–130
15. Thike AA, Chng MJ, Fook-Chong S, Tan PH (2001) Immunohistochemical expression of hormone receptors in invasive breast carcinoma: correlation of results of H-score with pathological parameters. Pathology 33:21–25
16. United Kingdom National Coordinating Group for Breast Screening Pathology (2001). Guidelines for non-operative diagnostic procedures and reporting in breast cancer screening. NHSBSP Publication, Sheffield
17. Zafrani B, Aubriot MH, Mouret E, De Cremoux P, De Rycke Y, Nicolas A, Boudou E, Vincent-Salomon A, Magdelenat H, Sastre-Garau X (2000) High sensitivity and specificity of immunohistochemistry for the detection of hormone receptors in breast carcinoma: comparison with biochemical determination in a prospective study of 793 cases. Histopathology 37:536–545

Virchows Arch (2002) 441:328–334
DOI 10.1007/s00428-002-0646-5

ORIGINAL ARTICLE

Peter Regitnig · Angelika Reiner · Hans-Peter Dinges
Gerald Höfler · Elisabeth Müller-Holzner
Sigurd F. Lax · Peter Obrist · Margaretha Rudas
Franz Quehenberger

Quality assurance for detection of estrogen and progesterone receptors by immunohistochemistry in Austrian pathology laboratories

Received: 30 October 2001 / Accepted: 17 February 2002 / Published online: 17 September 2002

Abstract Steroid hormone receptors are important prognostic and predictive factors in breast carcinomas. Thus their determination is of essential importance. The aims of this study were to assess the quality of the immunohistochemical assays, and to assess the interlaboratory and interobserver variability performed by different laboratories in Austria. Ten unstained slides for interlaboratory variability evaluation and ten immunohistochemically prestained slides for interobserver variability evaluation from breast carcinomas known to show different degrees of steroid hormone receptor expressions were sent to 32 surgical pathology laboratories in Austria (participation rate 97%). The participants were requested to perform their in-house immunohistochemistry (IHC) technique for estrogen receptors (ERs) and progesterone receptors (PRs) on the unstained slides. All slides were evaluated by estimating percentage and intensity of stained nuclei semiquantitatively. From these data the Reiner, Remmele and the Allred scores were calculated. A less than 10% cut-off level was chosen as threshold for positive cases. Regarding the series of prestained slides, both sensitivity and specificity were very high (>96.88%); false-positive and -negative rates were low (<3.31%). Interobserver variability showed moderate multirater kappa values concerning the ER (Reiner score: kappa=0.57) and PR scores (Reiner score: kappa=0.53). The agreement among observers was better for negative cases than positive cases. In-house slides representing interlaboratory variability showed fair to moderate kappa values concerning the ER and PR scores (kappa for ER Reiner score=0.41; PR=0.32). In this slide series, sensitivity and specificity were high (>82.2%) and false-positive or -negative rates were low in ER cases (<3.03) and moderately low in PR cases (17.46%). These results demonstrate that variability is higher when participants use their own staining method. In more detailed analysis, the automated IHC techniques showed an advantage over manual techniques concerning interlaboratory variability. There exists no difference in reproducibility with respect to scoring systems for steroid hormone receptor estimation.

Keywords Steroid hormone receptor status · Estrogen- and progesterone receptor · Immunohistochemistry · External quality assurance · Reproducibility · Breast carcinoma

P. Regitnig (✉) · G. Höfler · S.F. Lax
Institut für Pathologie der Karl-Franzens Universität Graz,
Auenbruggerplatz 25, 8036 Graz, Austria
e-mail: peter.regitnig@kfunigraz.ac.at
Fax: +43-316-3853432

A. Reiner
Pathologisch–Bakteriologisches Institut Donauspital am SMZO Wien, Austria

H.-P. Dinges
Institut für Pathologie LKH Klagenfurt, Austria

E. Müller-Holzner
Universitätsklinik für Frauenheilkunde und Geburtshilfe Innsbruck, Morphologisches Labor, Austria

P. Obrist
Institut für Pathologie der Universität Innsbruck, Austria

M. Rudas
Institut für Klinische Pathologie der Universität Wien, Austria

F. Quehenberger
Institut für medizinische Informatik der
Karl-Franzens Universität Graz, Austria

Introduction

The importance of determining estrogen receptor (ER) and progesterone receptor (PR) status of breast carcinoma as a prognostic factor and for treatment decisions is well known and widely accepted [4]. The ER and PR status is one of the most important predictive factors in breast carcinoma. Patients with steroid hormone receptor-positive breast carcinomas have a significant advantage in overall and recurrence-free survival receiving adjuvant endocrine treatment [8, 14]. Yet patients with ER-negative breast carcinomas do not benefit from this treatment. Therefore the correct determination of the

steroid hormone receptor status is critical. Nowadays, in Austria, the most widely used method for determination of ER and PR is immunohistochemistry (IHC) using monoclonal antibodies. However, the American Society of Clinical Oncology (ASCO) states in the 1997 recommendation for the use of tumor markers in breast cancer and the 2000 update that the IHC technique is not validated enough [2]. In other studies with large patient cohorts, comparing enzyme immunoassays and IHC, the immunohistochemical method was concluded to be more specific. This was due to the possibility of discarding nonrelevant positivity related to intraductal normal or neoplastic cells expressing hormone receptor and because it was easier, safer, less expensive and had an equivalent or better ability to predict response to adjuvant endocrine therapy [7, 10, 25].

IHC is especially advantageous over biochemical assays for small tumor samples which will become more frequent in the future because of higher detection rates of small carcinomas with improved mammography and the use of mammotome technique [24]. Considering the clinical importance and the frequent use of IHC for ER and PR analysis, it is important that quality assurance procedures are in place. With this consideration in mind, an external quality assurance project was started in Austrian hospitals under the patronage of the Austrian Society of Pathology. The aims were to assess the quality of the different IHC assays carried out by different laboratories and to assess the interlaboratory and interobserver variability. An additional goal was to find techniques that should be suggested as reliable assays for receptor determination.

Materials and methods

The study was considered to be carried out among the 33 departments of pathology in public hospitals in Austria, and the participation was voluntary. All participating laboratories were coded for further data entry and to guarantee anonymity. The study center was in the Department of Pathology of the Donauspital – SMZO, Vienna, Austria. Briefly, the study consisted of two parts. Two sets of histological slides were sent out. One set contained unstained slides to be stained using the in-house method for ER and PR IHC; the other set of slides was immunostained for ER and PR to be analyzed according to a standardized semiquantitative method that was explained on a form.

Case selection

The study material consisted of four invasive ductal carcinomas (not otherwise specified), two mixed lobuloductal carcinomas, one invasive lobular carcinoma, one adenoidcystic and one cribriform carcinoma. All cases were selected from the surgical pathology archive of the Department of Pathology of the Donauspital – SMZO, Vienna, Austria. Formalin-fixed and paraffin-embedded tissue blocks containing tumor and, if possible, adjacent non-neoplastic breast tissue were selected from each case. For the whole study only slides from one tissue block per case were used to keep differences in tissue fixation as minimal as possible. The tissue had been fixed in 7.5% buffered formalin. Each of the ER and PR slide series contained two receptor-negative and three receptor-positive cases.

Interlaboratory staining series

The interlaboratory staining series (ILSS) contained ten unstained histological slides. Five of these slides each marked with ER or PR, respectively, should have been stained either for ER or PR using the in-house routine laboratory technique. After immunostaining, the slides had to be evaluated by a pathologist of the participating laboratory who was familiar with the particular staining technique and staining intensity. The results for ER and PR IHC had to be reported in a standardized form containing both staining intensity and percentages of positive cells for each case and had to be returned together with the immunostained slides to the study center. A second form was requested to be filled out concerning the immunohistochemical method used.

Interobserver staining series

A second set of slides of the same breast carcinomas was stained in the laboratory of the study center using an automated IHC staining system [interobserver staining series (IOSS); Ventana Nexes, Ventana Medical Systems, Strasbourg, France] and sent to all participants for evaluation. This automated staining technique was chosen as reference ER and PR receptor detection technique, particularly since the laboratory of the study center had participated previously in a trial for the reproducibility of steroid hormone receptor IHC in the European Commission Working Group for Breast Screening Pathology. Using this particular automated staining protocol, highly concordant results with other participants of this trial had been achieved (publication in preparation).

Automated staining protocol for reference series

From 7.5% buffered formalin-fixed, paraffin-embedded tissues, 4-µm-thin sections were cut and mounted on precoated slides and stored at 58°C overnight. After dewaxing in xylene, slides were heated for antigen retrieval in 200 ml 0.001 M ethylene diamine tetraacetic acid (EDTA) buffer at pH 8.0 in a microwave oven at 300 W for 17 min. The primary antibody for ER (6F11, Ventana) was applied without further dilution at 37°C for 28 min and the PR antibody (1A6, Ventana) was incubated at 37°C for 32 min. For detection of antibody binding, the Ventana-Basic diaminobenzidine (DAB) kit was used. Endogenous peroxidase was blocked with Ventana's blocking solution for 4 min; DAB was used as chromogen for 8 min. The sections were counterstained with methylene blue.

Scoring

The participants were asked to determine the staining intensity (no staining, weak, moderate or strong) and the percentages of stained tumor cells in steps of 10% from 0% to 100% with an additional <10% step. Intermediate values were not allowed. For all the evaluations, standardized forms were used. From these data, three published and widely used immunoreactive scores, the Reiner score [15, 16], Remmele score [18, 19] and the Allred score [1, 7] were automatically calculated. These scores are also known as "Quick-scores". All these scoring systems combine the intensity and the number of positive cells in different ways for calculation of the score. A summary for calculation and score interpretation is given in Table 1. Since all scoring systems use different thresholds to determine whether a case is positive or negative, we chose a cut-off level for negative cases of less than 10% positive nuclei with weak staining intensity, which is equal to a Reiner score 2, Remmele score 1 and Allred score 2. This cut-off is clinically validated in some studies [3, 5, 13]. Nevertheless, until today there is uncertainty about the correct cut-off point for immunohistochemical detection of ER and PR in breast carcinomas [22].

Table 1 Different semiquantitative scoring systems for estrogen and progesterone receptors used in this study

Reiner score				
Intensity	+	Percentage	=	Score (0–7)
4 categories	No=0 Weak=1 Moderate=2 Strong=3	5 categories	0%=0 <10%=1 10% to 50%=2 51% to 80%=3 >80%=4	Score interpretation 0–2=negative 3=weakly positive 4–5=moderately positive 6–7=highly positive
Remmele score				
Intensity	x	Percentage	=	Score (0–12)
4 categories	No=0 Weak=1 Moderate=2 Strong=3	5 categories	0%=0 <10%=1 10%–50%=2 51%–80%=3 >80%=4	Score interpretation* 0–2=negative ≥3=positive
Allred score				
Intensity	+	Positive cells (percentage)	=	Score (0–8)
4 categories	No=0 Weak=1 Moderate=2 Strong=3	6 categories	0 cells (0%)=0 1 out of 100 cells (1%)=1 1 out of 10 cells (10%)=2 1 out of 3 cells (30%)=3 2 out of 3 cells (60%)=4 3 out of 3 cells (100%)=5	Score interpretation 0–2=negative ≥3=positive

* Remmele et al. did not give a cut-off value in his original works [18, 19, 20], but in a later statement [17] he wrote that 0–2 points should be classified as negative

Table 2 Kappa interpretation according to Landis and Koch

Kappa value	Strength of agreement
<0	Poor
0.00–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost perfect

Statistics

Excel 97 (Microsoft, Redmond, USA) was used for data input and score calculation. STATA 6.0 (Stata Corporation, Tex., USA) was used for statistical analysis of the results. Sensitivity and specificity were calculated for positive and negative test results as the main factors for treatment decisions. For these calculations the golden standard was estimated through evaluation of the reference series by a panel of four experts who also reevaluated the ILSS.

Multirater kappa was chosen as a measure of interobserver and interlaboratory variability for intensity, percentages of immunoreactive cells, resulting scores and for comparison of different laboratory techniques. Multirater kappa is widely used and hence comparable in medical studies dealing with interobserver variability. Kappa values were calculated according to Fleiss [6] for each group separately and for all groups (overall kappa). Overall kappa is the summary of the agreement across all observers, adjusted for the level of agreement that would be expected to occur solely by chance. Kappa was interpreted according to Landis and Koch (Table 2) [12]. Nevertheless multirater kappa harbors some disadvantages when used for three or more categories, as the number of possibilities that theoretically can be chosen by observers is not included in the calculation and therefore different estimation systems are difficult to compare. For example, intensity has 4 possible categories, whereas percentage has 12 possible categories. To reduce this statistical test effect, subgroups of percentages were combined to five groups according to percentage combination used in the Reiner and Remmele scoring systems.

Results

Participation rate

Of the 33 surgical pathology laboratories established in Austrian hospitals, 32 participated in this external quality assurance program. The only non-participating laboratory specializes in pulmonary and orthopedic pathology and thus does not analyze ER and PR in breast tissue. Thus, we had a very high participation rate of 97%.

Interobserver variability

To evaluate the interobserver variability, a standard set of prestained slides (IOSS) was used. Therefore the only variables that can cause variations in the results are perception and interpretation. Sensitivity and specificity were very high; false-positive and -negative rates were low (Table 3). One error occurred in one laboratory by exchanging one PR and ER slide from the same case (same case number, one slide stained for ER and one for PR). In order

Table 3 Sensitivity and specificity for prestained slides (interobserver staining series)

	Sensitivity	Specificity	False-positive rate	False-negative rate
Estrogen receptor	99.0%	100%	0%	1.0%
Progesterone receptor	96.8%	96.9%	3.1%	3.1%

Table 4 Overall kappa values for prestained slides (interobserver staining series)

	Overall kappa for estrogen receptors	Overall kappa for progesterone
Intensity	0.63	0.57
Percentage (grouped)	0.74	0.74
Reiner score	0.57	0.53
Remmele score	0.58	0.52
Allred score	0.57	0.51

Table 5 Sensitivity and specificity for laboratory stained slides (interlaboratory staining series)

	Sensitivity	Specificity	False-positive rate	False-negative rate
Estrogen receptor	97.8%	97.0%	3.0%	2.0%
Progesterone receptor	92.5%	82.5%	17.5%	7.5%

Table 6 Overall kappa values for laboratory stained slides (interlaboratory staining series)

	Overall kappa for estrogen receptor	Overall kappa for progesterone receptor
Intensity	0.46	0.35
Percentage (grouped)	0.52	0.46
Reiner score	0.41	0.32
Remmele score	0.41	0.31
Allred score	0.42	0.36

to demonstrate interobserver scoring variability in more detail, overall kappa values for ER and PR are summarized in Table 4. Interobserver variability showed moderate kappa values concerning the ER and PR scores, whereas the ER was slightly better. Better agreement among observers was seen in ER and PR negative cases with standard deviations (SD) of 0.35 and 0.39, respectively, in Reiner Score, whereas ER and PR positive cases had SD ranging from 0.46 to 1.28. This disagreement resulted mainly from different intensity estimations (ER kappa= 0.63; PR kappa=0.57) rather than percentage estimations (ER kappa=0.74; PR kappa=0.74).

Laboratory comparison

In-house IHC techniques, intensity and percentage estimation on pre-cut unstained slides (ILSS) were used to evaluate the interlaboratory reproducibility. Six cases (9.6%) in the ER series and four (6.4%) in the PR series were not evaluated by different laboratories due to technical problems. Two cases in the ER series and seven cases in the PR series were incorrectly scored as negative. Nevertheless, sensitivity and specificity were high, and false-positive or -negative rates were low in ER cases and moderately low in PR cases (Table 5). Overall kappa values are summarized in Table 6. Interlaboratory variability showed fair to moderate kappa values concerning the ER and PR scores, whereas the ER was better than PR. Only the ER negative cases had, as in the IOSS (prestained series), a smaller SD (0.63–0.71) than positive cases (SD: 0.76–1.53) in Reiner scores. In the PR series, no difference in SD for negative and positive cases was visible (SD positive cases: 0.62–1.59; negative cases: 0.78–1.78). Results achieved in the IOSS relative to the ILSS are shown in Fig. 1.

Comparison of methods

Overall four groups with different staining techniques could be distinguished. 14 laboratories used manual techniques that sometimes differed markedly, 31 used a Ventana autostainer, 26 used a Dako autostainer (Glostrup, Denmark) and one laboratory used the LabVision autostainer (Newmarket, England). The latter was not included in the following statistical analysis and results because it could not form its own group. Overall kappa values concerning staining results are shown in Table 7. The automated techniques (Ventana and Dako) were advantageous over manual techniques used in this study concerning interlaboratory consistency. The Dako autostainer had a substantial interlaboratory agreement in ER intensity, whereas the Ventana autostainer was better with respect to PR. Nevertheless, as shown by the nonparametric repeated-measures analysis, there were no statistically significant differences in estimation of percentage or of intensity between different staining techniques.

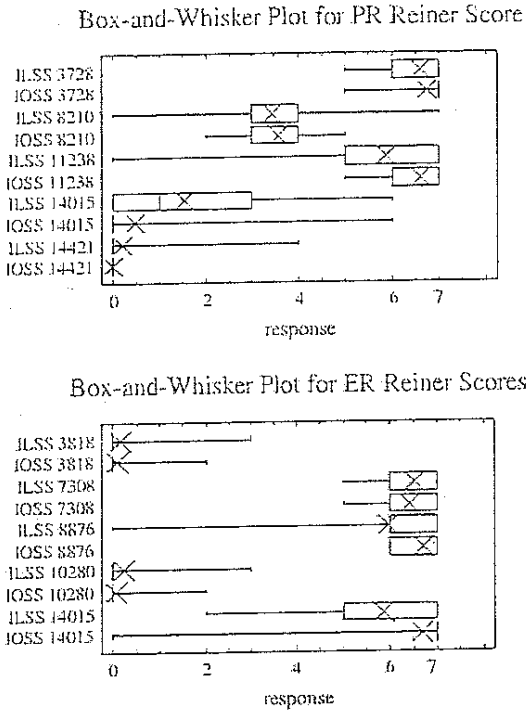

Fig. 1 Box and Whisker plots for Reiner scores for progesterone receptor (PR) and estrogen receptor (ER) estimation by 32 participants of prestained slides (interobserver staining series, *IOSS*) and by laboratory stained slides (interlaboratory staining series, *ILSS*). On the *y-axis*, the study case numbers and the slide series are designated. For each case this plot divides the data into four equal areas of frequency. A *box* encloses the middle 50% (1st to 3rd interquartiles), where the mean is represented as X. *Horizontal lines*, called whiskers, extend from each end of the box. The lower (*left*) whisker is drawn from the lower quartile to the smallest result; the other whisker is drawn from the upper quartile to the highest result

Table 7 Overall kappa values for laboratory stained series (ILSS) comparing different staining techniques. Best results are printed bold. *ER* estrogen receptor, *PR* progesterone receptor

Staining technique		Ventana n=11	Dako n=6	manually n=14
Intensity	ER	0.44	**0.62**	0.44
	PR	**0.39**	0.25	0.35
Percentage (grouped)	ER	0.49	**0.55**	0.53
	PR	**0.51**	0.44	0.40
Reiner score	ER	0.41	**0.45**	0.40
	PR	**0.39**	0.34	0.28
Remmele score	ER	0.42	**0.45**	0.38
	PR	**0.36**	0.31	0.26
Allred score	ER	0.42	**0.45**	0.40
	PR	**0.36**	0.32	0.25

Discussion

Immunohistochemical assays for ER and PR have largely replaced the biochemical ligand-binding-assays in Austria as in many other countries. The four main criteria required for technical validation of laboratory assays are specificity, sensitivity, reproducibility and the possibility to be interpreted in a uniform manner by different laboratories. In this study, an external quality control program concerning all technical criteria was performed using immunohistochemically prestained slides from breast carcinomas to test interobserver variability (IOSS) and unstained formalin-fixed, paraffin-embedded tissue sections to test interlaboratory variability (ILSS). As shown by Rhodes et al. [21] external material is an accurate indicator of in-house laboratory performance. For both series (IOSS and ILSS), specificity and sensitivity were high with respect to the decision if a tumor was steroid hormone receptor positive or negative with a less than 10% weak staining intensity cut-off level, which is equal to a Reiner score 2, Remmele score 1 and Allred score 2. That means treatment decisions based on the pathologist's steroid hormone receptor evaluation are correct in an essential percentage. This is a very important issue since many Austrian breast carcinoma patients participate in various clinical trials of the Austrian Breast Cancer and Colorectal Study Group (ABCSG). Several of these studies use the steroid hormone receptor status for patient randomization [9].

Nevertheless laboratories and observers showed considerable variability in estimation of intensity and percentage of receptor-related immunostaining. The interlaboratory variability in scoring the unstained PR series was unexpected. For better results in the future, more efforts toward standardization of laboratory techniques are necessary. A first step toward this aim was reviewing the slides of all participants and giving feedback to all laboratories on their performance. After the return of slides and forms to the coordinating laboratory and data analysis, each participant received data sheets showing the results of all laboratories in comparison with their own results and demonstrating the opinions of four experts to whom the slides were circulated for reevaluation. The panel of reviewers consisted of seven pathologists with special interest and expertise in breast pathology and evaluation of ER and PR IHC, respectively. Each case was presented in two data sheets: one containing intensity and the other one percentage of reactivity (Fig. 2).

As shown in a study by Rhodes et al. microwave antigen retrieval was the main cause of poor and variable results in immunohistochemical detection of steroid hormone receptors [23]. In their study, extension of heating time resulted in significant improvement regardless of all other variables in the immunohistochemical protocol. Today there is no standardized assay established for the immunohistochemical detection of ER and PR in Austria. Regarding the laboratory methods used by the participants, we could identify six different methods either with or without automation for either receptor, which gave the most reproducible results (Table 8). This means that the detailed staining results achieved with these assays were exactly in concordance with the majority results of participants. These protocols were forwarded to participants in order to suggest possibilities for improve-

Table 8 Best methods for immunohistochemical detection of estrogen receptors and progesterone receptors. Multibuffer: 5 g Titriplex, 2.5 g Tris buffer, 3.5 g potassium-citrate, Aqua dest. 1000 ml, pH 7.8

Best methods for estrogen receptors			
Method	Peroxidase/AEC	Peroxidase/AEC	Peroxidase/AEC
Autostainer	No	Techmate HORIZON	Ventana NEXES
Antigen retrieval	Microwave:3 min at 700 W+8 min at 400 W+4 min at 300 W	Microwave:6+7 min at 600 W	Microwave 3×10 min at 450 W
Buffer	Citrate buffer pH 6	Citrate buffer pH 6	Multibuffer
Primary antibody	1D5 (Dako)	1D5 (Dako)	6F11 (Ventana)
Dilution and incubation	1:25, 30 min	Prediluted, 25 min	Prediluted, 32 min/37° C
Blocking	3% H2O2, 5 min	Blocking solution from kit 2, 30 min	Inhibitor solution from kit 4 min/37° C
Detection system	ChemMate-kit (Dako)	ChemMate-kit (Dako)	Ventana AEC basic detection kit
Chromogen	AEC 5 min	AEC 5 min	AEC 8 min/37°C
Best methods for progesterone receptors			
Method	Peroxidase/ DAB	Peroxidase/DAB	Peroxidase/AEC
Autostainer	No	Ventana ES	Ventana NEXES
Retrieval method	80°C overnight	Microwave 30 min at 160 W	Microwave 3×10 min at 450 W
Buffer	Citrate buffer pH 6	Citrate puffer pH 6	Multibuffer (see below)
Primary antibody	Polyclonal A0098 (Dako)	1A6 (Ventana)	1A6 (Ventana)
Dilution and incubation	1:100, 25 min	Prediluted, 32 min/37° C	Prediluted, 32 min/37° C
Blocking	Blocking solution from kit 7 min	Inhibitor solution from kit 4 min/37° C	Inhibitor solution from kit 4 min/37° C
Detection system	ChemMate-kit (Dako)	Ventana DAB basic detection kit	Ventana AEC basic detection kit
Chromogen	DAB 1–5 min	DAB 8 min/37° C	AEC 8 min/37° C

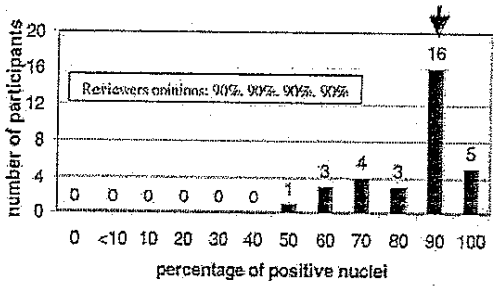

Fig. 2 Example of feedback data sheet sent to all participating laboratories. Information given is percentage of estrogen receptors (ER) in a case with ER positivity in primarily unstained slide series (ILSS). *Columns* show results of all participants, *numbers above columns* show the respective number of participants' opinions, and the *arrow* points to the participant result. In addition, four independent reviewer opinions on the participant's slide are noted

ments. Overall automated staining protocols gave the most reliable results.

Interobserver variability showed moderate to substantial kappa, which is quite promising for correct staining interpretation by pathologists. The exchange error of one laboratory was a mistake, which might be caused by using unfamiliar forms. The participants had no problems in identifying receptor-negative cases. The difficulties concerning correct classification of positive cases are less for estimating percentage of cells than for the intensity of staining.

Kappa values in different scoring systems (Reiner, Remmele or Allred) were nearly equal. This was obviously due to the fact that all scores were results of either multiplying or adding the same factors, i.e., intensity and percentage of positively stained tumor cells. Only the Allred scoring system differs in estimating the amount of positively stained tumor cells. As our results showed, semiquantitative estimation of percentage is not a major problem for experienced pathologists. There exists no difference for reproducibility with respect to scoring systems for steroid hormone receptor estimation.

Major interobserver differences were seen in the estimation of staining intensity. This problem cannot easily be explained. Hypothetically this could be caused by unfamiliar staining intensities in the IOSS compared with the familiar in-house intensities. In this case, kappa values in the ILSS should be better than in the IOSS, which was not true. This shortcoming may be due to the nature of human perception itself and thus may not be eliminated in conventional microscopy. Several image analysis systems exist that could help to overcome this problem at least partly [11, 18]. They are, however, expensive and the analysis is time consuming and therefore not available in every routine laboratory. Moreover, breast carcinoma is a very common disease and therefore methods are needed that may be used widely.

We conclude that the major factors influencing reproducibility of steroid hormone receptor IHC are staining intensity and estimation of the percentage of positively stained tumor cells. Both factors depend mainly on the immunohistochemical technique used. To achieve better assay concordance, we suggest laboratories interchange stained and unstained slides with a reference laboratory to ensure correct laboratory techniques and evaluation. Furthermore, standardization of immunohistochemical protocols should be achieved and regularly negative, slightly positive and strong positive controls should be used to guarantee internal quality assurance in each laboratory determining ER and PR expression for breast carcinoma patients. Further trials are under progress to improve the quality standards concerning IHC associated with breast pathology in Austria.

Acknowledgements We are indebted to all pathologists who contributed to this Austrian quality assurance study. The authors are further grateful to Mrs. Marina Engel for her diligent work on producing all the slides and for providing assistance to participating laboratories.

References

1. Allred DC, Harvey JM, Berardo M, Clark GM (1998) Prognostic and predictive factors in breast cancer by immunohistochemical analysis. Mod Pathol 11:155–168
2. Bast RC, Jr. Ravdin P, Hayes DF, Bates S, Fritsche H, Jr. Jessup JM, Kemeny N, Locker GY, Mennel RG, Somerfield MR (2001) 2000 update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology. J Clin Oncol 19:1865–1878
3. De Mascarel I, Soubeyran I, MacGrogan G (1995) Immunohistochemical analysis of estrogen receptors in 938 breast carcinomas. Concordance with biochemical assay and prognostic significance. Appl Immunohistochem 3:222–231
4. Early Breast Cancer Trialists' Collaborative Group (1998) Tamoxifen for early breast cancer: an overview of the randomised trials. Lancet 351:1451–1467
5. Ferno M, Andersson C, Fallenius G, Idvall I (1996) Oestrogen receptor analysis of paraffin sections and cytosol samples of primary breast cancer in relation to outcome after adjuvant tamoxifen treatment. The South Sweden Breast Cancer Group. Acta Oncol 35:17–22
6. Fleiss JG (1981) Statistical methods for rates and proportions. Wiley, New York, NY, pp 225–232
7. Harvey JM, Clark GM, Osborne CK, Allred DC (1999) Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. J Clin Oncol 17:1474–1481
8. Jaiyesimi IA, Buzdar AU, Decker DA, Hortobagyi GN (1995) Use of tamoxifen for breast cancer: twenty-eight years later. J Clin Oncol 13:513–529
9. Jakesz R, Hausmaninger H, Haider K, Kubista E, Samonigg H, Gnant M, Manfreda D, Tschurtschenthaler G, Kolb R, Stierer M, Fridrik M, Mlineritsch B, Steindorfer P, Mittlbock M, Steger G (1999) Randomized trial of low-dose chemotherapy added to tamoxifen in patients with receptor-positive and lymph node-positive breast cancer. J Clin Oncol 17:1701–1709
10. Kell DL, Kamel OW, Rouse RV (1993) Immunohistochemical analysis of breast carcinoma estrogen and progesterone receptors in paraffin-embedded tissue, correlation of clones ER1D5, and 1A6 with cytosol-based hormone receptor assay. Appl Immunohistochem 1:275–281
11. Kohlberger PD, Breitenecker F, Kaider A, Losch A, Gitsch G, Breitenecker G, Kieback DG (1999) Modified true-color computer-assisted image analysis versus subjective scoring of estrogen receptor expression in breast cancer: a comparison. Anticancer Res 19:2189–2193
12. Landis JR, Koch GG (1977) The measurement of observer agreement for categorial data. Biometrics 33:159–174
13. Perischuk LP, Feldman JG, Kim YD, Braithwaite L, Schneider F, Braverman AS, Axiotis C (1996) Estrogen receptor immunocytochemistry in paraffin embedded tissues with ER1D5 predicts breast cancer endocrine response more accurately than H222Sp gamma in frozen sections or cytosol-based ligand-binding assays. Cancer 77:2514–2519
14. Pritchard KI (2000) Current and future directions in medical therapy for breast carcinoma: endocrine treatment. Cancer 88:3065–3072
15. Reiner A, Spona J, Reiner G, Schemper M, Kolb R, Kwasny W, Fugger R, Jakesz R, Holzner JH (1986) Estrogen receptor analysis on biopsies and fine-needle aspirates from human breast carcinoma. Correlation of biochemical and immunohistochemical methods using monoclonal antireceptor antibodies. Am J Pathol 125:443–449
16. Reiner A, Neumeister B, Spona J, Reiner G, Schemper M, Jakesz R (1990) Immunocytochemical localization of estrogen and progesterone receptor and prognosis in human primary breast cancer. Cancer Res 50:7057–7061
17. Remmele W (1997) Pathologie, 2nd edn, vol 4. Springer, Berlin Heidelberg New York, p 259
18. Remmele W, Schicketanz KH (1993) Immunohistochemical determination of estrogen and progesterone receptor content in human breast cancer. Computer-assisted image analysis (QIC score) vs. subjective grading (IRS). Pathol Res Pract 189:862–866
19. Remmele W, Stegner HE (1987) Recommendation for uniform definition of an immunoreactive score (IRS) for immunohistochemical estrogen receptor detection (ER-ICA) in breast cancer tissue. Pathologe 8:138–140
20. Remmele W, Hildebrand U, Hienz HA, Klein PJ, Vierbuchen M, Behnken LJ, Heicke B, Scheidt E (1986) Comparative histological, histochemical, immunohistochemical and biochemical studies on oestrogen receptors, lectin receptors, and Barr bodies in human breast cancer. Virchows Arch 409:127–147
21. Rhodes A, Jasani B, Balaton AJ, Miller KD (2000) Immunohistochemical demonstration of oestrogen and progesterone receptors: correlation of standards achieved on in house tumours with that achieved on external quality assessment material in over 150 laboratories from 26 countries. J Clin Pathol 53:292–301
22. Rhodes A, Jasani B, Barnes DM, Bobrow LG, Miller KD (2000) Reliability of immunohistochemical demonstration of oestrogen receptors in routine practice: interlaboratory variance in the sensitivity of detection and evaluation of scoring systems. J Clin Pathol 53:125–130
23. Rhodes A, Jasani B, Balaton AJ, Barnes DM, Anderson E, Bobrow LG, Miller KD (2001) Study of interlaboratory reliability and reproducibility of estrogen and progesterone receptor assays in Europe. Documentation of poor reliability and identification of insufficient microwave antigen retrieval time as a major contributory element of unreliable assays. Am J Clin Pathol 115:44–58
24. Vargas HI, Agbunag RV, Khaikhali J (2000) State of the art of minimally invasive breast biopsy: principles and practice. Breast Cancer 7:370–379
25. Zafrani B, Aubriot MH, Mouret E, De Cremoux P, De Rycke Y, Nicolas A, Boudou E, Vincent-Salomon A, Magdelenat H, Sastre-Garau X (2000) High sensitivity and specificity of immunohistochemistry for the detection of hormone receptors in breast carcinoma: comparison with biochemical determination in a prospective study of 793 cases. Histopathology 37:536–545

BCU 7 | 2004

Editor's Note

Team in need of a coach

Every medical oncology fellow quickly learns about interdisciplinary cancer care, but thank God for the American College of Surgeons' mandate for tumor boards, because without them, we might be strangers. Personally, I don't like to think about any surgeon, radiation oncologist or medical oncologist not regularly attending one of these valuable meetings. However, the truth is that we really don't report to anyone, and our collaboration is pretty much voluntary.

This issue of our audio series attempts to demonstrate how critical it is that interdisciplinary team members talk to each other. We begin with the local control guys, and Pat Borgen and Frank Vicini comment on a plethora of surgical and radiation therapy research issues that profoundly affect systemic management decisions.

For example, Dr Vicini is the principal investigator of a critical NSABP-RTOG randomized clinical trial evaluating partial breast irradiation (PBI). This historic collaboration between two premier collaborative clinical trial groups will provide much-needed answers about PBI, albeit many years from now. In the interim, the pace at which this accelerated and patient-friendly treatment strategy permeates into the nonprotocol management algorithm utilized in the community treatment setting is anyone's guess.

While we wait for definitive research results, patients should seek input from every team member regarding the advisability of PBI and which technique is preferable. Pat Borgen cautions us that local control may have much more of an impact on long-term survival than previously recognized, and one might imagine that PBI could either have a deleterious effect (if it results in suboptimal local tumor control) or could be a more effective modality (because treatment can be implemented prior to chemotherapy).

With an increasing number of patients receiving taxane-based adjuvant regimens that can take up to six months to complete, earlier radiation therapy could have a potential antitumor advantage.

From a quality of life perspective, avoiding six weeks of daily treks for radiation therapy is appealing, particularly after the physical and emotional trauma of adjuvant chemotherapy. However, patients will surely want to know what their medical oncologist has to say on this issue before they opt for an unproven treatment modality.

Input from Craig Allred, the pathologist for the interdisciplinary team collaborating on this issue of *Breast Cancer Update*, is unfortunately very disheartening. I have nothing personal against pathologists or Craig, who is a really nice man, but if Adam Brufsky's interview provides ample documentation that contemporary systemic therapy of breast cancer is essentially target-driven, then Craig's comments leave us wondering if we have the ability to measure the most critical targets every oncologist must consider — ER, PR and HER2 status. (My apologies to Phillip Roth for that very long sentence.)

I keep expecting some rebel breast cancer patient advocacy group to stage a massive protest at the NCI to demand that pathologists provide impeccable ER, PR and HER2 assays. At the present time, however, women are going to continue to relapse unnecessarily or receive suboptimal palliative care because we can't get their pathology right. Even if recent history tells us that our usually capable nation is not totally effective in military intelligence gathering, we should be able to at least gather accurate

information for the war on cancer.

Maybe we need more than ACOS-mandated tumor boards. Maybe we need someone to rally and guide the entire team — including nurses, pharmacists, radiologists, psychologists, social workers and others — and take a deep breath, and really figure out how to work together better so patients can receive the very best care we have.

— Neil Love, MD
NLove@ResearchToPractice.net

Select publications

Barb Williams

Archives of Pathology and Laboratory Medicine: Vol. 124, No. 7, pp. 945–951.

The Total Test Approach to Standardization of Immunohistochemistry

Clive R. Taylor, MD, DPhil

From the Department of Pathology and Laboratory Medicine, Keck School of Medicine, University of Southern California, Los Angeles.

Accepted January 7, 2000

Histopathology continues to serve as the mainstay for diagnostic and therapeutic decisions in almost all forms of neoplasia. Yet, as previously described,[1] surgical pathologists in >routine practice have problems recognizing many of the more subtle criteria, and even experts are far from infallible.[2] It has been written that "All pathologists would probably agree that the interpretation of histologic sections is subjective, often extending, through a spectrum of possibilities"[3] and that "the final morphologic diagnosis often is hedged with uncertainty."[4] In some circumstances, the hedges loom exceedingly tall, so tall that even though "we stand on the shoulders of giants,"[5] we can scarcely see over them.

It has been argued[1] that rigorous scientific studies of the accuracy, reliability, and reproducibility of histopathologic diagnoses are rare but nonetheless disconcerting and that significant disagreement between pathologists exceeds 25% for difficult but clinically important diagnostic distinctions. With regard to malignant lymphomas, even the experts are seen to encounter difficulties: "It seems to me that pathologists are getting into difficulties in the area of non-Hodgkin's lymphoma. Not only do we clinicians have trouble in understanding pathologists, but at times they seem unable to understand each other."[6] Studies of diagnostic consensus for malignant lymphomas describe a level of agreement as low as 35% for practicing pathologists and 50% for experts applying the recently derived REAL classification. Attempts to predict prognosis by using histologic criteria for the grading of tumors are fraught with even greater difficulty.[1] The fact that morphologic criteria still constitute the 'gold standard' for diagnosis in many conditions reflects not only the success of pathologists in coming to terms with subjectivity but also in some measure the lack of an alternative and more accurate diagnostic

method. A serious and less well-recognized corollary is that in most cases we lack a reliable means of validating the surgical pathology diagnosis on which therapy is predicated.

SPECIAL STAINS Return to TOC

It may be argued that special stains evolved to meet just this need. The first special stains were variations of the basic biological dyes. Specific histochemical stains soon evolved based on the in situ identification of active enzymes within cells by the application of colorogenic substrates.[7] Many histochemical methods were, however, technically exacting, both in terms of tissue preparation and staining protocol, a situation that led to recognition of the critical importance of appropriate positive and negative controls for interpretation of the findings.

Although the methods were challenge enough, the reagents themselves were often poorly characterized and somewhat variable from manufacturer to manufacturer and lot to lot. In the United States, the Biologic Stain Commission (BSC) was founded in 1944 as a nonprofit corporation in large part to address the problem of standardization of chemical stains.[8] The BSC achieved considerable success with regard to biological dyes, but with the advent of immunohistochemistry, special stain technology was about to take a new leap forward, with great opportunities and new challenges in standardization.

THE ADVENT OF IMMUNOHISTOCHEMISTRY FOR ROUTINE FORMALIN PARAFFIN SECTIONS Return to TOC

In 1978 immunohistochemistry was only 4 years old. Nonetheless, imbued with the enthusiasm of relative youth and fortified by a mere 6 years of hands-on experience in the field, I was able to write of "the potential value of immunohistochemical methods as an immediate aid to diagnosis, and more importantly as a tool for the reshaping and redefining of current histologic criteria."[4] At that time, the whole field of immunohistochemistry as applied to routine formalin paraffin sections was encompassed in a score or so of manuscripts from around the world. By 1986 the literature had surpassed the 1000 mark, and I was moved to add, "Widespread application of immunohistological methods, by virtue of their great specificity, inevitably will transform histopathology from something resembling an art into something more closely resembling a science."[9] Although the utility of immunohistochemistry is accepted today by most pathologists, my optimism as to the scientific application of the method appears to have been somewhat misplaced in view of the lack of standardization and reproducibility among different laboratories. Immunohistochemistry has become an established method in surgical pathology—every (well, almost every) large academic center in the world uses the method routinely, some staining 10 000 or more slides per year. Immunohistochemistry is also offered by reference laboratories and even by many small community hospitals, where it is applied to the occasional case. But therein lies the rub; enormous variability has developed in terms of the reagents available, the detection methods used, and the interpretation and reporting of immunohistochemical findings. For these reasons, among others, the members of the BSC began, in the 1980s, to turn its attention to immunohistochemistry, sponsoring a series of workshops that were attended by pathologists, manufacturers, and representatives of the Food and Drug Administration.[10,11]

THE COMPLETE IMMUNOHISTOCHEMICAL STAIN Return to TOC

Immunohistochemistry is technically complex, and no aspect of this complexity can be ignored, from the moment of collecting the specimen to issuance of the final report. In this context, anatomic pathologists should be prepared to follow the fine example of their colleagues in clinical pathology and should broaden their responsibility to cover all aspects of the immunohistochemical assay.[1,12,13] The complete immunohistochemical stain (Table 1 ⊃=) begins with the identification of a need for the performance of immunohistochemical studies and progresses through procurement of the specimen, identification of the appropriate stain, selection of the proper reagents and protocol, and correct performance of the protocol with the appropriate controls. It concludes with an evaluation of the results and a written report. An attendant goal of long-term follow-up (outcome analysis) remains just that, a goal.

Clinical Question and Test Selection

Selection of the appropriate stain for a particular diagnostic problem should be an integral part of the process. Such a selection cannot be made in isolation from the clinical situation (diagnostic problem) on the one hand or the capabilities and limitations of the immunohistochemical laboratory on the other, including those limitations imposed by the experience of the staff and the type of tissue (frozen or fixed) available for study. The blunderbuss approach of ordering a panel of 10 or more stains on every suspected large cell anaplastic tumor can no longer be justified economically or academically. For each stain used in a particular case, the pathologist should ask, and answer, the question, "What will a positive/negative result add to the diagnosis?" If the answer is not known in the field of pathology or to the individual pathologist, then the stain should not be performed, because it will not add to the diagnostic equation.

In the face of the diagnostic conundrum that cannot be resolved by orthodox morphologic criteria, the pathologist has 2 main options. The first and more common modus operandi is to select stains from a panel that, in the experience of the pathologist, have proven to be of value in a particular diagnostic area (eg, anaplastic tumors, malignant lymphoma). The second approach uses an algorithm with sequential panels of selected stains that introduces a certain degree of logic and thought into the process but may compromise the turnaround time due to sequential staining.[13] Once the appropriate stains have been selected, the determinations of whether such stains can be performed successfully on the tissue available, whether the necessary reagents are at hand, and whether the detection methods are tried and true must be made; all these determinations will be readily apparent from review of the within-run (same day) and run-run (day-to-day) control records, which are an intrinsic part of an established quality assurance program. Although obvious, it can scarcely be overemphasized that without detailed knowledge of the performance characteristics of each primary antibody, it is not possible to interpret the final result, even in the presence of controls that perform as expected.

There is also an important ongoing shift in emphasis in the use of immunohistochemical stains from a primary focus on cell and tissue markers as an aid to the recognition and classification of tumors to the demonstration of cell products, receptors, or oncogenes of possible prognostic value, and the identification of infectious agents in situ.[13,14] The precision required for this latter approach places new and more exacting demands on our ability to perform immunohistochemistry or in situ hybridization in a reproducible and standardized manner.

Specimen Acquisition and Management

Aspects of tissue handling also cannot be ignored. Pathologists have sought to undo some of the adverse affects of formalin fixation either by the use of controlled enzymatic digestion or more recently by the antigen retrieval technique.[13,15,16] The latter method (sometimes known as heat-induced epitope

retrieval or as unmasking) subjects sections to microwave heating in the presence of a retrieval solution that may serve to stabilize or postfix the antigens present. Although vigorous heating in a microwave oven offends basic instincts, the method has contributed to the reproducibility of immunostaining by providing a more uniform presentation of antigens in tissue sections than is otherwise present following inconsistent fixation in formalin. However, one significant caveat has been issued,[1,16] namely that "as different antigen retrieval approaches are explored and propagated, there is a danger that different laboratories each will adopt a different procedure, producing varying degrees of restoration of antigenicity, that will add yet another variable to the overall process." Unfortunately this warning has come to pass.[16] Clearly, the preference is for anatomic pathologists to adopt more uniform and rigorous procedures for the fixation and processing of tissues, extending from the moment that the specimen is removed from the body to the end of the embedding process, particularly controlling the total time in fixative and the composition of the fixative itself. The plea also was made that "if we cannot have uniform fixation, let us at least strive towards a more uniform approach to 'unfixation,' utilizing standardized and well described antigen retrieval procedures."[1] The pressure for consistency in staining has grown even further because of the increasing focus on prognostic markers, where greater stringency is required.

Again, in the clinical laboratory, we collect a specimen in the right tube with the right preservative and the right anticoagulant, and we process it in the prescribed manner, all without thinking. In anatomic pathology, we largely fail to accomplish any of the analogous steps, again without thinking.

The National Committee for Clinical Laboratory Standards currently is developing detailed guidelines that address the issues of fixation, reagent selection, choice of protocol, and other aspects of the complete immunohistochemistry test. The document is in the final stages of approval, and publication is expected within the year.

Technology and Methodology

Although it is perhaps not practical to attempt to standardize a single protocol across all laboratories, because of the great variability in specimen procurement, fixation, and processing, it certainly should be possible to standardize protocols within a single laboratory. The goal is to ensure run-to-run reproducibility, evaluated against standard control sections from day to day and week to week. Achieving this goal requires careful selection of reagents, rigorous use of controls, and strict adherence to the protocol for preparation of slides, application of reagents to slides, and incubation times. Today, it is increasingly clear that, just as with complex procedures in the clinical laboratory, consistency of performance can only be achieved by automation of the score or more separate steps that constitute an immunohistochemical stain.[1] The Model T Ford immunostainer pioneered by David Brigati more than a decade ago has given way to a wide range of automated immunostainers, the best of which will outperform even the best technologist in terms of consistency during a sustained period. Of course, such instruments should not be viewed as a replacement for a skilled technologist but rather as an adjunct technology that frees the technologist to focus on issues of quality assurance and interpretation.

The issue of selection of appropriate high-quality antibodies and reagents was addressed directly by the BSC in conjunction with the Food and Drug Administration. The result was the publication of a set of guidelines for package inserts.[11] This document suggested not only a standard outline for manufacturers to follow in the testing and marketing of reagents but also provided a series of recommendations for positive and negative controls that could be followed by both manufacturers and performing laboratories (Table 2 ⊃). By adherence to these guidelines, manufacturers will provide pathologists with more detailed information on the specificity and working conditions of antibodies than currently is available.

With regard to detection procedures, the BSC concluded that it was unable to recommend a standardized staining protocol from the many variations in current use. Establishing a single universal protocol would seem to be impossible, recognizing that the different laboratories have different needs, and probably is not desirable, providing that all laboratories use appropriate controls. Here again is a spin-off advantage derived from the adoption of automated methods in that, just as in the clinical laboratory, automation tends to force compliance in terms of reagent choice and protocol, placing the onus of reagent qualification on the manufacturer, with a resulting increase in quality of product from the better manufacturers. Conversely, pathologists should recognize that any departure from the manufacturer's protocol or reagents effectively voids the responsibility of the manufacturer and places responsibility for validation totally on the performing laboratory.

Analytical Issues

The experience and training of the technologist who is performing the stain clearly are critical to this process. Again, following the model of the clinical laboratory, one major factor in ensuring reproducibility may be the growth of automation, with the extended capability for consistency and control that is inherent in the automated process. The availability of automated immunostainers has proven of overall benefit in many smaller laboratories, if not some larger ones, that do not have the luxury of highly skilled staff who are experienced in reagent titration and quality control. It is certainly our experience that the use of an automated stainer increases the reproducibility and reliability of a wide variety of immunohistochemical stains. Increasingly, other investigators report similar experience, although not all agree on the choice of system; hands-on demonstrations of the system(s) that most closely fit need and budget are necessary before initiating a purchase. With any automated system, there is a training period during which inconsistencies and difficulties may be encountered. However, as these are resolved, the end results far exceed the quality that is achievable by even the best technologist on a day-to-day basis.

Quality Assurance.—The College of American Pathologists (CAP) defines quality assurance as a "process of assuring that all pathology services involved in the delivery of patient care have been accomplished in a manner appropriate to maintain excellence."[1] As such, quality assurance is more a mind set than a system, but it does incorporate the much more precise concept of quality control, which is the "aggregate of processes and techniques so derived as to detect, reduce and correct deficiencies in the analytic process."

As described elsewhere,[1] "anatomic pathology changed little in the 100 years preceding 1970. Sequestered in a technologic limbo, it remained relatively untouched by the new methodologies and automated systems that revolutionized the clinical laboratory. The histology laboratory performing only a few simple stains escaped the rigors of quality assurance in general and quality control in particular. To dip a slide in hematoxylin for a few minutes, then briefly differentiate it in alcohol, until it looks 'about right' to the technologist and 'makes the pathologist happy' may suffice an H&E [hematoxylin-eosin] stain, but applied to immunohistochemistry it is a recipe for disaster."

In seeking to improve the reliability and reproducibility of immunohistochemistry, we would be well advised to take a 20-year-old page from our colleagues in the clinical immunology laboratory.[17] Immunofluorescence assays for antinuclear antibodies do not differ from immunohistochemistry in any substantial way in terms of immunologic principles or performance characteristics. Even the end point is somewhat similar, expressed as the degree of staining of certain tissue components, and here also interlaboratory inconsistencies posed similar problems.[1] However, there are differences in stringency of definition of the end point in that strict criteria exist for interpretation of immunofluorescence assays in relation to reference standards (controls) that have been available commercially for many years. Also,

defined protocols for quality control, storage, and utilization of the various reagents are part of routine day-to-day practice in the clinical laboratory, encouraged to a large degree by clinical laboratory certification programs that are intrinsic to hospital certification but have intruded meaningfully on anatomic pathology only recently (modified guidelines of the Laboratory Accreditation Program of CAP). If we are to elevate immunohistochemistry to a level of reproducibility that matches comparable techniques in the clinical laboratory, then we must subject immunohistochemistry to a similar degree of scrutiny, including all its component parts and each stage of the procedure.[1,12,13]

CAP Certification and Proficiency Testing Programs.—The initiative of the BSC prompted revision and expansion of the Check List for Immunohistochemistry used by inspectors during the CAP laboratory certification process. The criteria for certification of an immunohistochemistry laboratory were expanded from 5 to 15 items, producing a much more comprehensive set of requirements for the successful performance of immunohistochemical stains. Many aspects of this expanded checklist have been incorporated in the forthcoming National Committee for Clinical Laboratory Standards guidelines, which should prove invaluable to all laboratories performing immunohistochemistry in a diagnostic setting. In addition, CAP conducts an Immunohistochemistry and Proficiency Testing Program, in which participating laboratories receive unstained paraffin sections of representative cases with instructions to perform a panel of immunohistochemical stains and controls. Following performance of the stains, the laboratories report their findings to the CAP Cell Markers Program, where results are collated and compared. A detailed critique is provided to each subscriber. The program, therefore, serves as an external proficiency test of the performance and interpretation of immunohistochemical stains on external tissues; it does not purport to measure the efficacy of internal specimen handling and fixation procedures, a function that can only be met by internal quality control programs.

Results

Validation and Reporting.—A critical problem for immunohistochemistry has been the lack of universal controls or reference standards. Interestingly, a similar problem was encountered early in the evolution of quality assurance procedures in the clinical laboratory.[18] For serum assays, it was solved by establishing large standardized serum pools and making these available to both manufacturers and laboratories. In addition, CAP "check sample" programs established other pools, samples of which were provided to the growing number of laboratories participating in proficiency testing programs. The test results from these different laboratories served to provide extensive validation of these pools as additional reference standards. An analogous CAP "check sample" program exists for immunohistochemistry, but the pool of reference materials (paraffin blocks of characterized tumors) is of limited supply, insufficient to serve as a generally available standard.

The BSC has debated the development of a reference standard or a series of standards that could be made available for general use.[1] However, as noted herein, the solution is not so simple with regard to characterized paraffin blocks that cannot be pooled and are available in limited amounts: no 2 tumors are antigenically identical. Improved multitissue blocks[19] provide an interim solution. Multitissue reference standard blocks could even be prepared from the residual material of cases used in the CAP survey program, thereby providing material that already has been evaluated and validated by several hundred different laboratories. But such multitissue blocks, even though they contain minute slivers of tissue, are not inexhaustible. The BSC has proposed the possibility of developing infinite amounts of standard reference materials in the form of multitissue blocks composed of artificial tissues, including human tumor cell lines or human tumor heterotransplants in mice with severe combined immunodeficiency disorders.[1]

The Immunohistology Report.—In response to a request from the BSC, the Association of Directors

of Anatomic and Surgical Pathology published a proposal for the organization and content of an immunohistology report.[20] The essential elements of this proposal are incorporated into Table 3. Adherence to this proposal would do much to introduce consistency and thoughtfulness into the field. The interpretation and significance of the findings should be presented in the context of the overall differential diagnosis. Interpretation of the presence of specific positive staining, or lack thereof, is, of course, a complex issue. It is a function of the performance and examination of the proper controls and the experience of the laboratory performing the stains, especially the pathologist responsible for evaluating the stained slides. A number of difficulties remain into this area, including some of the same problems that intrude in surgical pathology in general. Just how brown (or red) is positive? If we accept "weakly positive" as a valid finding, then how do we define "weakly negative"? What percentage of positive cells constitutes a positive tumor? How many swallows make a summer?[21]

Interpretation and Significance

A second aspect to interpretation relates to the way we arrive at an opinion regarding the significance of a particular set of staining results in relation to the diagnosis or prognosis of the patient. Here one can do no better than to refer once more to the principles established in the process of test selection.

These are principles that are more readily expounded than accomplished. The literature has expanded far beyond the thousand papers written by 1986 to a number exceeding that in a single year. Most surgical pathology publications now incorporate immunohistochemical findings to some degree. To keep pace with the relevant literature is a challenge that it is impossible to meet in broad perspective. Even within some specialty fields, such as hematopathology, the relevant literature on CD marker counts is still climbing, with another conference due in 2000. The journal literature can scarcely stay abreast with the productivity of the immunologists and molecular engineers, and even if it could, the reader cannot. "As a rule disease can scarcely keep pace with the itch to scribble about it."[22] Specialist journals, such as *Applied Immunohistochemistry and Molecular Morphology*,[23] partly meet the need of currency that textbooks cannot provide. The prototypic Web site, pioneered by Dennis Frisman,[24] may provide precedent for the future but suffers from selectivity of content and places great demands on the Web master for maintenance of currency.

Appropriateness and Cost-effectiveness

Last but not least in considering the complete immunohistochemical stain, there are the issues of appropriateness, outcome analysis, cost, and cost-effectiveness. In these areas, there is little hard information. The accuracy of surgical pathology alone has already been found wanting in certain critical areas (eg, grading, borderline lesions of breast, dysplasias, lymphomas).[1] A number of studies have shown that immunohistochemistry is helpful. In one early study of more than 100 anaplastic tumors,[25] the H&E diagnosis of carcinoma or lymphoma was revised in approximately 50% of cases following basic immunohistochemical studies, providing dramatic evidence of a contribution to patient care (of 43 carcinomas, 27 were in reality lymphomas). In a separate study of 200 consecutive cases,[26] it was concluded that immunohistochemistry contributed to the diagnosis in almost 50% and was confirmatory in the remainder. In a subsequent analysis of 557 poorly differentiated round cell or spindle cell tumors that could not be diagnosed on the basis of H&E morphologic structure, immunohistochemistry "provided a definitive diagnosis in 70% of the former and 92% of the latter."[27] Other analyses of difficult cases seen in consultation practices around the country, including our experience at the University of Southern California of approximately 20 000 cases during 25 years, have been strongly supportive of the need, even the necessity, for appropriate immunohistochemical studies.

Cost-effectiveness must be viewed against this background. Reagents for an individual immunohistochemical stain are not inexpensive for some proprietary antibodies; furthermore, the salaries of laboratory personnel and professional fees for interpretation by a trained pathologist must also be taken into account. A panel of 2 or 3 immunohistochemical stains that distinguishes a lymphoma from a carcinoma may save thousands of dollars that may have been incurred by additional investigations, extended hospital stays, or inappropriate therapy. As an emerging example, the detection of micrometastases by immunohistochemical methods in bone marrow or possibly lymph nodes of patients with breast cancer offers potential benefits to a large patient population with significant cost savings.[14]

Although appropriateness criteria, as proposed by RAND (Santa Monica, Calif), for such medical procedures as coronary angiography and coronary artery bypass surgery have yet to be developed for immunohistochemistry,[12] the currently available data strongly support immunohistochemistry as a rapid and cost-effective technique that has become an essential part of surgical pathology practice. Concerted action with regard to standardization and quality assurance, coupled with progressive automation of the procedure and perhaps eventually even the reading of basic results by image analysis, will certainly enhance the value still further.

ONGOING ACTION Return to TOC

Much of this material, although written specifically for immunohistochemistry, applies equally to in situ hybridization methods and reagents used therein. On the technical side, all laboratories that perform immunohistochemical staining for diagnostic support are encouraged to participate in the CAP Laboratory Certification and Immunohistochemical Proficiency Testing Programs. A proposed format for reporting immunohistochemical results exists[20]; we would be wise to follow it. Finally, the National Committee for Clinical Laboratory Standards is expected to release quality assurance guidelines for immunohistochemistry in the near future, adding further to the momentum for overall quality improvement and consistency.

Collectively, these processes almost certainly will elevate standards, but to a degree they are peripheral to the problem. The performance and practice of immunohistochemistry are aspects of the practice of medicine. It is the province of the pathologist, and rightly it is our responsibility. If, as pathologists, we do not accept this responsibility, either by conscious choice or benign neglect, then we must expect others to establish and monitor standards of performance. If this occurs in immunohistochemistry, now that we have been alerted to the problem, we know who to blame!

"We have met the enemy, and he is us." (W. Kelly, Pogo cartoon, 1970).

Acknowledgments

My thanks are owed to Myrna Cisneros for preparation of the manuscript.

References Return to TOC

1. Taylor CR. An exaltation of experts: concerted efforts in the standardization of immunohistochemistry. *Appl Immunohistochem* 1993;1:232–243.

2. van Ginneken AM, van der Lei J. Understanding differential diagnostic disagreement in pathology.

Proc Annu Symp Comput Appl Med Care 1991:99–103.

3. Hensen DE. Studies in observer variation. *Arch Pathol Lab Med* 1991;115:991–992.

4. Taylor CR. Immunohistological approach to tumor diagnosis. *Oncology* 1978;35:189–197.

5. Newton I. If I have seen further than other men, it is by standing on the shoulders of giants. *Letter to R. Hooke* February 5, 1675.

6. Aisenberg AC. In case rounds of the Massachusetts General Hospital: case 30. *N Engl J Med* 1977;297:206

7. Mann G. *Physiologic Histology.*. Oxford, England: Oxford University Press; 1902.

8. Mowry RW. Report from the president: the Biological Stain Commission: its goals, its past and its present status. *Stain Technol* 1980;55:1–7. [PubMed Citation]

9. Taylor CR. *Immunomicroscopy: A Diagnostic Tool for the Surgical Pathologist.*. Philadelphia, Pa: WB Saunders Co; 1986.

10. Taylor CR. Quality assurance and standardization in the immunohistochemistry: a proposal for the annual meeting of the Biological Stain Commission, June 1991. *Biotech Histochem* 1992;67:110–117. [PubMed Citation]

11. Taylor CR. Report of the Immunohistochemistry Steering Committee of the Biological Stain Commission: proposed format: package inserts for immunohistochemistry products. *Biotech Histochem* 1992;67:323 [PubMed Citation]

12. Hilborne LH, Nathan LE. Quality of assurance in an era of cost containment. *Am J Clin Pathol* 1991;96(suppl 1):S6–S9.

13. Taylor CR, Cote RJ. *Immunomicroscopy: A Diagnostic Tool for the Surgical Pathologist.*. Philadelphia, Pa: WB Saunders Co; 1994.

14. Taylor CR, Cote RJ. Immunohistochemical markers of prognostic value in surgical pathology. *Histol Histopathol* 1997;12:1039–1055. [PubMed Citation]

15. Shi S-R, Key ME, Kalra KL. Antigen retrieval in formalin-fixed, paraffin-embedded tissues: an enhancement method for immunohistochemical staining based on microwave oven heating of tissue sections. *J Histochem Cytochem* 1991;39:741–748. [PubMed Citation]

16. Shi SR, Cote RJ, Chaiwun B. et al. Standardization of immunohistochemistry based on antigen retrieval technique for routine formalin-fixed tissue sections. *Appl Immunohistochem* 1998;6:89–96.

17. Feigenbaum P, Medsger TA, Kraines G, Fries JF. The variability of immunologic tests. *J Rheumatol* 1982;9:408 [PubMed Citation]

18. Dorsey DB. The evolution of proficiency testing in the U.S.A. In: *Proceedings of the Second National Conference on Proficiency Testing.* Bethesda, Md: Information Services; 1975.

19. Miller RT. Multitumor "sandwich" blocks in immunohistochemistry: simplified method of

preparation and practical uses. *Appl Immunohistochem* 1993;1:156–159.

20. Banks PM. Incorporation of immunostaining data in anatomic pathology reports. *Am J Surg Pathol* 1992;16:808 [PubMed Citation]

21. Heywood J. One swallowe maketh not sommer. *A dialogue containing the Proverbs in the English tongue*:1546

22. Mayow J. (1640–1679), *De Rachitide,* Pt. V.

23. *Applied Immunohistochemistry & Molecular Morphology*.. Philadelphia, Pa: Lippincott Williams & Wilkins; 1999.

24. Frisman D. Immunohistoquery. Available at: http:/immunoquery.com.

25. Gatter KC, Alcock C, Heryet A. et al. Clinical importance of analyzing malignant tumours of uncertain origin with immunohistochemical techniques. *Lancet* 1985;1:1302 [PubMed Citation]

26. Leong AS-Y, Wright J. The contribution of immunohistochemical staining in tumor diagnosis. *Histopathology* 1987;11:1295

27. Leong AS, Wannakrirot P. A retrospective analysis of immunohistochemical staining in identification of poorly differentiated round cell and spindle cell tumors—results, reagents and costs. *Pathology (Australia)* 1992;24:254–260.

Tables Return to TOC

Table 1. The Total Test*

Elements of Testing Process	Quality Assurance Issues	Responsibili
Clinical question and test selection	Indications for immunohistochemistry, selection of stain(s)	Surgical pathologist, s cian
Specimen acquisition and management	Specimen collection, fixation, processing, sectioning	Pathologist or technol
Technology and methodology	Reagents, protocols, sensitivity, specificity	Pathologist or technol
Analytical issue	Qualifications of staff in/a laboratory and interlaboratory proficiency testing of procedures	Pathologist or technol
Validation and reporting	Criteria for positivity and negativity in relation to controls	Pathologist or technol
Interpretation and significance	Experience and qualifications of pathologist, proficiency testing of interpretational aspects, diagnostic and prognostic significance, appropriateness and correlation with other data	Surgical pathologist a

* Data are from Taylor.[1]

Table 2. Types and Purposes of Daily Quality Control Materials for Immunohistochemistry*

Type of Control	Antigen (Analyte)	Antibody (Reagent)	Purp
Positive	1. Nonpatient tissue or cells containing antigen to be detected and quantitated 2. Known expected result 3. (*a*) Fixed or processed in same way as patient sample (*b*) Fixed or processed in manner shown to preserve antigen under analysis	Antibody reagent of the kit constituted in same way as intended for patient sample	1. Control of analysis 2. Training of ance of p Comparis tiation of n 3. (*a*) Validat analysis, i and proces (*b*) Valida analysis es processing val laborat
Negative control (specific)	1. Tissues or cells expected to be negative by antibody of kit) 2. Processed in same way as patient sample 3. May be portion of patient sample	Antibody reagent of the kit constituted in same way as intended for patient sample	Detection of a body cross or cellular c
Negative control (nonspecific)	1. Patient tissue with components that are the same as tissue to be studied 2. Processed in same way as patient sample	1. Diluent (as used with antibody) *without* antibody OR 2. Antibody not specific for antigen or interest in same diluent as used with kit antibody	Detection of a ground stair

* Data are from Taylor[1] and Taylor and Cote.[33]

Table 3. The Immunohistochemistry Report*

1. Patient demographics and specimen identification data
2. Reference to the diagnostic problem (i.e., differential diagnosis)
3. Nature of specimen analyzed (frozen, fine needle aspiration, paraffin section, and fixative)†
4. Statement of all stains used with details of all primary antibodies (designate specificity and clone where appropriate)‡
5. Findings both positive and negative for all stains; sufficient details of patterns and controls to justify the interpretation§
6. The immunohistochemistry report should not stand alone but should be integrated into the final surgical pathology report

* Data are from Taylor.[1]

† Details of fixation and detection systems should be kept on record in the laboratory wher they were performed but should be incorporated in the report, where they are an essential pa of the interpretative process (eg, whether enzyme digestion or antigen retrieval was used or B versus formalin fixation for certain leukocyte antigens).

‡ for example anti–T cell is not acceptable; report should read T-cell antibody UCHL- (CD45RO). Anti-pan keratin should read anti-keratin cocktail (AE-1 + Cam 5.2 + 34βE12 + 35βH11).

§ Detailed control records should be retained in the laboratory where they were performe Findings may also be stated in the report, where they are contributory. For example, in a brea tumor that is negative for estrogen receptor, the presence of residual normal breast epitheliu that shows positivity is significant. Similarly, in an anaplastic large cell tumor that shows absenc of staining for leukocyte common antigen (CD45), the presence of intermingled leukocyte con mon antigen–positive small lymphocytes should be described.

Presented at the College of American Pathologists Conference XXXV: Solid Tumor Prognostic Factors: Which, How and So What?, Chicago, Ill, June 10–13, 1999.

Reprints: Clive Taylor, MD, DPhil, Department of Pathology and Laboratory Medicine, University of Southern California, 2011 Zonal Ave, HMR 204, Los Angeles, CA 90033-1054.

George Tilley

From: George Tilley
Sent: Monday, October 24, 2005 3:09 PM
To: 'Hassen, Philip'
Cc: Bob Williams (robert.williams@easternhealth.ca)
Subject: RE: ER/PR testing

Phil, the media associated with this seems to have leveled off, but needless to say individual patients are anxious to reaffirm their results. I do know that we have asked the Canadian Association of Pathologists to put this issue on the agenda for their November Board meeting. Also our Chief of Laboratory Medicine is in the process of making contact with the individuals that Bob Bell referenced. So it is very much "a wait and see" approach.

George

George Tilley
President/Chief Executive Officer
Eastern Health
c/o Corporate Office
Waterford Bridge Rd.
A1E 4J8
Tel: 709-777-1330
Fax: 709-777-1302

From: Hassen, Philip [mailto:phassen@cpsi-icsp.ca]
Sent: Monday, October 24, 2005 1:08 PM
To: George Tilley
Subject: RE: ER/PR testing

George

How are things going given the issues?

Phil

From: George Tilley [mailto:George.Tilley@hccsj.nl.ca]
Sent: Thursday, October 20, 2005 12:47 PM
To: Hassen, Philip
Cc: robert.williams@easternhealth.ca
Subject: FW: ER/PR testing

Hi Phil,

Sorry I wasn't able to connect with you earlier. I didn't connect with Bob Bell

until late last evening. He offered some advice put there appears to be a gap in terms of anyone who owns this issue. Everyone seems to be concurring that this is a big problem, but who should take the lead is unclear to say the least. CBC TV on The National did a story on this last night and they interviewed a Phd from British Columbia which suggested a need to follow up on this through the country. I am curious whether this story got picked up where you are.

In the meantime, the only plan we have at this time is to write the Canadian Association of Pathologists. Other suggestions would be welcomed. I have attached my email to Bob Bell.

George

George Tilley

President/Chief Executive Officer

Eastern Health

c/o Corporate Office

Waterford Bridge Rd.

A1E 4J8

Tel: 709-777-1330

Fax: 709-777-1302

From: George Tilley
Sent: Thursday, October 20, 2005 4:08 PM
To: Robert Bell (robert.bell@uhn.on.ca)
Cc: Bob Williams (robert.williams@easternhealth.ca)
Subject: ER/PR testing

Hi Bob,

Just wanted to thank you again for calling me back last evening. I appreciated the opportunity to talk through this difficult issue with you and get your advice.

There appears to be growing body of evidence and opinion that there are many questions inherent with this ER/PR test. Even for me as a non clinician, it raises eye brows.

As a member of the Board of the Canadian Patient Safety Institute I want to ensure that, if there are learning's for others from our experience, we will share it. Having said that, I understand that Ontario already has a comprehensive accreditation program in place for its labs.

Our Laboratory Clinical Chief will be contacting the two individuals you referenced to also see what insight they can offer in terms of national follow. It appears that there is a gap in terms of a national entity who can take the lead with this issue, so we will likely have to take a shot gun approach to the follow up and hope that there is someone who can keep it moving.

In the meantime we will write the Canadian Association of Pathologists and others that we may subsequently identify. I will also seek Phil's help at the CPSI.

George

George Tilley

President/Chief Executive Officer

Eastern Health

c/o Corporate Office

Waterford Bridge Rd.

A1E 4J8

Tel: 709-777-1330

Fax: 709-777-1302