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

Anthony Rhodes, PhD,[1] Bharat Jasani, PhD,[2] Andre J. Balaton, MD,[3] Diana M. Barnes, DSc,[4] Elizabeth Anderson, PhD,[5] Lynda G. Bobrow, FRCPath,[6] and Keith D. Miller, FIBMS[1]

Key Words: Immunohistochemistry; Estrogen receptors; Progesterone receptors; Technical validation; Microwave antigen retrieval

Abstract

Immunohistochemical assays for estrogen receptors (ERs) and progesterone receptors (PRs) have not been surveyed for technical validity. In the present study, the reliability of the immunohistochemical assay for ER and PR was evaluated using data from 105 laboratories participating in external quality assessment (EQA) during a 2-year period. Technical variables associated with reliable immunostaining were analyzed. The efficiency of the antigen retrieval step was identified as the single most important contributory factor influencing the overall reproducibility of the assays. Reliable assays were found in 24 (36%) of 66 laboratories participating in continual EQA, including the majority of centers known to have clinically validated results. Inadequate assay sensitivity, with subsequent weak staining, was the main cause of poor and variable results by laboratories using microwave antigen retrieval; too short a heating time was identified as the principal contributory factor. Extension of the heating time resulted in significant improvement regardless of all other variables in the immunohistochemical protocol. Continual participation in EQA is an effective means for identifying and ameliorating variables that influence the reliability of immunohistochemical assays for predictive markers, thereby assisting in technical validation and standardization.

The immunohistochemical assays performed for estrogen receptors (ERs) and, to a lesser extent, progesterone receptors (PRs) are unique in their role in the management of breast cancer, and there is worldwide acceptance that assays for ERs and PRs provide valuable information to aid in the selection of patients for endocrine treatment.[1-3] More recently, other assays, such as those that detect HER-2/neu overexpression, are beginning to assume clinical relevance in terms of their predictive value for immunotherapeutic management of invasive breast carcinoma.[4-11]

The US Food and Drug Administration (FDA) classifies the immunohistochemical assays for ER, PR, and HER-2/neu as class II devices; the criteria for this classification include the generation of a stand-alone report sent directly to the clinician.[12] Because of the predictive importance of these stand-alone reports, it is paramount to ensure that the reliability (specificity, sensitivity, and reproducibility) of the tests on which they are based is adequately high to minimize false-positive or false-negative results. Antibodies to ER, PR, and HER-2/neu are the first generation of immunohistochemical markers for which results are evaluated semiquantitatively for reporting purposes, and any error is likely to be further compounded by the use of differing methods of semiquantitative analysis and threshold values to interpret positive and negative results.[13] In this context, the level of risk for an immunohistochemical assay, as stipulated by Robinowitz,[14] is the probability of harm incurred by the patient from a given error in diagnosis. The risk levels for the results obtained by immunohistochemical assays for ER, PR, and HER-2/neu are considered generally higher than those suspected for markers that are mere adjuncts to histopathologic interpretation (FDA, class I devices) and that do not indicate a specific line of treatment to the oncologist or clinician treating the patient.[12]

The recent publicity about potential benefits of trastuzumab (Herceptin) therapy for patients with breast carcinomas that overexpress the HER-2/*neu* gene has further highlighted the importance of using accurate and validated predictive or prognostic markers, especially when they are evaluated semiquantitatively.[8,10] Four of the principal criteria required for technical validation of such assays are that they should be specific, sensitive, reproducible, and interpreted in a uniform manner in different laboratories.[6,8,15] Until recently, few independent data have been available to assist in the technical validation of immunohistochemical assays for ER and PR, and it is perhaps too early to provide any such data on the reproducibility and reliability of the HER-2/neu assays.

Over the years, an extensive bank of quality assurance (QA) data has been established to provide technical validation of the biochemical ligand-binding assays (LBAs) performed for ERs and PRs, ranging from information on their reproducibility and standardization of the technique to variation in the distribution and frequency of receptor-positive breast carcinomas in different laboratories.[16-24] In practice, the principal use of these assays has been as predictive markers, clinically validated by studies spanning more than 20 years and involving large numbers of patients in randomized clinical trials. These studies have shown that the LBA is sufficiently sensitive, specific, and reproducible to reliably identify subsets of patients with significantly different risk factors associated with recurrence, survival, or treatment response.[6,25,26] Consequently, there can be little doubt of the technical and clinical validity of the LBA to determine the ER and PR status of women with breast cancer.

The same, however, cannot be said for the immunohistochemical assays for ER and PR, despite the fact that they have largely replaced the biochemical LBA in many parts of the world. The immunohistochemical assays in general do not meet the guidelines for validation as prognostic or predictive markers, mainly because of their lack of technical validation.[6,15] To assist in the validation process, 3 recent studies evaluated the sensitivity of the assays for ER and PR using different scoring systems and established the frequency of receptor-positive breast carcinomas occurring in large numbers of laboratories.[13,27,28] In addition, data from one of these studies showed that variation in fixation and tissue preparatory methods as used by more than 150 laboratories were not significant contributory factors preventing optimal demonstration of ERs and PRs on routine material.[27] This suggests that the interlaboratory differences in assay sensitivity found in these studies may be caused by inefficiencies of the immunologic reagents, the immunostaining methods, or both. However, present data are limited on the technical aspects of the assay associated with reproducible results. While the monoclonal antibodies used in the tests may undergo stringent QA, in both production and marketing, there is little standardization in the methods used for conducting the assay or the way in which the results are evaluated.[6] On the other hand, it is unreasonable to expect laboratories to change current working practices in the absence of scientific evidence to support a proposed guideline or standardized method.

The current recommendations are that laboratories should perform rigorous clinical validation of the immunohistochemical assays themselves or should follow procedures of laboratories known to have performed such validation.[15] The first option is not available to laboratories beginning use of the assay. The second option is available by participation in an established external quality assessment (EQA) program,[27] which allows a laboratory to gauge its assay sensitivity against that achieved by a large number of other laboratories, some of which will have clinically validated their results. Data accumulated in the process of continual multilaboratory participation in EQA hold the potential to provide important information about not only the interlaboratory sensitivity and specificity of these tests but also the reproducibility of results. Analysis of these results may in turn help identify variables associated with reproducible assays and assist in their technical validation and standardization. If, in the future, the clinical reliability of assays to markers such as ER and HER-2/neu is to be assured, an established mechanism by which to technically validate them is imperative.

The aims of the present study were to use EQA data accumulated during a 2-year period to evaluate the degree of reproducibility of the assays for ERs and PRs among a large number of participating laboratories and to identify variables that are most closely associated with the maximum reliability, as defined by consistently high scores in national EQA. In addition, we sought to identify the main contributory factor responsible for suboptimal results and to study whether adjustment of this factor could bring about substantial interlaboratory improvement in the demonstration of ERs by immunohistochemical assay.

Materials and Methods

Determination of Interlaboratory Reproducibility

Laboratory Participation

Between August 1996 and September 1998, 8 United Kingdom National External Quality Assessment Scheme for Immunocytochemistry (UK NEQAS-ICC) ER and PR assessment runs took place. At 4 of these assessments, participating laboratories were requested to demonstrate ERs

alone; at 1, ERs and PRs; and at 3, PRs alone. Laboratories not stocking an antibody to PRs demonstrated ERs instead at all 8 assessment runs **Table 1**. The number of laboratories participating ranged from 105 for the first assessment run to 212 for the last assessment run of the series. For the purposes of establishing the reproducibility of the immunohistochemical assay for ER and PR, only the data from laboratories participating in all 8 assessment runs were included for analysis.

Assessment Procedure

At each assessment run, participating laboratories were sent 2 unstained tissue sections from composite tumor blocks fixed for 24 hours in 10% neutral buffered formalin and paraffin processed, typically comprising 3 infiltrating ductal carcinomas (IDCs) with differing levels of ER and PR expression (Table 1). The tumors had been analyzed previously by immunohistochemistry and biochemically by the LBA. Analysis by immunohistochemistry included staining by the organizing laboratory of every 25th section for ER or ER and PR to ensure that all slides contained similar proportions of invasive tumor. Participants were asked to demonstrate ER and/or PR and to return the best stained slide(s) with a completed questionnaire detailing the main technical variables used when staining the slides. At each of the assessment runs, the quality of immunohistochemical results achieved by each participant on its own in-house tumors also was evaluated. For this, each laboratory was requested to submit 2 unstained tissue sections from the same tissue block as the in-house tumor(s). On receipt, these were stained by the UK NEQAS-ICC organizing laboratory for the same receptor, coded, and filed alongside the in-house slides stained by the participant before assessment.

Assessment of Slides

Assessment of slides was achieved as described previously.[27] Briefly, an expert panel of 4 comprising a combination of consultant pathologists and biomedical and clinical scientists assessed the quality of the immunohistochemical result independently on a single-blind basis, with each assessor awarding marks out of a possible total of 5 for each of the coded slides. The 4 individual marks awarded then were added to give a total mark out of a possible total of 20. Marks were awarded by comparison of the proportion and intensity of tumor nuclei staining in the participant's slide with that achieved on replicate sections of the same cases by the UK NEQAS-ICC organizing center. A total mark of 13 or more of 20 indicated acceptable immunostaining quality

Table 1
Breast Tumors Circulated by UK NEQAS-ICC for Assessments Between August 1996 and September 1998

Run No.	Date	No. of Laboratories	Tumor Type	Receptor	Level of Expression: Immunohistochemical*	Level of Expression: Cytosol†
34‡	August 1996	105	IDC	Estrogen	90%-100%, +++	248
			ILC		75%-80%, ++	29
			IDC		0%	0
36	January 1997	118	IDC	Progesterone/estrogen§	90%, +++	ER, 66; PR, 307
38‡	April 1997	175	IDC	Estrogen	90-100%, +++	122
			IDC		90%, ++	41
			IDC		0%	7
39‡	August 1997	176	IDC	Progesterone/estrogen§	99%-100%, +++	ER, 309; PR, 1,582
			ILC		75%, ++	ER, 11; PR, 21
			IDC		1%, +	ER, 7; PR, 6
40‡	December 1997	192	IDC	Estrogen	99%, +++	15
			IDC		99%, +++	11
			IDC		90%, ++	12
41	April 1998	178	IDC	Estrogen and progesterone‖	ER, 75%-80%, ++; PR, 0%	ER, 10; PR, 2
42‡	June 1998	205	IDC	Estrogen	100%, +++	70
			IDC		90%, ++	41
			IDC		15%, +	65
43‡	September 1998	212	IDC	Progesterone/estrogen§	90%, +++	ER, 13; PR, 66
			IDC		75%, ++	ER, 7; PR, 8
			IDC		1%, +	ER, 11; PR, 12

ER, estrogen receptor; IDC, infiltrating ductal carcinoma; ILC, infiltrating lobular carcinoma; UK NEQAS-ICC, United Kingdom National External Quality Assessment Scheme for Immunocytochemistry.
* ER/PR expression of tumors is described in terms of proportion of invasive nuclei staining (%) and average staining intensity (+, ++, +++).
† Values are expressed in fmol/mg protein.
‡ Composite blocks containing more than one tumor were used.
§ Participants not stocking an antibody to PR were asked to demonstrate ER instead. On these runs, tumors were chosen that showed similar expression for ER and PR, as determined by immunohistochemical testing at the organizing laboratory.
‖ On this assessment, participants were asked to demonstrate both ER and PR.

and a pass at assessment; a mark of 10 to 12 of 20 was considered suboptimal and borderline, while a total mark of 9 or less of 20 was considered unacceptable quality and a failure at assessment. One of the main criteria for unacceptable staining quality was the failure to detect 10% or more of receptor-positive tumor nuclei in a tumor that has been shown by the UK NEQAS-ICC organizing center to clearly express more than 10% of ER- or PR-positive nuclei. To ensure assessor concordance, on the marking of the 20th slide, all scores were read out in turn, and when there was a difference more than 1 point between any of the assessors' individual marks, the respective slide was reviewed until consensus was reached between the conflicting assessors.

Technical Variables of Laboratories With Reproducible Staining vs Those for All Other Participating Laboratories

The main technical variables—antigen retrieval method, antibody clone and supplier, type of detection system and supplier, and use of automated immunostainers—as stated in returned questionnaires were recorded for all laboratories participating during the 2-year period. The percentage of use of a particular antibody clone, antigen retrieval method, detection system, or automation was established at each of the 8 runs for laboratories shown to have reproducible staining. This was compared with the percentage of use by all other laboratories at each of the runs.

Statistics

The chi-square goodness-of-fit test was used to determine whether the observed proportion of participants failing to achieve reproducible immunohistochemical assays for ER and PR was significantly higher than would be expected by chance. The mean proportion of laboratories achieving reproducible staining using a particular technical variable was compared with the mean proportion of all other laboratories using the same variable during the 2-year period. The Mann-Whitney *U* rank sum test was used to test for differences between the 2 groups for each of the main technical variables studied.

Contribution of Microwave Antigen Retrieval Heating Time to Interlaboratory Reproducibility of the Immunohistochemical ER Assay

Tumors Circulated at Assessment

Three IDCs fixed in 10% neutral buffered formalin and paraffin processed and with relatively high, zero, and low ER expression **Image 1** were dissected and made into a total of 12 composite tissue blocks. Each of the 12 blocks comprised a representative piece of tissue from each tumor and some benign glandular breast tissue for the internal positive control. Sections from these tissue blocks were circulated to each participating laboratory for 2 consecutive assessment runs (officially UK NEQAS-ICC runs 46 and 47, which for simplicity have been termed the first and second assessment runs).

Participant Instructions for Low Assay Sensitivity During the First Assessment

Participants identified as having weak staining for ER at the first assessment run and who used the main form of antigen retrieval (microwave oven) were advised at the second run, in addition to using exactly the same protocol as previously on 1 slide, to perform extended antigen retrieval times on additional slides supplied by the UK NEQAS-ICC. Additional times of 5, 10, and 15 minutes were advised for an original heating time of less than 15 minutes and 5 and 10 minutes if the original time was 15 minutes or more. From the first run, the main primary antibody conditions (dilution, incubation time, and temperature) for laboratories achieving acceptable staining were identified as being 1:50 for 60 minutes at room temperature for clone 1D5 (Dako A/S, Glostrup, Denmark) and 1:40 for 60 minutes at room temperature for clone 6F11 (Novocastra Laboratories, Newcastle-upon-Tyne, England). Participants in the study group whose antibody conditions were different from these were asked to use these values at each of the antigen retrieval times for the second run, in addition to their usual set of antibody conditions. Participants were requested to label their stained slides with their participant code number and a letter code (a-h), according to the antigen retrieval and antibody changes instigated. To prevent undue bias in the assessment of the additional slides, each was given a temporary participation code number. Therefore, the assessors were unable to distinguish between slides submitted as part of the regular assessment run and those submitted after implementation of the recommended changes. All but 1 member of the assessment panel evaluated the slides at both the first and second runs of the study. All were experienced and had been involved in UK NEQAS-ICC assessments of ER and PR immunohistochemical assays for a number of years.

Statistics

Comparison of the median scores and the interquartile ranges (IQRs) of the scores of participants in the microwave antigen retrieval study group at the first and second assessment runs was achieved by the use of box plots. The 2-sample Wilcoxon matched-pair signed rank test was used to compare the distribution of scores at the first and second assessment runs. The independent-samples *t* test was used to compare the original microwave antigen retrieval heating times used by the study group with the heating times giving the best result for the same study group at the second assessment run. The Pearson correlation coefficient (*r*) was used to

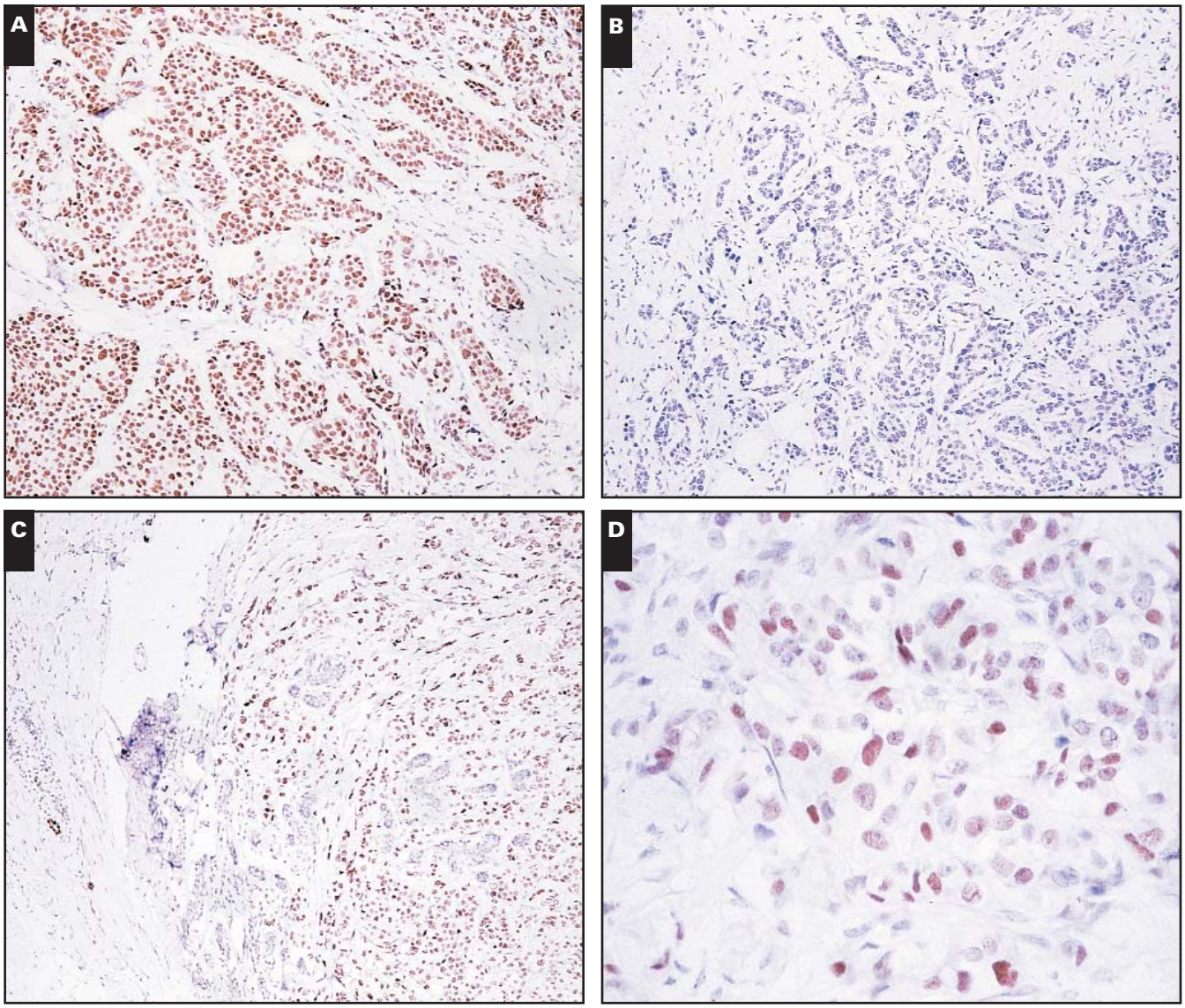

Image 1 Optimal demonstration by the organizing laboratory of the 3 tumors used in the microwave antigen retrieval study (clone 1D5 [Dako, Ely, England], pressure-cooker pretreatment, and *Streptomyces* avidin biotin complex detection). **A**, The relatively high estrogen receptor (ER)-expressing infiltrating ductal carcinoma (IDC) (×100). **B**, The ER-negative IDC (×100). **C**, The low ER-expressing IDC (×100). **D**, The low ER-expressing IDC (×400).

test for association between the antigen retrieval time giving the best result for the study group and the power output (measured in watts) of the microwave ovens used by these laboratories. Last, the chi-square goodness-of-fit test was used to compare the proportion of participants achieving acceptable staining for ERs at the 2 assessment runs.

Results

Determination of Interlaboratory Reproducibility

One hundred and five laboratories participated in the first of the 8 assessment runs used to determine assay reproducibility (Table 1), and 66 participated in all subsequent runs, of which 61 (92%) of 66 were based in Europe. They consisted of laboratories from major cancer centers or university hospitals (n = 29), general hospitals (n = 30), private clinics or laboratories (n = 6), and a biotechnology company (n = 1). The median number of assessments for which this group of laboratories achieved acceptable staining was 6, with 11 laboratories achieving acceptable staining for all 8 assessments and 13 achieving acceptable staining for all except 1 run. Of 5 laboratories that participated in all 8 runs and that are known, through published clinical studies (expert laboratories), to have validated their immunohistochemical assays, 1 achieved acceptable staining at 8 of 8

assessments, 3 at 7 of 8 assessments, and 1 at 6 of 8 assessments. If acceptable staining at all or all but 1 assessment run is used to define reproducible acceptable standards, then only 24 (36%) of 66 participating laboratories achieved reproducible results, compared with the 33 (50%) of 66 expected on the basis of the null hypothesis (chi square = 4.909; P = .027).

Specificity of the Immunohistochemical Assay for ERs and PRs

No tumor that was defined and circulated as ER negative was judged to be ER positive by immunohistochemical assay at any of the assessments in the ensuing period. The same applied for the immunohistochemical assessment for PR, except for 2 assessments, (runs 39 and 43) when 1 (0.6%) of 176 and 1 (0.5%) of 212 participating laboratories, respectively, stained more than 10% of invasive nuclei in a tumor that, according to the immunohistochemical assays of the organizing laboratory and expert laboratories, had a PR-negative status. This discrepancy did not relate to nonspecific staining of tumors, which typically is seen in the results of approximately 4% of laboratories at assessment,[13] which was readily distinguishable from true nuclear staining because it was accompanied by varying degrees of excessive staining of cytoplasm and connective tissue and the staining of nuclei of nontumor cells, such as lymphocytes and fibroblasts.

Technical Variables Associated With Reproducible Assays

Laboratories identified as having reproducible immunohistochemical assays used the same main technical variables as all other participating laboratories during the 2-year period. However, the percentage of use differed to varying degrees **Table 2**. Of particular relevance was the finding that a mean of 54% (95% confidence interval [CI], 50%-58%) of laboratories with reproducible assays used the pressure cooker, with only a mean of 34% (95% CI, 28%-40%) using the microwave oven. In comparison, the percentage of use by other laboratories during this period was almost reversed, with a mean of only 26% (95% CI, 23%-30%) using the pressure cooker and a mean of 60% (95% CI, 54%-66%) using the microwave oven (Table 2).

Table 2
Main Technical Variables for Immunohistochemical Receptor Assays*

	Proportional Use by Laboratories (%)					
	With Reproducible Assays		Others			
Variable	**Mean**	**95% CI**	**Mean**	**95% CI**	**Mann-Whitney *U***	***P* (2-Tailed)**
Antigen retrieval						
Sodium citrate buffer†	81	73-91	84	80-88	30.000	.833
Microwave oven	34	28-40	60	54-66	0.000	.001
Pressure cooker	54	50-58	26	23-30	0.000	.001
Main primary antibodies						
Estrogen receptor						
Clone 1D5, Dako, Glostrup, Denmark	81	78-83	67	62-73	0.000	.008
Clone 6F11, Novocastra, Newcastle-upon-Tyne, England	17	14-20	19	12-26	7.000	.242
Progesterone receptor						
Clone 1A6, Novocastra	64	60-68	46	40-52	0.000	.021
Clone 1A6, other suppliers‡	12	0-26	20	6-34	6.000	.564
Detection systems						
Avidin-biotin systems						
Dako	59	54-66	53	48-58	15.000	.072
Vector	32	26-36	12	10-14	0.000	.001
Other detection systems§	9	6-12	35	30-40	0.000	.001
Automation						
None	65	62-69	63	59-68	20.500	.221
Full‖	18	14-24	18	14-22	28.000	.672
Partial¶	14	10-20	16	14-18	31.000	.913

CI, confidence interval.
* Comparison between laboratories with sensitive and reproducible assays and all other laboratories participating during the same 2-year period.
† Sodium citrate buffer, pH 6.0-6.2; 0.01- to 0.2-mol/L concentration.
‡ Use of other progesterone receptor clones ranged from only 4% (CI, 0%-8%) to 12% (CI, 4%-18%).
§ Includes avidin-biotin systems from various suppliers.
‖ Includes Optimax, Biogenex (San Ramon, CA); Horizon, Dako; Techmate 500, Dako; Immunostainer, Lab Vision (Newmarket, England); Cadenza, Life Sciences International (Basingstoke, England); Ventana ES, Ventana Medical Systems (Strasbourg, France); and NEXES, Ventana.
¶ Sequenza, Life Sciences.

Microwave Antigen Retrieval Heating Time Limits Interlaboratory Reproducibility of the Immunohistochemical ER Assay

First and Second Runs of the Microwave Heating Time Study

A total of 226 laboratories participated in the first and second assessment runs, of which 139 (61.5%) and 137 (60.6%) of 226 achieved acceptable staining, respectively. Of the laboratories achieving an unacceptable result (score, 12 or less of 20) at the first run, 81 (86%) of 94 received a low score because of weak staining, ie, the proportion of tumor nuclei stained was considerably less than that stained by the UK NEQAS-ICC organizing laboratory (Image 1). Owing to the relatively low proportion of ER-positive nuclei in the low ER-positive tumor, suboptimal (weak) staining inevitably resulted in fewer than 10% of the nuclei being stained. The remaining 13 laboratories (14%) failed at assessment owing to excessive background or cytoplasmic staining.

Of the 81 laboratories failing at the first assessment owing to weak staining, 47 (58%) used a microwave oven, 20 (25%) a stand-alone steel pressure cooker heated by a hot plate, 10 (12%) a plastic pressure cooker in a microwave oven, and 4 (5%), other forms of antigen retrieval, eg, autoclave or overnight incubation in an oven at relatively low temperatures.

A total of 29 (62%) of 47 laboratories identified as failing at the first assessment owing to weak staining and using a microwave oven for antigen retrieval voluntarily stained and submitted additional slides at the second assessment following the directions suggested. An additional 5 laboratories, while submitting additional slides, were excluded from further analysis because of excessive nonspecific staining on all slides (n = 1), almost total loss of tissue from all slides (n = 1), and implementation of changes other than those recommended (n = 3).

Antigen Retrieval Buffers, Antibodies, and Detection Systems

The main technical variables used by laboratories in the study group and all other laboratories are shown in **Table 3**.

Comparison of Scores and Quality of Staining for Microwave Antigen Retrieval at Different Heating Times

The median score achieved by participants in the study group at the first assessment run was 9 with an IQR of 11 to 8 **Figure 1**. When the same set of participants implemented the same method at the second assessment run, the median score increased to 12 (IQR, 14-7). The Wilcoxon signed rank test, however, showed that this improvement was not statistically significant **Table 4**. These scores were used as the baseline set of scores representing the participants' choice of technical variables as used at the first and second assessment runs.

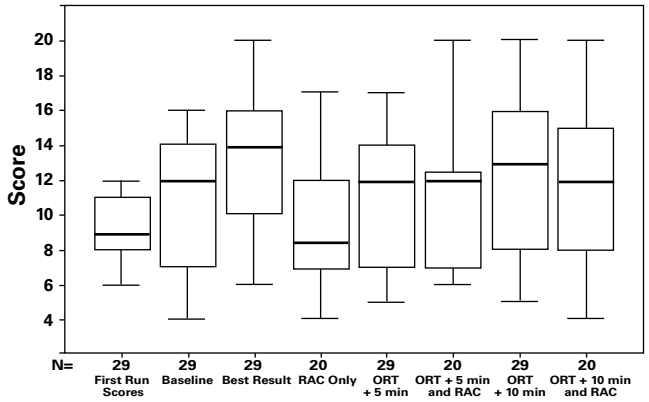

Figure 1 Comparison of the scores given to 29 laboratories for the quality of estrogen receptor demonstration, using different microwave antigen retrieval heating times on duplicate tissue sections containing 3 infiltrating ductal carcinomas. Table 4 compares these scores using the Wilcoxon signed rank test. Baseline scores are the second run scores using same conditions as the first run. The best result is the best following United Kingdom National External Quality Assessment Scheme for Immunocytochemistry recommendations. ORT, original microwave antigen retrieval time; RAC, change to recommended primary antibody conditions. Note: Owing to the small sample, data are not presented for laboratories returning slides stained with ORT + 15 minutes.

When the extended antigen retrieval time, the recommended antibody conditions giving the best results, or both were implemented, the median score of the study group increased to 14 (IQR, 16-10; Figure 1). These scores were significantly higher than those of the baseline slides (Table 4). The only single change that resulted in significantly higher scores compared with the control slides was the addition of 10 more minutes to the in-house antigen retrieval time. Recommended changes in antibody dilution, time of incubation, or both, on their own or when accompanying increased antigen retrieval times, did not result in scores significantly different from those for the control slides. **Image 2** shows the results for 1 participant in the microwave study group. The staining for ER by this laboratory on the first and second runs achieved scores of 10 of 20 on both occasions using the participant's chosen technique and a microwave heating time of 15 minutes. These low scores were given because the immunostaining failed to adequately demonstrate sufficient invasive nuclei in the low ER-expressing tumor. However, an extension of the heating time by 10 minutes (total time, 25 minutes) resulted in adequate staining of the low ER-expressing tumor and a

Table 3
Main Technical Variables Used at the 2 Assessment Runs Involving the Microwave Heating Time Study*

Technical Step	All Laboratories (n = 236)	Study Group (n = 29)
Antigen retrieval system		
Microwave oven	106 (44.9)	29 (100)
Pressure cooker	90 (38.1)	—
Pressure cooker in a microwave	30 (12.7)	—
Other	10 (4.2)	—
Antigen retrieval buffer		
Sodium citrate, pH 6.0-6.2; 0.01- 0.2-mol/L concentration†	125 (52.9)	23 (79)
Sodium citrate, pH 7.0; various molarities	42 (17.8)	—
Sodium citrate, pH 8.0; various molarities	3 (1.3)	—
EDTA, pH 8.0; 0.01-mol/L concentration	11 (4.7)	—
EDTA, pH 7.0; various molarities	4 (1.7)	—
EDTA, pH 8.0; no molarity given	1 (0.4)	—
EDTA, pH 7.0; no molarity given	—	1 (3)
Vector unmasking solution†	15 (6.4)	1 (3)
Sodium citrate/EDTA, pH 6.0-6.2†	2 (0.8)	1 (3)
Dako Target Retrieval, pH 9.9†	5 (2.1)	1 (3)
Tris-HCl, pH 9.6; no molarity given	1 (0.4)	—
Antigen retrieval buffer, pH 6.0; no molarity given†	—	1 (3)
Incomplete data	27 (11.4)	1 (3)
Primary antibody		
Dako clone 1D5	152 (64.4)	19 (66)
Novocastra clone 6F11	61 (25.8)	7 (24)
Other suppliers of clones 1D5 and 6F11	23 (9.7)	3 (10)
Detection system		
Avidin-biotin based systems (various suppliers)	197 (83.5)	24 (83)
Dako Envision	15 (6.4)	3 (10)
Incomplete data	24 (10.2)	2 (7)
Automation		
None	125 (53.0)	15 (52)
Life Sciences Sequenza	45 (19.1)	5 (17)
Biogenex Optimax	15 (6.4)	2 (7)
Dako Techmate 500	14 (5.9)	5 (17)
Lab Vision immunostainer	7 (3.0)	1 (3)
Ventana NEXES	9 (3.8)	—
Dako Techmate Horizon	8 (3.4)	—
Ventana ES	5 (2.1)	—
Life Sciences Cadenza	5 (2.1)	1 (3)
Incomplete data	3 (1.3)	—

* Data are given as number (percentage). For locations of manufacturers, see Table 2.
† Substantial improvement (a score of 13 or more out of 20) at the second run of the study was seen by participants using extended antigen retrieval times and these buffers.

score of 15 (Image 2). In addition, the nuclei of normal glands, the relatively high ER-expressing tumor included in the same composite block for these runs, and the participant's own in-house tumor all showed greater clarity of staining when using the extended heating time of 25 minutes. It is important to note that the other tumor included in the composite block, which had been shown by the organizing laboratory and the immunohistochemical assays of 236 laboratories to be completely negative for ER at both runs (Image 1), remained ER negative and free of background staining when using all the extended antigen retrieval times.

Comparison of Microwave Antigen Retrieval Times

The mean antigen retrieval time of the original methods used by laboratories in the study group was 18 minutes (95% CI, 16-19 minutes). This time is significantly lower than the mean antigen retrieval time of 26 minutes (95% CI, 24-28 minutes), giving the best result at the second assessment run for the study group ($t = –6.597$; $P < .001$).

Correlation of Microwave Oven Power Output With Optimal Antigen Retrieval Time

There was no significant correlation between wattage of the microwave ovens used by the study group (mean, 800 W; 95% CI, 760-840 W) and the antigen retrieval times used by the same laboratories to obtain the best result at the second assessment run (Pearson $r = –0.134$; $P = .504$).

Influence of Type of Buffer, pH, and Buffer Molarity on Microwave Antigen Retrieval

Substantial improvement (a score of 13 or more of 20) for ER staining was seen at the second assessment after extended antigen retrieval times with many of the buffer types of varying pH and molarity used by the study group (Table 3).

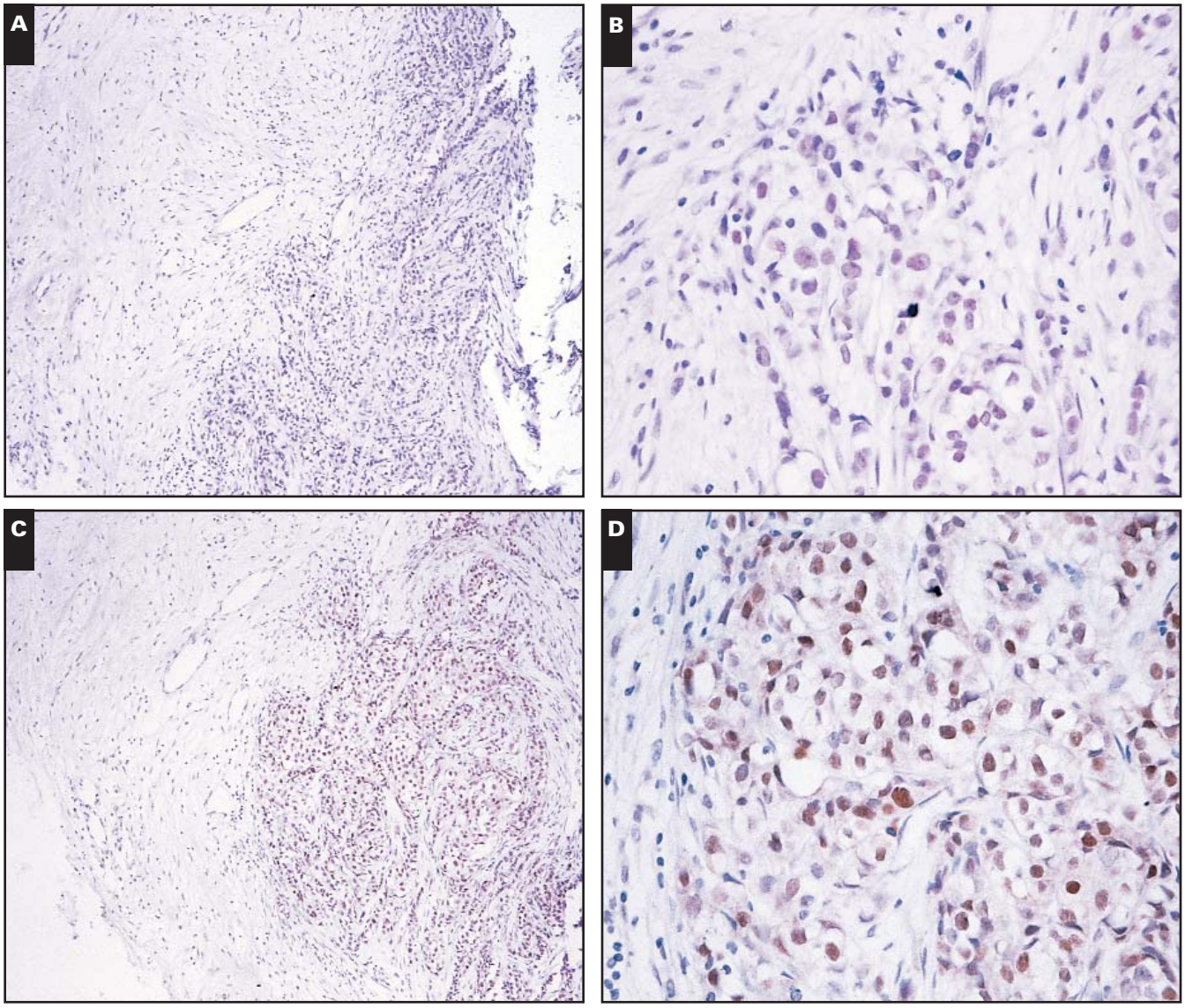

Image 2 Demonstration of the low estrogen receptor (ER) expressing infiltrating ductal carcinoma (IDC) used in the microwave antigen retrieval study by 1 laboratory in the second run using 2 heating times, the 1D5 clone (Dako, Glostrup, Denmark), and *Streptomyces* avidin biotin complex detection. **A**, Heating time, 15 minutes (×100). **B**, Heating time, 15 minutes (×400). **C**, Heating time, 25 minutes (×100). **D**, Heating time, 25 minutes (×400).

Table 4
Comparison of Scores in Figure 1 Using the Wilcoxon Signed Rank Test

Scores	Wilcoxon Z	*P* (2-Tailed)
First run scores vs baseline scores, which were the second run (same conditions) (n = 29)	–3.558*	.066
Best result following UK NEQAS-ICC recommendations vs baseline scores (n = 29)	–3.558†	<.001
Change to recommended primary antibody conditions only vs baseline scores (n = 20)	–1.30*	.193
Original microwave antigen retrieval time plus		
5 min vs baseline scores (n = 29)	–0.228†	.820
5 min with change to recommended primary antibody conditions vs baseline scores (n = 20)	–0.990†	.322
10 min vs baseline scores (n = 29)	–2.445†	.014
10 min with change to recommended primary antibody conditions vs baseline scores (n = 20)	–1.757†	.079

UK NEQAS-ICC, United Kingdom National External Quality Assessment Scheme for Immunocytochemistry.
* Based on positive ranks.
† Based on negative ranks.

Comparison of Study Group Scores for 2 Assessment Runs With Scores for Participating Laboratories Not in the Study Group

The proportion of 29 laboratories in the microwave antigen retrieval study group that achieved acceptable immunostaining (a score of 13 or more of 20) for ER at the first assessment run was zero. Despite this, the proportion achieving acceptable staining at the second assessment run increased to 19 (66%) of 29, of which 18 (62%) did so using the recommended extended heating times. In comparison, the proportion of all other laboratories achieving acceptable immunostaining decreased from 139 (72.4%) of 192 for the first assessment run to 116 (60.4%) of 192 at the second assessment run. Of these laboratories not in the study group and that used microwave antigen retrieval, 49 (80%) of 61 achieved acceptable staining at the first assessment, but only 31 (51%) of 61 did so at the second assessment run. Wilcoxon signed rank test results showed that the scores achieved by the microwave study group were significantly higher at the second assessment run than the first assessment run ($Z = –4.359$; $P = .001$), while there was no significant difference in the scores achieved by all other laboratories at the runs ($Z = –1.461$; $P = .144$).

According to the aforementioned findings, if extended antigen retrieval times had not been used at the second run, only 52.6% (137 – 18 = 119/226) of the total participating laboratories would have been expected to achieve acceptable staining. The observed number of laboratories achieving acceptable staining, ie, 137 (60.6%) of 226, was significantly higher (chi square = 5.267; $P = .022$).

Discussion

Validation of Assay Standards

To accurately assess the results achieved by different laboratories participating in EQA, it is essential to first validate the standards against which optimal sensitivity is defined. A number of institutions participating in the UK NEQAS-ICC program for hormonal receptors produced results that have been validated clinically. These laboratories have confirmed a positive relationship between positive receptor status determined by immunohistochemical assay and a favorable response to endocrine therapy, with the immunohistochemical assay shown to be as efficient or more efficient than biochemical testing.

It was established previously in an earlier study that these laboratories consistently achieved the expected result when participating in the UK NEQAS-ICC program for hormonal receptors. That is, at each assessment during a 4-year period, these laboratories consistently stained the expected proportion of ER- and PR-positive nuclei in low, medium, and high receptor-positive tumors and were awarded numeric scores of 13 or more of 20.[27] These same tumors, which had been tested initially using the biochemical LBA and the immunohistochemical assay, were used in the present study to establish the reproducibility of the assay in participant laboratories. Of these, 19 (95%) of 20 were similarly receptor positive or negative with either assay, using a threshold value of 10% or greater of invasive tumor nuclei stained by immunohistochemistry and 10 fmol/mg protein or greater with the LBA as designating receptor-positive status. While the use of any threshold value is arbitrary, a 10% threshold has been used in several studies that correlated immunohistochemical receptor assay results with clinical and biochemical values.[29-33] It also was the threshold value most commonly used routinely by laboratories participating in the UK NEQAS-ICC during the period studied.[13] Also of importance in the present study was the reproducibility of the UK NEQAS-ICC scoring system to evaluate the quality of immunohistochemical assays; this was ensured at assessment by checking assessor concordance after every 20 slides.

Reproducibility of Assays

The proportion of laboratories achieving reliable and reproducible staining was surprisingly low, although it includes all but one of the participating centers known to have clinically validated their results. Of laboratories that participated in all 8 assessments during the 2-year period, only 24 (36%) of 66 achieved acceptable staining at all or all but one of the runs on the UK NEQAS-ICC and their own in-house tumors. More than half of these 24 laboratories provided data on the frequency of ER-positive cases reported in their own laboratories, as described in an earlier study.[28] All are included in the group defined in that study as having high immunohistochemical assay sensitivity and a mean frequency of ER-positive breast carcinomas of 77% (95% CI, 75%-79%). Consequently, the ability of these laboratories to achieve reproducible staining of optimal sensitivity and specificity seems to be assured.

Technical Variables Associated With Reliable and Reproducible Assays

The results of the present study show that expert laboratories with reproducible assays by and large used the same main technical variables as used by all other participating laboratories during the 2-year evaluation period. However, the percentage of use of these differed to varying degrees for each of the main variables studied.

It seems highly unlikely that the batches of immunologic reagents used by the expert laboratories were different from those used by the general population of laboratories during the 2-year period, particularly because the expert

laboratories are based in several countries. Similarly, it is unlikely that even within the same expert laboratory the same batch of reagent was used throughout the 2-year period. Consequently, batch-to-batch variation in reagent quality may be eliminated as the underlying reason for unacceptable results at assessment. This is in keeping with the stringent QA instigated by the producers of the clones and detection systems used by the laboratories performing the immunohistochemical assay. Recent FDA regulations require proof of such stringent QA, particularly for markers such as ER and PR that, unlike the majority of antibodies used in routine pathology, now require premarket notification.[12,34] The greater use of the Dako 1D5 clone than the Novocastra 6F11 clone by laboratories achieving reproducible staining and other participating laboratories probably is due to a large extent to the fact that the 1D5 clone[35] has been on the market considerably longer than the 6F11 clone.[36]

When the technical variables used by the expert laboratories were studied, the most important finding was that a greater proportion used pressure cookers for antigen retrieval. In contrast, the main form of antigen retrieval used by the other participating laboratories was the microwave oven. However, this is not to say that the microwave oven or other forms of heating apparatus, used with the appropriate buffer, were not effective per se, since 3 of 4 laboratories shown to have reproducible staining in the present study and that had clinically validated results in previous studies used a microwave oven routinely. It is for historic reasons that the microwave is the most frequently used piece of equipment to effect heat-induced epitope retrieval. Much of the initial research by Shi et al[37] and Cattoretti et al[38] was performed using microwave ovens, and many other laboratories based their assay development on the work of these authors and the equipment they used. Understandably laboratories, particularly those with a high workload and, therefore, little time to do research and development, are likely to be reluctant to change unless it can be categorically shown that another form of equipment is superior. The greater use of the pressure cooker than the microwave oven in laboratories achieving reliable and reproducible staining suggested, however, that some feature of this form of antigen retrieval makes it particularly effective for ER and PR detection. The widely used clones to ER (1D5, 6F11) and PR (1A6) as used by participants in the present study, which reflect the use in most routine pathology departments, necessitate effective heat-induced epitope retrieval, as they will not work on routinely prepared tissues without it.[35,36]

Relationship Between Antigen Retrieval Time and Temperature

Probably the most salient difference between the pressure cooker and the majority of other heating appliances used for antigen retrieval, including the microwave oven, is the maximum temperature reached. Pressure cooking results in temperatures of 115°C or higher, ie, superheating, whereas the microwave oven does not raise the temperature of an aqueous buffer above 100°C, although this temperature is reached more quickly.[39] Recent studies have shown a consistent correlation between antigen retrieval time and the temperature required to achieve optimal immunohistochemical results, with longer antigen retrieval times being required for lower temperatures and vice versa.[39-44] As temperatures substantially higher than 100°C cannot be achieved with the standard microwave oven, extension of the heating time seems to be the only practical way of increasing the efficiency of antigen retrieval with this piece of equipment.

Extension of Microwave Heating Time Results in Significant Improvement

It seemed, therefore, reasonable to hypothesize that some of the poor staining for ERs and PRs seen in laboratories using the main form of antigen retrieval (ie, the microwave oven) may have been due to too short a heating time for the maximum temperature maintained during the antigen retrieval step. Hence, the second part of the present study investigated the effects on the sensitivity of the ER assays by extending microwave antigen retrieval heating times by laboratories initially achieving only low sensitivity when participating in national EQA. The null hypothesis tested was that these recommendations would have no significant effect on the distribution of scores. This null hypothesis, however, can be rejected because the scores after extension of the heating time (+5, +10, or +15 minutes) were significantly higher ($P < .001$) than those obtained by the same laboratories using the original heating times.

While staining of participants' own in-house tumors with differing antigen retrieval times was not a feature of the present study, 7 of 29 laboratories in the group performed this additional staining. Quite importantly, 5 of 7 of these laboratories also achieved better staining on their own in-house slides using the recommended extended antigen retrieval times.

Two important findings were that the technical specificity of the assays was not affected by extended antigen retrieval times (negative tumors remained negative regardless of the heating time) and that section detachment at the prolonged heating times occurred in only 1 laboratory, although the laboratory also experienced section loss with the original microwave heating time.

Antigen retrieval time was the main limiting factor resulting in weak staining in the study group. Change of the primary antibody conditions with or without extended antigen retrieval failed to produce significant improvement in scores compared with baseline. If anything, this tended to

result in lower scores than if the participants' own choice of antibody conditions was used. This finding emphasizes the importance of each laboratory identifying its own ideal conditions of dilution, incubation time, and temperature for primary antibody use and not assuming that those recommended by others will be optimal when used in a different laboratory.

It is noteworthy that since the microwave antigen retrieval study, 3 additional runs have taken place as part of the continual process of assessment. Of the laboratories that have participated in all these runs and that achieved improved results with extended microwave antigen retrieval heating times, 14 (88%) of 16 have continued to achieve acceptable results on the tumors circulated at assessment and their own in-house tumors.

A number of factors in the microwave antigen retrieval step, apart from insufficient temperature or heating time, potentially can influence antigen retrieval efficiency and, therefore, result in weak staining, eg, type of buffer and its pH and power output (wattage) of the microwave oven. Of these, buffer type and pH are the most frequently cited.[39, 42-47]

In the present study, microwave antigen retrieval heating time was studied irrespective of the buffer or pH. Although 79% of laboratories used a standard citrate buffer at a pH of 6.0 to 6.2, improvement was seen with increasing antigen retrieval time regardless of the type of buffer or its pH (Table 3). For example, the laboratory indicated in Table 3 as using the Dako Target Retrieval buffer at pH 9.9 initially used a heating time of 10 minutes, which resulted in a score of 11 of 20 at the first run and 14 of 20 at the second run. However, extension of the heating times to 15, 20, and 25 minutes resulted in higher scores of 17, 16, and 16, respectively, at the second run.

Shi et al[47] recommended that an antigen retrieval solution at high pH should be used for most of the antibodies used in surgical pathology. However, this recommendation is based on studies that used relatively short antigen retrieval times of 10 minutes in a microwave oven to determine the relationship between staining intensity and pH and type of buffer. The authors report a "V-form" response for ER and MIB1 if staining intensity is plotted against buffer pH, with high staining intensity occurring at low and high pH and relatively weaker staining at intermediate pH values.[47] The authors subsequently recommended the use of Tris-HCl buffer at pH 1 or 10 over the more commonly used sodium citrate buffer, which at pH 6.0 is intermediate in this pH range.[44,47] However, the microwave heating times giving the best results in the present study (mean, 26 minutes; 95% CI, 24-28 minutes), are approximately 2.5 times greater than the time used by Shi et al.[47] Similarly, in a study by Balaton et al[48] involving 14 French breast cancer centers using microwave antigen retrieval and citrate buffer, extension of the heating time to 25 minutes was required to allow all laboratories to achieve optimal staining for ERs on a series of breast carcinomas fixed under varying conditions. Previously, Rhodes et al[27] showed that the organizing laboratory was 90% to 100% efficient in optimally assaying tumors for ERs and PRs submitted for assessment during a 3-year period from up to 212 participating laboratories, with equivalent or greater sensitivity than that achieved by the laboratories from which they were submitted, using a standard antigen retrieval protocol and a stainless steel pressure cooker and citrate buffer at pH 6.0. All these findings raise the important question about whether a noticeable difference in staining intensity is seen for ERs and PRs when using citrate buffer at pH 6.0 and Tris-HCl buffer at pH 10 or 1, after the heating times or temperature have been optimized.

For laboratories currently obtaining weak staining for ERs or PRs at assessment and using a standard microwave oven, we recommend that these laboratories use pressure cooking antigen retrieval, as prevalently used by the expert centers, or extend the microwave heating time by increments of 5 minutes until optimal staining is achieved. This will be evident when the maximal numbers of invasive tumor nuclei are demonstrated reliably with the maximal staining intensity. Excessive microwave heating beyond this point will be apparent by the following: a decline in staining intensity and clarity, very poor nuclear morphologic features, or nonspecific nuclear staining of lymphocytes and fibroblasts. The use of a microwave oven with a relatively high power output, as used by laboratories in the present study (mean, 800 W; 95% CI, 760-840 W), is recommended. In addition, the reproducibility of the staining using the same heating time will be assisted by the following: (1) a rotating stage in the microwave oven to help to prevent hot and cold spots; and (2) keeping the microwave oven contents (buffer container and volume, number of slides) constant between runs, as this greatly influences the heating efficiency of the microwave oven. Other steps to include in an effective optimization program for the immunohistochemical demonstration of ERs and PRs are listed in **Table 5**.

Microwave antigen retrieval heating time seems to be the single most important factor contributing to the overall reliability of the immunohistochemical assays for ERs and PRs in routine practice. Multilaboratory participation in EQA is an effective means by which to identify and ameliorate variables influencing the reliability of assays for predictive markers and thereby assist in their technical validation and standardization.

From the [1]United Kingdom National External Quality Assessment Scheme for Immunocytochemistry and the Department of Histopathology, University College London Medical School,

Table 5
Main Factors Contributing to Optimization of Immunohistochemical Demonstration of Estrogen and Progesterone Receptors on Routinely Fixed and Paraffin-Processed Breast Carcinoma Specimens

- Delay in fixation results in destruction of receptor epitopes.[49,50] To guard against this destruction, excised tumors should be immediately sliced and placed in buffered formalin for 24 hours before tissue processing.
- Use of control sections from a composite block comprising tumors exhibiting zero, low, and high estrogen receptor and progesterone receptor expression, plus some benign breast tumor with normal ducts.
- Sections taken on slides coated with 3-aminopropyltriethoxysilane, or a similar adhesive, to prevent detachment during the heating stage. Sections should be dried overnight at 37°C and then for 2 hours at 58°C to ensure maximum antigen preservation.
- Optimization of antigen retrieval heating time using an antigen retrieval buffer and equipment of choice. Increase heating time until optimal staining is achieved, using increments of 5 minutes (microwave oven), 1 minute* (stainless steel pressure cooker heated by hot plate), or 2 minutes* (plastic pressure cooker in a microwave oven).
- Use of primary antibodies that are known to have undergone stringent quality assurance during production and that have been validated in clinical studies; optimization of dilution by testing the conditions recommended in the package insert and titrations either side of this dilution, eg, for a recommended dilution of 1:50, test at 1:30, 1:40, 1:50, 1:60, and 1:70.
- Use of an avidin-biotin based detection system† or one with equivalent sensitivity.
- Use of a light nuclear counterstain that provides good color contrast with the chromogen of choice.
- If introducing automation, close comparison of the sensitivity and specificity achieved with the manual method on the same cases before adoption of the automated method.
- Participation in a recognized external quality assessment program to ensure that maximal sensitivity and specificity are being achieved on in-house and externally prepared tumor specimens.

* Time at full pressure.
† If excessive cytoplasmic staining predominates, a biotin blocking step should be introduced.[51,52]

London, England; [2]Immunocytochemistry and Molecular Pathology Unit, University of Wales College of Medicine, Cardiff, Wales; [3]Centre de Pathologie, Bievres, France; [4]Hedley Atkins ICRF Breast Pathology Laboratory, Guy's Hospital, London, England; [5]Clinical Research Department, Christie Hospital NHS Trust, Manchester, England; [6]University of Cambridge Clinical School of Medicine, Addenbrookes Hospital, Cambridge, England.

Address reprint requests to Dr Rhodes: UK NEQAS-ICC Manager, Dept of Histopathology, University College London Medical School, Rockefeller Bldg, University St, London WC1E 6JJ, England.

Acknowledgments: We thank all the laboratories participating in UK NEQAS-ICC, without whom this study would have been impossible.

References

1. Nicholson RI, McClelland RA, Gee JMW. Steroid hormone receptors and their clinical significance in cancer. *J Clin Pathol*. 1995;48:890-895.
2. Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet*. 1998;351:1451-1467.
3. Wishart GC, Harnettt AN, Purushotham AD. Oestrogen receptor status: no longer an optional extra. *Breast*. 1998;7:154-155.
4. Press MF, Hung G, Godolphin W, et al. Sensitivity of HER-2/neu antibodies in archival tissue samples: potential source of error in immunohistochemical studies of oncogene expression. *Cancer Res*. 1994;54:2771-2777.
5. Seidman AD, Beselga J, Yao T-J, et al. HER2/*neu*-over-expression and clinical taxane sensitivity: a multivariate analysis in patients with metastatic breast cancer (MBC) [abstract]. *Proc Am Soc Clin Oncol*. 1996;15:104.
6. Allred DC, Harvey JM, Berado M, et al. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol*. 1998;11:155-168.
7. 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 cisplatin in patients with HER2/neu-overexpressing metastatic breast cancer refractory to chemotherapy treatment. *J Clin Oncol*. 1998;16:2659-2671.
8. Allred DC, Swanson PE. Testing for erbB-2 by immunohistochemistry in breast cancer [editorial]. *Am J Clin Pathol*. 2000;113:171-175.
9. Jacobs TW, Gown AM, Yaziji H, et al. HER-2/neu protein expression in breast cancer evaluated by immunohistochemistry: a study of interlaboratory agreement. *Am J Clin Pathol*. 2000;113:251-258.
10. Tubbs RR, Stoler MH. The quality of HER-2/*neu* predictive immunohistochemistry: something FISHy [editorial]? *Mod Pathol*. 2000;13:1-3.
11. Jimenez RE, Wallis T, Tabasczka P, et al. Determination of HER-2/*neu* status in breast carcinoma: comparative analysis of immunohistochemistry and fluorescent in situ hybridization. *Mod Pathol*. 2000;13:37-45.
12. Taylor CR. FDA issues final rule for classification and reclassification of immunochemistry reagents and kits [editorial]. *Am J Clin Pathol*. 1999;111:443-444.
13. Rhodes A, Jasani B, Barnes DM, et al. 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;53:125-130.
14. Robinowitz M. Revised comments on the final ruling for IHCS [editorial]. *Appl Immunohistochem*. 1998;6:116-117.
15. Harvey JM, Clark GM, Osbourne CK, et al. 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.

16. Breast Cancer Co-Operative Group. Standards for the assessment of hormone receptors in human breast cancer: report of a workshop on September 29, 1972, at the Antoni van Leeuwenhoek-Huis, Amsterdam. *Eur J Cancer.* 1973;9:379-381.
17. Breast Cancer Co-Operative Group. Revision of the standards for the assessment of hormone receptors in human breast cancer: workshop, held on 16-17 March, 1979, in the Netherlands Cancer Institute. *Eur J Cancer.* 1980;16:1513-1515.
18. Koenders A, Thorpe SM. Standardization of steroid receptor assays in human breast cancer, I: reproducibility of estradiol and progesterone receptor assays. *Eur J Cancer Clin Oncol.* 1983;19:1221-1229.
19. Thorpe SM, Koenders A. Standardization of steroid receptor assays in human breast cancer, III: selection of reference material for intra- and inter-laboratory quality control. *Eur J Cancer Clin Oncol.* 1986;22:939-944.
20. Koenders A, Thorpe SM. Standardization of steroid receptor assays in human breast cancer, IV: long-term within- and between-laboratory variation of estrogen and progesterone receptor assays. *Eur J Cancer Clin Oncol.* 1986;22:945-952.
21. Koenders A, Thorpe SM. Standardization of steroid receptor assays in human breast cancer, II: samples with low receptor content. *Eur J Cancer Clin Oncol.* 1983;19:1467-1472.
22. Thorpe SM. Immunological quantitation of nuclear receptors in human breast cancer: relation to cytosolic estrogen and progesterone receptors. *Cancer Res.* 1987;47:1830-1835.
23. Blankenstein MA, Benraad TJ. Multilaboratory assessment of tissue prognostic factors in breast cancer: the EORTC Receptor Study Group experience. *Can J Oncol.* 1992;3(suppl):58-64.
24. Romain S, Laine Bidron C, Martin PM, et al. Steroid receptor distribution in 47,892 breast cancers: a collaborative study of 7 European laboratories. *Eur J Cancer* 1995;31A:411-417.
25. Clark GM. Prognostic and predictive factors. In: Harris J, Lippman ME, Morrow M, et al, eds. *Diseases of the Breast.* Philadelphia, PA: Lippincott-Raven; 1996:461-485.
26. Osbourne CK. Receptors. In: Harris JR, Hellman S, Henderson IC, et al, eds. *Breast Diseases.* Philadelphia, PA: Lippincott; 1991:301-325.
27. Rhodes A, Jasani B, Balaton AJ, et al. 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.* 2000;53:292-301.
28. Rhodes A, Jasani B, Balaton AJ, et al. Frequency of oestrogen and progesterone receptor positivity by immunohistochemical analysis in 7,016 breast carcinomas: correlation with patient age, assay sensitivity, threshold value and mammographic screening. *J Clin Pathol.* 2000;53:688-696.
29. Pellicer EM, Sundblad A. Evaluation of antibodies to oestrogen receptors [letter]. *Appl Immunohistochem.* 1994;2:141.
30. De Mascarel I, Soubeyran G, MacGrogan J, et al. Immunohistochemical analysis of estrogen receptors in 938 breast carcinomas: concordance with biochemical assay and prognostic significance. *Appl Immunohistochem.* 1995;3:222-231.
31. 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;77:2514-2519.
32. 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.
33. 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;74:1120-1125.
34. Swanson PE. Labels, disclaimers, and rules (oh, my!): analyte-specific reagents and practice of immunohistochemistry. *Am J Clin Pathol.* 1999;111:445-448.
35. 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;55:651-654.
36. Bevitt DJ, Milton ID, Piggot N, et al. New monoclonal antibodies to oestrogen and progesterone receptors effective for paraffin section immunohistochemistry. *J Pathol.* 1997;183:228-232.
37. 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.
38. Cattoretti G, Becker MHG, Key G, et al. Monoclonal antibodies against recombinant parts of the Ki-67 antigen (MIB 1 and MIB 3) detect proliferating cells in microwave processed formalin fixed paraffin sections. *J Pathol.* 1992;168:357-363.
39. Suurmeijer AJH, Boon ME. Pretreatment in a high-pressure microwave processor for MIB1 immunostaining of cytological smears and paraffin tissue sections to visualize the various phases of the mitotic cycle. *J Histochem Cytochem.* 1999;47:1015-1020.
40. Lucassen PJ, Ravid R, Gonatas NK, et al. Activation of the human supraoptic and paraventricular nucleus neurons with aging and in Alzheimer's disease as judged from increasing size of the Golgi apparatus. *Brain Res.* 1993;632:105-113.
41. Kawai K, Serizawa A, Hamana T, et al. Heat-induced antigen retrieval of proliferating cell nuclear antigen and p53 protein in formalin-fixed, paraffin-embedded sections. *Pathol Int.* 1994;44:759-764.
42. Pileri SA, Roncador G, Ceccarelli C, et al. Antigen retrieval techniques in immunohistochemistry: comparison of different methods. *J Pathol.* 1997;183:116-123.
43. Shi SR, Cote RJ, Yang C, et al. Development of an optimal protocol for antigen retrieval: a "test battery" approach exemplified with reference to the staining of retinoblastoma protein (pRB) in formalin fixed paraffin sections. *J Pathol.* 1996;179:347-352.
44. Shi SR, Cote RJ, Chaiwun B, et al. Standardization of immunohistochemistry based antigen retrieval technique for routine formalin-fixed tissue sections. *Appl Immunohistochem.* 1998;6:89-96.
45. 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;174:301-307.
46. Morgan JM, Navabi H, Jasani B. Role of calcium chelation in high temperature antigen retrieval at different pH values. *J Pathol.* 1997;182:233-237.

47. Shi SR, Inman A, Young L, et al. Antigen retrieval immunohistochemistry under the influence of pH using monoclonal antibodies. *J Histochem Cytochem.* 1995;43:193-201.
48. Balaton AJ, Mathieu M-C, Le Doussal V. Optimization of heat induced epitope retrieval for estrogen receptor determination by immunohistochemistry on paraffin sections: results of a multicentric study. *Appl Immunohistochem.* 1996;4:259-263.
49. von Wasielewski R, Mengel M, Nolte M, et al. Influence of fixation, antibody clones, and signal amplification on steroid receptor analysis. *Breast J.* 1998;4:33-40.
50. Group for Evaluation of Prognostic Factors Using Immunohistochemistry in Breast Cancer (GEFPICS-FNCLCC). Recommandations pour l'évaluation immunohistochimique des récepteurs hormonaux sur coupes en paraffine dans les carcinomes mammaires: mise à jour 1999. *Ann Pathol.* 1999;19:336-343.
51. Wood GS, Warnke R. Suppression of endogenous avidin-binding activity in tissues and its relevance to biotin-avidin detection systems. *J Histochem Cytochem.* 1981;29:1196-1204.
52. Miller RT, Kubier P, Reynolds B, et al. Blocking of endogenous avidin-biotin activity in immunohistochemistry: the use of skimmed milk as an effective and economical substitute for commercial biotin solutions. *Appl Immunohistochem Molecul Morphol.* 1999;7:63-65.