Guideline on Hormone Receptor Testing in Breast Cancer

S. Nofech-Mozes, E. Vella, S. Dhesy-Third, and W. Hanna

A Quality Initiative of the Program in Evidence-Based Care (PEBC), Cancer Care Ontario (CCO)

Report Date: April 8, 2011

The full Evidence-based Series (EBS) 22-1 is comprised of 3 sections and is available on the CCO website (http://www.cancercare.on.ca)
PEBC Pathology & Laboratory Medicine page at:
https://www.cancercare.on.ca/toolbox/qualityguidelines/clin-program/pathlabebs/

Section 1: Guideline Recommendations
Section 2: Evidentiary Base
Section 3: EBS Development Methods and External Review Process

For further information about this series, please contact:

Dr. Wedad Hanna, Professor, Department of Laboratory Medicine and Pathobiology
Faculty of Medicine, University of Toronto
Fax: 416-480-4271   Email: Wedad.Hanna@sunnybrook.ca

For information about the PEBC and the most current version of all reports, please visit the CCO website at http://www.cancercare.on.ca/ or contact the PEBC office at:
Phone: 905-527-4322, ext. 42842     Fax: 905-526-6775     E-mail: ccopgi@mcmaster.ca

Journal Citation (Vancouver Style):

Guideline Citation (Vancouver Style): Nofech-Mozes S, Vella E, Dhesy-Third S, Hanna W. Guideline on hormone receptor testing in breast cancer. Toronto (ON): Cancer Care Ontario; 2011 Apr 8. Program in Evidence-based Care Evidence-Based Series No.: 22-1.
Evidence-Based Series 22-1: Section 1

Guideline on Hormone Receptor Testing in Breast Cancer:
Guideline Recommendations

S. Nofech-Mozes, E. Vella, S. Dhesy-Thing, and W. Hanna

A Quality Initiative of the Program in Evidence-Based Care (PEBC), Cancer Care Ontario (CCO)

Report Date: April 8, 2011

PURPOSE
The overall purpose of this guideline is to improve the quality and accuracy of hormone receptor (HR) testing and its utility as a prognostic and predictive marker for invasive and in situ breast cancer.

QUESTIONS
1. Clinical Validity of Immunohistochemistry (IHC)
   Can IHC reliably determine the levels of expression of the HR pathway and potentially correlate with the clinical outcome compared to other assays (dextran-coated charcoal [DCC] or ligand-binding assay [LBA], enzyme immunoassay [EIA], enzyme-linked immunosorbent assay [ELISA], and flow cytometry)?

2. Optimizing IHC
   How should HR testing be performed optimally to assess true HR status? This includes evaluating the effect of the following:
   
   - Preanalytic variables (i.e., variables of testing involving the collection, fixation, and storage of samples)
   - Analytic variables (i.e., variables associated with the method of testing itself)
   - Thresholds to define results
   - Postanalytic variables (i.e., variables associated with handling of the results, such as reporting)

3. Quality Assurance of IHC
   What parameters should be used to assess the proficiency of an individual laboratory performing HR-status testing?
4. Clinical Validity of the Oncotype DX Assay
   Can Oncotype DX reliably determine the levels of expression of the HR pathway?

TARGET POPULATION
   Tissue from adult female patients with primary or metastatic invasive or in situ breast cancer was the focus, since most of the literature includes tissue from female patients. However, the recommendations in this guideline would also apply to tissue from male breast cancer patients.

INTENDED USERS
   Any personnel involved in HR testing and interpretation and any health care provider involved in the management of breast cancer patients.

CONTEXT
   This guideline is not intended to advise clinicians when to perform HR testing for invasive or in situ breast cancer but to improve the quality and accuracy of HR testing. Limited data, based on the retrospective analysis of the National Surgical Adjuvant Breast and Bowel Project (NSABP) B24 study suggest that estrogen receptor (ER) expression is an important predictor of the response to tamoxifen in patients with ER-positive ductal carcinoma in situ (DCIS) (1). The results for ER-negative DCIS were inconclusive (1). Although HR testing for DCIS is not mandatory, some oncologists are seeking the HR status in DCIS when considering the benefit of adjuvant endocrine treatment according to the Cancer Care Ontario treatment guidelines for DCIS (2). If DCIS is tested for HR status, this guideline should be used. In cases of multifocal invasive carcinoma with different histologic types and/or grades, separate HR testing should be conducted.

RECOMMENDATIONS AND KEY EVIDENCE
1. Clinical Validity of IHC
   Can IHC reliably determine the levels of expression of the HR pathway and potentially correlate with the clinical outcome, compared to other assays (DCC/LBA, EIA, ELISA, and flow cytometry)?
   - IHC should be used instead of DCC/LBA, EIA, ELISA, or flow cytometry.
     In 20 out of 22 studies that compared IHC to another test and included patient outcomes, ER and/or progesterone-receptor (PR) using IHC was found to predict patient response to endocrine therapy and/or provide prognostic data such as overall survival, disease-free survival/interval, progression-free survival, metastasis-free interval, and recurrence/relapse-free survival at least as well as or better than the reference assays of DCC or EIA or ELISA (3-24). IHC has certain advantages over older forms of HR testing, which include the requirement of smaller amounts of tissue; the ability to conduct testing on formalin-fixed, paraffin-embedded (FFPE) tissue (allowing for histological examination); and the storage and retrieval of archived, stained slides for retrospective analysis.

2. Optimizing IHC
   How should hormone receptor testing be performed optimally to assess the true HR status?

Preanalytical Variables
Core Biopsies
   - Core biopsies may be used to assess ER and PR status in the primary tumour prior to neoadjuvant therapy or in the case of metastatic disease. However, as they are
derived from only a small sample of a larger tumour where normal ducts and lobules are frequently not present and in view of the heterogeneity in tumour HR expression, it is preferable to test the tumour in the surgical excision specimen. Core biopsies may also be preferentially assessed when the surgical specimen is inadequate due to inadequate fixation or lack of an invasive component.

- If the surgical specimen is adequately fixed and internal and external controls are concordant, it is not mandatory to repeat testing on negative tumours.
  - These recommendations differ from the American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP) recommendations where core biopsies are preferred (25-27). The potential lack of normal ducts and lobules and tumour heterogeneity are reasons the Working Group preferred excisional specimens. ASCO/CAP preferred core biopsies, because typically cores are immediately placed in formalin, minimizing cold ischemic time, and therefore may be exposed to more uniform and consistent tissue fixation (25-27). In the case of inadequate fixation of surgical specimens, the Working Group preferred core biopsies as well.
  - Eighteen comparative studies found concordance values greater than 83% (median 95%) for ER and greater than 69% (median 88.5%) for PR from core biopsies against standard surgical specimens (28-45).
  - Variability between cores and whole tumours may be dependent on intratumoural heterogeneity or be affected by treatment (33,42).

**Tissue Microarray (TMA)**

- TMA should not be used for diagnostic purposes. The test may be used for quality assurance and as a research tool. At least two cores should be removed from each tumour.
  - Seven comparative studies found concordance values greater than 95% (median 97%) for ER and greater than 85% (median 93.25%) for PR using TMAs against standard whole sections (46-52). Six retrospective studies found prognostic values of HR using TMAs (7,51,53-56). Three studies found good agreement between cores (46,49,55). However, three other studies suggested there may be heterogeneity in the amount of tumour tissue between cores and that cores might be lost (47,50,57). The Working Group recommended against the use of TMAs for diagnostic purposes for several reasons, including the heterogeneity in the cores, the requirement of specialized skill and reading, the possibility for errors with low-positive tumours, and the delay in HR testing caused by waiting for TMAs to be constructed from multiple patients.

**Fine Needle Aspiration (FNA)**

- FNA is not the preferred method to assess HR status; however, FNA may be used to assess ER or PR status on cell blocks using IHC, especially in metastatic disease where a core biopsy may not be a possibility.
- FNAs should be fixed in neutral buffered formalin (NBF), using cell blocks from pellets
- If not fixed in formalin, the protocol should be validated and have appropriate controls
- In cases where samples are HR negative, at least 100 nuclei should have been scored from FNA samples, when available.
  - One of the limitations with using FNA is that the distinction between in situ and invasive components of a malignant aspirate is often impossible to make.
Sixteen comparative studies found concordance values greater than 86% (median 94.5%) for ER and greater than 73% (median 80.2%) for PR from FNA against the standard of IHC on paraffin sections (58-73). Two studies showed a high degree of reliability using FNAs (59,74).

The Working Group preferred formalin fixation to maintain consistency with fixative recommendations for surgical sections (please see below).

One study suggested at least 100 nuclei should be used as a criterion to reject technically suboptimal slides for scoring FNAs (71).

**Frozen Sections**

- Paraffin sections should be used instead of frozen sections. However, frozen sections can be used to validate the quality of fixation of paraffin sections.
- The concordance between frozen and paraffin sections among 21 studies was over 72% (median 91.4%) for ER and over 84% (median 90.8%) for PR (9,21,68,75-92). One study found that ER assessed in paraffin, frozen samples or using DCC were similar in predicting response to endocrine therapy (75), whereas another study found that ER assessed in paraffin sections was a better predictor of endocrine response than ER assessed in frozen samples (21).
- It is common practice to use paraffin sections, because histological and HR assessment can be determined on the same block. In addition, tumours that are small and/or non-palpable may be difficult to sample for frozen section analysis.

**Fixation Type**

- Ten percent NBF should be used as a fixative for optimal results.
- The formulation of NBF should be confirmed with each lot and if different should be validated.
- Other fixatives may be used if they have been validated against NBF.
- Ethanol should not be used as a fixative.
  - Two studies, using paraffin sections or TMAs, suggest that 10% NBF as a fixative achieves the best results and that ethanol should be avoided as a fixative (93-94). Furthermore, NBF is commonly used in many laboratories. If laboratories use other fixatives, then their results should be validated against tissue fixed in NBF. (See the internal quality assurance [QA] section below.)

**Fixation Time**

- Place tissue into fixative as soon as possible following removal (preferably immediately and not more than one hour afterwards).
- Avoid freezing and cryosectioning before fixing the sample.
- Fix for 24 hours for optimal results.
- With proper antigen retrieval, tissue can be fixed for approximately six to 72 hours, which includes the time the tissue is loaded on the machine to the time that alcohol is introduced.
- Underfixation is more critical than overfixation; however, shorter or longer fixation times do not exclude the specimen for HR testing if proper validation of such protocols is well documented.
  - Two comparative studies found the best results were achieved when tissue was fixed for 24 hours in NBF, immediately after surgery (93-94). Poorer results were found when tissue was frozen and cryosectioned before fixing or when there was a delay in onset of fixation of 12 hours (93). Another study found that paraffin
sections from tissue that had been fixed one to four hours after removal had poorer concordance with frozen tissue than paraffin sections from tissue fixed within a few minutes after removal (76). The Working Group chose to use a conservative estimate of a one-hour threshold for cold ischemic time to fixation.

- Two comparative studies found that tissues fixed for as little as four hours in one study and six to eight hours in another study could achieve similar results to tissue fixed for longer periods of time (76, 95-96). Again, the Working Group chose a conservative time of six hours as a minimum requirement for the duration of fixation as well as to be consistent with the HER2/neu-testing guidelines (98). However, this is a controversial issue and any future updates of this guideline should align with any changes made to the ASCO/CAP HR guidelines or HER2/neu-testing guidelines.

- Four comparative studies showed that, with proper antigen retrieval, tissue fixed for several days will continue to immunoreact for HR (80, 89, 93, 97). The longest duration of fixation used was 48 hours for PR, compared to frozen sections, and 96 hours for ER, compared to tissue fixed for 24 hours. The Working Group chose 72 hours for both ER and PR as maximum requirements for the duration of fixation in order to be consistent with HER2/neu testing recommendations (98). In addition, longer fixation does not exclude the specimen for HR testing.

**Tissue Processing**

- Surgical specimens received in the pathology laboratory should be oriented, inked for surgical margin assessment, and carefully sliced at 5-10 mm intervals before being placed into formalin. If the tissue samples are to be further sectioned and placed into tissue cassettes at a later time, gauze pads or paper towels should be placed in between tissue slices to assist with the penetration of the formalin. If a gross tumour is easily identifiable, a section including a portion of the tumour and normal breast tissue from around the tumour can be placed in the same cassette and immediately fixed at the time of the initial gross evaluation. This action will ensure good tissue fixation and that the normal breast tissue, acting as an internal positive control, will be fixed in the same manner as the tumour. For samples obtained remotely from the grossing laboratory, pathologists, in collaboration with the personnel at the remote locations, should ensure the sample is bisected at the level of the tumour and promptly placed in fixative prior to its refrigeration and transport. Although this is less optimal than the rapid gross examination of fresh samples by the pathologist, it is preferable to uncut, fixed or unfixed tissue stored in the refrigerator.

- The block samples from the specimen should be 3-5 mm in thickness. Similarly, the appropriate thickness of the sections from the paraffin block should be 3-5 µm.

- If the laboratory is remote from the site of the surgery then the surgeon should ink and slice the sample and place it in formalin.

- No published studies were found to inform these recommendations. Therefore, these recommendations were developed through consensus with the Working Group, based on common practices in anatomic pathology laboratories.

**Storage of Slides**

- Use freshly cut sections for optimal results.

- Two comparative studies using TMAs showed better HR detection was found with freshly cut sections compared with slides stored for longer periods of time (99-100). As well, one study found that longer storage decreased HR associations with
tumour-specific survival (101). Several studies suggested possible methods to preserve antigenicity (paraffin coating and protection from light and heat) (99,102-103); however, the Working Group decided that, for diagnostic purposes, freshly cut sections should be recommended.

**Analytical Variables**

**Controls**

- Controls should include positive and negative breast cancer cases, plus a low positive case if possible.
- The control tissue should be fixed and processed in the same manner as the patient samples.
- Controls should be run with each HR IHC batch run; however, on slide external controls for automated processes are preferred.
- Internal controls of normal breast elements must be evaluated when present on the section. In the majority of breast cancer resection specimens, 5-10% of luminal epithelial cells will express ER in normal breast elements. If normal breast elements are completely ER negative, a false negative result should be considered, and another block should be tested or ER expression may be measured by another method.
- Other types of tissue such as the endometrium or myometrium, or uterine cervix, could be used for external controls. For progesterone receptors, hyperplastic benign prostatic tissue could be used for controls. When selecting controls, attention should be made to the variation in the level of expression during the menstrual cycle in premenopausal women that would give a broad level of expression for HR.
- The controls should use tissue. Cell lines may be used as controls but not in isolation. Cell lines lack stroma and are not subjected to the same processing methods.
  - No articles were found to inform these recommendations. Therefore, these recommendations were developed through consensus with the Working Group based on common practices in their laboratories.

**Antigen Retrieval**

- Antigen retrieval may be used to recover immunoreactivity from tissue; however, the method, reagents and duration must be optimized for each antibody.
- Stringent compliance to validated standard operating procedures developed in assay validation must be adhered to, and quality control (QC) documentation must be in place.
- Any modifications must be validated.
  - Eleven comparative studies were highly variable in their choice of antibody and antigen retrieval method (80,92,104-112). Since no specific recommendations for antigen retrieval could be derived from the evidence, the Working Group recommended compliance with validated standard operating procedures used in the laboratory, either following package inserts or validating against a reference standard (see internal QA section below).

**Reagents**

- Stringent compliance to validated standard operating procedures developed in assay validation must be adhered to, and QC documentation must be in place.
Any modifications must be validated.
- Due to the lack of evidence in this area, these recommendations were developed through consensus with the Working Group based on common practices and to be consistent with recommendations developed for HER2/neu-testing (98).

Antibodies
- Currently approved antibodies by Health Canada that have a Class II medical device license may be used.
- The Working Group chose to recommend antibodies approved by Health Canada, based on the suggestion from external reviewers. Also, the Quality Management Program - Laboratory Services in Ontario recognizes all tests approved and licensed by Health Canada.

Antibody Detection
- Antibody incubation and detection must be optimized for the specific antibodies used and must be validated with every lot of antibody.
- Although there were comparative studies that identified differences between antibody detection systems for various antibodies (93,112-115), the Working Group did not want to specifically endorse any type of detection system as each protocol needs to be validated or the package inserts need to be followed, regardless of the detection system used (see internal QA section below).

Automated versus Manual Staining
- Correctly operated and validated automated staining protocols and equipment should be used for all steps in the process; however, validated manual staining may be used.
- Maintenance and service records should be regularly updated and filed in the laboratory.
- There were two comparative studies that found a higher percentage of ER positivity using immunostainers compared to manual staining (116-117). One study found automated staining was more reliable than manual staining but did not find a statistically higher percentage of ER positivity with immunostainers (118). The Working Group decided that automated staining, if available, is recommended, but validated methods should be used.

Image Analysis
- Validated image analysis systems may be used, particularly for low-positive cases, to reduce subjectivity and improve interobserver reliability.
- Pathologists must supervise all image analyses.
- IHC using image analysis in 29 comparative studies had good reliability and was found to be highly concordant and correlated with manual estimation, but there were no prospective trials with patient outcomes to convince the Working Group that image analysis should replace manual estimation (6,11,13,53-55,76,119-140).

Thresholds to Define Results
- Only nuclear (not cytoplasmic) staining should be scored.
- There are three categories for staining:
  - Positive: ≥10% staining for ER or PR
  - Low positive: 1% to 9% staining for ER or PR
Negative: < 1% staining for ER and PR

- A statement of intensity can be reported as weak, moderate, or strong; however, this is more important if the tumour is low positive with weak intensity.
  - Different scoring methods were shown to be comparable, but the percent of cells stained positive is the simplest method and is correlated with patient outcomes (5,7,10-11,18,21,53,56,135,141-144). Moreover, six studies found that the proportion of positive staining was at least as good a predictive or prognostic indicator, or better, than the intensity of stained cells or a combination score of the proportion and intensity of stained cells (12-13,16,20,144-145).
  - Eight studies suggest that ER values as low as 10% and possibly as low as 1% can have predictive or prognostic value (9,12,15-16,20,23,143-144). Furthermore, other studies have shown the percentage of PR positivity, to as low as 1%, provides additional predictive or prognostic value (5,19,22,56,143,146), independent of the ER values (20,144). However, all of these studies are retrospective and mainly evaluate the prognostic value of HR in patients treated with endocrine therapy. They do not directly assess the predictive ability of HR at different cutoffs by examining the interaction between patients that receive endocrine therapy versus those patients who do not receive endocrine therapy at various cut points. Since most patients receive endocrine therapy, these studies would require large sample sizes. Therefore, in order to be maximally sensitive and based on the retrospective evidence, the Working Group decided to use 1% positivity for ER or PR as a threshold for defining low-positive samples. However, since there is high interlaboratory variability in assessing low-positive samples resulting in high false-negative rates (94,147), and the nature of the evidence to support a 1% threshold is retrospective, the Working Group felt that clinical judgment should be used when assessing patients with low-positive HR tumours for endocrine therapy, especially when chemotherapy might not be a tolerable option for some patients. Although the Working Group suggested that staining intensity could be included, they felt there was sufficient evidence not to make it a mandatory requirement in the evaluation of HR positive and negative status. Reporting intensity was more valuable in low-positive samples where clinical judgment might be necessary.

Postanalytical Variables

After the standard background patient information (see below), results should be stated clearly and bolded, and the additional information about the quality of the test should follow in synoptic format.

- The following items should be reported.
  To avoid lengthy and exhaustive reports, the Working Group makes a distinction between elements that need to be included in the formal pathology report and items that need to be collected and kept in the laboratory without necessarily being in the final report. Asterisked items should be included in the printed final report, visible to the physician. Other information should be available in the laboratory records.
  - *Standard background patient information
  - *Specimen identification (case and block number)
  - *Specimen site and type
  - Specimen fixative type (*if not 10% NBF)
  - Time to fixation (if available)
  - Duration of fixation (if available)
  - Antibody clone and vendor
Method used (test and vendor)
- Image analysis method (if used)
- Adequate controls
- Adequacy of sample for evaluation
- *Results
  - Percentage of positively stained cells for ER and PR
- *Interpretation
  - Low positive 1-9% for ER or PR
  - Positive ≥ 10% for ER or PR
  - Negative < 1% for ER and PR
  - Not interpretable
  - Intensity of staining: weak, moderate or strong

- *Sample exclusion criteria to perform or interpret a HR IHC assay include the following:
  - Tissue fixed using other than 10% NBF unless validated
  - When controls were inappropriate
  - Core needle biopsies with
    - Edge or retraction artifact involving entire core
    - Crush artifact (thin-gauge vacuum-extraction needle samples)
    - Tissues where controls exhibit unexpected results
  - If the assay is negative, then:
    - Look at the histology and grade; some cancer types such as classic lobular and tubular carcinoma or low-grade tumours are only rarely ER negative.
    - For cases where there is incongruence between HR testing and histology and/or grade, repeat the assay or have another laboratory repeat the assay
  - Controlled decalcified specimens with EDTA can be used. When possible, tease out fragments of the tumour so that no decalcification is required.
  - In samples with only DCIS, the diagnosis of DCIS should be mentioned and scored.
  - In tumours that contain both the in situ and invasive components, only the invasive component should be tested and scored.

Reporting elements not supported by evidence described in the answers to question 2, were developed through consensus with the Working Group, based on commonly used practices. These reporting elements are also in keeping with the recommendations for HER2/neu testing (98).

3. Quality Assurance of IHC
What parameters should be used to assess the proficiency of an individual laboratory performing HR-status testing?

Caseload
- The evidence to guide exact caseloads per laboratory or pathologist is not available; however, laboratories could use the recommendations for HER2/neu testing, although this is not mandatory. These cases should be carried out by experienced pathologists at laboratories participating in IHC external QA accreditation programs.
- Appropriate training is recommended for pathologists who report on HR status. The appropriate training may be part of a residency or fellowship program, mentorship with an experienced pathologist, or a formal didactic course. Cases should include
at least 20 negative and at least 20 positive specimens, with some being weakly positive. When possible, cases should be reviewed with an experienced pathologist and at least a 90% concordance should be achieved; any discordance should be assessed on a dual-head microscope. In a challenge set of 40, less than three discordant cases are considered acceptable (148).

- Staff should be encouraged to show each other borderline cases; experienced pathologists should be consulted in cases with low-positive staining or weak intensity.
- These activities could be tracked in pathology computer systems.
- The number of tests performed by each pathologist should be considered to ensure competency.
- Test volume should be addressed in conjunction with the laboratory’s adherence to strict QC and QA practices.
- Medical laboratory technologists who perform Class II IHC testing should undergo appropriate training.

**Internal QA**

- Initial test validation should take place together with ongoing QC and equipment maintenance.
- Initial and ongoing education, training, and competency assessment of laboratory personnel should also be implemented and recorded.
- The use of standardized operating procedures, including routine use of control materials, should be enforced, and modified procedures should be revalidated.
- Ongoing competency assessment and education of pathologists should take place.
- When validating a new antibody, at least 40 samples should be tested with known results. Those should include 20 positive, of which at least five are low positive, and 20 negative cases. However, if the laboratory has little experience with performing HR testing, doubling the sample size is advisable (148).
- An assay concordance rate of 90% for ER-positive or PR-positive tumours and a 95% concordance rate for ER-negative or PR-negative tumours should be achieved.
- Adequate validation should be ensured, preferably by using a selection that includes approximately 30% unequivocally positive cases, approximately 30% low positive, and approximately 40% negative.
- Validation documentation must be kept for at least five years.
- Any modification to preanalytical, analytical, or postanalytical procedures requires additional validation to ensure accurate performance.
- At least semiannual trend analysis should be performed by internal audit for each institution and preferably for each pathologist; approximately 70% of samples should be ER positive and approximately 50% of samples should be PR positive; this may vary depending on the referred patient population.

**External QA**

**HR-specific external QA**

- Laboratories are required to participate in at least one external proficiency program assessing analytical and postanalytical components (such as the Canadian IHC Quality Control Group, the Ontario External Quality Assessment program organized by the Quality Management Program - Laboratory Services in Ontario, the United Kingdom National External Quality Assessment Service [UK NEQAS], or CAP) with at least two testing events (mailings) annually (149).
- Unsatisfactory performance results will be addressed according to the regulations of the accreditation program.

IHC external QA
- Each laboratory should be accredited to perform IHC and follow standard operating procedures (according to the Ontario Laboratory Accreditation [http://home.qmpls.org/external/index.html] requirements).
- Onsite inspection for IHC should take place every other year, with an annual requirement for self-inspection.
- A review of the laboratory validation, procedure, QA results and processes, results, and reports for IHC should be put into place.

All recommendations addressing question 3 were developed, through consensus with the Working Group, to be consistent with HER2/neu testing recommendations (98) and the ASCP/CAP recommendations (148). The percentages of ER and PR positivity for periodic trend analysis were derived from the mean frequency observed across 71 laboratories, using a 10% threshold in the UK_NEQAS (150).

4. **Clinical Validity of Oncotype DX**
   Can Oncotype DX reliably determine the levels of expression of the HR pathway?

- **Oncotype DX can accompany IHC results but should not replace them.**
  - Fourteen studies investigating Oncotype DX have evaluated the predictive validity of the recurrence score in ER-positive breast cancer, and none were found that correlated the expression of the HR-related genes to clinical outcome (151-164). The Working Group decided there was insufficient evidence to suggest that HR assessed with Oncotype DX is a better predictive or prognostic indicator than HR assessed using IHC.

**FUTURE RESEARCH**
Future research that would provide valuable information for these recommendations would include studies that validated the score derived from the HR-related genes assessed by Oncotype DX, prospective trials that included patient outcomes comparing image analysis with manual estimation, and studies comparing 1% versus 10% thresholds as predictive HR markers.

**Funding**
The PEBC is a provincial initiative of Cancer Care Ontario supported by the Ontario Ministry of Health and Long-Term Care through Cancer Care Ontario. All work produced by the PEBC is editorially independent from its funding source.

**Copyright**
This report is copyrighted by Cancer Care Ontario; the report and the illustrations herein may not be reproduced without the express written permission of Cancer Care Ontario. Cancer Care Ontario reserves the right at any time, and at its sole discretion, to change or revoke this authorization.

**Disclaimer**
Care has been taken in the preparation of the information contained in this report. Nonetheless, any person seeking to apply or consult the report is expected to use independent medical judgment in the context of individual clinical circumstances or seek out the supervision of a qualified clinician. Cancer
Care Ontario makes no representation or guarantees of any kind whatsoever regarding the report content or use or application and disclaims any responsibility for its application or use in any way.

Contact Information

Dr. Wedad Hanna, Professor, Department of Laboratory Medicine and Pathobiology  
Faculty of Medicine, University of Toronto  
Fax: 416-480-4271 Email: Wedad.Hanna@sunnybrook.ca

For information about the PEBC and the most current version of all reports, please visit the CCO website at http://www.cancercare.on.ca/ or contact the PEBC office at:  
Phone: 905-527-4322, ext. 42842 Fax: 905-526-6775 E-mail: ccopgi@mcmaster.ca
REFERENCES


RECOMMENDATIONS - page 14


107. Huang Z, Zhu W, Meng Y, Xia H. Development of new rabbit monoclonal antibody to progesterone receptor (Clone SP2): no heat pretreatment but effective for paraffin


Guideline on Hormone Receptor Testing in Breast Cancer:
Evidentiary Base

S. Nofech-Mozes, E. Vella, S. Dhesy-Thind, K.L. Hagerty, P.B. Mangu, S. Temin, and W. Hanna

A Quality Initiative of the
Program in Evidence-Based Care (PEBC), Cancer Care Ontario (CCO)

Report Date: April 8, 2011

QUESTIONS
1. Clinical Validity of Immunohistochemistry (IHC)
   Can IHC reliably determine the levels of expression of the hormone receptor (HR) pathway and potentially correlate with the clinical outcome compared to other assays (dextran-coated charcoal [DCC] or ligand-binding assay [LBA], enzyme immunoassay [EIA], enzyme-linked immunosorbent assay [ELISA], and flow cytometry)?

2. Optimizing IHC
   How should HR testing be performed optimally to assess true HR status? This includes evaluating the effect of the following:
   - Preanalytic variables (i.e., variables of testing involving the collection, fixation, and storage of samples)
   - Analytic variables (i.e., variables associated with the method of testing itself)
   - Thresholds to define results
   - Postanalytic variables (i.e., variables associated with handling of the results, such as reporting)

3. Quality Assurance of IHC
   What parameters should be used to assess the proficiency of an individual laboratory performing HR-status testing?

4. Clinical Validity of Oncotype DX
   Can Oncotype DX reliably determine the levels of expression of the HR pathway?
Hormone Receptor Testing In Breast Cancer: A Systematic Review of the Literature

INTRODUCTION

Testing for the presence of hormone receptors (estrogen and progesterone receptors [ER and PR]) has become the standard of care in the treatment of women with breast cancer. The results of such testing help to direct therapy, and therefore, it is critical that these tests are accurate and reliable. This ensures that women who might benefit from endocrine therapy are provided with treatment and that those who are unlikely to derive benefit do not receive these treatments and are not exposed to unnecessary side effects.

Originally, DCC LBA was the method most commonly used to assess ER and PR. Other assays such as EIA and ELISA then became available. By the late 1980s and early 1990s, IHC of formalin-fixed paraffin-embedded (FFPE) specimens began to replace the DCC assay as IHC has distinct advantages over DCC. These include the need for smaller amounts of tissue, the ability to conduct testing on FFPE tissue (obviating the need for fresh/frozen tissue), the ability to correlate staining with histology, and the storage and retrieval of archived slides for later testing.

Unfortunately, concomitant with the widespread uptake of IHC across North America and Europe is the wide variability in how different laboratories perform the assay and, therefore, the results that they obtain. Because of these issues, there is a need for evidence-based guidance related to ER and PR testing with IHC. The evidence from this systematic review has been used to inform the recommendations developed in Ontario and by the American Society of Clinical Oncology (ASCO) in partnership with the College of American Pathologists (CAP) (1-4).

Additionally, although the main focus of this systematic review is on the performance of IHC, we have also included all papers on the 21-gene assay signature know as Oncotype DX. Oncotype DX is a relatively recent technology that uses reverse-transcription polymerase chain reaction (rt-PCR) for messenger ribonucleic acid (mRNA) quantitation of a panel of 16 genes plus five controls and was tested and validated retrospectively using clinical trial-annotated FFPE samples from breast cancer clinical trials in ER-positive disease. The assay is performed in a Clinical Laboratory Improvement Amendments-certified central laboratory in the United States, although it has not been reviewed under the new Food and Drug Administration In Vitro Diagnostic Multivariate Index Assays (IVDMIA) Guidance.

This systematic review was developed, therefore, to evaluate the existing evidence concerning the following aspects of IHC:

Clinical Validity of IHC
Can IHC reliably determine the levels of expression of the HR pathway and potentially correlate with the clinical outcome compared to other assays (DCC/LBA, EIA, ELISA, and flow cytometry)? While IHC testing is widely used, the opinion was that all other issues described below were only worth investigating if IHC was confirmed as being clinically valid.

Optimizing IHC
How should HR testing be performed optimally to assess true HR status? This includes evaluating the effect of:

- Preanalytic variables (i.e., variables of testing involving the collection, fixation, and storage of samples)
- Analytic variables (i.e., variables associated with the method of testing itself)
- Thresholds to define results

EVIDENTIARY BASE - page 2
Postanalytic variables (i.e., variables associated with handling of the results, such as reporting)

The goal of optimizing IHC is to adjust the variables described above to minimize the costs (e.g., time, resources), while maintaining the same or better accuracy of the test.

Quality Assurance of IHC
What parameters should be used to assess the proficiency of an individual laboratory performing HR status testing?

Clinical Validity of Oncotype DX
Can Oncotype DX reliably determine levels of expression of HR pathway?

METHODS
Evidence was selected and reviewed by one member of the Program in Evidence-Based Care (PEBC) staff and four members of the ASCO staff. Abstracts were initially reviewed by four individuals. After reviewing a set of 200 abstracts, a kappa (k) coefficient was calculated. Any discrepancies were resolved among the four individuals, and another set of 200 abstracts were reviewed. We considered a k coefficient value of greater than 0.7 as good agreement (5). After a third set of 200 abstracts, a k of 0.72 was achieved, and the remaining abstracts were reviewed by one of the four individuals. Full-text review was completed by any two of the four reviewers. Data was extracted for each publication by one of four reviewers and was independently confirmed.

This systematic review is a convenient and up-to-date source of the best available evidence on HR testing in breast cancer. The body of evidence in this review primarily comprises comparative retrospective data. That evidence forms the basis of the American recommendations developed by ASCO in partnership with CAP, and the Ontario recommendations developed separately by the PEBC (Appendix 1) (1-4). The PEBC is supported by the Ontario Ministry of Health and Long-Term Care through Cancer Care Ontario. All work produced by the PEBC is editorially independent from its funding source.

Literature Search Strategy
MEDLINE (Ovid, 1990 - May 2008) was searched using a disease-specific medical subject heading (MeSH) term (“breast neoplasms”) and marker-specific MeSH terms (“receptors, estrogen”, “receptors, progesterone”). EMBASE (Ovid, 1990 - 2008 week 23) was similarly searched using a disease-specific Excerpta Medica Tree term (“breast tumor”) and marker-specific terms ((?estrogen or progesterone or hormone) and receptor?).tw). Comments, letters, editorials, notes, errata, short surveys, news, newspaper articles, and patient education handouts were excluded. The Cochrane Database of Systematic Reviews (2nd Quarter, 2008) was searched using disease-specific text words ((breast or mammar$) and (cancer? or carcinoma? or neoplasm? or tumo?$ or malignan$).mp) and marker-specific terms ((?estrogen or progesterone or hormone) and receptor?.mp). The complete search strategies for each database are detailed in Appendix 2. Abstracts from the San Antonio Breast Cancer Symposium (SABCS) (http://www.sabcs.org/SymposiumOnline/index.asp#abstracts; 2000-present) were also searched, as were those from the ASCO Annual Meeting (2000-2008; http://www.asco.org/ASCO/Abstracts+%26+Virtual+Meeting/Abstracts).

An environmental scan of the non-indexed evidence was also performed. The environmental scan was comprised of two parallel processes, one a targeted search of international guideline developers and key organizations (August 6, 2008) and the other an
untargeted search to identify previously unknown sources of evidence (September 2-3, 2008). A listing of the organizations that were examined in the targeted search is given in Appendix 3. For the untargeted search, the Google™ online Internet search engine was used with the keywords “hormone/estrogen receptor” and “test” and “quality and/or breast” and “guideline”, “oncotype DX”, “hormone/estrogen receptor immunohistochemistry” and “guideline/standard”. The bibliographies of retrieved articles were also searched for relevant reports.

Study Selection Criteria

Inclusion Criteria

All study designs including randomized controlled trials, cohort designs, cases series, evaluation studies, comparative studies, prospective studies, and retrospective studies from 1990 onward comparing IHC in paraffin-embedded, female primary or metastatic invasive or in situ breast cancer sections with another assay (DCC, EIA, ELISA, or flow cytometry) were included. In addition, all studies from 1990 onwards that investigated preanalytic variables, analytic variables, postanalytic variables, or laboratory proficiency strategies in IHC in paraffin-embedded female breast cancer sections were included. Since Oncotype DX is a novel technique with few published articles, all studies from 1990 onwards were included. All studies had to report quantifiable data. Studies that linked test performance to any clinical outcome were specifically sought, but clinical outcome data was not required for inclusion. All studies were required to be comparative, in the sense that the same tissue sample have to have been tested by more than one method, or more than one variation of IHC, and the results of both tests reported and compared (see summary of outcomes assessed, below). As well, all studies from 1990 that examined quality assurance (QA) strategies, or proficiency testing, or individual or institutional training for IHC were included.

Systematic reviews, consensus statements, and practice guidelines from 1990 onward were included if they addressed HR testing in female breast cancer using IHC in paraffin-embedded sections or Oncotype DX. A cutoff date of 1990 was chosen as this was the time when IHC began to come into common use.

Exclusion Criteria

Studies that included samples only from males were excluded. Studies measuring ER beta or using cell lines were excluded, as were studies comparing IHC with mRNA expression other than Oncotype DX. Publications in a language other than English were not eligible due to a lack of translation resources. Non-systematic reviews, case studies, letters, editorials, and commentaries were excluded.

Summary of Outcomes Assessed

The primary clinical outcome of interest was the correlation between HR status and benefit from endocrine therapy; however, articles with any patient outcomes were included. There was an a priori expectation that few studies would report on clinical outcomes, especially in studies that addressed the issues of optimizing IHC and the QA of IHC.

Other outcomes of interest included sensitivity, specificity, and overall accuracy. Where diagnostic tests had binary outcomes, the most common outcome measure reported is the percent concordance between the assay being evaluated and another test, or the same test with altered parameters. Concordance can refer to overall concordance (performance on combined negative and positive cases), sensitivity (performance on positive cases only), or specificity (performance on negative cases only). It is necessary to mention that concordance of assays does not guarantee accuracy (i.e., how close the measured values are to a supposed true value). Evaluating the accuracy of a test requires its comparison to a gold standard,
which does not exist at present. No assay such as IHC, DCC, EIA, ELISA, or flow cytometry is perfectly accurate for identifying all patients expected to benefit or not benefit from hormone therapy. Furthermore, there is no standardized single method for IHC detection of HRs in breast cancer tissue. One of the goals of this systematic review is to develop evidence that can be used to establish standards for IHC.

The correlation between HR IHC scores was considered if the scoring method of IHC was not dichotomized. Additionally, measures of precision of IHC such as the inter-rater reliability between laboratories or pathologists were included.

Synthesizing the Evidence

The data were not pooled because there was considerable heterogeneity in preanalytic and analytic variables such as tissue collection, the antibody used, and the method of fixation between studies.

RESULTS

Literature Search Results

See Tables 1 and 2 for a synopsis of the number of studies identified. Briefly, 17,093 abstracts were identified from the search of electronic databases. After the abstract review, 863 papers were retrieved for full-text review. Of these, 225 met our inclusion criteria. Since HR status determines endocrine-treatment decisions, when a study reported concordance data or concordance could be calculated, that information was extracted. If a study did not report concordance data but reported correlation, those data were extracted instead.

Twenty abstracts from the SABCS were considered for inclusion. Three abstracts were identified from 2005, eight from 2006, and nine from 2007 (6-25). Data were not extracted from abstracts, because no new data beyond that found when searching other electronic databases were identified.

Two consensus guidelines that addressed HR testing in breast cancer, using IHC, were found through the environmental scan. The recommendations from the National Breast and Ovarian Cancer Centre and the Australian Cancer Network, as well as the British Columbia Cancer Agency, did not describe the processes used in the development of the statements, although the documents provided reference lists (26-27). Since these documents were not evidence-based and the consensus processes were not described in detail, these recommendations were not used to inform the recommendations found in (1-4).

One existing systematic review met the inclusion criteria for the current systematic review; it was developed by the Johns Hopkins Evidence-Based Practice Center (EPC) for the Agency for Healthcare Research and Quality (AHRQ) and was identified from the environmental scan (28). This report evaluated the analytical and clinical validity of Oncotype DX, Mammaprint, and the two-gene (H:I) assay.
Table 1. Search criteria and studies identified.

<table>
<thead>
<tr>
<th>Search Method</th>
<th>MEDLINE and EMBASE Combined*</th>
<th>SABCS Annual Conference Proceedings</th>
<th>Environmental scan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search criteria</td>
<td>OVID</td>
<td>See Appendix 2</td>
<td>See Appendix 3</td>
</tr>
<tr>
<td>Number of articles identified by electronic search</td>
<td>17093</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Number of articles marked for possible inclusion after title/abstract review</td>
<td>863</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Number of articles included in review</td>
<td>179</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

*Because Cochrane Database of Systematic Reviews did not yield any relevant results, they were not included.

Table 2. Studies eligible for inclusion in this report.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Number of Studies</th>
<th>Summary of Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immunohistochemistry vs. other tests</td>
<td>22</td>
<td>Table 3</td>
</tr>
<tr>
<td>Preanalytic variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core biopsy</td>
<td>19</td>
<td>Table 4</td>
</tr>
<tr>
<td>TMA</td>
<td>13</td>
<td>Tables 3 or 4</td>
</tr>
<tr>
<td>FNA</td>
<td>16</td>
<td>Table 4</td>
</tr>
<tr>
<td>Frozen sections</td>
<td>21</td>
<td>Table 4</td>
</tr>
<tr>
<td>Fixation type</td>
<td>6</td>
<td>Table 5</td>
</tr>
<tr>
<td>Fixation time</td>
<td>8</td>
<td>Table 5</td>
</tr>
<tr>
<td>Storage of slides</td>
<td>8</td>
<td>Table 6</td>
</tr>
<tr>
<td>Analytic variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antigen retrieval</td>
<td>11</td>
<td>Table 7</td>
</tr>
<tr>
<td>Antibodies</td>
<td>55</td>
<td>Tables 3, 4, 7 or 8</td>
</tr>
<tr>
<td>Antibody detection</td>
<td>5</td>
<td>Table 7</td>
</tr>
<tr>
<td>Automated vs. manual staining</td>
<td>3</td>
<td>Table 7</td>
</tr>
<tr>
<td>Manual scoring</td>
<td>32</td>
<td>Tables 3, 4 or 8</td>
</tr>
<tr>
<td>Image analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AQUA</td>
<td>5</td>
<td>Table 4 or 10</td>
</tr>
<tr>
<td>CAS</td>
<td>8</td>
<td>Table 3 or 10</td>
</tr>
<tr>
<td>Photoshop</td>
<td>3</td>
<td>Table 10</td>
</tr>
<tr>
<td>SAMBA</td>
<td>1</td>
<td>Table 3</td>
</tr>
<tr>
<td>Other</td>
<td>13</td>
<td>Table 10</td>
</tr>
<tr>
<td>Laboratory Proficiency</td>
<td>15</td>
<td>Table 11</td>
</tr>
<tr>
<td>Oncotype DX/21-gene assay</td>
<td>14</td>
<td>Table 12</td>
</tr>
</tbody>
</table>

Abbreviations: AQUA, automated quantitative analysis; CAS, computer-assisted image analysis system; FNA, fine needle aspiration; SAMBA, Système d’Analyse Microphotométrique à Balayage Automatique; TMA, tissue microarray; vs., versus.

Study Design and Quality
For studies examining the clinical validity or clinical threshold of IHC compared to other tests, no randomized controlled trials were located, although a number of studies obtained specimens (slides, tissue blocks) collected during the course of a randomized trial to
be used for retesting. For this reason, most of the evidence reviewed is based on retrospective reviews, case series, and non-randomized prospective studies.

Of those studies designed to assess the technical aspects of IHC, most changed one or two IHC parameters and evaluated the concordance between the test results of the modified paradigm and a standard laboratory protocol. One of the limitations of these studies is the underlying assumption that the standard laboratory protocols have been validated against clinically validated assays. Only a few of the studies contain patient outcomes, and these studies are described in more detail.

There were several factors that affected the quality of the included studies. Samples collected prospectively would be subjected to less bias than samples assessed retrospectively and, if such a situation could be determined, it was recorded in the evidence tables. Only a few studies reported clearly whether the pathologists were blinded to the HR status of tissue, and this information was recorded in the evidence tables. When sample sizes were small for a single study (n<10), this was noted in the evidence summaries and also in the tables. The following evidence summaries highlight the best available evidence located in this review, with respect to the topics covered. The evidence provided context and some direction for the development of recommendations found in (1-4).

OUTCOMES
Can IHC reliably determine the levels of expression of the HR pathway and potentially correlate with the clinical outcome compared to other assays (DCC/LBA, EIA, ELISA, and flow cytometry)?

Although 88 studies in total compared the performance of IHC against DCC, EIA, ELISA or flow cytometry, only 22 of those studies included patient outcomes (Table 3) (29-50). Since IHC is the current standard of practice, the goal of this review was to confirm clinical validity. Therefore, it was decided to extract data only from those studies that directly linked the performance of IHC to patient outcomes. With the exception of two studies (31,48), ER and/or PR using IHC was consistently found to predict patient response to endocrine therapy and/or provide prognostic data such as overall survival, disease-free survival/interval, progression-free survival, metastasis-free interval, and recurrence/relapse-free survival at least as well as or better than the reference assays of DCC or EIA or ELISA (29-30,32-47,49-50).

The 66 studies that did not report on clinical outcomes are included in Appendix 4 for reference but are not discussed further in this review, unless they provide information that addresses another IHC question.
Table 3. Studies that compared IHC to DCC, EIA, ELISA, or flow cytometry and included patient outcomes.

<table>
<thead>
<tr>
<th>Author</th>
<th>No. of Patients (eligibility)</th>
<th>Intervention (outcome)</th>
<th>Original (cutoff)</th>
<th>Assay (cutoff)</th>
<th>Retrospective Assay (cutoff)</th>
<th>Outcome According to Biomarker</th>
<th>Blind</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pertschuk</td>
<td>74 patients</td>
<td>Patients received endocrine therapy for a minimum of 3 months, and any favourable response was maintained for at least 3 months</td>
<td>DCC (positive if ≥3 fmol/mg)</td>
<td>1D5 (positive if ≥10% positive cells)</td>
<td>IHC best predictor of clinical endocrine response (r, 0.57; 95% CI, 0.51-0.73, p&lt;0.001), DCC did not correlate (r, 0.002, 95% CI, -0.23-0.23, p&lt;0.99); IHC correctly predicted endocrine response in 29 of 30 discordant cases with DCC (p&lt;0.001)</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Regan</td>
<td>571 (premenopausal), 976 (postmenopausal)</td>
<td>Specimens from International BC Study Group Trials VIII and IX. Trials compared adjuvant chemoendocrine treatment with endocrine treatment alone in pre- and postmenopausal patients with lymph node negative BC</td>
<td>Specimens from International BC Study Group Trials VIII and IX</td>
<td>ER 1D5, PgR 1A6</td>
<td>Using 10% cut-point, ER status (for all patients) and PR status (for postmenopausal patients) predicted DFS similar to extraction assays; among premenopausal patients, PR IHC status could predict response to endocrine therapy (c index=0.60 vs. 0.51, p=0.003), unlike extraction assays</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Alberts</td>
<td>316 patients (tissue blocks)</td>
<td>Patients with lymph node positive BC from two prospective trials of adjuvant chemotherapy</td>
<td>DCC (positive if ≥10 fmol/mg)</td>
<td>ERTDS (scored 0 to 3; ≥1+ (&gt;0% and &lt;10%) considered positive)</td>
<td>ER by DCC better than ER by IHC as prognostic for DFS and OS; review of discordant cases indicates IHC may more accurately reflect ER status of malignant cells in some patients</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Andersen</td>
<td>349 patients; 309 IHC assessed</td>
<td>Patients were subset of 1700 patients in Danish BC Coop Group RCT of adjuvant TAM, median f/u 86 months</td>
<td>DCC (positive if ≥10 fmol/mg)</td>
<td>ER monoclonal antibodies (any nuclear staining above control considered positive)</td>
<td>IHC: ER+ significantly longer DFS (p&lt;0.001) and survival (p&lt;0.001) than ER-. DCC assay showed similar magnitude. No significant difference in DFS (p=0.52) or survival (0.54) between patients who received TAM and controls. Same was true for receptor-defined subgroups regardless if ER was estimated by IHC or DCC.</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Barnes</td>
<td>170 patients</td>
<td>First-line TAM in metastatic BC (51% response rate)</td>
<td>DCC; 74% ER positive (≥20 fmol/mg); response rate, 58%</td>
<td>ER ID5, NCL-PGR; 31% to 69% ER positive (various IHC scoring methods including a histoscore (cutoff 100), quick score, and category score (neg, weak, mod, strong); response rate, 64% to 69%;</td>
<td>IHC superior for predicting duration of response and type of response; for ER status, histoscore (chi square 30.31 p&lt;0.001), quick score (chi square 45.76 p&lt;0.001), category score (chi square 49.22 p&lt;0.001) and DCC (chi square 11.91, p&lt;0.001) gave significant results when used to predict duration of response; 43/71 (61%) ER+/PR- responded, 33/46 (72%) ER+/PR+ responded; Cox multivariate analysis</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>No. of Patients (eligibility)</td>
<td>Intervention (outcome)</td>
<td>Original Assay (cutoff)</td>
<td>Retrospective Assay (cutoff)</td>
<td>Outcome According to Biomarker</td>
<td>Blind</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------</td>
<td>----------------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>Biesterfeld 1997(34)</td>
<td>N=111; 108 for IHC analysis</td>
<td>Patients with unilateral invasive breast cancer; node+ postmenopausal or premenopausal or ER+ given therapy; postmenopausal ER+ given TAM; f/u mean 6.3 +/- 0.7 yrs</td>
<td>DCC (cutoff ER &gt;10 fmol/mg, PR &gt;20 fmol/mg)</td>
<td>ER 6F11, PR CLA 02/1 (manual ER &gt;0% and PR&gt;10% positivity)</td>
<td>showed ER expression related to time to progression (Chi-square=41.8, p&lt;0.0001 using quick score)</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Chebil 2003(35)</td>
<td>402 (ER); 394 (PR)</td>
<td>Patients with stage II tumours participating in two prospective RCTs (South Sweden BC Group); treated with TAM for 2 yrs; med f/u almost 6 yrs</td>
<td>EIA (positive if ≥25 fmol/mg)</td>
<td>ER 1D5 or 6F11, PR Polyclonal or 1A6; positive if ≥10% positivity</td>
<td>For ER: DCC+/IHC+ or DCC-/IHC- had significantly better PFS than DCC-/IHC- (p=0.001, p=0.007, respectively); better PFS for DCC-/IHC+ than DCC+/IHC- (p=0.10). For PR: DCC+/IHC- significantly better PFS than DCC-/IHC- (p=0.03) but other associations weaker</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>de Mascarel 1995(36)</td>
<td>938 patients</td>
<td>Consecutive patients with distance-metastasis-free invasive ductal carcinoma not otherwise specified, median f/u 108 months</td>
<td>DCC (cut points 10, 50 or 100 fmol/mg)</td>
<td>ER1D5 (cut points 5%, 10% or 50% positive cells)</td>
<td>For node positive group, ER positivity correlated with better outcomes, regardless of test or cutoff point used; although with multivariate analysis ER failed to reach significance (regardless of assay or cutoff level)</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Elledge 2000(37)</td>
<td>205 patients with blocks (original n=349, all ER positive by DCC)</td>
<td>Patients from SWOG 8228, ER+ metastatic BC treated with TAM 10 mg twice a day (n=56) or 10 mg/m² twice a day (n=149); 9 yrs median f/u</td>
<td>DCC (positive if ≥3 fmol/mg)</td>
<td>ER-6F11, PR-KD68 (Allred score ≥2)</td>
<td>Relationship between higher IHC ER or PR and increasing response to tamoxifen (ER p=0.001; PR p=0.03), time to treatment failure (ER p=0.003; PR p=0.007), and OS (ER p=0.006; PR p=0.03); In logistic regression models, only ER by IHC retained significance for predicting TAM response (p=0.02), along with menopausal status. TTF and OS significantly longer for patients with higher ER or PR IHC scores. Low, intermediate and high ER or PR categories showed similar differences in RRs by DCC or IHC.</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Esteban 1994(50)</td>
<td>n=250</td>
<td>FFPE tissue blocks of stage I and II BC who had partial or radical mastectomy; f/u median 64 months (54 to 214.3)</td>
<td>DCC (cut points of 10 or 20 fmol/mg)</td>
<td>ER: 1D5; used SAMBA image analysis - mean optical density (MOD)</td>
<td>With IHC ER, patients could be divided into three risk groups (stratified on MOD values), based on their OS (p=0.018) and DFS (p=0.038); using DCC, ER status</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>No. of Patients (eligibility)</td>
<td>Intervention (outcome)</td>
<td>Original Assay (cutoff)</td>
<td>Retrospective Assay (cutoff)</td>
<td>Outcome According to Biomarker</td>
<td>Blind</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>-------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>Ferno 1996(38)</td>
<td>98 samples</td>
<td>Consecutive patients with stage II tumours participating in two prospective RCTs investigating effect of TAM treatment (South Sweden BC Group); med f/u 38 months</td>
<td>EIA (positive if ≥25 fmol/mg) or isoelectric focusing in polyacrylamide gels (positive if ≥10 fmol/mg)</td>
<td>1D5 (positive if &gt;10% positivity)</td>
<td>ER+ patients had better outcome after adjuvant treatment than ER- (p=0.003 for DCC, p=0.004 for IHC).</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Fisher 2005(39)</td>
<td>n=1891 in initial phase III RCT, n=402 in this study</td>
<td>NSABP B-09, patients with lymph node positive BC</td>
<td>DCC (positive if ≥10 fmol/mg)</td>
<td>1D5, 1A6; IHC scored by 2 ind observers using %, intensity, and any-or-none algorithms</td>
<td>interobserver agreement ER % 87%, intensity 87%, all-or-none 89%, PR % 84%, intensity 86%, all-or-none 79%; Multivariate analysis-all methods for scoring ER predicted better prognosis for OS in patients with unfavourable lymph node status at 5 and 10 yrs; univariate analysis-ER+ score obtained with all methods significantly related to OS at 5 and 10 yrs; Results less consistent for PR scores and DFS and RFS</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Grabau 2000(40)</td>
<td>164 (ER); 132 (PR)</td>
<td>Random sample of patients with invasive ductal carcinoma drawn from consecutive patients treated for primary, operable, invasive breast cancer, Median f/u 8.3 yrs</td>
<td>DCC (positive if &gt;0 fmol/mg)</td>
<td>ERT1D5, KD68; (positive if &gt;0% positivity)</td>
<td>receptor positive fared better than negative in all cases; classification based on IHC allowed better discrimination of patients than classification based on DCC</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Harvey 1999(41)</td>
<td>1,982 patients</td>
<td>Patients with primary BC; 26% received endocrine therapy and 13% received combined chemo-endocrine therapy; median f/u 65 months</td>
<td>DCC (positive if ≥3 fmol/mg)</td>
<td>6F11 (Allred score &gt;2 or 1% to 10% weakly positive cells)</td>
<td>Multivariate analysis of patients receiving adjuvant endocrine therapy alone, IHC-DCC at predicting improved DFS (HR/p=0.474/0.0008 and 0.707/0.3214, respectively); tests equivalent at predicting OS (0.379/0.0001 and 0.381/0.0003, respectively)</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Iwase 2003(42)</td>
<td>489 patients</td>
<td>Patients with breast cancer undergoing operations; 77 treated with endocrine therapy</td>
<td>EIA ER positive if &gt;13 fmol/mg, PR positive if &gt;10 fmol/mg</td>
<td>ER1D5, PR636 (Allred score &gt;2 or 1% to 10% weakly positive cells)</td>
<td>IHC (p&lt;0.0001) better at predicting relapse-free survival than EIA (p=0.0009), tumours ER+ or PR+ were significantly related to response to endocrine therapy</td>
<td>Unclear</td>
<td></td>
</tr>
<tr>
<td>MacGrogan 942</td>
<td>Consecutive patients with</td>
<td>DCC ER positive if KD68; Percent</td>
<td>IHC &gt;=50% had better OS than &lt;50%; no</td>
<td></td>
<td></td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>Intervention (outcome)</td>
<td>Original (cutoff)</td>
<td>Retrospective (cutoff)</td>
<td>Outcome According to Biomarker</td>
<td>Blind</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------------------------</td>
<td>------------------</td>
<td>-----------------------</td>
<td>--------------------------------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996(43)</td>
<td>metastasis-free breast invasive ductal carcinomas, median f/u 117.9 months</td>
<td>&gt;10 fmol/mg, PR positive if &gt;15 fmol/mg</td>
<td>positivity (0-100%) and intensity (1-3)</td>
<td>relation found for DCC between OS and quantity of PR; in node-negative, IHC-PR&lt;10% only independent predictor for OS using multivariate analysis and DCC-PR &lt;15 fmol/mg significant predictor in node-positive group</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thomson 2002(46)</td>
<td>Survival was analyzed for premenopausal women with stage II BC randomized to ovarian ablation or chemotherapy (CMF) according to their ER status; median f/u 13.9 yrs</td>
<td>DCC originally done in 270 patients by three Scottish labs who participated in UK NEQAS</td>
<td>determined retrospectively using IHC on original tumour blocks from 1996 (quick score)</td>
<td>significant interaction between IHC quick score and treatment showed ovarian ablation more beneficial for patients with + quick score, patients with quick score of 0 had higher risk of death with ovarian ablation; this relationship was not observed with DCC results</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ogawa 2004(45)</td>
<td>Patients received adjuvant endocrine therapy; median f/u: 38 months</td>
<td>EIA ER cut point 13 fmol/mg, PR cut point 10 fmol/mg</td>
<td>ER 1D5 and PR MAB429 (Allred ≥3)</td>
<td>IHC (ER p=0.0007 and PR p=0.0087) at cutoff score of 3 was more predictive of disease-free survival rate than EIA (ER p=0.037 and PR p=0.0258); DFS following endocrine therapy was most significant using cutoff of proportion score 3 (= 10% positive stain) for both ER1D5 and PR (MAB429); multivariate analysis, adjusting for tumour size and lymph node status, shows ER and PR (cutoff PS 3) are independent</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mohsin 2004(44)</td>
<td>362 patients received endocrine therapy only, 593 received no adjuvant therapy</td>
<td>DCC (positive if ≥5 fmol/mg)</td>
<td>PgR1294 (Allred score ≥2 or 1% to 10% weakly positive cells)</td>
<td>neither PS or IS were associated with patient outcome; for patients with no adjuvant therapy: best cutoffs for disease-free and overall survival were TS=4 (p=0.053, adjusted p=0.371) and TS=5 (p=0.003, adjusted p=0.021), respectively; patients who received adjuvant endocrine therapy: best cutoff for disease-free and overall survival were TS=2, with p-values 0.0003 (adjusted p=0.0021) and 0.0002 (adjusted p=0.0014), respectively; IHC stronger predictor of disease-free survival (HR, 0.546, p=0.0034) and overall survival (HR, 0.595, p=0.0040) than DCC (DFS HR, 0.673, p=0.0534, OS)</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>No. of Patients (eligibility)</td>
<td>Intervention (outcome)</td>
<td>Original Assay</td>
<td>Assay</td>
<td>Retrospective Assay</td>
<td>Assay</td>
<td>Outcome According to Biomarker</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------</td>
<td>---------------------------------------------------------------------------------------</td>
<td>----------------</td>
<td>-------</td>
<td>---------------------</td>
<td>-------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Younes 1996(47)</td>
<td>Group 1: n=65 (frozen and DCC performed); Group 2: n=91 (IHC and DCC performed)</td>
<td>Patients were treated for invasive breast cancer</td>
<td>frozen: NCL-ER-LH1; DCC</td>
<td>1D5 (% positive cancer cells and H-score)</td>
<td>On frozen tissue (LH1), H-score &gt;85.5 (p=0.0341) or &gt;82% positive cells (p=0.0275) had better survival but not with FFPE tissue (1D5 with different patients); Survival analysis showed no significant cutoffs for IHC or DCC; significant difference in survival with cutoff of 25 fmol/mg using DCC, but not for any cutoffs using IHC</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Cheang 2006(48)</td>
<td>n=4150; SP1 vs. 1D5 n=4105</td>
<td>Retrospective cohort of 4,150 tested centrally using DCC; median f/u 12.4 yr; tissue frozen before FFPE then TMAs (1 core/patient) constructed</td>
<td>DCC (positive if &gt;1 fmol/mg)</td>
<td>1D5 or SP1 (positive ≥ 1% positivity)</td>
<td>Using TMA, 1D5 vs. SP1: 90%; for BC-specific survival: SP1 p=4.78 10⁻¹³, 1D5 p=1.65 10⁻⁷, DCC p=6.08 10⁻¹⁶; for RFS: SP1 p=3.98 10⁻⁸, 1D5 p=3.71 10⁻⁸, DCC p=9.25 10⁻¹⁶; for patients receiving TAM, multivariate analysis (with age, tumour size, grade, and lymphovascular and nodal status), SP1 better independent prognostic factor than 1D5. Among patients with discrepant ER results, the 8% of patients who were SP1 positive/1D5 negative showed good outcomes, and the 2% SP1-negative/1D5 positive had poor outcomes; SP1 is 8% more sensitive than 1D5 using DCC as gold standard</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Horti 2007(49)</td>
<td>IHC n=486, DCC n=51, EIA n=385</td>
<td>Paraffin-embedded surgical specimens from The Cancer Institute Hospital of the Japanese Foundation for Cancer Research, Tokyo, Japan (1982 to 1993); 659 had postoperative TAM therapy; median f/u 154 months</td>
<td>DCC or EIA (no cut points given)</td>
<td>1D5 (cut point 10%)</td>
<td>IHC: no difference between ER- and ER-groups for overall RFS (p=0.67307), significant difference for 5-year RFS (p=0.04045); DCC/EIA: no difference between ER+ and ER- for overall RFS (p=0.2088) or 5-year survival (p=0.62722)</td>
<td>NR</td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** BC, breast cancer; CI, confidence interval; CLA, conjugated linoleic acid; DCC, dextran-coated charcoal; DFS, disease-free survival; EIA, enzyme immunoassay; ELISA, enzyme-linked immunosorbent assay; ER, estrogen receptor; FFPE, formalin-fixed paraffin-embedded; FNA, fine needle aspiration; f/u, follow-up; HR, hormone receptor; IHC, immunohistochemistry; ind, independent; IS, intensity score; NR, not reported; NSABP, National Surgical Adjuvant Breast and Bowel Project; neg, negative; OS, overall survival; PFS, progression-free survival; PR, progesterone receptor; PS, proportion score; RCT, randomized controlled trial; RFS, recurrence-free survival; SAMBA, Système d'Analyse Microphotométrique à Balayage Automatique; SWOG, Southwest Oncology Group; TAM, tamoxifen; TMA, tissue microarray; TS, total score; TTF, time to treatment failure; UK NEQAS, United Kingdom National External Quality Assurance Schemes; vs., versus; yr(s), year(s).
How should HR testing be performed optimally to assess true HR status?

For tissue collection strategies, IHC performed on FFPE surgical sections was considered the reference standard. Because of the heterogeneity across studies, the concordance values are summarized as minimum and median values across those studies with two comparisons, for example core biopsy versus surgical resection. When there were multiple comparisons, for example different fixation times, results that were common across the studies were noted.

Preanalytic Variables

Specimen Collection Methods

Core Biopsy

Eighteen articles examined IHC ER or PR values taken from core biopsies and compared them to IHC on surgical specimens used as the standard (Table 4) (51-68). Fourteen papers examined ER and found concordance values greater than 83% (median 95%) (51,53,55-58,60,62-68). Twelve papers investigating PR observed concordance rates greater than 69% (median 88.5%) (51,55-58,60,62-66,68). Although some studies found HR status was nearly identical between biopsies and surgical specimens (53,55,59,63) or had fewer positive ER and/or PR cases on biopsies compared to surgical specimens (55-57,65), most studies found higher ER and/or PR positive rates in core biopsies compared to surgical specimens (51,54,56-58,60,62,64,66-68). This finding may be attributed to sampling from the periphery of the tumour or better fixation of the procedures of core biopsies. One study detected higher ER positivity at the periphery than at the centre for whole tumours (t=6.6; p=0.001), but not core biopsies (t=0.3827) (54). Another study found high concordance rates between core biopsies taken from different parts of the tumour (ER 100%, PR 85%) (68). Similarly, a study assessing the receptor status of individual foci in multifocal invasive ductal breast cancer (although on mastectomy specimens) found no major differences in ER status from different foci (69). Variability between cores and whole tumours may also be dependent on the size of the tumour or may represent tumour heterogeneity; Cavaliere et al (2005) found significant differences between core and excisional biopsies for T1 tumours (ER p<0.01, PR p<0.001), but not T2 tumours (61). Furthermore, when core biopsies were compared with paired surgical resections, preoperative chemotherapy resulted in a significant shift to more HR negativity compared to a no-treatment group in one study (ER p=0.02, PR p=0.0005) (57) but not another (59).

Tissue Microarray (TMA)

Seven articles reported comparisons between IHC on TMAs and the reference standard on whole sections (Table 4) (70-76). There was over 95% (median 97%) agreement between the two methods for ER (70-76) and over 85% (median 93.25%) agreement for PR (70-72,74-76). Good agreement was found between multiple cores; one study demonstrated 99% concordance between two cores (73), and another study showed pairwise correlations between five cores were generally greater than 0.75 for both ER and PR (77). The chance of two to ten microarray cores correctly representing whole sections revealed that two cores were comparable to whole sections in more than 95% of ER and PR cases (70). However, another study found that 47% of ER and 28% of PR cases were positive in only one out of three cores (78). Furthermore, there might be missing or inappropriately selected cores. One study found 15/157 (10%) ER and 13/157 (8%) PR cores were missing or had no invasive tumour (71). Likewise, another study found 23/237 (9.7%) had at least one core out of four missing, and 15/237 (6.3%) had a least one core out of four that did not contain tumour tissue (75).

Six retrospective studies using TMAs found HR levels correlated with patient outcomes (Tables 3 or 4). ER and PR status was positively associated with tumour-specific survival when
evaluated from single or multiple TMAs or large sections (p<0.0015) (72). Likewise, disease-specific survival or overall survival was correlated with ER expression using TMAs in three studies (48,79-80). In another study using automated quantitative analysis (AQUA) image analysis, the highest ER score among five cores was most prognostic of disease-specific survival (p=0.0003), whereas for PR, minimum (p=0.0006), maximum (p=0.0001), and average (p=0.0006) scores were equally significant in predicting disease-specific survival (77). One study demonstrated the predictive value of PR in ER positive cases when HRs were tested on TMAs prepared from the retrospective analysis of samples from a randomized controlled trial comparing tamoxifen treatment with no adjuvant treatment in premenopausal women. This study found improved recurrence-free survival on tamoxifen treatment with increasing ER (>10%) and PR (>75%) expression and improved overall survival with tamoxifen treatment with increasing PR (>75%) expression (81).

**Fine-Needle Aspiration (FNA)**

There were sixteen studies that compared IHC on FNA samples to the reference standard of IHC on paraffin sections (Table 4) (82-97). Fourteen articles examining ER (83-91,93-97) showed concordance values greater than 86% (median 94.5%), and ten studies examining PR (84-87,90-91,94-97) demonstrated concordance values greater than 73% (median 80.2%). There were two studies examining the reliability of using FNA. One study showed no change in the ER or PR score between repeat FNAs (98), and the other study showed interobserver reproducibility of 100% for ER and 86.8% for PR using cell blocks prepared by fine-needle aspiration (87). Tafjord et al (2002) found discordant results between pathology reports and consensus scores from re-evaluations could be attributed to cytopathologists accepting technically suboptimal slides for scoring (97). They recommend using at least 100 nuclei when scoring FNAs.

**Frozen Sections**

There were 21 studies that compared the standard of IHC on frozen sections to IHC on paraffin sections (Table 4) (29,36,83,99-116). Among articles examining ER, fourteen articles had concordance values between frozen and paraffin sections over 72% (median 91.4%) (29,36,83,100,102-103,106-109,111,113-115), and two studies had correlations greater than 0.7 (105,110). For those articles investigating PR, six studies demonstrated concordance values over 84% (median 90.8%) (99,101,107-108,112,116) between frozen and paraffin sections, and a correlation of 0.74 was found in one study (104).

Andersen et al (1990) found that ER status derived from frozen samples or DCC were better prognostic indicators of outcome survival (frozen p=0.000016, DCC p=0.000095) and disease-free survival (frozen p=0.00017, DCC p=0.000093), but ER status assessed from paraffin sections were also significant (OS p=0.00084, DFS p=0.017) (106). As well, all three techniques were similar in predicting response to adjuvant endocrine therapy (overall survival: paraffin p=0.0042, frozen p=0.0012, DCC p=0.0027; disease-free survival: paraffin p=0.025, frozen p=0.0039, DCC p=0.031) (106). Pertschuk et al (1996) found ER assessed in paraffin sections with 1D5 to be a better predictor of endocrine response (r, 0.57; 95% confidence interval (CI), 0.51 to 0.73; p<0.001) than ER assessed in frozen samples with H222, although ER levels in frozen samples were significantly correlated with endocrine response (r, 0.34; 95% CI, 0.16 to 0.51; p<0.001) (29). In addition, among discordant cases between paraffin and frozen tissue, paraffin sections correctly predicted endocrine response in 16 of 21 cases (p<0.02) (29).
<table>
<thead>
<tr>
<th>Author</th>
<th>Comparison</th>
<th>N</th>
<th>Patients/Specimens</th>
<th>Kit/Antibody</th>
<th>Results/Conclusion</th>
<th>Blind</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Core biopsy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garimella 2007(69)</td>
<td>Comparison between foci</td>
<td>N=18</td>
<td>patients had multifocal breast cancer</td>
<td>ER 6F11 and PR 312</td>
<td>Each tumour focus has same ER status for all patients. For 16/18 patients, all foci had same PR status.</td>
<td>NR</td>
</tr>
<tr>
<td>Smyczek-Gargya 2002(55)</td>
<td>large-core needle biopsy vs. surgical specimen</td>
<td>n=35</td>
<td>patients received treatment before surgery</td>
<td>PgR (Dianova, 1:50), ER (Immunotech, 1:100)</td>
<td>Core biopsy vs. surgical specimens: ER 97%, PR 89%</td>
<td>no</td>
</tr>
<tr>
<td>Usami 2007(65)</td>
<td>core needle biopsy vs. surgical excision</td>
<td>n=111 women (112 lesions)</td>
<td>no preoperative treatment</td>
<td>ER 6F11 and PR clone 6</td>
<td>Needle core biopsy vs. surgical excision: ER 95%, PR 88%</td>
<td>yes</td>
</tr>
<tr>
<td>Wood 2007(66)</td>
<td>core biopsy vs. excision specimen</td>
<td>n=100; assessable n=95 for ER, n=93 for PR</td>
<td>retrospective review of consecutive patients with core biopsies and surgical specimens; cores reassessed</td>
<td>ER 1D11 and PR PGR636</td>
<td>Core biopsy vs. excisional biopsy: ER 95.8%, PR 90.3%</td>
<td>yes</td>
</tr>
<tr>
<td>Zidan 1997(51)</td>
<td>core biopsy vs. excision biopsy</td>
<td>n=30; assessable n=26</td>
<td>excision taken approx 2 wks following core biopsy from consecutive cases of palpable breast carcinoma</td>
<td>1D5 and NCL-PGR</td>
<td>Core biopsy vs. excisional biopsy: ER 93%, PR 69%</td>
<td>yes</td>
</tr>
<tr>
<td>Al Sarakbi 2005(58)</td>
<td>core needle biopsy vs. surgical excision</td>
<td>n=95 for ER and 93 for PR from 93 patients</td>
<td>retrospectively studies consecutive cases of breast cancer; excision 2-3 wks after biopsy; no preoperative treatment</td>
<td>DAKO mAb</td>
<td>Core biopsy vs. surgical excision: ER 93%, PR 92%</td>
<td>no</td>
</tr>
<tr>
<td>Anania 1997(52)</td>
<td>surgical biopsy vs. large core needle biopsy</td>
<td>n=196, 187 women and 3 men; assessable n=70</td>
<td>retrospectively selected first 70 cases of cancer with T value inferior to 15mm</td>
<td>not given</td>
<td>Core biopsy vs. surgical biopsy: ER r=0.80, PR r=0.75</td>
<td>No</td>
</tr>
<tr>
<td>Arens 2005(59)</td>
<td>core biopsy (CB) vs. excision</td>
<td>n=55</td>
<td>n=25 experimental (received preoperative chemotherapy) CB and whole tumours, n=30 control CB and whole tumours</td>
<td>1D5 and PgR636</td>
<td>change between core biopsy vs. excisional biopsy: ER p=0.85, PR p=0.42</td>
<td>yes</td>
</tr>
<tr>
<td>Badoual 2005(60)</td>
<td>core biopsy vs. surgical specimens</td>
<td>n=468</td>
<td>no preoperative treatment</td>
<td>ER1D5 and PR10A9</td>
<td>Core biopsy vs. surgical excision: ER 90.3%, PR 89.3%</td>
<td>yes</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>Patients/Specimens</td>
<td>Kit/Antibody</td>
<td>Results/Conclusion</td>
<td>Blind</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------------------------------</td>
<td>-------</td>
<td>---------------------------------------------</td>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Burge 2006(63)</td>
<td>core biopsy vs. surgical excision</td>
<td>87</td>
<td>retrospective review; no systemic therapy</td>
<td>ER 6F11 and PR 636</td>
<td>Core biopsy vs. surgical excision: ER 95%, PR 89%</td>
<td>yes</td>
</tr>
<tr>
<td>Cahill 2006(64)</td>
<td>core biopsy vs. surgical specimen</td>
<td>95</td>
<td>consecutive patients over 3 yrs</td>
<td>not given</td>
<td>Core biopsy vs. surgical excision: ER 99%, PR 95%</td>
<td>no</td>
</tr>
<tr>
<td>Cavaliere 2005(61)</td>
<td>core biopsy vs. excision</td>
<td>68</td>
<td>no preoperative treatment</td>
<td>1D5 and PR 1A6</td>
<td>Core biopsy vs. excisional biopsy: ER T1a,b,c 69.04% (p&lt;0.01), T2 50% (NS), T1 + T2 61.76% (p&lt;0.05), T1c 67.74% (P&lt;0.05); PR T1a,b,c 73.80% (p&lt;0.001), T2 61.53% (NS), T1 + T2 69.11% (p&lt;0.01), T1c 74.19% (p&lt;0.01)</td>
<td>yes</td>
</tr>
<tr>
<td>Connor 2002(56)</td>
<td>core needle biopsy vs. surgical specimens</td>
<td>44</td>
<td>no preoperative treatment</td>
<td>not given</td>
<td>Core biopsy vs. surgical specimens: ER 98%, PR 82%</td>
<td>no</td>
</tr>
<tr>
<td>Douglas-Jones 2001(54)</td>
<td>core biopsy vs. excision</td>
<td>51</td>
<td>fresh sections cut from core and excised tumour, and placed on same slide; mean microscopic fields (0.4mm): 18 (range11-27)/core and 16.5 (range 9-25) for excised tumours</td>
<td>6F11</td>
<td>Core biopsy vs. excisional biopsy: ER r=0.876; periphery to centre: whole tumour (t=-6.6, p=0.001), core biopsies (t=-0.3827)</td>
<td>no</td>
</tr>
<tr>
<td>Hodi 2007(67)</td>
<td>core biopsy vs. excision biopsy</td>
<td>379</td>
<td>retrospective review of consecutive patients with core biopsies and surgical specimens</td>
<td>1D5</td>
<td>Core biopsy vs. excisional biopsy: ER 98.8%</td>
<td>no</td>
</tr>
<tr>
<td>Jacobs 1998(53)</td>
<td>core biopsy vs. excision</td>
<td>56; assessable n=54</td>
<td>retrospectively selected consecutive patients with breast cancer; median time to excision 18 days (range, 4-47 days); no preoperative treatment</td>
<td>1D5</td>
<td>Core biopsy vs. excisional biopsy: ER 100%</td>
<td>yes</td>
</tr>
<tr>
<td>Mann 2005(62)</td>
<td>core biopsy vs. surgical specimens</td>
<td>100</td>
<td>retrospective review of consecutive patients with core biopsies and surgical specimens; no systemic therapy</td>
<td>ER 6F11 and PR PgR316</td>
<td>Core biopsy vs. surgical specimens: ER 86%, PR 83%</td>
<td>yes</td>
</tr>
<tr>
<td>Sutela 2008(68)</td>
<td>core biopsy vs. surgical specimen</td>
<td>197</td>
<td>no preoperative treatment; included patients with invasive breast carcinoma in at least two containers (1st</td>
<td>ER 6F11 and PR M3569</td>
<td>Core biopsy vs. surgical specimens: ER 83%, PR 88%; across 3 containers: ER 100%, PR 85%</td>
<td>no</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>Patients/Specimens</td>
<td>Kit/Antibody</td>
<td>Results/Conclusion</td>
<td>Blind</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------------------------------------</td>
<td>------</td>
<td>------------------------------------------------------------------------------------</td>
<td>--------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Taucher 2003(57)</td>
<td>core biopsy vs. surgical specimen</td>
<td>n=722</td>
<td>core in A, 2B3 in B, rest in C and in surgical sample</td>
<td>H222 and KD68</td>
<td>Core biopsy vs. surgical specimens: no treatment: ER 91%, PR 80%; preoperative treatment: ER 86%, PR 79%; significant shift from ER+ to ER- and PR+ to PR- due to primary chemotherapy (p=0.02, p=0.0005 respectively)</td>
<td>yes</td>
</tr>
<tr>
<td>Moeder 2007(77)</td>
<td>comparison between cores</td>
<td>n=669</td>
<td>archival breast tissue; f/u median 8.3 yr (range 2.4 - 41.5); 5 cores/case; used AQUA; used 10-yr Kaplan-Meier survival plots; cutpoints ER 10, PR 13</td>
<td>ER 1D5 and PR 636</td>
<td>In general, pairwise correlations between 5 cores were greater than 0.75 for ER and PR</td>
<td>no</td>
</tr>
<tr>
<td>Henriksen 2007(76)</td>
<td>whole sections vs. TMA</td>
<td>n=70</td>
<td>2mm cores from archival primary tumour tissue from RCT phase III trial that were HR pos or unknown</td>
<td>ER: 1D5 and PR: 16</td>
<td>cores vs. whole sections: 96% (ER) and 93% (PR)</td>
<td>no</td>
</tr>
<tr>
<td>Parker 2002(73)</td>
<td>Intra- or interlab reliability; TMA vs. whole section</td>
<td>n=58</td>
<td>TMA derived from 29 invasive breast cancers retrospectively; 5 labs (4 community, 1 university) evaluated TMAs</td>
<td>Lab 4: 1D5; other labs: 6F11</td>
<td>ER status agreement across 5 labs: k=0.84; lab 4 agreement with others: k=0.64-0.77; interobserver agreement for lab 4 across 5 observers: overall k=0.76; for ER status, TMA (2 cores/case) vs. whole sections: 96%, concordance between 2 cores: 99%</td>
<td>Yes at university lab</td>
</tr>
<tr>
<td>Quraishi 2007(78)</td>
<td>number of cores showing ER/PR expression using IHC; evaluating tissue microarrays (TMAs) using IHC</td>
<td>n=32</td>
<td>3 1mm cores/case from archival tissue; 2 normal ductal epithelium, 8 DCIS, 19 invasive ductal carcinoma, 3 metastatic ductal carcinoma</td>
<td>mouse monoclonal antibodies to human ER and PR</td>
<td>Many tumours showed ER/PR expression in only 1 out of 3 cores; 15/32 (47%) for ER and 9/32 (28%) for PR showed 1 positive core</td>
<td>no</td>
</tr>
<tr>
<td>Sapino 2006(75)</td>
<td>tissue microarray vs. full section</td>
<td>36 TMAs prepared from 237 cases; n=20 first cases for TMA vs. sections</td>
<td>consecutive cases of breast cancer; excluded tumour &lt;1cm</td>
<td>ER: 6F11 and PR: 1A6</td>
<td>4 cores/case yielded 100% (ER) and 85% (PR) concordance with full sections; 2/22 ER/PR negative TMA cases were positive on full sections; 23/237 (9.7%) cases had only 1 core inappropriately selected and 2/237 (0.8%) cases had 2 inappropriate cores; 6/237 (2.5%) cases had 3 or 4 cores missing</td>
<td>no</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>Patients/Specimens</td>
<td>Kit/Antibody</td>
<td>Results/Conclusion</td>
<td>Blind</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------------------------</td>
<td>-------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>---------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Torhorst 2001(72)</td>
<td>TMA vs. large sections</td>
<td>n=611; included n=553 in analysis</td>
<td>retrospective review of patients treated for primary breast cancer f/u 65.8 months (range 1 - 151); 4 cores/case</td>
<td>ER ID5 and PR 1A6</td>
<td>The frequency of ER positivity was virtually the same when measured from a single TMA section as compared to entire sections; For PR, the concordance was slightly lower (88%) and three samples from each tumour were required to achieve the same level of positivity as large section analyses; ER and PR status positively associated with tumour-specific survival when evaluated from single or multiple TMAs or large sections (p&lt;0.0015)</td>
<td>NR</td>
</tr>
<tr>
<td>Zhang 2003(74)</td>
<td>using TMA vs. full sections for IHC</td>
<td>n=97, assessable n=93</td>
<td>breast cancer samples</td>
<td>ER-1D5 and NCL-PGR, 1A6</td>
<td>With single cores, concordance with full sections was 97% (ER) and 98% (PR)</td>
<td>no</td>
</tr>
<tr>
<td>Camp 2000(70)</td>
<td>microarray vs. whole-section</td>
<td>n=38, samples assessable ER: n=36, PR: n=35</td>
<td>sequential selection of cases with adequate tissue from 1932 to 1999; 10 cores samples/case</td>
<td>monoclonal antibodies to ER and PR: DAKO</td>
<td>35/36 ER and 34/35 PR average stain/case agreed with whole section; 3/8 ER and 2/4 PR had discrepant results between cores taken from periphery vs. centre; 2 disks were comparable to whole sections in more than 95% of ER and PR cases; archival tissue up to 68 yrs old (PR) and 58 yrs old (ER) was suitable for IHC</td>
<td>no</td>
</tr>
<tr>
<td>Gillett 2000(71)</td>
<td>microarray vs. whole-section</td>
<td>n=253, samples assessable n=157</td>
<td>archival primary breast cancer, no DCIS cases</td>
<td>ER: 1D5 and PR: PR88</td>
<td>15/157 (10%) ER and 13/157 (8%) PR cores were missing or had no invasive tumour; sections vs. multicores: ER 95%, PR 93.5%; discrepant samples generally had lower status in cores</td>
<td>no</td>
</tr>
<tr>
<td>Camp 2002(79)</td>
<td>Compared survival of patients with tumours with high (top 25%) versus low (bottom 25%) ER expression</td>
<td>n=340</td>
<td>constructed TMAs from 345 FFPE cases of node-positive breast-carcinoma; f/u 60 months</td>
<td>1D5; used AQUA image analysis</td>
<td>prognostic information: AQUA RR=2.44 (p=0.0003), manual RR=2.06 (p=0.002)</td>
<td>unclear</td>
</tr>
<tr>
<td>McCabe 2005(80)</td>
<td>various antibody concentrations on outcome survival</td>
<td>n=250</td>
<td>TMAs constructed from samples from patients with invasive BC; follow-up time ranged from 2.4 months to 41.5 years (median = 8.3 years)</td>
<td>ID5 diluted 1:100 or 1:1000; used AQUA image analysis</td>
<td>Using 1D5 (1:100 or 1:1000), ER expression associated with increased survival regardless of antibody concentration; (1: 100); Median survival of patients with high ER at 5 yr was 88.7% (78.3% - 99.2%), at 10 yr was 73.2% (58.1% - 88.3%), and at 20 yr was 63.1% (46.3% - 79.9%), Median survival of patients</td>
<td>unclear</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>Patients/Specimens</td>
<td>Kit/Antibody</td>
<td>Results/Conclusion</td>
<td>Blind</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
<td>---</td>
<td>--------------------</td>
<td>-------------</td>
<td>--------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Stendahl 2006(81)</td>
<td>IHC - retest in TMA for further subclassification</td>
<td>500</td>
<td>premenopausal patients; 474 ER assessed, 456 PR assessed</td>
<td>6F11 and PR16; fractions of positive tumour cell nuclei were subgrouped as 0 (0-10%), 1 (11-50%), 2 (51-75%), and 3 (76-100%)</td>
<td>Gradually increasing TAM effect with tumours &gt;10% ER+ nuclei; PR &gt; 75% predicted a significant treatment effect for RFS (RR, 0.42 [0.25-0.70], p=0.001) and OS (RR, 0.49 [0.28-0.84], p=0.010), irrespective of ER status</td>
<td>yes</td>
</tr>
<tr>
<td>Nizzoli 2000(85)</td>
<td>IHC from surgical specimens vs. ICC from FNA</td>
<td>104</td>
<td>patients with primary breast carcinoma; IHC on archival tissue; FNA fixed in methanol and acetone</td>
<td>IHC: ER-1D5 and PgR-hPRa2/hRa3; ICC: ER-H222 and PgR-KD68</td>
<td>FFPE vs. FNA: ER 89%, PR 78%</td>
<td>No</td>
</tr>
<tr>
<td>Pinder 1995(82)</td>
<td>FNA using cytoblocks vs. resected tissue sections</td>
<td>n=50; ER: 34</td>
<td>FNA specimens prepared using a Cytoblock technique and fixed in formal saline; IHC performed on parallel tissue sections</td>
<td>Dako anti-estrogen</td>
<td>FFPE vs. FNA: r=0.806</td>
<td>no</td>
</tr>
<tr>
<td>Vesoulis 2004(89)</td>
<td>FNA vs. surgical specimen</td>
<td>n=118</td>
<td>retrospective review of adenocarcinomas</td>
<td>ER 6F11</td>
<td>FFPE vs. FNA: 98.3%</td>
<td>yes</td>
</tr>
<tr>
<td>Schmitt 1995(83)</td>
<td>fine-needle aspiration (FNA)-1D5 vs. formalin-fixed (FF)-1D5 vs. frozen sections (FS)</td>
<td>n=31</td>
<td>specimens of primary breast cancer; fixed in absolute ethanol; 1-28 days between cytologic sampling and surgical excision</td>
<td>FNA and FF: 1D5; FS: H222</td>
<td>For ER, frozen vs. paraffin: 100%; frozen vs. FNA: 96.8%; FF vs. FNA: 96.8%</td>
<td>no</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N/Specimens</td>
<td>Patients/Specimens</td>
<td>Kit/Antibody</td>
<td>Results/Conclusion</td>
<td>Blind</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------------</td>
<td>-------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>---------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Sianesi 2000</td>
<td>IHC vs. ICC</td>
<td>37; 128 patients over 5 yr period; FNA fixed in methanol and acetone</td>
<td>ER H222, PR KD68</td>
<td>FFPE vs. FNA: ER 89.6%, PR 76.9%</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>Zoppi 2002</td>
<td>FNA vs. FFPE</td>
<td>n=101</td>
<td>FFPE from biopsies; FNA fixed in alcohol</td>
<td>1D5 and 1A6</td>
<td>FNA vs. FFPE: ER 94%, PR 73%</td>
<td>no</td>
</tr>
<tr>
<td>Bajetta 1998</td>
<td>FNA vs. surgical specimen</td>
<td>n=32</td>
<td>FNA and FFPE from post-TAM samples; FNA fixed in formalin</td>
<td>H222 and KD68</td>
<td>FNA vs. FFPE: ER rho=0.66, p&lt;0.001, PR rho=0.84, p&lt;0.001</td>
<td>no</td>
</tr>
<tr>
<td>Briffod 2000</td>
<td>cytoblocks vs. tissue samples</td>
<td>n=55</td>
<td>cytoblocks prepared by fine-needle cytopuncture, without aspiration; fixed in neutral formal; interobserver reproducibility of cytoblocks results assessed by 2 independent observers</td>
<td>ER 1D5 and PR 1A6</td>
<td>FFPE vs. FNA: ER 96.3%, PR 81.8%; interobserver reproducibility: ER 100%, PR 93%; primary tumour vs. node metastases: ER 94.7%, PR 86.8%</td>
<td>no</td>
</tr>
<tr>
<td>Cano 2003</td>
<td>FNA-SP1 vs. FF-6F11 vs. FF-SP1; FF-1A6 vs. FF-SP2</td>
<td>39 women and 1 man</td>
<td>patients had primary breast cancer, no antigen retrieval with SP1 and SP2; FNA fixed in absolute ethanol</td>
<td>FNA: ER SP1; surgical specimens: ER SP1 and PR SP2, ER 6F11 and PR 1A6</td>
<td>FFPE vs. FNA: ER 95%</td>
<td>no</td>
</tr>
<tr>
<td>Jayaram 2005</td>
<td>FNA vs. FFPE</td>
<td>n=77</td>
<td>prospectively collected samples from palpable breast masses; FNA fixed in methanol</td>
<td>Dako (Glostrup, Denmark)</td>
<td>FNA vs. FFPE: ER 86%</td>
<td>no</td>
</tr>
<tr>
<td>Klorin 2003</td>
<td>FNA vs. FFPE</td>
<td>n=83</td>
<td>prospectively collected samples from women who were suspected of having malignant tumour; FNA fixed in absolute ethanol</td>
<td>1D5 and 1A6</td>
<td>FNA vs. FFPE: ER 91.6%, PR 90.4%</td>
<td>no</td>
</tr>
<tr>
<td>Konofaos 2006</td>
<td>ICC on FNAs vs. IHC on FFPEs</td>
<td>n=119</td>
<td>used ThinPrep processed smears from FNA; FNA fixed in Cytolyt; consecutive breast cancer samples; no hormone treatment at least 2 months prior to FNA, no</td>
<td>PR M3569 and ER M7047</td>
<td>FFPE vs. FNA: ER 99.15%, PR 96.64%</td>
<td>yes</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>Patients/Specimens</td>
<td>Kit/Antibody</td>
<td>Results/Conclusion</td>
<td>Blind</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------------</td>
<td>---------</td>
<td>------------------------------------------------------------------------------------</td>
<td>--------------</td>
<td>------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Leung 1999(84)</td>
<td>surgical specimens vs. FNA</td>
<td>n=142</td>
<td>patients with breast carcinoma; FNA postfixed with Cytospray</td>
<td>ER-1D5, PR-1A6</td>
<td>Surgical specimens vs. FNA: ER 98%, PR 89%</td>
<td>no</td>
</tr>
<tr>
<td>Makris 1997(95)</td>
<td>FNA vs. FFPE</td>
<td>n=50</td>
<td>patients with primary breast carcinoma; FNA fixed in methanol and acetone</td>
<td>H222 and KD68</td>
<td>FNA vs. FFPE: ER 91.5%, PR 75.5%</td>
<td>no</td>
</tr>
<tr>
<td>Makris 1999(98)</td>
<td>repeat FNAs</td>
<td>non-treatment, control group, n=20 repeat FNAs; assessable ER: n=16 pairs, PR n=17 pairs</td>
<td>2-week interval between repeat FNAs among consecutive patients with palpable lumps; fixed in methanol and acetone</td>
<td>ER: H222 and PR: KD68</td>
<td>No change in ER or PR score between repeat FNA</td>
<td>no</td>
</tr>
<tr>
<td>Suthipintawong 1997(96)</td>
<td>FNA vs. FFPE</td>
<td>n=14</td>
<td>malignant breast tumours; FNA fixed in 10% buffered formalin</td>
<td>1D5 and KD68</td>
<td>FNA vs. FFPE: ER 92.9%, PR 78.6%</td>
<td>no</td>
</tr>
<tr>
<td>Tafjord 2002(97)</td>
<td>FNA vs. FFPE; pathology reporting vs. consensus re-evaluation</td>
<td>pathology reports: ER n=70, PR n=59; consensus re-evaluation: ER n=63, PR n=45</td>
<td>archival samples, FNA fixed in acetone/formalin</td>
<td>6F11 and 1A6</td>
<td>FNA vs. FFPE: pathology reports: ER 88.6%, PR 64.4%, consensus re-evaluation: ER 96.8%, PR 91.1%; cytopathologists accepted technically suboptimal slides as valid for scoring; recommend using at least 100 nuclei for evaluation</td>
<td>no</td>
</tr>
<tr>
<td>Frozen sections</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muller-Holzner 1993(99)</td>
<td>frozen vs. paraffin</td>
<td>n=34</td>
<td>invasive ductal breast carcinoma</td>
<td>KD 68</td>
<td>frozen vs. paraffin: PR 94%</td>
<td>yes</td>
</tr>
<tr>
<td>Nedergaard 1995(100)</td>
<td>FFPE (ID5) vs. frozen (H222) vs. FFPE (H222)</td>
<td>n=101 FFPE, 83/101 = frozen, 51/83 = DCC</td>
<td>primary breast cancer where fresh frozen tissue had been analysed</td>
<td>frozen: H222; FFPE: ID5/H222</td>
<td>For ER, frozen vs. paraffin (H222): 80 %, frozen vs. paraffin (1D5): 89%</td>
<td>no</td>
</tr>
<tr>
<td>Ozzello 1991(101)</td>
<td>PR frozen specimens, paraffin-</td>
<td>125 samples; FFPE: 52 tumours;</td>
<td>primary breast cancer samples</td>
<td>PR: JZB39 and KD68</td>
<td>For PR, DCC vs. frozen: 77.2%; frozen vs. paraffin: 85.4%</td>
<td>no</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>Patients/Specimens</td>
<td>Kit/Antibody</td>
<td>Results/Conclusion</td>
<td>Blind</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------------------------</td>
<td>-------</td>
<td>---------------------------------------------------------</td>
<td>--------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>embedded specimens, stored imprint and cryostat sections</td>
<td>n=41 paraffin vs. frozen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paterson 1990(102)</td>
<td>IHC (method 1: frozen section, method 2: Carson's buffered formalin, method 3: methacarn fixative) vs. DCC</td>
<td>n=70; assessed n=32 (frozen vs. formalin)</td>
<td>prospective; received fresh tissue</td>
<td>Monoclonal rat IgG to human ER (Abbott)</td>
<td>For ER, DCC vs. frozen: 95%, r = 0.87; frozen vs. formalin: 78%</td>
<td>yes</td>
</tr>
<tr>
<td>Pertschuk 1996(29)</td>
<td>frozen vs. paraffin; includes patient outcomes</td>
<td>n=74</td>
<td>patients received endocrine therapy for a minimum of 3 months and any favourable response was maintained for at least 3 months</td>
<td>frozen: H222, paraffin: 1D5</td>
<td>frozen vs. paraffin: ER 72%; paraffin best predictor of clinical endocrine response (r=0.57; 95% CI 0.51-0.73, p&lt;0.001), frozen a significant predictor (r=0.34, 95% CI 0.16-0.51, p&lt;0.001); among 21 discordant cases, paraffin correctly predicted endocrine response in 16 cases (p&lt;0.02)</td>
<td>yes</td>
</tr>
<tr>
<td>Raymond and Leong 1990(103)</td>
<td>ICC: frozen vs. paraffin sections; ICC vs. DCC</td>
<td>n=32 (includes 1 male)</td>
<td>prospective; received fresh tissue; 30 infiltrating ductal carcinoma, 1 infiltrating lobular carcinoma, 1 DCIS</td>
<td>ER-ICA/H222</td>
<td>For ER, frozen vs. paraffin: 94%</td>
<td>No</td>
</tr>
<tr>
<td>Scharl 1990(104)</td>
<td>FFPE vs. snap frozen; FFPE vs. DCC</td>
<td>n=170 FFPE vs. DCC; n=82 FFPE vs. snap frozen</td>
<td>retrospectively obtained FFPE blocks of tumour</td>
<td>PR antibody M60-10 (mPR1)</td>
<td>For PR, frozen vs. FFPE: r = 0.74</td>
<td>No</td>
</tr>
<tr>
<td>Snead 1993(105)</td>
<td>ICC vs. IHC</td>
<td>n=94</td>
<td>consecutive patients with advanced metastatic breast cancer treated with hormone therapy as primary treatment</td>
<td>ICC and IHC: Abbott H222</td>
<td>For ER, frozen vs. paraffin: r = 0.8</td>
<td>No</td>
</tr>
<tr>
<td>Andersen 1990(106)</td>
<td>ICC and DCC and ER-PAR (paraffin embedded)</td>
<td>n=130; n=109 frozen vs. paraffin, n=116 DCC vs. frozen, n=115 DCC vs. paraffin</td>
<td>human breast cancer specimens from postmenopausal (&gt;5 years) high-risk (defined as a tumour &gt; 5 cm in diameter and/or positive lymph nodes and/or skin or fascial invasion) patients; f/u median 75</td>
<td>H222SP2γ</td>
<td>For ER, DCC vs. frozen: 87%; frozen vs. paraffin: 82%, DCC vs. paraffin: 75%; ER status using paraffin or frozen were predictive of DFS (p=0.017 and p=0.00017 respectively) and OS (p=0.00084 and 0.000016 respectively); for DCC DFS p=0.00093, OS p=0.0000095; all three assays could predict response to endocrine therapy (radiotherapy vs. radiotherapy and...</td>
<td>no</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>Patients/Specimens</td>
<td>Kit/Antibody</td>
<td>Results/Conclusion</td>
<td>Blind</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------------------------</td>
<td>------</td>
<td>--------------------------------------------------------</td>
<td>---------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Cavaliere</td>
<td>frozen section (FS) vs. paraffin section (PS) vs. EIA</td>
<td>n=115</td>
<td>fresh or refrigerated (1-4hrs) primary breast cancers and one recurrent sample; used CAS image analysis</td>
<td>FS and EIA: ER H222 and PR KD68; PS: ER 1D5 and PR 1A6</td>
<td>For ER, frozen vs. paraffin: 83.4%; For PR, frozen vs. paraffin: 87.8%</td>
<td>Yes</td>
</tr>
<tr>
<td>de Mascarel</td>
<td>frozen vs. paraffin sections</td>
<td>n=103</td>
<td>consecutive cases of distant-metastasis-free primary invasive ductal cancer</td>
<td>IHC: ID5; frozen: H222</td>
<td>For ER, frozen vs. paraffin: 88.3%</td>
<td>NR</td>
</tr>
<tr>
<td>Elias</td>
<td>EIA vs. frozen sections</td>
<td>n=48</td>
<td>breast cancer tissue</td>
<td>ER H222 and PR KD68</td>
<td>For ER, EIA vs. frozen: 93.8%; For PR, EIA vs. frozen: 91.7%, frozen vs. paraffin: 93.8%</td>
<td>yes</td>
</tr>
<tr>
<td>Fee</td>
<td>frozen vs. paraffin</td>
<td>n=100; assessable n=97</td>
<td>archival tissue of infiltrating ductal and lobular breast cancer that had frozen sections analyzed</td>
<td>frozen: H222; paraffin: ER-1D5</td>
<td>For ER, frozen (H222) vs. paraffin (1D5): 100% (10% cutoff), r = 0.87</td>
<td>yes</td>
</tr>
<tr>
<td>Goulding</td>
<td>Dako ID5 FFPE vs. Abbott H222 FFPE vs. Abbott H222 frozen</td>
<td>n=90</td>
<td>case series of primary breast cancers treated with hormonal therapy</td>
<td>ID5 or H222</td>
<td>For ER, frozen (H222) vs. paraffin (1D5): r = 0.7 using H-score, r = 0.6 using % positivity</td>
<td>yes</td>
</tr>
<tr>
<td>Ibarra</td>
<td>H222 frozen vs. antibody cocktail vs. each antibody in cocktail; antibody cocktail normally fixed vs. overfixed</td>
<td>n=56</td>
<td>consecutive postmenopausal patients; samples fixed at RT in 10% phosphate-saline-buffered formalin for 90 min or 1, 3, or 7 days</td>
<td>ER mAb cocktail: 1D5 and LH1</td>
<td>H222 (frozen) vs. LH1: concordance 51/56 (91%), sensitivity 88%; H222 (frozen) vs. 1D5: concordance 53/56 (95%), sensitivity 93%; H222 (frozen) vs. (LH1 &amp; 1D5); concordance 55/56 (98%), sensitivity 98%</td>
<td>no</td>
</tr>
<tr>
<td>Masood</td>
<td>frozen vs. DCC, H222 on FFPE vs. H222 on frozen sections</td>
<td>62</td>
<td>archival breast cancer tissue that had DCC and frozen section results</td>
<td>H222</td>
<td>For ER, DCC vs. frozen: 94%; frozen vs. paraffin: 85%</td>
<td>no</td>
</tr>
<tr>
<td>Masood</td>
<td>DCC vs. frozen vs. FFPE</td>
<td>n=116</td>
<td>archival breast cancer tissue that had DCC and frozen results</td>
<td>PR KD68</td>
<td>For PR, DCC vs. frozen: 90%; frozen vs. paraffin: 84%</td>
<td>No</td>
</tr>
<tr>
<td>Miller</td>
<td>DCC vs. FFPE (D75) vs. frozen (D75) vs. FFPE</td>
<td>n=67, FFPE (D75)</td>
<td>fresh breast cancer biopsies</td>
<td>D75 or H222</td>
<td>For ER, frozen (D75) vs. paraffin (D75): 100%</td>
<td>no</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>Patients/Specimens</td>
<td>Kit/Antibody</td>
<td>Results/Conclusion</td>
<td>Blind</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------------------------------</td>
<td>-------------------</td>
<td>-------------------------------------------------</td>
<td>--------------</td>
<td>--------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td></td>
<td>(D75) vs. FFPE (H222)</td>
<td>vs. frozen (D75)</td>
<td>n=59, FFPE (H222) n=40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stylianodou, Papadimitriou 1992(115)</td>
<td>frozen vs. EIA; frozen vs. paraffin sections</td>
<td>n=86; included 83 frozen (72 EIA and frozen), 68 FFPE</td>
<td>breast cancer specimens obtained at surgery</td>
<td>H222</td>
<td>For ER, frozen vs. paraffin: 95.38%; EIA vs. frozen: 84.72%</td>
<td>yes</td>
</tr>
<tr>
<td>Von Boguslawsky 1994(116)</td>
<td>microwave in citrate vs. nothing</td>
<td>n=25</td>
<td>unfixed breast cancer tissue</td>
<td>PR; KD68</td>
<td>For PR, frozen vs. paraffin (with antigen retrieval): 24/25 (96%)</td>
<td>yes</td>
</tr>
</tbody>
</table>

**Abbreviations:** AQUA, automated quantitative analysis; BC, breast cancer; CI, confidence interval; DCC, dextran-coated charcoal; DCIS, ductal carcinoma in situ; DFS, disease-free survival; EIA, enzyme immunoassay; ER, estrogen receptor; FFPE, formalin-fixed paraffin-embedded; FNA, fine needle aspiration; HR, hormone receptor; ICC, immunocytochemistry; IHC, immunohistochemistry; k, kappa; NR, not reported; NS, not significant; OS, overall survival; PR, progesterone receptor; RCT, randomized controlled trial; RFS, recurrence free survival; RT, room temperature; TAM, tamoxifen; TMA, tissue microarray; vs., versus.
Fixation Issues

Fixation type

Comparative studies examining fixation type or fixation time can be found in Table 5. Two studies, using paraffin sections or TMAs, suggest that 10% neutral buffered formalin as a fixative achieves the best results and that ethanol should be avoided as a fixative (117-118). Four studies found other fixatives were associated with formalin (102,119-121). Methacarn was concordant with formalin (100%, n=6) (120) and showed similar rates of ER positivity to formalin (methacarn: 53%, formalin: 59%) (102). An alcohol-based fixative (100%, n=62) (121) and RCL2®, a new formalin-free fixative suitable for the molecular analysis of breast tissue, (100%, n=6) (120) were concordant with formalin. Similarly, a laboratory using acetic acid as a fixative had comparable proportions of positive cases to laboratories using formalin (119).

Fixation time

Two studies comparing various fixation types and duration using paraffin sections or TMAs found that fixation for 24 hours, immediately after surgery, in neutral buffered formalin (NBF) achieved the most intense and highest percentage of positively stained HR cells (117-118). With paraffin sections, poor results were observed with freezing plus cryosectioning before fixation and the worst results with a 12-hour delay in the onset of fixation (117). Similarly, another study showed poorer ER concordance with frozen tissue when there was a delay in fixing paraffin sections of up to four hours compared to a few minutes (107).

With proper antigen retrieval, tissue that has been fixed for several days will continue to immunoreact for ER and PR (104,111,117,122). Underfixation is more critical than overfixation, and one study suggests that the minimum fixation time for reliable IHC ER results is six to eight hours, regardless of the type or size of the specimen (123). Likewise, two studies mentioned previously found that tissue fixed for six hours in NBF was 92% and 80% ER immunoreactive relative to 24 hours in NBF for paraffin sections and TMAs, respectively (117-118). Furthermore, another study found that fixation for as little as four hours could achieve similar results to tissues fixed for longer periods (124).
### Table 5. Studies that compared different fixation types or times.

<table>
<thead>
<tr>
<th>Author</th>
<th>Comparison</th>
<th>N</th>
<th>Fixation</th>
<th>Kit/Antibody</th>
<th>Antigen Retrieval</th>
<th>Results/Conclusion</th>
<th>Blind</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nassiri</td>
<td>alcohol-based fixative, automated formalin-free microwave processing vs. formalin fixative, VIP processing</td>
<td>paired tissue sections from n=62 specimens</td>
<td>alcohol-based fixative: universal molecular fixative, containing methanol and polyethylene glycol; formalin fixative: 10% neutral buffered formalin; all within 30 min; duration: less than 24 hours for 32 cases, 24-48 hours for 18 cases, 48-72 hours for 12 cases</td>
<td>pharmDX kit (Dako): per kit instructions or 1DS</td>
<td>Monoclonal rat IgG to human ER (Abbott)</td>
<td>Formalin fixation for 48 hrs resulted in PR detection but staining was less intense than cryostat sections</td>
<td>yes</td>
</tr>
<tr>
<td>Paterson</td>
<td>IHC (method 1: frozen section, method 2: Carson's buffered formalin, method 3: methacarn fixative) vs. DCC</td>
<td>n=70; assessed: method 1 n=42, method 2 n=46, method 3 n=47</td>
<td>Method 1: fixation on slides in 3-7% formaldehyde: PBS, methanol, acetone; Method 2: Carson's buffered formalin; Method 3: methacarn; duration: Method 1: NR; Method 2: 24-48 hours; Method 3: 4-24 hours</td>
<td>Monoclonal rat IgG to human ER (Abbott)</td>
<td>Methods 2 &amp; 3: DNase pretreatment</td>
<td>Alcohol based molecular fixative vs. formalin r=0.83, 100% concordant</td>
<td>yes</td>
</tr>
<tr>
<td>Scharl</td>
<td>FFPE vs. snap frozen</td>
<td>170; 82 additional snap frozen cases</td>
<td>2 fresh, PR-rich breast cancer biopsies were divided into four: 1 was snap frozen and 3 were fixed in 5% formalin at RT for 12, 24, and 48 h; In each case two trypsinized paraffin sections and if available two fixed cryostat sections were incubated for IHC</td>
<td>Mi 60-10 (mPR1)</td>
<td>FFPE: digested with trypsin</td>
<td>Formalin fixation for 48 hrs resulted in PR detection but staining was less intense than cryostat sections</td>
<td>no</td>
</tr>
<tr>
<td>Arber</td>
<td>effect of various lengths of fixation on ER and PR IHC results</td>
<td>33</td>
<td>10% neutral buffered formalin; duration 24 hours to up to 154 days (range of maximum fixation, 7 to 154 days) with a mean/median time of maximum formalin fixation of 53/42 days</td>
<td>1D5 and PR1A6</td>
<td>HIER buffer (pH 5.5-5.7; Ventana)</td>
<td>ER/PR staining will continue to immunoreact with antigen retrieval for up to 57 days</td>
<td>no</td>
</tr>
<tr>
<td>Cavaliere</td>
<td>frozen section (FS) vs. paraffin section (PS)</td>
<td>n=115</td>
<td>subdivided into 47 received fresh (5-10 min) and 68 refrigerated (1-4 hrs) samples from nearby hospital; primary breast cancers and one recurrent sample; used CAS image analysis</td>
<td>FS: ER H222 and PR KD68; PS: ER 1D5 and PR 1A6</td>
<td>microwaved in citrate buffer</td>
<td>For ER, FS vs. fresh: 91.5%, FS vs. refrigerated: 77.9%; For PR, FS vs. fresh: 93.6%, FS vs. refrigerated: 83.9%</td>
<td>yes</td>
</tr>
<tr>
<td>Delfour</td>
<td>neutral</td>
<td>n=6</td>
<td>NBF stored at RT considered gold</td>
<td>ER 6F11 and water bath,</td>
<td>100% between Methacarn,</td>
<td></td>
<td>no</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>Fixation</td>
<td>Kit/Antibody</td>
<td>Antigen Retrieval</td>
<td>Results/Conclusion</td>
<td>Blind</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------------------------------------------------</td>
<td>---------</td>
<td>--------------------------------------------------------------------------</td>
<td>--------------</td>
<td>--------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>2006(120)</td>
<td>buffered formaldehyde 4% (NBF) vs. methacarn vs. RCL2 fixatives</td>
<td></td>
<td>standard; two fixatives tested in this report were a methacarn solution (60% (v/v) methanol, 30% chloroform, and 10% glacial acetic acid) and RCL2 solution stored at 4°C; onset ~ 15min; duration: NBF: 24 hrs; methacarn or RCL2: overnight</td>
<td>PR: 636</td>
<td>using citrate or EDTA buffers, when needed</td>
<td>RCL2 and NBF for ER (5 pos 1 neg) and PR (4 pos 2 neg)</td>
<td></td>
</tr>
<tr>
<td>Goldstein 2003(123)</td>
<td>study 1 (prospective): 3 vs. 6 vs. 8 vs. 10 vs. 12 hrs vs. 1 vs. 2 vs. 7 days of fixation; study 2 (retrospective): needle core biopsy vs. resection specimens</td>
<td>24 study 1: 10% neutral-buffered formalin; tissue temporarily stored in 100% cold ethanol</td>
<td>1D5</td>
<td>25 or 40 minutes with EDTA buffer</td>
<td>mean Q scores significantly different for 3, 6, and 8 hrs (p&lt;0.001); Q score plateau after 6 to 8 hrs; mean fixation time for needle core biopsy and resection specimens with similar results was 6.3 hr (p=0.01); minimum fixation time for reliable IHC ER results is 6 to 8 hrs, regardless of type or size of specimen</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>Ibarra 1995(111)</td>
<td>antibody cocktail vs. each antibody in cocktail; antibody cocktail normally fixed vs. overfixed</td>
<td>56</td>
<td>fixed at RT in 10% phosphate-saline-buffered formalin for 90 min or 1, 3, or 7 days</td>
<td>ER mAb cocktail: 1D5 and LH1</td>
<td>microwaved in citrate buffer; time prolonged up to a minimum of 20 min for overfixed tissue</td>
<td>Number of ER positive cases decreased by 20, 26, and 35% after 7 days vs. 90 min of fixation with the ER- cocktail, ER-1D5, and ER-LH1, respectively; antigen retrieval could recover some of the loss (1D5 87%, LH1 81%, cocktail 97%)</td>
<td>no</td>
</tr>
<tr>
<td>Jensen 1995(124)</td>
<td>different fixation times</td>
<td>25, 9 were allowed to fix at different times</td>
<td>10% formalin within 15 min; duration: two to four hours, four to 24 hours, 24 to 48 hours, and 48 to 166 hours, randomly assigned</td>
<td>ER (Dako, Glostrup, Denmark)</td>
<td>microwave oven</td>
<td>No correlation between storage intervals (2-4 hrs, 4-24 hrs, 24-48 hrs, 48-166 hrs) and mean % of ER</td>
<td>no</td>
</tr>
<tr>
<td>Lee 2002(119)</td>
<td>comparing different fixation and processing methods from 7 labs and tested</td>
<td>n=447 from 11 different hospitals; assessed: n=420 cases from seven hospitals</td>
<td>1 Bayer VIP 2000: 10% formal saline; 2 Leica TP 1050:10% neutral buffered formal saline; 3 Shandon hypercenter XP: 10% formal saline; 4 Leica TP1050: 10% formal saline; 5 Shandon Pathcentre: 10% neutral buffered</td>
<td>ER: 6F11; used automated immunostainer</td>
<td>microwaving</td>
<td>Formal saline or neutral buffered formal saline or acetic were comparable; no significant differences in proportion of positive cases in each category (neg, mod pos, strongly pos) across 7</td>
<td>yes</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>Fixation</td>
<td>Kit/Antibody</td>
<td>Antigen Retrieval</td>
<td>Results/Conclusion</td>
<td>Blind</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
<td>---</td>
<td>----------</td>
<td>--------------</td>
<td>-------------------</td>
<td>--------------------</td>
<td>-------</td>
</tr>
<tr>
<td>W. R. von Wasielewski 1998(117)</td>
<td>various fixation types and times</td>
<td>3 cases divided into 11 portions variously fixed and processed</td>
<td>fixed in 10% formalin or ethanol for various durations</td>
<td>ER: 1D5 and 6F11 PR: 1A6 and polyclonal antiserum (PR, Dako)</td>
<td>heat-induced epitope retrieval (microwaved in citrate buffer) with and without a tyramine amplification technique</td>
<td>Best result with 10% NBF, worst results with ethanol fixation; Best results with 24 hr incubation for ER and PR; fixation for up to 4 days had little influence on ER or PR detection; poor results with freezing plus cryosectioning before fixation, especially for ER; worst results with 12 hr delay in onset of fixation for ER and PR, 6 hrs was 92% and 85.3% immunoreactive relative to 24 hrs in NBF for ER and PR, respectively</td>
<td>yes</td>
</tr>
<tr>
<td>von Wasielewski 2002(118)</td>
<td>various fixation types and times using TMAs</td>
<td>6 cases of breast carcinoma received immediately after surgical resection and divided into 22 portions</td>
<td>fixed in 10% formalin or ethanol for various durations</td>
<td>1D5</td>
<td>microwave epitope retrieval with citrate buffer</td>
<td>For TMA, NBF had best results, pH of formalin had little influence on results; ethanol gave poor results; For TMA, 24 hrs in NBF had best results; up to 72 hrs in NBF did not alter ER detection; worst results with simulation of cryosectioning before fixation or delayed onset of fixation, followed by ethanol fixation; 6 hrs was 80% immunoreactive relative to 24 hrs in NBF</td>
<td>yes</td>
</tr>
</tbody>
</table>

**Abbreviations:** CAS, computer-assisted image analysis system; DCC, dextran-coated charcoal; EDTA, ethylenediaminetetraacetic acid; ER, estrogen receptor; FFPE, formalin-fixed paraffin-embedded; hr(s), hour(s); IHC, immunohistochemistry; mAB, monoclonal antibody; mod, moderately; NBF, neutral buffered formalin; neg, negative; NEQAS, National External Quality Assurance Schemes; NR, not reported; PBS, phosphate buffered saline; pos, positive; PR, progesterone receptor; RT, room temperature; TMA, tissue microarray; vs., versus.
Storage of slides

Eight studies were identified that examined optimal methods of storing unstained cut slides for IHC analysis (Table 6) (125-132). Storing slides at room temperature generally resulted in the loss of HR antigenicity (127-130). Using TMAs, slides stored for six months showed weaker HR associations with tumour-specific survival compared to slides stored for one week (131). Several studies indicated that storing slides at low temperatures (125,132), protection from light (132), paraffin coating (129), and storage in a nitrogen chamber (129) are possible ways to preserve antigenicity in slides. However, it appears that the best results are obtained from freshly cut sections, especially with TMAs (129-130).

Table 6. Studies that compared different storage methods.

<table>
<thead>
<tr>
<th>Author</th>
<th>Comparison</th>
<th>N</th>
<th>Storage</th>
<th>Kit/Antibody</th>
<th>Results/Conclusion</th>
<th>Blind</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bromley 1994(125)</td>
<td>variable storage lengths in variable storage conditions</td>
<td>n=4</td>
<td>Slides were stored under 4 sets of conditions: 1. RT 2. 4 °C, 3. Abbott Medium 4. Sucrose 4 °C; At intervals of 1 wk, 2 wks, 4 wks, 8 wks, and 4 mos</td>
<td>Abbott kit</td>
<td>Best preservation found in slides stored in 10% sucrose/PBS at 4°C for 4 wks</td>
<td>no</td>
</tr>
<tr>
<td>Hendricks 1996(126)</td>
<td>stored for 1-2 days vs. greater than 365 days</td>
<td>n=17</td>
<td>Slides were stored at RT for either 1-2 days (short interval storage) or greater than 365 days (long interval storage)</td>
<td>ER1D5</td>
<td>no difference in intensity or proportion of cell staining between different storage intervals; paraffin slides can be stored at RT for as long as 1 yr using ER1D5</td>
<td>yes</td>
</tr>
<tr>
<td>Jacobs 1996(127)</td>
<td>fresh vs. stored for 12 wks slides</td>
<td>n=9</td>
<td>Slides stored at RT for 12 wks and freshly cut slides from the same paraffin blocks</td>
<td>ER1D5</td>
<td>Intensity of ER staining decreases on slides stored at RT for 12 wks (p=0.0001); 1 scored as neg at 12 wks</td>
<td>yes</td>
</tr>
<tr>
<td>Bertheau 1998(128)</td>
<td>1 yr vs. 3 mos vs. recent unstained slides</td>
<td>n=6</td>
<td>recut blocks and stored for 3 mos or 1 yr in drawers under safe light conditions at RT</td>
<td>monoclonal antibody Novocastra</td>
<td>Decreased ER staining in 5/8 3-mo slides and 5/6 1 yr slides compared to slides &lt;7 days old</td>
<td>yes</td>
</tr>
<tr>
<td>DiVito 2004(129)</td>
<td>paraffin-coating of slide and/or storage in a nitrogen dissection chamber (stored for 3 mos) vs. freshly cut slides; TMAs stored at 2 vs. 6 vs. 30 days</td>
<td>n=200 (100 node pos and 100 node-neg breast cancer TMAs)</td>
<td>Slides stored at RT in ambient air for 2, 6, or 30 days or in desiccating nitrogen chamber for 3 mos (some slides were paraffin coated) or paraffin coated with or without antioxidants (BHT) for 5 mos</td>
<td>mouse anti-ER antibody (DAKO)</td>
<td>For TMA, statistically significant decrease in ER detection at 6 days (p=0.0001) and 30 days (p=0.0001) (not 2 days, p=0.1121) compared to recent sections; paraffin coating and nitrogen storage of slides improved ER detection; addition of antioxidant with paraffin did not improve retention</td>
<td>no</td>
</tr>
<tr>
<td>Fergenbaum 2004(130)</td>
<td>stored vs. fresh TMAs</td>
<td>n=125 (single TMA block from 125 invasive cancers from case-control study)</td>
<td>TMA sections stored at RT for 6 mos or cut and stained the same day</td>
<td>NCL-ER-6F11/2 and PR NCL-PGR-312</td>
<td>stored vs. fresh ER 88%, PR 90.3%; stored vs. fresh % of pos cases: ER 59% vs. 59% (p=1.0), PR 56.3% vs. 61.4% (p&lt;0.01); stored vs. fresh median %: ER 27% vs. 38% (p=0.001), PR 30% vs. 70% (p&lt;0.0001)</td>
<td>yes</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>Storage</td>
<td>Kit/Antibody</td>
<td>Results/Conclusion</td>
<td>Blind</td>
</tr>
<tr>
<td>-------------</td>
<td>------------------------------------------------</td>
<td>--------------</td>
<td>--------------------------------</td>
<td>-----------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Mirlacher</td>
<td>6 mos storage vs. 1 wk storage</td>
<td>n=522</td>
<td>f/u mean 90 mos (range 3-152); stored at 4°C for 6 mos or RT for 1 wk</td>
<td>ER 6F11 and PR PGR312</td>
<td>For TMA, slides stored for 6 mos at 4°C showed lower positive ER (65% to 46%, p&lt;0.0001) and PR (33% to 18.5%, p&lt;0.0001) cases and weaker ER (p=0.09) and PR (p=0.11) associations with tumour-specific survival than freshly cut sections</td>
<td>yes</td>
</tr>
<tr>
<td>Blind 2008</td>
<td>Paraffin coating slides vs. no coating; various oxidative treatments</td>
<td>2 TMA blocks cut into sections</td>
<td>paraffin study: sections stored at RT for 1, 2, 4, 7, 10, 14, 21, 30, 45 and 60 days and covered with paraffin or stored at RT and uncovered or were fresh TMA; oxidation study: FFPE ER-positive block sections mounted on silane-coated slides and subjected to (a) pretreatment in a H₂O₂ bath (chemical oxidation), (b) storage under dry heat (thermal oxidation), or (c) exposure to ultraviolet (UV) irradiation (photo-oxidation)</td>
<td>ER-6F11 and PR PgR 636</td>
<td>For TMAs, paraffinization resulted in highly variable ER results possibly due to incomplete paraffin removal; paraffin sections exposed to UVA for more than 4 days and dry heating of more than 10 days resulted in reduction of ER antigenicity; chemical oxidation had little influence on intensity (degraded tissue however); PR was reduced after 7 days of UVA irradiation or 4 wks of dry heating</td>
<td>no</td>
</tr>
</tbody>
</table>

**Abbreviations**: BHT, 2,6-di-tertbutyl-4-methyl-phenol; ER, estrogen receptor; FFPE, formalin-fixed paraffin-embedded; f/u, follow-up; mo(s), month(s); neg, negative; pos, positive; PR, progesterone receptor; RT, room temperature; TMA, tissue microarray; UVA, ultraviolet A; vs., versus; wk(s), week(s); yr(s), year(s).

**Analytic Variables**

**Antigen Retrieval**

Comparative studies investigating antigen retrieval methods can be found in Table 7. Antigen retrieval increases immunoreactivity compared to no antigen retrieval (116,133-134) and is an effective method to recover immunoreactivity from overfixed tissue (111). Antigen retrieval does not appear to be required for some of the newer, more sensitive antibodies such as SP1 and SP2 (134-136). However, the method (wet autoclave, pressure cooker, steamer, microwave, enzymatic digestion, water bath) (116,133,137-139) reagents (citrate buffer, ethylenediaminetetraacetic acid [EDTA]) (140), and duration (138,141) used for antigen retrieval need to be optimized for each antibody. Furthermore, one study found that retention of sections during antigen retrieval might be influenced by the type of slide, the section thickness, and the use of adhesives. The maximum retention of sections stained for PR was obtained using Superfrost Plus slides with Mayer albumen adhesive (138).

**Antibodies**

Thirty-two articles that reported the type of IHC antibody used in paraffin-embedded sections included patient outcomes, using the following monoclonal antibodies: ER1D5,
ER6F11, ERH222, ERM7047, ERSP1, ERAER311, ERLH1, ERLH2, PR1A6, PRKD68, PR636, PR312, PR16, PR1294, or PRcla02/1 (Tables 3, 4, 7, or 8) (29-45, 48-50, 80-81, 99, 110, 142-149). Use of these antibodies with their specific protocols was correlated with such patient outcomes as response to endocrine therapy, overall survival, disease-free survival/interval, progression-free survival, metastasis-free interval, recurrence/relapse-free survival, and time-to-treatment failure. Specifically, when the antibodies 1D5, 6F11 and SP1, 1A6, 636, KD68, 1294, 429 and 312 were used, HR levels were correlated with patient response to endocrine therapy using 1% or 10% thresholds (29-30, 33, 35, 37-38, 41-42, 44-45, 48, 81, 142-144, 147). Of these antibodies, the most well validated from six large studies (greater than 1000 samples) were 1D5, 6F11, SP1 for ER and 1A6, 1249, and 312 for PR (30, 41, 44, 48, 142, 144).

There were 27 articles that included comparisons between more than one IHC antibody using paraffin sections against other IHC antibodies using paraffin sections or against the reference standard of either DCC or IHC on frozen sections (Tables 3 and 7) (35, 48, 88, 110-111, 114, 121, 134-137, 149-164). Among the 20 ER studies including concordance values for more than one ER IHC antibody used in paraffin sections, 1D5, H222, LH1, CC4-5, 6F11, ER88, D75, AER314, AER320, and SP1 demonstrated greater than 90% concordance with a comparator (48, 88, 111, 114, 134-135, 149-158, 160-162, 164). As well, the combination of 1D5 and LH1 was found to be 98% concordant with H222 on frozen sections (111), and the combination of 1D5 with ER-2-123 was 99% concordant with 6F11 (161).

Antibody Detection

There were five studies that compared different antibody detection systems (Table 7) (117, 140, 165-167). Labelled streptavidin biotin (LSAB) as the method of detection is more sensitive and generated superior scores than did EnVision Plus (dextran polymer technology) when 1D5 and 6F11 antibodies were tested (140, 165). However, a novel immunohistological method called the ImmunoMax method, using a monoclonal antibody from Novocastra, detected more EIA-positive ER cases than LSAB (167). In another study, a tyramine amplification technique increased the concordance of 6F11 and polyclonal PR antiserum with
DCC (117). When using H222, Pronase predigestion with a biotin-labelled secondary antibody and an avidin-alkaline phosphatise conjugate had a higher concordance with DCC than did sections digested with DNase and a peroxidise-antiperoxidase detection system (166).

Automated Versus Manual Staining

Table 7 shows the studies that compared automated versus manual staining. When autostainers were compared to manual staining, one study detected more intense ER staining and a higher percentage of ER positivity with autostainers (168), and another study observed fewer negative ER cases with autostainers among ER-PR+ samples (169). One study found automated staining protocols gave the most reliable results between laboratories; however, intensity and percent positivity were not statistically significant between the different staining techniques (170).
Table 7. Studies that compared different antigen retrieval methods, or antibodies (with patient outcomes or comparing multiple antibodies), or antigen detection systems or automated versus manual staining.

<table>
<thead>
<tr>
<th>Author</th>
<th>Comparison</th>
<th>N</th>
<th>Patients/Specimens</th>
<th>Kit/Antibody</th>
<th>Antigen Retrieval/Assay Platform</th>
<th>Results/Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balaton 1996(141)</td>
<td>various heat-induced epitope retrieval (HIER) times</td>
<td>15 centres selected 10 breast cancer specimens; DCC values not available for 1 centre</td>
<td>5 samples negative with &lt;10 fmol/mg and 5 samples positive with &gt;50 fmol/mg and &gt;30% immunostained cells</td>
<td>1D5</td>
<td>duration of HIER varied from 1 to 30 min with 5-min intervals in the microwave oven and from 0 to 5 min (time under pressure) with 1-min intervals in the pressure cooker</td>
<td>Variability among optimum heating times among 14 labs; each lab needs to optimize own heating time</td>
</tr>
<tr>
<td>Bier 1995(137)</td>
<td>comparing AR methods</td>
<td>n=32 frozen sections and 32 corresponding FFPE sections</td>
<td>frozen sections taken from randomly selected breast carcinomas; corresponding sections were fixed in 4% formalin for 16-96 h</td>
<td>frozen: H222; FFPE: H222 or 1D5.26 or LH1 or CC4-5</td>
<td>H222: enzymatic protease digestion; LH1, CC4-5, or 1D5.26: microwave (MW) irradiation in plastic Coplin jars; alternatively serial sections for each antibody received wet autoclave pretreatment</td>
<td>ER status (10% cutoff): H222 protease 22/32 (69%), auto 27/32 (89%); LH1 MW 28/32 (88%), auto 28/32 (88%); CC4-5 MW 31/32 (97%), auto 31/32 (91%); 1D5 MW 31/32 (97%), auto 30/31 (94%)</td>
</tr>
<tr>
<td>Frost 2000(133)</td>
<td>no AR (AR) vs. microwave AR (MAR) vs. low-temperature AR (LTAR)</td>
<td>n=10</td>
<td>10 blocks/case containing invasive carcinoma and/or DCIS</td>
<td>ER88 and PR88</td>
<td>3 slides for each case: no AR, or MAR in citrate buffer, or LTAR with enzymatic predigestion in trypsin</td>
<td># positive cases (dilution 1:30): ER DCIS no AR 1, MAR 4, LTAR 7; ER invasive carcinoma no AR 1, MAR 7, LTAR 9; PR DCIS no AR 6, MAR 6, LTAR 5; PR invasive carcinoma no AR 4, MAR 7, LTAR 6</td>
</tr>
<tr>
<td>Ibarra 1995(111)</td>
<td>H222 frozen vs. antibody cocktail vs. each antibody in cocktail; antibody cocktail normally fixed vs. overfixed</td>
<td>n=56</td>
<td>consecutive postmenopausal patients; samples fixed at RT in 10% phosphate-saline-buffered formalin for 90 min or 1, 3, or 7 days</td>
<td>ER mAb cocktail: 1D5 and LH1</td>
<td>microwaved in citrate buffer; time prolonged up to a minimum of 20 min for overfixed tissue</td>
<td>H222 (frozen) vs. LH1: concordance 51/56 (91%), sensitivity 88%; H222 (frozen) vs. 1D5: concordance 53/56 (95%), sensitivity 93%; H222 (frozen) vs. (LH1 &amp; 1D5): concordance 55/56 (98%), sensitivity 98%; Number of ER positive cases decreased by 20, 26, and 35% after 7 days vs. 90 min of fixation with the ER cocktail, ER-1D5, and ER-LH1, respectively; AR could recover some of the loss (1D5 87%, LH1 81%, cocktail 97%)</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>Patients/Specimens</td>
<td>Kit/Antibody</td>
<td>Antigen Retrieval/Assay Platform</td>
<td>Results/Conclusion</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Huang 2005(135)</td>
<td>SP1 vs. 1D5</td>
<td>n=61</td>
<td>5 (3 ER+, 2 ER-) invasive breast carcinoma TMAs and 56 (32 ER+, 24 ER-) tissue blocks</td>
<td>1D5 or SP1</td>
<td>1D5: microwaved in citrate buffer SP1: preheated and nonpreheated</td>
<td>1D5 vs. SP1 (with or without heat pretreatment): 100%</td>
</tr>
<tr>
<td>Huang 2006(136)</td>
<td>SP2 vs. 1A6</td>
<td>n=107</td>
<td>TMAs from 64 cases of breast carcinoma and 43 tissue blocks</td>
<td>SP2 or 1A6</td>
<td>1A6: microwaved in citrate buffer SP2: preheated and nonpreheated</td>
<td>Using TMA, 1A6 vs. SP2(without heat pretreatment): 98%; 1A6 vs. SP2(with heat pretreatment): 100%</td>
</tr>
<tr>
<td>Mote 1998(138)</td>
<td>autoclave pretreatment vs. microwave pretreatment; slide type, section thickness and use of adhesives</td>
<td>n=3 tumours, n=1 myometrium; sections: histogrip n=56, silane n=39, superfrost plus n=77 (no adhesive n=13, Mayer albumen adhesive n=56); no adhesive 2 um n=30, 4 um n=37, Mayer albumen 2 um n=71, 4 um n=10</td>
<td>FFPE sections stored at RT</td>
<td>primary mouse anti-human PR monoclonal antibody or primary mouse anti-human PR monoclonal antibody that detects PR A and PR B (hPRa 7), or PR B alone (hPRa 6)</td>
<td>microwaved or autoclaved in sodium citrate solution for 20 min</td>
<td>stronger staining for autoclave compared to microwave (p=0.009); % section retention: no difference between silane and histogrip coated slides, more retained with superfrost plus slides (p=0.0001); Mayer albumen adhesive on superfrost plus slides had best retention and better than with no adhesive (p=0.001); slides (all types) with Mayer albumen adhesive cut at 2 µm had better retention than those cut at 4 µm (p=0.003)</td>
</tr>
<tr>
<td>Neves 2005(139)</td>
<td>pressure cooker vs. microwave oven vs. steamer vs. water bath</td>
<td>n=80</td>
<td>42 ER+ and 38 ER- cases of invasive ductal carcinoma</td>
<td>1D5</td>
<td>4 different methods for antigen recovery: pressure cooker, microwave oven, steamer, and water bath</td>
<td># positive: Electronic pressure cooker 42, microwave oven 28, steamer 30, and water bath 28</td>
</tr>
<tr>
<td>Rossi 2005(134)</td>
<td>SP1 vs. 6F11; SP2 vs. 636; AR vs. no AR</td>
<td>n=38</td>
<td>archival breast carcinoma</td>
<td>ER SP1 or 6F11 and PR SP2 or 636</td>
<td>microwaved in citrate buffer or no AR used</td>
<td>6F11 (with AR) vs. 6F11 (no AR): 83%; SP1 (with AR) vs. SP1 (no AR): 100%; 636 (with AR) vs. 636 (no AR): 0%; SP2 (with AR) vs. SP2 (no AR): 100%; For ER, 6F11 (with AR) vs. SP1 (with or without AR): 100%, for PR, 636 (with AR) vs. SP2 (with or without AR): 100%</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>Patients/Specimens</td>
<td>Kit/Antibody</td>
<td>Antigen Retrieval/Assay Platform</td>
<td>Results/Conclusion</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------------------------------</td>
<td>-----</td>
<td>---------------------------------------------</td>
<td>--------------</td>
<td>----------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Vassallo 2004(140)</td>
<td>2 AR methods and 3 detection systems</td>
<td>n=20</td>
<td>consecutive ER+ invasive ductal carcinoma; paraffin blocks 100% ER positive by IHC</td>
<td>1D5 or 6F11</td>
<td>2 AR solutions 1) citrate buffer, 10 mM, pH 6.0 or 2) Tris-EDTA buffer, pH 8.9; 3 detection systems: EnVision (EV), EnVision Plus (EP), or labelled streptavidin biotin (LSAB)</td>
<td>EDTA performed better than citrate buffer for 6F11-EV (p=0.0016) and 6F11-EP (p=0.0033) groups; LSAB had best scores and fewer negative cases followed by EnVision Plus; no other differences found</td>
</tr>
<tr>
<td>Von Boguslawsky 1994(116)</td>
<td>microwave in citrate vs. nothing; frozen vs. paraffin</td>
<td>n=25</td>
<td>unfixed breast cancer tissue; fixed in 10% buffered formalin 12-48 hr</td>
<td>PR: KD68</td>
<td>microwaved in citrate acid or nothing</td>
<td>For PR, frozen vs. paraffin (with AR): 24/25 (96%), frozen vs. paraffin (without AR): 17/25 (68%)</td>
</tr>
<tr>
<td>Antibodies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arihiro 2007(162)</td>
<td>BioGenex vs. DAKO vs. Ventana</td>
<td>n=89</td>
<td>archival consecutive cases of breast carcinoma</td>
<td>ER88, 1D5, and 6F11 or PR88, PgR636, and PR16</td>
<td>BioGenex: autoclaved in citra plus; DAKO: water bath with Target retrieval solution; Ventana: heated with retrieval solution</td>
<td>No difference in % of positive ER or PR cases between ER88, 1D5, and 6F11 or PR88, PgR636, and PR16; distribution of IS different across techniques for ER and PR; distribution of PS and TS different across techniques for ER only</td>
</tr>
<tr>
<td>Bevitt 1997(152)</td>
<td>NCL-ER-6F11 vs. ER1D5</td>
<td>n=55</td>
<td>sequential breast carcinomas</td>
<td>NCL-ER-6F11 or ER1D5</td>
<td>microwaved in citrate buffer</td>
<td>1D5 vs. NCL-ER-6F11: 91%</td>
</tr>
<tr>
<td>Cano 2003(88)</td>
<td>FF-6F11 vs. FF-SP1; FF-1A6 vs. FF-SP2</td>
<td>39 women and 1 man</td>
<td>fresh surgical specimens of primary breast carcinoma</td>
<td>ER SP1 or 6F11, PR SP2 or 1A6</td>
<td>6F11 and 1A6: Dako target retrieval solution</td>
<td>For ER, SP1 vs. 6F11: 100%; for PR, SP2 vs. 1A6: 100%</td>
</tr>
<tr>
<td>Ellis 2008(171)</td>
<td>Central evaluation of posttreatment ER status from randomized double-blind trial</td>
<td>228</td>
<td>postmenopausal women with ER+ stage II and II BC in the P024 neoadjuvant endocrine therapy trial (letrozole/TAM for 4 mos before surgery)</td>
<td>1D5 (threshold Allred score &gt;2 or 1% to 10% weakly positive cells)</td>
<td>Boiling in citrate buffer</td>
<td>The 16 patients whose tumours converted from ER+ to ER- posttreatment had worse RFS (HR of relapse = 2.4, 95% CI =1.0 to 5.3, p=.03) and BCSS (HR of breast cancer death = 4.3, 95% CI = 1.6 to 11.7, p=.002) than patients with ER+ tumours after treatment; pretreatment Allred scores were similar for both ER- and ER+ posttreatment tumours (p=.2)</td>
</tr>
<tr>
<td>Goulding</td>
<td>Dako ID5 FFPE</td>
<td>n=90</td>
<td>case series of</td>
<td>1D5 or H222</td>
<td>1D5: microwaved in</td>
<td>1D5 vs. H222: r=0.7 using H-</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>Patients/Specimens</td>
<td>Kit/Antibody</td>
<td>Antigen Retrieval/Assay Platform</td>
<td>Results/Conclusion</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------------</td>
<td>------------</td>
<td>---------------------------------------------------------</td>
<td>-----------------------</td>
<td>----------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1995(110)</td>
<td>vs. Abbott H222 FFPE vs. Abbott H222 frozen</td>
<td></td>
<td>primary breast cancers treated with hormonal therapy; f/u 6 mos</td>
<td>citrate buffer H222: pronase digestion</td>
<td>score, r = 0.6 using % positivity; Using H-score cutoff of 50, ER status predicting response to tamoxifen therapy: sensitivity 1D5: 90%, H222: 67%; specificity 1D5: 51%, H222: 62%</td>
<td></td>
</tr>
<tr>
<td>Hendricks and Wilkinson 1993(151)</td>
<td>DCC vs. H222 or ERID5</td>
<td>n=20</td>
<td>primary breast carcinoma with DCC values (14 ER+ 6 ER-, &gt;10 fmol/mg)</td>
<td>H222 or ERID5</td>
<td>H222: digested with pronase; ERID5: microwaved with citrate buffer</td>
<td>DCC vs. H222: concordance 8/20 (65%), sensitivity 57%, specificity 83%; DCC vs. ER1D5: concordance 16/20 (80%), sensitivity 93%, specificity 50%</td>
</tr>
<tr>
<td>Huang 1996(153)</td>
<td>AER311 vs. 1D5</td>
<td>n=97</td>
<td>invasive ductal and invasive lobular carcinoma with DCC values</td>
<td>AER311 or 1D5</td>
<td>boiled in citrate buffer using a pressure cooker</td>
<td>1D5 vs. AER311: 77%</td>
</tr>
<tr>
<td>Huang 1999(156)</td>
<td>AER311 vs. 1D5, AER314 vs. AER320</td>
<td>n=10</td>
<td>tissue blocks ER+ by DCC</td>
<td>AER311 or 1D5, AER314 or AER320</td>
<td>boiled in citrate buffer using a pressure cooker</td>
<td>1D5 vs. AER311: 5/10 had ≥ 50 H-score difference, 2/10 AER311 had values less than 50, all other scores for both antibodies were over 50 (concordance 80%); AER314 vs. AER320: 4/10 had ≥ 50 H-score difference, all scores were over 50 (concordance 100%)</td>
</tr>
<tr>
<td>Ibrahim 2008(163)</td>
<td>UK NEQAS Scheme for ICC breast HR module to assess PR</td>
<td>n=281; n=278 for comparison between antibodies</td>
<td>slides consisted of 3 infiltrating ductal breast carcinomas, previously classified as a high PR expresser, a moderate to low PR expresser, and a negative tumour; participants asked to return best of two stained sections for PR for central review</td>
<td>SP2 used by 30/278 (10.8%) of the participants</td>
<td>various</td>
<td>SP2 showed false-positive and nonspecific staining on previously established PR-tumour; 20% of participants using SP2 passed (score ≥12/20)</td>
</tr>
<tr>
<td>Kaplan 2005(160)</td>
<td>1D5 vs. 6F11</td>
<td>n=592 from 586 patients</td>
<td>archival cases of primary breast carcinoma</td>
<td>1D5 and 6F11</td>
<td>heated with citrate buffer</td>
<td>1D5 vs. 6F11: 97.5%</td>
</tr>
<tr>
<td>Kleer</td>
<td>DCC vs.</td>
<td>n=29 used DCC, 29 archival cases of manual IHC</td>
<td>manual IHC with H222:</td>
<td>H222 vs. CC4-S: 95%; CC4-S vs. yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>Patients/Specimens</td>
<td>Kit/Antibody</td>
<td>Antigen Retrieval/Assay Platform</td>
<td>Results/Conclusion</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------------------------------</td>
<td>---------------</td>
<td>---------------------------------------------------------------------</td>
<td>-----------------------------------</td>
<td>----------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1999(157)</td>
<td>automated CC4-5 vs. automated 6F11 vs. manual H222</td>
<td>manual IHC with H222 and automated IHC with CC4-5; n=21 manual H222 vs. automated CC4-5; n=35 used automated IHC using 6F11 and CC4-5</td>
<td>breast cancer with DCC and IHC values; 35 archival consecutive cases of breast cancer</td>
<td>with H222; automated CC4-5 or 6F11</td>
<td>pronase; automated CC4-5 or 6F11: microwaved in citrate buffer</td>
<td>6F11: 97%</td>
</tr>
<tr>
<td>Love 2002(147)</td>
<td>Retrospective analysis of RCT</td>
<td>709 cases, 66% assessable, f/u 3.6 years</td>
<td>RCT of premenopausal women with operable breast cancer received either adjuvant oophorectomy and TAM or after observation when metastatic disease developed</td>
<td>Cocktail of 6F11 and 1D5, PR1294 (threshold Allred score &gt;2 or 1% to 10% weakly positive cells)</td>
<td>Boiling in pressure cooker with TrisHCl buffer</td>
<td>Positive ER or PR status associated with improved DFS (ER p=0.001) in adjuvant treatment group, negative HR status associated with no benefit from adjuvant treatment; using multivariate Cox model, ER positivity related to DFS and OS</td>
</tr>
<tr>
<td>Miller 1993(114)</td>
<td>DCC vs. FFPE (D75) vs. frozen (D75) vs. FFPE (H222)</td>
<td>DCC vs. FFPE (D75) n=67, FFPE (D75) vs. frozen (D75) n=59, FFPE (H222) n=40</td>
<td>fresh biopsies</td>
<td>frozen: D75 FFPE: D75 or H222</td>
<td>FFPE: ficin predigestion</td>
<td>FFPE (D75 vs. H222): r = 0.79, 100% concordant</td>
</tr>
<tr>
<td>Muller-Holzner 1993(99)</td>
<td>relationship between PR content and response to TAM; PR in primary tumour vs. corresponding lymph node metastases</td>
<td>n=20</td>
<td>Retrospective review of Relapsing breast carcinoma patients treated with tamoxifen</td>
<td>KD 68</td>
<td>With or without trypsin or pronase digestion</td>
<td>significant correlation between PR status and response to TAM (p&lt;0.0001); PR negative generally showed tumour progression and PR positive showed no change or remission with TAM; PR was generally lower in metastases compared to primary tumour (p&lt;0.001)</td>
</tr>
<tr>
<td>Nassiri 2008(121)</td>
<td>alcohol-based fixative, automated</td>
<td>n=62</td>
<td>paired tissue sections from 62 specimens</td>
<td>pharmDX kit (Dako) or 1D5</td>
<td>pharmDX kit (Dako): per kit instructions; monoclonal antibody</td>
<td>pharmDX vs. 1D5; r = 0.88 p&lt;0.05</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>Patients/Specimens</td>
<td>Kit/Antibody</td>
<td>Antigen Retrieval/Assay Platform</td>
<td>Results/Conclusion</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------------------------------------------------------------------</td>
<td>-----</td>
<td>-----------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>----------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td></td>
<td>formalin-free microwave processing vs. formalin fixative, VIP processing; pharmDX Kit (Dako) vs. 1D5</td>
<td></td>
<td></td>
<td></td>
<td>(clone 1D5): S1699 DakoCytomation, Carpinteria, CA</td>
<td></td>
</tr>
<tr>
<td>Nedergaard</td>
<td>FFPE (1D5) vs. FFPE (H222)</td>
<td>n=215</td>
<td>retrospective samples from unilateral primary breast carcinoma</td>
<td>1D5 or H222</td>
<td>1D5: microwaved in citrate buffer; H222: pronase digestion</td>
<td>1D5 vs. H222: 77%</td>
</tr>
<tr>
<td>1996(164)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nichols</td>
<td>CC4-5 automated immunostainer vs. ER1D5 manual staining vs. DCC</td>
<td>n=103</td>
<td>archival sequential cases of invasive breast cancer</td>
<td>CC4-5 or ER1D5</td>
<td>microwaved in citrate buffer</td>
<td>ER1D5 vs. CC4-5: 99%</td>
</tr>
<tr>
<td>1996(155)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O'Keane</td>
<td>H222 vs. 17-beta estradiol vs. DCC</td>
<td>17-beta estradiol n=50, H222 n=38</td>
<td>archival paraffin sections; 33 ER+ and 17 ER- using DCC (&gt;10 fmol/mg)</td>
<td>H222 and polyclonal antibody to 17-beta estradiol</td>
<td>H222: enzymatic pretreatment (DNase)</td>
<td>DCC vs. anti-estradiol antibody: concordance 39/50 (78%), sensitivity 0.76, specificity 0.82; DCC vs. H222: concordance 32/38 (84%), sensitivity 0.93, specificity 0.56</td>
</tr>
<tr>
<td>1990(150)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phillips</td>
<td>comparison of assays</td>
<td>n=256; ER n=212, PR n=204</td>
<td>254 infiltrating breast cancer and 2 ductal carcinoma in situ specimens; TMAs constructed</td>
<td>6F11 vs. (1D5 and ER-2-123); PgR 1294 (DAKO or Allred protocol)</td>
<td>pressure cooker (DAKO method)</td>
<td>Using TMA, 6F11 (Allred) vs. (ER-2-123 &amp; 1D5 (DAKO)): 99%; PgR1294 (Allred) vs. PgR1294 (DAKO): 99%</td>
</tr>
<tr>
<td>2007(161)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Press</td>
<td>Comparison of IHC antibodies vs. DCC</td>
<td>59; (39 were PR+ by DCC)</td>
<td>2 multituum blocks were prepared from biopsies of primary breast cancer</td>
<td>PgR636 or PgR1294 or PR 88 or PR 10A9 or KD 68 or JZB 39 or PR-2C5 or 1A6 or PR 4-12 or hPrA2 or hPrA3 or hPrA8</td>
<td>microwave or steamer</td>
<td>DCC vs. PgR636: 90%; DCC vs. PgR1294: 88%; DCC vs. PR88: 85%; DCC vs. KD88: 83%; DCC vs. 10A9: 85%; DCC vs. PR-2C5: 83%; range for all antibodies: 52-90%</td>
</tr>
<tr>
<td>2002(159)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rost</td>
<td>EVG F9 vs.</td>
<td>n=25</td>
<td>FFPE blocks</td>
<td>EVG F9 or</td>
<td>microwave exposure</td>
<td>H222 vs. EVG F9: 16/25 (64%)</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>Patients/Specimens</td>
<td>Kit/Antibody</td>
<td>Antigen Retrieval/Assay Platform</td>
<td>Results/Conclusion</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
<td>-----</td>
<td>-------------------------------------</td>
<td>--------------</td>
<td>----------------------------------</td>
<td>------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>2000(158)</td>
<td>H222</td>
<td>n=119, 1D5 vs. H222: n=27</td>
<td>fresh untreated primary breast carcinoma, cutoff H-score &gt;20</td>
<td>H222</td>
<td>IHC: 1D5 and H222 EIA: D547 and H22</td>
<td>1D5: microwave treatment, H222: enzymatic predigestion</td>
</tr>
<tr>
<td>Saccani 1994(154)</td>
<td>IHC vs. EIA; 1D5 vs. H222</td>
<td>n=119, 1D5 vs. H222: n=27</td>
<td>fresh untreated primary breast carcinoma, cutoff H-score &gt;20</td>
<td>H222</td>
<td>IHC: 1D5 and H222 EIA: D547 and H22</td>
<td>1D5 vs. H222: r = 0.91, concordance 20/27 (74%)</td>
</tr>
<tr>
<td>Santeusanio 2000(149)</td>
<td>prognostic value of different antibodies</td>
<td>n=46 FFPE blocks</td>
<td>Adjuvant chemotherapy was given to node-positive patients, either pre- or postmenopausal, with negative ER status; postmenopausal patients with positive ER status received tamoxifen</td>
<td>H222</td>
<td>H222 EIA: D547 and H22</td>
<td>microwaved in citrate buffer</td>
</tr>
<tr>
<td>Cohen 1991(166)</td>
<td>Pronase vs. DNase vs. frozen vs. DCC</td>
<td>n=68</td>
<td>archival FFPE samples of breast carcinomas</td>
<td>H222</td>
<td>Pronase vs. DNase</td>
<td>Pronase (FFPE) vs. DCC: 75%, DNase (FFPE) vs. DCC: 60%, frozen vs. DCC: 88%</td>
</tr>
<tr>
<td>Goldstein 2007(165)</td>
<td>various incubation times and detection systems</td>
<td>n=22; 6 sets from each 22 tissue blocks</td>
<td>100% ER positive by IHC</td>
<td>1D5</td>
<td>1) 12 hour incubation at 1:800 and a supersensitive, labelled streptavidinbiotin detection system (LSAB) chromogen detection system (12 h-Standard) 2) 2-hour incubation at 1:50 using supersensitive LSAB (2 h-SS) 3) 2-hour incubation at 1:50 using a polymer (EnVision Plus for 30 min) (2 h-Env)</td>
<td>12 h-standard: 0% neg; 2 h-SS: 9% neg; 2 h-Env: 23% neg</td>
</tr>
<tr>
<td>Sumiyoshi 2001(167)</td>
<td>LSAB (FFPE) vs. ImmunoMax (IM) (FFPE) vs. EIA</td>
<td>n=75</td>
<td>50 negative and 25 positive cases based on EIA</td>
<td>monoclonal antibody (Novocastra)</td>
<td>microwave in citrate buffer</td>
<td>LSAB vs. EIA: 85%, IM vs. EIA: 85%; more positive cases detected by LSAB than IM or EIA</td>
</tr>
<tr>
<td>W.R. von</td>
<td>various</td>
<td>n=88</td>
<td>cases of archival</td>
<td>ER: 1D5 and heat-induced epitope</td>
<td>A tyramine amplification</td>
<td>yes</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>Patients/Specimens</td>
<td>Kit/Antibody</td>
<td>Antigen Retrieval/Assay Platform</td>
<td>Results/Conclusion</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------------------</td>
<td>--------------------------</td>
<td>--------------------</td>
<td>-------------------------------</td>
<td>----------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Wasielewski</td>
<td>detection methods, IHC vs. DCC</td>
<td></td>
<td>breast carcinoma</td>
<td>6F11 PR: 1A6 and polyclonal antiserum (PR, Dako)</td>
<td>retrieval (microwaved in citrate buffer) with and without tyramine amplification technique</td>
<td>technique increased concordance of 6F11 (from 80/88 (91%) to 82/88 (93.2%)) and polyclonal PR antiserum (73.9% to 82.9%) with DCC</td>
</tr>
<tr>
<td>Biesterfeld</td>
<td>automated immunostainer vs. manual staining</td>
<td>n=5 tissues; 15 sections/case; 25 sections/staining method</td>
<td>FFPE all ER+</td>
<td>NCL-ER 6F11</td>
<td>immunostainers were Optimax Plus and the Mark 5 HSS; used consecutively due to limited availability of immunostainers</td>
<td>Using 6F11, % positivity and intensity were higher with immunostainers than manual staining; % positivity for one immunostainer was significantly different from manual staining (P&lt;0.0001), but not for the other immunostainer; % positivity significantly different between 2 immunostainers (p=0.0143); staining intensity of immunostainers were significantly different from manual staining (p&lt;0.0001, p=0.0048)</td>
</tr>
<tr>
<td>Navani</td>
<td>manual vs. automated</td>
<td>n=44; 7/44 deplete of tumour, therefore, n=37</td>
<td>FFPE blocks from two different hospitals; ER-PR+ (with manual staining) invasive breast carcinomas</td>
<td>1D5 or ER88</td>
<td>ER88: i1600 Biogenex Automated Immunostainer</td>
<td>1D5 (manual): 100% negative; ER88 (automated) 24% negative</td>
</tr>
<tr>
<td>Regitnig</td>
<td>automated immunostainer vs. manual staining</td>
<td>automated techniques (Ventana n=11 labs, DAKO n=6 labs), manual technique n=14 labs</td>
<td>results from quality assurance program in Austria; each lab had to stain 5 slides for ER and 5 for PR</td>
<td>various</td>
<td>Ventana or DAKO autostainers vs. manual staining</td>
<td>Overall automated staining protocols gave the most reliable results (higher k values observed for either Ventana or DAKO compared to manual for each scoring method)</td>
</tr>
</tbody>
</table>

**Abbreviations:** BC, breast cancer; DCC, dextran-coated charcoal; DCIS, ductal carcinoma in situ; DFS, disease-free survival; EDTA, ethylenediaminetetraacetic acid; EIA, enzyme immunoassay; ER, estrogen receptor; FFPE, formalin-fixed paraffin-embedded; f/u, follow-up; HR, hormone receptor; ICC, immunocytochemistry; IHC, immunohistochemistry; IS, intensity score; k, kappa; LSAB, labelled streptavidin-biotin; mAB, monoclonal antibody; mos, months; NR, not reported; OS, overall survival; PR, progesterone receptor; PS, proportion score; RCT, randomized controlled trial; RFS, relapse-free survival; RT, room temperature; TAM, tamoxifen; TMA, tissue microarray; TS, total score; UKNEQAS, United Kingdom National External Quality Assurance Schemes; vs., versus.
Thresholds to Define Results

Manual Scoring

There are two primary parameters for scoring: proportion score (PS, percentage of positive neoplastic cells) and intensity score (IS), as well as a combination of the two (Table 8). Studies that compared different types of manual scoring and thresholds can be found in Tables 3, 4, or 9. Four studies demonstrated substantial associations between different scoring methods. EIA was correlated with cell counting (r=0.78), a modified immunoreactive score (r=0.77), and the H-score (r=0.75) (172). The H-score was also highly associated with modified additive and multiplicative quick scores (rho=0.887 and 0.892) (173). As well, there was strong agreement for ER and good agreement for PR between the H-score and the percentage of positive staining (ER k=0.879, PR k=0.662) (174). Furthermore, the percentage of positive staining, using a scale of 0 to 3+, was correlated with the Allred score (r=0.74, p<0.001) (175).

Three studies showed a high reliability when comparing different scoring methods. One study found an intraclass correlation for ER or PR of 0.98 across three pathologists (176). As well, each pathologist’s score was strongly correlated with scores derived from image analysis (0.77-0.78) (176). Another study showed increased k values when using a binary scoring system (k=0.73) compared to scores with a scale from 0 to 3+ (k=0.45) (177) for ER. In comparison, Fisher et al (2005) found high interobserver agreement for ER (87%-89%) and PR (79%-86%) using percent positivity, intensity, or binary scoring methods (39).

There were two studies that showed differences in scoring methods (162,178). For ER, 3.2% cases were categorized as borderline cases using the Allred method, but only 0.9% samples were considered uncertain cases using the J-score (178). In another study, the percentage of positive cells, separated into categories using 1% and 10% as dividers (ER k=0.67, PR k=0.72), showed better agreement across five different staining techniques than did Allred’s total score (ER k=0.34, PR k=0.45); however, Allred’s PS (ER k=0.48, PR k=0.60) followed by the IS (ER k=0.44, PR k=0.49) had better agreement than did the total score (162). Moreover, three studies found that Allred’s PS was a more significant predictive indicator than was the IS or the total score for ER (45,49,143) but not for PR (44). Another study found that all the methods used for scoring ER (percent positivity, intensity, or all-or-none algorithms) predicted a better prognosis for overall survival in patients with unfavourable lymph node status at five and ten years (39). Similarly, three scoring methods, PS (cutoff 10%), IS (cutoff moderate or greater intensity), and an aggregate score combining PS and IS, were found to significantly predict recurrence-free survival among patients who had received tamoxifen (38). Taken together, these studies suggest the percentage of nuclear staining or a binary score, either positive or negative, are adequate and possibly superior to an intensity score or combinations of percentage and intensity scores based on reliability and predictive or prognostic validity.

Although the binary score may be an adequate scoring method, the percentage of stained cells may provide valuable predictive and prognostic information to oncologists when considering treatment strategies. Thirteen studies described the relationship between HR levels in paraffin sections and patient outcomes (29,33,37,43,48,50,79,81,142-146). Overall survival (37,43,50,79,146), disease-free survival (50,144), recurrence/relapse-free survival (81,146), 5-year survival (145), breast cancer-specific survival (48), time-to-treatment failure/progression (33,37), response to endocrine therapy (29,37,143), and time-to-recurrence (142) were all positively associated with ER levels. Overall survival (37,43), disease-free survival (144), time-to-treatment failure (37), response to endocrine therapy (37,143), and time-to-recurrence (142) were positively related to PR levels. These studies suggest that patients with higher HR levels will have a higher probability of positive outcomes, which may influence oncologists’ treatment decisions.
Fifteen articles examined cutpoints for optimizing predictive or prognostic validity (30,33,36,38,41,44-47,49,81,143-145,148). Several retrospective studies suggest that ER values as low as 10% (36,38,45,49) and possibly as low as 1% (41,46,143-144) can have a predictive or prognostic value, depending on the antibody, staining procedure, and patient outcome evaluated. Three studies suggesting 10% as a threshold examined the relationship between ER status and the prognosis of patients treated with endocrine therapy. Differences in 5-year recurrence-free survival at a cutoff of 10% but not 1% and overall and 5-year survival at a cutoff of 33% were found for patients with primary breast cancer who had surgery and postoperative tamoxifen monotherapy (49). Disease-free survival was most significant for ER using a PS cutoff of three (corresponding to 10% stained cells) for patients who received endocrine therapy (45). Among patients who were treated with tamoxifen, there was a significant difference in recurrence-free survival between patients with ER-positive versus ER-negative tumours at a 10% cutoff (recurrence rate 16% vs. 38%, p=0.0041) (38).

Three studies suggesting 1% cutpoints investigated ER status as a prognostic indicator among patients treated with endocrine therapy. Harvey et al (2008) assessed tissue from 1,982 patients who received adjuvant endocrine therapy, either alone or in combination with chemotherapy (41). The best cutoff for disease-free survival was a score of greater than two, which corresponds to as few as 1% to 10% weakly stained cells, even after adjustments for multiple significance testing (p<0.01) (41). Notably, in the study conducted by Harvey et al (1999), there were only 67 cases with a score of 2, and it is not clear how many of the responders were closer to 10% and how many were closer to 1% (41). For patients with metastatic breast cancer who received first-line endocrine therapy on relapse, a correlation was found between their ER status and their response to endocrine therapy at 10% (p=0.011) and 1% (p=0.034) thresholds or response to tamoxifen therapy at a 10% threshold (p=0.030) (143). Furthermore, patients with 1% or more ER levels had better survival after relapse (p=0.0005). Locally assessed ER-positive and/or PR-positive tumours reassessed centrally as HR negative had poorer disease-free survival compared to locally and centrally assessed ER-positive and/or PR-positive tumours, using either 10% or 1% cutoffs (144). One study examined ER as a predictive indicator by reporting its interaction with treatments. A significant interaction between ER levels and treatment (p=0.001) showed ovarian ablation was more beneficial for women with a positive quick score (corresponding to as few as 1% cells stained), whereas women with a quick score of 0 (no nuclear staining) had an increased overall risk of death (hazard ratio, 2.19; 95% CI, 1.26 to 3.81) if they underwent ovarian ablation rather than receiving combination chemotherapy (46).

Studies have shown that the PR status provides additional predictive or prognostic value (33,44,144,148), independent of ER values (45,143), especially among premenopausal women (30,81). Again, the prognostic validity for PR in patients treated with endocrine therapy has been demonstrated with as few as 1% of stained nuclear cells in retrospective studies (30,44,143-144). Among patients who received adjuvant endocrine therapy, the best cutoff for both disease-free (adjusted p=0.0021) and overall (adjusted p=0.0014) survival was a total PR score of greater than two, which corresponds to greater than 1% weakly positive stained cells (44). For patients with metastatic breast cancer who received first-line endocrine therapy on relapse, a correlation was found between PR status and their response to endocrine therapy at a 1% threshold (p=0.044) or their response to tamoxifen therapy at 10% (p=0.021) and 1% thresholds (p=0.047) (143). Furthermore, patients with 1% or more PR levels had better survival after relapse (p=0.0008) (143). Locally assessed ER-positive and/or PR-positive tumours reassessed centrally as HR negative had poorer disease-free survival compared to tumours assessed locally and centrally as ER and/or PR positive, using either 10% or 1% cutoffs (144). PR status at a 1% cutoff discriminated disease-free survival among premenopausal patients treated with endocrine therapy (c index=0.60, p=0.003) (30).
Table 8. Different scoring methods reported.

<table>
<thead>
<tr>
<th>Scoring Method</th>
<th>Algorithm</th>
<th>Definitions</th>
<th>Range</th>
</tr>
</thead>
</table>
| Allred score (Allred 1993, 1998)(179-180) | PS + IS | PS: 0 = none, 1 = <1/100, 2 = 1/100 to 1/10, 3 = 1/10 to 1/3, 4 = 1/3 to 2/3, 5 = >2/3  
IS: 0 = none, 1 = weak, 2 = intermediate, 3 = strong | 0, 2-8 |
| H-score (McClelland, 1990)(181) | sum of PS at a given IS x IS | PS: 1-100%  
IS: 0 = none, 1 = weak, 2 = moderate, 3 = strong | 0-300 |
| J-score (Kurosumi 2007)(178) | PS | PS: 0 = none, 1 = ≥1%, 2 = >1% to <10%, 3 = ≤10%  
0 = negative, 1 or 2 = uncertain (equivocal), 3 = positive | 0-3 |
| Modified immunoreactive score (Umemura 2004) (172) | PS + IS | PS: 0 = none, 1 = <1%, 2 = 1 to 10%, 3 = 11 to 33%, 4 = 33 to 66%, 5 = 67 to 100%  
IS: 0 = none, 1 = weak, 2 = moderate, 3 = strong | 0, 2-8 |
| Modified additive quick score (Detre 1995)(173) | PS + IS | PS: 1 = 0 to 4%, 2 = 5 to 19%, 3 = 20 to 39%, 4 = 40 to 59%, 5 = 60 to 79%, 6 = 80 to 100%  
IS: 0 = none, 1 = weak, 2 = intermediate, 3 = strong | 1-9 |
| Modified multiplicative quick score (Detre 1995)(173) | PS x IS | PS: 1 = 0 to 4%, 2 = 5 to 19%, 3 = 20 to 39%, 4 = 40 to 59%, 5 = 60 to 79%, 6 = 80 to 100%  
IS: 0 = none, 1 = weak, 2 = intermediate, 3 = strong | 0-18 |
| Quick score (Leake et al, 2000)(182) | PS + IS | PS: 0 = none, 1 = <1%, 2 = 1 to 10%, 3 = 11 to 33%, 4 = 34 to 66%, 5 = 67 to 100%  
IS: 0 = none, 1 = weak, 2 = moderate, 3 = strong | 0-8 |

Abbreviations: IS = intensity score; PS = proportion score.
Table 9. Studies that compared different types of manual scoring and thresholds.

<table>
<thead>
<tr>
<th>Author</th>
<th>Comparison</th>
<th>N</th>
<th>Patients/specimens</th>
<th>Kit/Antibody</th>
<th>Scoring</th>
<th>Results/Conclusion</th>
<th>Blind</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dowsett 2008(142)</td>
<td>Central re-test of 2,006 FFPE tumour blocks</td>
<td>2,006</td>
<td>Patients postmenopausal with HR+ BC treated on monotherapy arms (anastrozole or TAM) of ATAC trial</td>
<td>6F11 and 312</td>
<td>ER: H-score cutoff&gt;1, PR % of pos stained cells cutoff 10%</td>
<td>Lower ER or PR levels related to shorter TTR for tamoxifen-treated patients (ER p=0.078; PR p=0.0012) and anastrozole-treated patients (ER p=0.0009; PR p=0.0001)</td>
<td>Yes</td>
</tr>
<tr>
<td>Horii 2007(49)</td>
<td>Prognostic value at various cutpoints</td>
<td>n=486; assessable: both invasive and intraductal cancer: n=364</td>
<td>Paraffin-embedded surgical specimens from The Cancer Institute Hospital of the Japanese Foundation for Cancer Research, Tokyo, Japan (1982 to 1993)</td>
<td>1D5</td>
<td>ER status diagnosed in entire cancer area (i.e. invasive and intraductal cancer area); Cutoffs set at 1%, 10% and 33% for proportion of positive cells and at total scores of 2-3, 3-4 or 4-5 for Allred score.</td>
<td>Significant differences between ER status and overall survival at a cutoff of 33% (p=0.03790) and with 5 yr survival at cutoffs of 10% (p=0.04045) or 33% (p=0.00045); significant difference was seen in overall and 5 year recurrence-free survival for a cutoff in total Allred score between 4 and 5 (p=0.01772 and p=0.00049, respectively)</td>
<td>NR</td>
</tr>
<tr>
<td>Kurosumi 2007(178)</td>
<td>Allred score vs. J-score</td>
<td>n=439</td>
<td>HR assessed from surgically-resected and needle core biopsies; no mention of blinding</td>
<td>ER: 1D5, PR: PR636</td>
<td>J-Score: 0: negative, 1 and 2: uncertain/equivocal (greater than 1% but less than 10%), 3: positive vs. Allred: endocrine therapy is considered likely to be effective when the Allred is ≥ 3; 'borderline': Allred=3 and 4</td>
<td>Using J-Score, 0.9% uncertain cases; using Allred, 3.2% borderline cases for ER</td>
<td>No</td>
</tr>
<tr>
<td>Peterson 2008(175)</td>
<td>Allred vs. qualitative assessment vs. IHC index using image analysis</td>
<td>n=16 women and n=1 man</td>
<td>Tissue blocks previously assessed with PET</td>
<td>1D5</td>
<td>Used Allred and another qualitative assessment (IHC Score range 0 to 3+); image analysis: Photoshop - calculated IHC index</td>
<td>Allred vs. image analysis: r=0.81; IHC Score vs. image analysis: r=0.73; IHC Score vs. Allred: r=0.74; interobserver variability for image analysis: r=0.99, p&lt;0.001</td>
<td>No</td>
</tr>
<tr>
<td>Sharangpani 2007(176)</td>
<td>image analysis scoring vs. manual scoring</td>
<td>n=134</td>
<td>Archival cases of breast cancer</td>
<td>ER 1D5 and PR 1A6</td>
<td>Image acquisition utilized either scanner (Aperio ScanScope T2 System,</td>
<td>Manual vs. image analysis: ER 85%, PR 81%; ICC for ER or PR = 0.98; each</td>
<td>Yes</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>Patients/specimens</td>
<td>Kit/Antibody</td>
<td>Scoring</td>
<td>Results/Conclusion</td>
<td>Blind</td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------------------------------------------</td>
<td>-----</td>
<td>-----------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Umemura 2004(172)</td>
<td>Cell-counting score (CCS) vs. modified immunoreactive score (IRS) vs. H score using a highly-sensitive procedure</td>
<td>n=46</td>
<td>Taken from 44 breast cancer samples with EIA values &lt;100 fmol/mg</td>
<td>1D5</td>
<td>CCS, IRS and H score; same sections stained by autostainer were used for evaluation. For CCS and H score, 500 cancer cells in same area of tumour were estimated</td>
<td>Using samples with &lt; 100 fmol/mg. EIA vs. cell counting: r=0.78, or modified IRS: r=0.77, or H-score: r=0.75</td>
<td>No</td>
</tr>
<tr>
<td>Arihiro 2007(162)</td>
<td>BioGenex (automated) vs. DAKO (manual and automated) vs. Ventana (automated)</td>
<td>89 FFPE</td>
<td>Archival consecutive cases of invasive ductal carcinoma of breast; 5 methods of HR evaluation</td>
<td>BioGenex: ER88, PR88; DAKO: 1D5, PgR636; Ventana: 6F11, PR16</td>
<td>Allred score and proportion of cells stained (0, &lt;1%, ≥1% or &lt;10%, or ≥10%); consensus obtained on discordant samples; pathologist blinded to patient characteristics</td>
<td>Intermethod k values using %: ER 0.67, PR 0.72; using Allred (TS) ER 0.34, PR 0.45, (PS) ER 0.48, PR 0.60 (IS) ER 0.44, PR 0.49</td>
<td>Yes</td>
</tr>
<tr>
<td>Aye 2001(174)</td>
<td>H-score vs. percentage positive score</td>
<td>n=150</td>
<td>FFPE from patients with invasive breast cancer</td>
<td>ER: ER1D5, Dako M7047, PR: PR1A6, Dako M3529</td>
<td>H-score or % stained tumour cells</td>
<td>H-score (50 cutoff) vs. 10% cutoff: ER k=0.879, PR k=0.662</td>
<td>No</td>
</tr>
<tr>
<td>Bejar 1998(177)</td>
<td>Image analysis scoring vs. pathologists’ manual scoring Retrospective analysis, other</td>
<td>n=20</td>
<td>Archival consecutive sections of ductal carcinoma</td>
<td>Monoclonal mouse antihuman ER antibody</td>
<td>Wscannary Image Analyzer software (Galai Corporation, Migdal HaEmek, Israel); manual: scored from 0 to 3+ or neg/pos</td>
<td>For ER, manual vs. image analysis (binary status): 79%; reliability: binary scoring system (k=0.73) scores ranging from 0 to 3+ (k=0.45)</td>
<td>No</td>
</tr>
<tr>
<td>Cowen 1990(145)</td>
<td>Best cutoff points for survival and DFI</td>
<td>141 paraffin slides</td>
<td>No tx given before excision and adjuvant tx immediately after excision was anti-hormone therapy (Tam or oophorectomy), 7</td>
<td>H222</td>
<td>Slides were scored 1 through 5 representing -, plus or minus -, +, + + or + + + + by the three authors independently, with consensus agreement</td>
<td>Scores between 3 and 4 showed greatest difference for survival to term (p=0.052), DFI (p=0.068) and survival to 5 yrs (p=0.018) and scores between 1 and 2 showed</td>
<td>Yes</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>Patients/specimens</td>
<td>Kit/Antibody</td>
<td>Scoring</td>
<td>Results/Conclusion</td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>----------------------------------</td>
<td>-----------</td>
<td>------------------------------------------------------------------------------------</td>
<td>--------------</td>
<td>--------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Detre 1995(173)</td>
<td>Quick score vs. H score</td>
<td>n=119;</td>
<td>FFPE taken from a cohort of 119 untreated patients studied previously</td>
<td>EIA: Abbott</td>
<td>H-score or quickscore (percent % and intensity either added or multiplied together)</td>
<td>greatest difference for 5-yr DFI (p=0.014); trend of higher ER score with 5-yr survival (p&lt;0.02)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>assessable n=96</td>
<td></td>
<td>H222; ER IHC: Dako 1D5</td>
<td></td>
<td>Additive quick score vs. H-score: rho=0.887; multiplicative quick score vs. H-score: rho=0.892</td>
<td></td>
</tr>
<tr>
<td>Jalava 2005(148)</td>
<td>Various cutpoints vs. survival</td>
<td>n=324;</td>
<td>Primary breast cancer patients</td>
<td>ER: M7047; PR: NCL-PgR</td>
<td>To find the optimal cutpoints for HR, all quartiles and 1%, 5%, 10%, 90% and 99% score values were tested in univariate analysis; cutpoints with the lowest p-values were considered most optimal for dividing patients into 2 gps differing in survival</td>
<td>For ER, 15% and 30% cutpoints gave highest significance with respect to survival (p=0.037 and p=0.0682, respectively); In the analysis of PR positivity, no statistically significant cutpoints were found among all patient groups; cutpoints differed in different patient groups</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>assessable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>n=324, PR: n=264</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>n=264</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lockwood 1999(146)</td>
<td>10% vs. 30% cutoff</td>
<td>n = 200</td>
<td>Randomly selected from a cohort of 736 patients previously studied; f/u median 65 (5-93) months</td>
<td>ER1D5</td>
<td>Colour video image analysis system (VideoPro 32; Leading Edge, Marion, South Australia)</td>
<td>Decreased risk of RFS and OS with higher % of ER (RFS: RR=0.99 (0.98-1.00) p=0.0043; OS: RR=0.99 (0.97-1.00) p=0.0199)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ER=186</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yamashita 2006(143)</td>
<td>Predictive value of Allred</td>
<td>75 patients</td>
<td>Patients had metastatic BC, received 1st-line therapy with endocrine therapy, tissue examined on relapse</td>
<td>1D5 and PR636</td>
<td>Allred</td>
<td>Significant correlation between positive ER expression and response to all endocrine tx was found at a total score cut off of ≥ 4 (Allred p=0.020), or a cutoff of 10% (p=0.011) or 1% (p=0.034; not for TAM only); for PR a total score cutoff of 5 (p=0.038), or 1% (p=0.044) or 10% (p=0.021; for TAM only) showed significant correlations; patients with 1% or more ER/PR positive cells had better survival after relapse (p=0.005 and p=0.0008, respectively); correlation between IHC ER and PR was not significant</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>Patients/specimens</td>
<td>Kit/Antibody</td>
<td>Scoring</td>
<td>Results/Conclusion</td>
<td>Blind</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>--------------</td>
<td>-------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Viale 2007(144)</td>
<td>Central (retrospective) vs. local (prospective during study) laboratory; also assess impact of PR on letrozole vs. TAM</td>
<td>3650 patients</td>
<td>3650 patients assigned to monotherapy arms assessable (total 8010 randomized)</td>
<td>1D5 and PR1A6</td>
<td>Percentage of staining cells; thresholds 1% or 10%</td>
<td>Central review confirmed 97% of tumours as HR+ (ER, PR &gt;=10%). Pts w/tumours reclassified centrally as ER- or HR- had poor DFS. Central review of ER showed superior prognostic discrimination (1% or 10% cutoffs) compared to local assessment; Among patients with centrally assessed ER+ tumours, letrozole had better DFS than TAM, irrespective of PR expression. Central review changed assessment of receptor status in a substantial proportion of patients.</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Abbreviations**: ATAC, Arimidex, Tamoxifen, Alone or in Combination; BC, breast cancer; DFI, disease-free interval; DFS, disease-free survival; EIA, enzyme immunoassay; ER, estrogen receptor; FFPE, formalin-fixed paraffin-embedded; gps, groups; HR, hormone receptor; ICC, immunocytochemistry; IHC, immunohistochemistry; k, kappa; NR, not reported; OS, overall survival; PET, positron emission tomography; PR, progesterone receptor; PS, proportion score; pts, patients; RFS, relapse free survival; RR, relative risk; TAM, tamoxifen; TS, total score; TTR, time to recurrence; tx, treatment; vs., versus.
Image Analysis

There were 29 comparative studies that included IHC results in paraffin sections assessed with image analysis (Tables 3, 4, or 10). HR values obtained from image analysis were compared to the standards of either IHC results assessed manually in paraffin sections or IHC results from frozen sections or DCC or EIA results. The image analysis systems used were CAS (39,50,107,151,183-186), AQUA (77,79-80,187-188), SAMBA (50), Photoshop (175,189-190) or other types of imaging systems (34,146,176-177,191-199). Image analysis is generally highly correlated with and concordant to manual estimation and has good reliability, but there are no prospective trials with patient outcomes to suggest that manual estimation should be replaced by image analysis. The seven studies that did include patient outcomes were retrospective designs. Two studies suggest that image analysis may provide higher prognostic significance compared to manual estimation; however, both methods yielded highly significant values (34,79). Two other studies examined the best cutpoints for image, and one found no differences in survival with varying antibody concentrations (80). Another study examined the optimal representation of TMA tissue spots as a function of disease-specific survival (77). The remaining study found differences in prognostic value between imaging systems; optimal ER splits for disease-free survival and recurrence-free survival were found with ACIS but not with the CAS percentage score (39).
Table 10. Studies that compared image analysis with manual scoring.

<table>
<thead>
<tr>
<th>Author</th>
<th>Comparison</th>
<th>N</th>
<th>patients/specimens</th>
<th>Kit/Antibody</th>
<th>image analysis</th>
<th>Results/Conclusion</th>
<th>Blind</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chung 2007(187)</td>
<td>AQUA scores vs. manual scores; slide-to-slide comparisons</td>
<td>29 slides from 11 cases (either one, two, or three blocks from the same case and one section studied per each block)</td>
<td>Archival tissue blocks of primary invasive ductal adenocarcinomas</td>
<td>1D5</td>
<td>To normalize our ER scores and allow slide-to-slide comparisons, 0.6 um histospots of representative breast cancer cases with known ER scores were assembled into a ‘gold standard array’ (GSA) and placed adjacent to each whole section</td>
<td>Manual vs. AQUA: 73%; AQUA scores between slides from different blocks from same cases: binary 75%, continuous 19%</td>
<td>NR</td>
</tr>
<tr>
<td>Gokhale 2007(191)</td>
<td>Manual scores vs. ACIS scores vs. Ariol SL-50 scores</td>
<td>n=64</td>
<td>FFPE sections from invasive breast carcinoma</td>
<td>Mouse monoclonal antibody against ER (Novacastra, UK)</td>
<td>The same slides were scanned by the 2 automated systems, ChromaVision ACIS and Applied Imaging Ariol SL-50</td>
<td>For ER (% positivity), manual vs. ACIS: 95%; manual vs. Ariol SL-50: 94%; ACIS vs. Ariol SL-50: 92%</td>
<td>NR</td>
</tr>
<tr>
<td>Kostopoulos 2007(192)</td>
<td>Image analysis vs. manual scoring</td>
<td>n=22</td>
<td>Breast biopsies</td>
<td>ER 6F11</td>
<td>colour-texture based image analysis using a light microscope (Axiostar Plus - ZEISS; Germany) and a colour video camera (DC 300F - LEICA; Germany); used MatLab</td>
<td>For ER, manual vs. image analysis: 86.4%</td>
<td>NR</td>
</tr>
<tr>
<td>Peterson 2008(175)</td>
<td>Allred vs. qualitative assessment vs. IHC index using image analysis</td>
<td>n=16 women and n=1 man</td>
<td>Tissue blocks previously assessed with PET; used Allred and another qualitative assessment (IHC Score range 0 to 3+)</td>
<td>1D5</td>
<td>Photoshop; calculated IHC index</td>
<td>Allred vs. image analysis: r=0.81; IHC Score vs. image analysis: r=0.73; IHC Score vs. Allred: r=0.74; interobserver variability for image analysis: r=0.99, p=0.001</td>
<td>no</td>
</tr>
<tr>
<td>Rothmann 2000(193)</td>
<td>Manual scoring vs. automated scoring</td>
<td>n=13</td>
<td>Archival series of cases of infiltrating ductal carcinoma</td>
<td>NexES (Ventana)</td>
<td>SpectraCube (Applied Spectral Imaging, Migdal HaEmek, Israel)</td>
<td>For ER status, manual vs. image analysis: 11/13 (85%)</td>
<td>NR</td>
</tr>
<tr>
<td>Sharangpani 2007(176)</td>
<td>Image analysis scoring vs. manual scoring</td>
<td>n=134</td>
<td>Archival cases of breast cancer</td>
<td>ER 1D5 and PR 1A6</td>
<td>Image acquisition utilized either scanner (Aperio ScanScope T2 System, Vista, CA or manual vs. image analysis: ER 88%, PR 81%; ICC for ER or PR = 0.98; each</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>patients/specimens</td>
<td>Kit/Antibody</td>
<td>image analysis</td>
<td>Results/Conclusion</td>
<td>Blind</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------------------------------------</td>
<td>-----</td>
<td>--------------------------------------------------------</td>
<td>-------------------------------------</td>
<td>---------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Sherman</td>
<td>Image analysis (AQUA) vs. manual scoring</td>
<td>n=842</td>
<td>From a subset of unselected in situ or invasive breast cancer with available FFPE blocks constructed 4 TMAs blocks with 2-fold representation per tumour</td>
<td>1D5 or PgR 636</td>
<td>AQUA</td>
<td>Using TMA, manual vs. AQUA: ER $r^2=0.80$, PR $r^2=0.86$</td>
<td>NR</td>
</tr>
<tr>
<td>Aziz</td>
<td>EIA vs. IHC</td>
<td>n=22; ER n=21, PR n=20</td>
<td>Tissue samples of breast carcinoma; one half EIA other half IHC; EIA cutoff &lt;10 fmol/mg</td>
<td>H222 and KD68</td>
<td>Videometric 150 image analysis system (American Innovision, San Diego, CA); cutoff %POS ≤4.9</td>
<td>EIA vs. IHC (ER &amp; PR): 97.5%</td>
<td>NR</td>
</tr>
<tr>
<td>Baddoura</td>
<td>DCC vs. H-score (manual) vs. image analysis (PNA and QIC SCORE)</td>
<td>n=66</td>
<td>Archival slides from primary breast infiltrating ductal carcinoma with previous DCC results (threshold 10 fmol/mg 34 ER+ and 32 ER-)</td>
<td>H222</td>
<td>CAS/200; QIC SCOREs (percentage of positive nuclear area staining (PNA) x % of positive stain/10) were calculated</td>
<td>DCC vs. H-score: 82%, DCC vs. QIC SCORE: 89%, DCC vs. PNA: 91%</td>
<td>no</td>
</tr>
<tr>
<td>Ballouk</td>
<td>CAS-200 vs. DCC</td>
<td>n=34 for IHC, DCC ER n=21, PR n=20</td>
<td>Archival tissue blocks from cases of mucinous carcinoma; DCC cutoff: &lt;10 fmol/mg</td>
<td>Monoclonal antibodies (Abbott) for ER and monoclonal CAS PR kit (Cell Analysis Systems)</td>
<td>CAS-200; analyzed 15 fields at random from each sample; cutoff: positive area and positive staining &gt; 10%</td>
<td>For mucinous carcinoma, DCC vs. CAS: ER 67%, PR 70%</td>
<td>no</td>
</tr>
<tr>
<td>Bejar</td>
<td>Image analysis scoring vs. pathologists’ manual scoring Retrospective</td>
<td>n=20</td>
<td>Archival consecutive sections of ductal carcinoma</td>
<td>Monoclonal mouse antihuman ER antibody</td>
<td>Wscannary Image Analyzer software (Galai Corporation, Migdal-HaEmek, Israel)</td>
<td>For ER, manual vs. image analysis (binary status): 79%; reliability: binary scoring system</td>
<td>no</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>patients/specimens</td>
<td>Kit/Antibody</td>
<td>image analysis</td>
<td>Results/Conclusion</td>
<td>Blind</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------------</td>
<td>---------</td>
<td>--------------------</td>
<td>-------------------------------</td>
<td>--------------------------------------</td>
<td>------------------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Biesterfeld (34)</td>
<td>IHC and DCC</td>
<td>N=111; 108 for IHC analysis</td>
<td>Unilateral invasive breast cancer with known DCC values (cutoff ER &gt;10 fmol/mg, PR &gt;20 fmol/mg); therapy given to node-pos postmenopausal, or premenopausal or ER-; postmenopausal ER+ given tamoxifen; f/u 6.3 +/- 0.7 yrs</td>
<td>ER 6F11 or PR Medac CLA 02/1</td>
<td>TV image analysis system CM-2 (Hund, Wetzlar, Germany); mean immunopositivity of all viewing fields under analysis (ERMEAN, PRMEAN) were calculated (cutoff ER &gt;0%, PR &gt;1.5%); manual cutoff: ER &gt;0%, PR &gt;10%</td>
<td>DCC vs. IHC (image analysis): ER 77%, PR 62%; manual vs. image analysis: ER 91%, PR 88%; univariate survival analysis: manual: ER p=0.0110, PR p=0.0084; image analysis: ER p=0.0253, PR p=0.0005; multivariate analysis using DCC PR status and either manual ER/PR status or image analysis ER/PR status: DCC PR/manual ER&amp;PR: p=0.0045; DCCPR/image analysis ER&amp;PR: p=0.0012</td>
<td>NR</td>
</tr>
<tr>
<td>Camp 2002(79)</td>
<td>Manual scoring vs. automated scoring</td>
<td>n=340</td>
<td>Constructed TMAs from 345 FFPE cases of node-positive breast-carcinoma; f/u 60 months</td>
<td>1D5</td>
<td>AQUA; compared survival of patients with tumours with high (top 25%) versus low (bottom 25%) ER expression</td>
<td>Manual vs. AQUA: r=0.884; intrarater reproducibility using 2 separate histospots from same tumour: manual r=0.732, AQUA r=0.824; prognostic information: AQUA RR=2.44 (p=0.0003), manual RR=2.06 (p=0.002)</td>
<td>unclear</td>
</tr>
<tr>
<td>Cavaliere 2001(107)</td>
<td>frozen section (FS) vs. paraffin section (PS)</td>
<td>n=115</td>
<td>primary breast cancers and one recurrent sample</td>
<td>FS: ER H222 and PR KD68; PS: ER 1D5 and PR 1A6</td>
<td>CAS 200 on both FS and PS; at least 30 random fields of the tumour at 40X were evaluated</td>
<td>Frozen vs. paraffin: ER 83.4%, PR 87.8%</td>
<td>yes</td>
</tr>
<tr>
<td>Charalambous 1993(199)</td>
<td>IHC vs. DCC</td>
<td>n=43, ER n=43, PR n=39</td>
<td>Retrospective cases of primary breast cancer</td>
<td>H222 and KD 68</td>
<td>Image analysis software (Leading Edge, Bellevue Heights, SA, Australia)</td>
<td>ER IHC vs. DCC: 81%, PR IHC vs. DCC: 87%</td>
<td>yes</td>
</tr>
<tr>
<td>Diaz 2004(195)</td>
<td>Image analysis score (QCA) vs.</td>
<td>n=70</td>
<td>Consecutive cases of FFPE from invasive</td>
<td>6F11</td>
<td>Image-analysis software package</td>
<td>For ER, manual vs. image analysis: 98%;</td>
<td>no</td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N</td>
<td>patients/specimens</td>
<td>Kit/Antibody</td>
<td>image analysis</td>
<td>Results/Conclusion</td>
<td>Blind</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
<td>---</td>
<td>--------------------</td>
<td>--------------</td>
<td>---------------</td>
<td>-----------------</td>
<td>-------</td>
</tr>
<tr>
<td>manual score</td>
<td>breast carcinoma</td>
<td></td>
<td></td>
<td>(QCA, Lake Bluff, IL); field most representative of ER staining was chosen by one pathologist, and this field was also used in manual scoring</td>
<td>interobserver agreement: manual k 0.89-0.94, image analysis k 0.78-0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Esteban 1994(50)</td>
<td>DCC vs. IHC (image analysis); different mean optical density thresholds</td>
<td>n=250</td>
<td>FFPE tissue blocks of stage I and II breast carcinoma who had partial or radical mastectomy; DCC (cutpoints ER 3 fmol/mg, PR 5 fmol/mg); f/u median 64 months (54 to 214.3)</td>
<td>ER: 1D5 and PR: (Nova Castra, Vector Labs, Burlingame CA)</td>
<td>CAS cutpoints ER 0.5%, PR 15%; SAMBA</td>
<td>DCC vs. CAS: ER r=0.71, PR r=0.54, concordance ER 91.7%, PR 86.7%; SAMBA: ER Mean optical density (MOD) cutoff of 5: best association with OS (p=0.006), cutoff of 10: best association with DFS (p=0.018)</td>
<td>NR</td>
</tr>
<tr>
<td>Fisher 2005(39)</td>
<td>Manual scoring vs. computer-assisted scoring</td>
<td>n=1891 in initial phase III RCT, n=402 in this study</td>
<td>Primary tumour with ER and some PR DCC values; IHC scored by 2 ind observers using %, intensity, and any-or-none algorithms</td>
<td>ER 1D5 and PR 1A6</td>
<td>CAS 200 measured both ER and PR; ACIS measured ER only</td>
<td>Various manual methods vs. CAS/ACIS: ER r=47-87%, PR r=64-75%; significant split for ACIS but not for CAS % score for correlation between ER and DFS or RFS</td>
<td>yes</td>
</tr>
<tr>
<td>Gibney 1998(186)</td>
<td>DCC vs. automated vs. manual scoring</td>
<td>n=110 for IHC; n=109 for DCC</td>
<td>Consecutive primary breast carcinomas with adequate tissue block and PR DCC results; DCC cutoff &lt;10 fmol/mg; IHC cutoff &gt;5%</td>
<td>Monoclonal anti-PR (Biogenex)</td>
<td>CAS 200</td>
<td>For PR, manual vs. CAS: 90%; DCC vs. CAS: 78.9%; DCC vs. manual: 83.5%</td>
<td>No</td>
</tr>
<tr>
<td>Gorczyca 1998(196)</td>
<td>Automated (LSC) vs. manual scoring</td>
<td>n=30</td>
<td>Archival sections of primary invasive breast carcinoma</td>
<td>ER 1D5</td>
<td>Laser Scanning Cytometer CompuCyte Inc, Cambridge, Massachusetts, USA)</td>
<td>For ER, manual vs. image analysis: r²=0.71; (using 10% cutoff) 87% concordance</td>
<td>yes</td>
</tr>
<tr>
<td>Hatanaka 2003(197)</td>
<td>Quantitative scoring (LI) vs. H score</td>
<td>n=20</td>
<td>Cases of primary breast carcinoma</td>
<td>ER 1D5 and PgR636; H-score cutoff ER 50, PR 30</td>
<td>WimROOF image processing software (Mitani Corp., Tokyo, Japan); labeling index (LI) calculated as a percentage of the</td>
<td>Manual vs. image analysis: ER (100%) PR (90%)</td>
<td>NR</td>
</tr>
</tbody>
</table>

EVIDENTIARY BASE - page 52
<table>
<thead>
<tr>
<th>Author</th>
<th>Comparison</th>
<th>N</th>
<th>patients/specimens</th>
<th>Kit/Antibody</th>
<th>image analysis</th>
<th>Results/Conclusion</th>
<th>Blind</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hendricks and Wilkinson 1993(151)</td>
<td>DCC vs. H222 or ERID5</td>
<td>n=20</td>
<td>FFPE tissue with DCC results (cutoff 10 fmol/mg 14 ER+ 6 ER-)</td>
<td>H222 or ERID5</td>
<td>CAS 200 for ERID5; staining expressed as positive nuclear area (PNA); average of 20 field counts/specimen</td>
<td>DCC vs. ER PNA: r=0.670, sensitivity and specificity not improved by image analysis</td>
<td>yes</td>
</tr>
<tr>
<td>Kohlberger 1999(198)</td>
<td>Image analysis scoring vs. semi-quantitative scoring</td>
<td>n=80 specimens</td>
<td>Specimens of patients with breast cancer</td>
<td>Super Sensitive ER/PR Complete Kit (Bio Genex)</td>
<td>True-color computer-assisted image analysis</td>
<td>For ER, manual (% positive) vs. image analysis: r=0.64</td>
<td>yes</td>
</tr>
<tr>
<td>Layfield 1996(185)</td>
<td>Automated stainer + image analysis vs. DCC</td>
<td>n=236</td>
<td>Archival FFPE tissue from cases of stage I/II breast carcinoma; DCC cutoff &lt;10 fmol/mg</td>
<td>Rabbit polyclonal antibody directed against sites present on the human ER antigen (Ventana Medical Systems) and PR 1A6</td>
<td>CAS 200; performed on 10 fields that contained DAB staining for each sample; cutoff &lt;20 fmol/mg for ER and &lt;10 fmol/mg for PR</td>
<td>DCC vs. CAS: ER 71%, PR 73%</td>
<td>no</td>
</tr>
<tr>
<td>Lehr 1997(189)</td>
<td>DCC vs. Photoshop, EIA vs. Photoshop</td>
<td>n=28</td>
<td>Archival tissue block of consecutive series of infiltrating ductal carcinoma</td>
<td>1D5 or PR: 1A6, PR88, 1A9, mPRI</td>
<td>Photoshop</td>
<td>DCC vs. Photoshop: ER r=0.70, PR r=0.86; EIA vs. Photoshop: ER r=0.76, PR r=0.80</td>
<td>NR</td>
</tr>
<tr>
<td>Lockwood 1999(146)</td>
<td>10% vs. 30% cutoff</td>
<td>n = 200 ER=186</td>
<td>Randomly selected from a cohort of 736 patients previously studied; f/u median 65 (5-93) months</td>
<td>ER1D5</td>
<td>Colour video image analysis system (VideoPro 32; Leading Edge, Marion, South Australia)</td>
<td>Greater predictive power with 30% cutoff than 10% cutoff for DFS and OS</td>
<td>NR</td>
</tr>
<tr>
<td>Mofidi 2003(190)</td>
<td>Percentage positive vs. digital H-scores vs. EIA</td>
<td>n=156</td>
<td>Prospective consecutive FFPE sections of primary breast carcinoma</td>
<td>1D5</td>
<td>Photoshop</td>
<td>Manual vs. Photoshop: r²=0.844; EIA vs. Photoshop: r²=0.748</td>
<td>NR</td>
</tr>
<tr>
<td>Moeder</td>
<td>Prognostic value of n=669; Archival breast tissue; ER 1D5 and AQUA</td>
<td>n=669;</td>
<td>Archival breast tissue;</td>
<td></td>
<td>For ER, the highest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>Comparison</td>
<td>N patients/specimens</td>
<td>Kit/Antibody</td>
<td>image analysis</td>
<td>Results/Conclusion</td>
<td>Blind</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>----------------------</td>
<td>--------------</td>
<td>----------------</td>
<td>--------------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>2007(77)</td>
<td>cores</td>
<td>assessable ER 585, PR 583</td>
<td>f/u median 8.3 yr (range 2.4 - 41.5); 5 cores/case; used 10-yr Kaplan-Meier survival plots; cutpoints ER 10, PR 13</td>
<td>PR 636</td>
<td>score was most prognostic of disease-specific survival (p=0.0003); for PR, minimum (p=0.0006), maximum (p=0.0001), and average (p=0.0006) scores were equally significant</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** AR, antigen retrieval; AQUA, automated quantitative analysis; DAB, diaminobenzidine; CAS, computer-assisted image analysis system; DCC, dextran-coated charcoal; DFS, disease-free survival; EIA, enzyme immunoassay; ER, estrogen receptor; FFPE, formalin-fixed paraffin-embedded; f/u, follow-up; ICC, immunocytochemistry; IHC, immunohistochemistry; k, kappa; min(s), minute(s); NR, not reported; OS, overall survival; PET, positron emission tomography; PR, progesterone receptor; RCT, randomized controlled trial; RFS, recurrence free survival; RR, relative risk; SAMBA, Système d'Analyse Microphotométrique à Balayage Automatique; TAM, tamoxifen; TMA, tissue microarray; vs., versus.
Postanalytic Variables

There were no studies that provided quantitative data on postanalytic variables such as reporting elements.

What parameters should be used to assess the proficiency of an individual laboratory performing HR status testing?

Laboratory proficiency can be broken down into two areas: internal quality assurance (QA) (e.g., initial validation, ongoing proficiency testing (PT) of the laboratory and of individual pathologists) and external quality assurance (e.g., accreditation program participation).

In the current review, fifteen articles were identified that presented information about PT/QA efforts in relation to ER-PR testing in breast cancer (Table 11) (73,118,163,170,200-210). The papers presented various quality assessment schemes for assuring inter and intra-laboratory reliability and reproducibility in histopathology.

The sensitivity of any IHC assay is determined by several parameters; these include the quality of tissue fixation and tissue preparation, the quality and concentration of the primary antibody, the power of the antigen retrieval, and the detection systems. With the advent of tissue microarrays, it is now possible to assess a hundred tissue cores on a single slide. Thus, multiple tissue samples can be analyzed at different laboratories and by different pathologists. Of all the PT and QA programs tested or described in the studies identified, all emphasized the importance of protocol validation, standardization of methods, and standardized cutoffs with accurate scoring.

Many groups have reported that there is a wide variation in testing protocols and HR positivity results between countries and institutions. The Royal College of Pathologists of Australasia (RCPA) Quality Assurance Program (QAP) in 2007 reported on an audit of laboratories in 2005 and 2006 (200). In terms of protocols used for reporting, the College found that these laboratories used a variety of cutoff values: for ER positivity, 15 used 1-4%, 7 used 5-9%, 22 used 10%, and 13 used “other.” Twenty-two laboratories provided no answer to this question. Overall, laboratories conducted satisfactory testing. However, there were a number of individual laboratories that did not meet target values, and this would have had an impact on treatment decision making. Based on previous large cohort studies examining the rates of ER/PR positivity, the expectation was that 95% of laboratories should have had an ER positivity rate between 73% and 84.6%, and a PR positivity rate between 53.1% and 75.9%. In the 2006 audit, ER-positive rates varied from 26% to 100%, PR-positive rates from 23% to 96%. Twenty-two of 67 laboratories had ER-positivity rates below 70%, while eight laboratories had PR-positivity rates below 50%.

The UK NEQAS has published the results of its ongoing quality assurance scheme (163,202-204,208). The comparisons of results achieved on tumours regularly processed by participating laboratories to results achieved with UK NEQAS tumour samples by these same laboratories shows a significant correlation between the two (203). Using data on 7016 breast cancers provided from 71 participating laboratories, investigators attempted to determine the frequency of receptor positivity and to investigate potential causes of any observed variation in the rates of detection. They found a significant correlation between the frequency of positivity and sensitivity of the IHC assay for both ER and PR. Overall, for laboratories using a 10% threshold value for receptor positivity, they reported that the mean frequency of positivity for ER and PR in laboratories with high sensitivity was 77% and 63%, respectively; for laboratories with low sensitivity, these frequencies were 72% and 51%, respectively (202). In a separate publication, the authors investigated interlaboratory variation in the detection of ERs, and found that, while more than eighty percent of laboratories successfully demonstrated ER positivity on medium- and high-expressing tumours, only thirty-seven
percent of laboratories could demonstrate ER positivity on low-expressing tumours. Based on this data, the authors concluded that breast cancers with low ER positivity showed a false negative rate of between 30% and 60% (204).

In the quality network of the German Society for Pathology and the Berufsverband Deutscher Pathologen, QA yearly trials, based on TMAs, are set up for different target molecules, including ER and PR. Up to 200 slides can be produced from one TMA block, assuring that all participants in the trial obtain almost identical material and that results among different laboratories become comparable. Tissue microarray slides with 20–30 tissue spots either negative or expressing ER at low, medium, or high levels are distributed among the participants. Whereas the majority of laboratories (>80%) usually succeed in demonstrating ER positivity in the medium and high-expressing tissue spots, less than 50% of participants obtain the correct results in tissue samples with low expression (118). Poor interlaboratory agreement usually is based on insufficient retrieval efficacy or suboptimal IHC. Interobserver variability, which has been tested in the trials by reviewing all immunostains, is in most instances not responsible for aberrant evaluations (118).

A study was carried out in 33 pathology departments in public hospitals in Austria and a reference laboratory in Vienna (170). The aims of the study were to assess the quality of the IHC assays, and to assess the interlaboratory and interobserver variability performed by different laboratories. Ten unstained slides for interlaboratory variability evaluation and ten IHC prestained slides for interobserver variability evaluation were sent out. The participants performed their in-house IHC technique for ERs and PRs on the unstained slides. All slides were evaluated by estimating percentage and intensity of stained nuclei semiquantitatively. The Reiner, Remmele, and Allred scores were calculated. A less than 10% cutoff level was chosen as threshold for negative 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 multitater k values concerning the ER (Reiner score: k=0.57) and PR scores (Reiner score: k=0.53). The agreement among observers was better for negative cases than for positive cases, whereas the set of unstained slides that were stained locally showed that interlaboratory variability achieved fair to moderate k values concerning the ER and PR scores (k 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%). Variability is higher when laboratories use their own staining method. Automated IHC techniques showed an advantage over manual techniques concerning interlaboratory variability. Investigators determined that to achieve better assay concordance, laboratories should interchange stained and unstained slides with a reference laboratory.

A study from Vancouver used TMAs (29 invasive breast cancers, 58 cores) to assess interobserver and interlaboratory variation in the determination of ER with IHC (73). TMAs were sent to five laboratories in British Columbia (one university hospital and four community hospitals). Four of the laboratories used a scoring system of 0 to 3+ (0 negative, all else positive), with a cutoff point for positivity ranging from 3% to 10% of tumour nuclei staining. The fifth laboratory scored cases as negative or positive (>3% tumour nuclei stained counted as positive, less than that, negative). For the four laboratories using the 0-3+ scoring system, overall k was 0.54. When the scores were translated to only negative or positive determinations, and the scores from the fifth laboratory were included, overall k rose to 0.84. (One laboratory using the 0-3+ scoring system had an overall lower rate of positive results, which resulted in lower k values.) As part of the study, one observer also scored the slides returned from the five laboratories, in order to assess laboratory variation without the confounding of variable observers. Results showed an overall k of 0.63 using the 0-3+ scoring system and 0.93 using just negative or positive determinations. For the laboratory with the
overall lower rate of positive values using the 0-3+ scoring system, interobserver agreement was assessed to attempt to determine whether the variability was due to technical factors or interpretation. The slide stained by this laboratory was independently reviewed by four other pathologists in the same laboratory; overall k for the five pathologists was 0.76. The authors concluded that this suggested methodologic differences were the main contributors to the variability noted. When examining the staining method used by this laboratory compared to the other four, they found that this laboratory used a different antibody (ID5 vs. 6F11 used by the others) and a different antigen retrieval method.

Table 11. Studies about proficiency testing or quality assurance.

<table>
<thead>
<tr>
<th>Ref</th>
<th>Country and Settings</th>
<th>Study Design</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lambkin 1998(206)</td>
<td>Ireland; 6 hospitals surveyed</td>
<td>Hospitals were surveyed regarding immunostaining protocols, Each lab (n=16) was supplied with 2 sets of 10 (formalin fixed (unbuffered, paraffin wax processed blocks) sections of BC to immunostain for ER by their method. Each slide was assessed for intensity of immunostaining, background, normal duct immunostaining (internal positive control) and % of tumour cells expressing the receptor and assigned an overall score of 1+ to 3+.</td>
<td>Many participants achieved a low score in at least one of their 10 sections. There were variations in the intensity of immunostaining, poor nuclear staining, and focal immunostaining, particularly with borderline positives.</td>
</tr>
<tr>
<td>Rhodes 2000(202)</td>
<td>UK NEQAS-ICC; 71 Local labs and Central lab</td>
<td>Questionnaire asked participants to submit data on the last 100 cases of HR testing by IHC</td>
<td>There was a significant positive correlation between frequency of HR+ and the sensitivity of the IHC assay, for both ER (rs=0.346; p=0.019; two tailed) and PR (rs=0.561; p=0.003; two tailed). The mean frequency of HR positivity for labs using the same 10% threshold value was 77% for ER (95% CI, 74% to 80%) in labs with high sensitivity and 72% (95% CI, 68% to 76%) for those with low assay sensitivity (p=0.025). For PR, the mean frequency of HR positivity for labs using the same 10% threshold value and having high assay sensitivity was 63% (95% CI, 57% to 69%), and 51% (95% CI, 38% to 65%) for labs with assays of low sensitivity (p=0.022). The mean frequency of ER positivity for labs serving hospitals and clinics where mammographic screening does and does not take place was 73.4% and 75.7%, respectively (p=0.302; two tailed).</td>
</tr>
<tr>
<td>Rhodes 2000(203)</td>
<td>UK; Local labs and Central lab</td>
<td>Investigate the sensitivity of IHC assays for ER and PR achieved by labs on breast tumours that were fixed and processed in their own department, and to compare this with the degree of sensitivity they achieve on tumours circulated as a part of an external quality</td>
<td>There was also a significant positive correlation between the two sets of scores High-expressing tumours (quick scores of 7 and 6) Participating lab n=78; % 51 and Organizing lab n=123; % 80.9 Level of concordance n=72; % 55.8; % of total 47.4; k value -0.091, p=0.043; Medium-expressing tumours (quick scores of 5 and 4) Participating lab n=42; % 27.6; Organizing lab n=19; % 12.5; Level of concordance n=7; % 13.0; % of total</td>
</tr>
<tr>
<td>Ref</td>
<td>Country and Settings</td>
<td>Study Design</td>
<td>Outcome</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Rhodes</td>
<td>UK NEQAS; 200 from 26 countries</td>
<td>To investigate interlaboratory variance in the IHC in detection of ER so as to determine the rate of false negatives that could adversely influence the decision to give adjuvant tamoxifen treatment.</td>
<td>Over 80% of laboratories were able to demonstrate ER+ on the medium and high expressing tumours, but only 37% of laboratories scored adequately on the low-expressing tumour. Spearman’s rank coefficient showed a highly significant positive correlation between the levels of sensitivity achieved by individual laboratories on the tumours of differing ER expression. 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. When the threshold values used by participants to designate a tumour as either ER+ or ER- were used, the proportion of assays which would have recorded the high, medium, and low expressing tumours as ER+ fell to 98.0%, 80.0%, and 32.8%, respectively (for all evaluations, p&lt;0.0001, two-tailed). Approximately one third of participants failed to demonstrate any tumour nuclei at all in the low expressor. 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).</td>
</tr>
<tr>
<td>Layfield</td>
<td>US; 300 - 200/300 questionnaire recipients from CAP. 4 per state. An additional 100/300 pathologists with expertise in the area of breast pathology from ER and PR.</td>
<td>Survey of 300 laboratories. 80 answers received.</td>
<td>Standardization techniques differed considerably among labs. Forty-nine (61%) labs performed the assay in-house, while the remainder sent the material out for assay. All responding labs performing their HR analysis in-house used the IHC technique. Forty-three (80%) labs 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 labs performing HR analysis in-house used a manual quantification technique. Four (8%) used computer assisted image analysis, and a single laboratory used...</td>
</tr>
</tbody>
</table>
Evidentiary Base

<table>
<thead>
<tr>
<th>Ref</th>
<th>Country and Settings</th>
<th>Study Design</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parker 2002(73)</td>
<td>Vancouver, Canada; n=29; 5 laboratories</td>
<td>External quality control measuring inter or intrarater reliability using tissue microarray to establish it as an effective and efficient tool for assessing interlaboratory variation in ER staining and interpretation.</td>
<td>Of the 29 cases, 6 cases were interpreted originally as negative for ER (score = 0), and 23 were interpreted as + (3 scored as 1+, 10 scored as 2+, and 10 scored as 3+). Assessment of interlaboratory agreement for the 4 laboratories whose pathologists applied the 0 to 3+ scoring system revealed an overall k of 0.54. Pairwise k analyses showed good concordance (k=0.63-0.69) between 3 of the laboratories (1, 2, and 3), but only fair concordance between these laboratories and laboratory 4 (k=0.41-0.46). Laboratory 4 reported lower intensity ER staining for several of the tumour cores compared with laboratory 1, resulting in a lower k value. This also was observed when laboratories 2 and 3 were compared with laboratory 4. Translation of the scores to - or + ER determinations and inclusion of the fifth laboratory resulted in excellent agreement (overall k=0.84), although laboratory 4 again demonstrated less agreement with the other laboratories (k=0.64-0.77). Laboratory 4 reported several tumour cores as negative for ER that laboratory 1 reported as +, resulting in a lower k value. The same was observed when laboratory 4 was compared with laboratories 2, 3, and 5.</td>
</tr>
<tr>
<td>von, Wasielewski 2002(118)</td>
<td>Germany; of 200 laboratories asked to take part in trial, 172 (86.0%) participated</td>
<td>External quality assurance (preanalytic variables (fixation, tissue sampling). Possible influence of fixation and tissue processing on ER immunostains (Study part A). Each participating laboratory received 1 unstained TMA slide for ER testing. For interlaboratory testing (Study part B), a new TMA system was chosen, enabling</td>
<td>More than 80% of the laboratories demonstrated ER positivity in the medium- and high-expressing tissue spots, but only about 43% succeeded with tissue spots with low expression. Poor interlaboratory agreement was based on insufficient retrieval efficacy as shown by additional tests using autoclave pretreatment. TMA technology has proved its suitability for large-scale IHC trials, giving rise to new dimensions in control assessment.</td>
</tr>
<tr>
<td>Ref</td>
<td>Country and Settings</td>
<td>Study Design</td>
<td>Outcome</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Regitnig 2002(170)</td>
<td>Austria; 33 depts of pathology in public hospitals; Study centre was Dept of Pathology, Donaupital, Vienna,</td>
<td>each of the 172 participating laboratories to analyze 30 tissue spots from the same specimen on 1 slide. Moreover, interobserver agreement (Study part C) was tested by comparing the self-assessment of each participating laboratory with that of a review of identical slides.</td>
<td>Sensitivity and specificity were high and false-positive or -negative rates were low in ER cases and moderately low in PR cases. Interlaboratory variability showed fair to moderate k 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). The automated autostainer techniques (Ventana and Dako) were advantageous over manual techniques used in this study concerning interlaboratory consistency.</td>
</tr>
<tr>
<td>Chebil 2003(205)</td>
<td>Sweden; 3 University and 6 County hospitals</td>
<td>Assess the quality of the IHC assays and the interlaboratory and interobserver variability performed by different labs in Austria. Additional goal was to find techniques that should be suggested as reliable assays for HR determination</td>
<td>The concordance was high, with a k value of 0.78 for ER and 0.72 for PR. Since some of the 22 pathologists considered one or several sections to be inevaluable, the evaluation comprised three groups (&lt;10% vs. &gt;10% vs. not evaluated). Although the % of positive cases varied among the 22 pathologists (ER: 57%-79%; PR: 41%-66%), the overall concordance was good for both ER and PR (k 0.78 and 0.72, respectively).</td>
</tr>
<tr>
<td>Wells 2004(209)</td>
<td>EORTC for BC Screening Pathology</td>
<td>To test the variability of ER testing using ICC, centrally stained and unstained slides;</td>
<td>Almost complete concordance among readers (k=0.95) in ER-negative tumours on the stained slide and excellent concordance among readers (k=0.82) on the slides stained in each individual laboratory. Tumours showing strong positivity were reasonably well assessed (k=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.</td>
</tr>
<tr>
<td>Miller 2007(208)</td>
<td>UK NEQAS: Local and Central Laboratories</td>
<td>Predictive and prognostic testing for breast cancer now (2007) an important part of cellular pathology. This information is intended to help those who are not achieving best s cores to improve their performance. It is a modular EQA system.</td>
<td>Review paper. Key point - Over the years when assessing the staining qualities of the submitted in house slides, there is clear evidence that the majority of labs do not fix and process their breast cancers adequately. This really has to be addressed if information provided by the scheme is to benefit those staining for hormonal receptors and HER-2 and the reliability of predictive testing is to be improved.</td>
</tr>
<tr>
<td>Francis 2007(200)</td>
<td>Australia/New Zealand; 55</td>
<td>An audit of labs reporting breast carcinomas was</td>
<td>A total of 55 labs returned information for the time periods in 2004 and 67 for 2005. For the</td>
</tr>
<tr>
<td>Ref</td>
<td>Country and Settings</td>
<td>Study Design</td>
<td>Outcome</td>
</tr>
<tr>
<td>-----</td>
<td>----------------------</td>
<td>--------------</td>
<td>---------</td>
</tr>
<tr>
<td></td>
<td>laboratories</td>
<td>performed in 2005 and 2006 to evaluate in-house results. Labs were asked to submit the HR and HER2/neu-status on each invasive breast carcinoma for the previous 6 month period up to a maximum of 100 cases.</td>
<td>2005 series there were 42 Australian sites, 9 New Zealand sites and 16 international labs. The results are very similar for both audits. For the IHBR06 audit, data were returned for a total of 4807/patients for ER/PR and 3980 patients for HER2/neu. However, 19 patients were reported as ER+ with no PR status, 13 patients were reported as ER2 with no PR status (28 had HER2/neu data) and 20 patients had HER2/neu status reported with no ER/PR data. Overall 59.0% of cases were ER+/PR+, 15.9% were ER+/PR2.2.4% were ER2/PR+ and 22.7% were ER2/PR2 for patients with complete data. In the 2006 audit, ER positive rates varied from 26% to 100%, PR positive rates from 23% to 96%.</td>
</tr>
<tr>
<td>Thomas 2008(210)</td>
<td>United Kingdom; 115 laboratories</td>
<td>Sloane Project - UK-wide audit. Reviewed the recorded cytonuclear grade, receptor status, scoring method, and negative/positive cutoff points for 1,684 patients</td>
<td>No significant difference in frequency of recording HR status in the three cytonuclear grade groups (48% high-grade, 43% intermediate-grade, and 44% low-grade DCIS had ER-status assessed). ER status recorded for 763 cases: 606 (79%) ER+, 161 (21%) ER-. ER positivity in low- and intermediate-grade DCIS was significantly more common than in high-grade DCIS (p&lt;0.001). Forty % of reporting laboratories used 1 method and same cutoff criterion throughout, 19% used one method but different cutoff criteria for different cases, 15% used 2 methods for different cases with equivalent cutoff points for each method, while 25% used 2 methods but non-equivalent cutoff points from case to case. The grade distribution of cases from laboratories that did not report HR status at all was not different from the overall distribution. 60 of 81 laboratories (74%) included information about ER scoring method used. Reporting practices were commonly inconsistent within individual laboratories. The distribution among the following 3 scoring methods: 49% Allred scoring, 32% simple % of positive cells, and 19% Histoscore.</td>
</tr>
<tr>
<td>Ibrahim 2008(163)</td>
<td>UK NEQAS; n = 550 from 50 countries; breast pathology module n=361 from 34 countries</td>
<td>UK NEQAS Scheme for ICC breast hormone receptor module in which participants were asked to demonstrate PRs. The slides consisted of 3 infiltrating ductal breast carcinomas, previously classified as a high PR expressor, a moderate to low PR expressor, and a negative tumour.</td>
<td>Overall Pass Rates—an assessment of 75 of the breast pathology module reported, there were 281 participants; of those, 209 (74.4%) submitted slides assessed as showing an acceptable level of staining and thus passed; 43 (15.3%) achieved a borderline result; and 29 (10.3%) failed the assessment. The median score was 15. The most commonly used antibody was the PR 636 mouse monoclonal antibody (DAKO), with 128 users (46.0%); next was PR-312 mouse monoclonal antibody (Novocastra labs), with 49 users (17.6%); third was the rabbit monoclonal SP2 antibody (Lab Vision/NeoMarkers), with 30 users (10.8%). Of the 2 rabbit monoclonal antibodies used by participants, the Ventana 1E2 and achieved an overall acceptable pass rate of 67%, in marked</td>
</tr>
</tbody>
</table>
Collins 2008(201)  US; 37 labs in US; various labs in Canada and Saudi Arabia  Pathology reports versus central lab testing for IHC/ICC  High level of concordance between ER results abstracted from pathology reports and results from repeat ER testing on TMA sections. Overall, ER status from path report and from TMA, 87.3% agreement (n=1615/1851; k 0.64, p<0.001). Path report ER positive, TMA negative, 5.4% (n=99/1851). Path report ER negative, TMA positive, 7.4% (n=137/1851). ER status from path report with biochemical assay and from TMA, 86.1% agreement (n=1302/1512); ER status from path report with IHC and from TMA, 92.3% agreement (n=310/336; k 0.78, p<0.001).

Abbreviations: BC, breast cancer; CAP, College of American Pathologists; CI, confidence interval; depts, departments; DCIS, ductal carcinoma in situ; EORTC, European Organization for the Research and Treatment of Cancer; EQA, External Quality Assessment; ER, estrogen receptor; HER2/neu, Human Epidermal growth factor Receptor 2; HR, hormone receptor; ICC, immunocytochemistry; IHC, immunohistochemistry; k, kappa; PR, progesterone receptor; SD, standard deviation; TMA, tissue microarray; UK NEQAS, United Kingdom National External Quality Assurance Schemes

Can Oncotype DX reliably determine levels of expression of HR pathway?

This test was originally developed to predict the risk of recurrence for women with ER-positive, node-negative, invasive breast cancer who took or were planning to take tamoxifen. The test analyzes mRNA using rt-PCR, and so differs from IHC, which determines protein expression. In addition, the HR status has usually been examined by IHC prior to Oncotype DX analysis. This assay includes 16 genes for which increased expression at the mRNA level is associated with breast cancer recurrence and five control genes. The HR-related genes are: SCUBE2, BCL2, ER, and PR.

The test is performed in only one central laboratory. Ordering physicians or institutions mail FFPE tissue to this laboratory. The report initially provided only the final Recurrence Score with three risk categories for risk of recurrence, but during 2008 and 2009, the individual readouts in mRNA units were added for the expression of the ER, PR, and HER2/neu genes. There were no published studies of individual ER, PR, or HER2/neu measurements from the Oncotype DX assay and correlation with clinical outcome.

The systematic review identified 14 publications examining Oncotype DX (Table 12). They can be roughly grouped into: development and validation of the 21-gene assay related to the prognostic outcome of patients treated with endocrine therapy (211-217), expansion of the original indication to also predicting the benefit of chemotherapy (218-221), and publications on other gene expression assays or predictive models that compared the results to those from the 21-gene assay (220,222-224).

An earlier systematic review (28) examined the evidence for three gene expression-based assays, and concluded that, compared to the other assays, Oncotype DX was the only one with evidence of a clinical utility to influence the selection of chemotherapy.
Table 12. Studies that included Oncotype DX.

<table>
<thead>
<tr>
<th>Author</th>
<th>Assay Parameter*</th>
<th>Main Conclusions</th>
<th>Role/evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cronin 2004(211)</td>
<td>Platform Technology</td>
<td>Concordance RT-PCR (92 gene assay) and IHC: ER ~94%, PR ~84%. Gene expression profile on paraffin versus frozen (48-gene assay) adjusted Pearson correlation r= .91, p&lt;.0001</td>
<td>A precursor of Oncotype DX</td>
</tr>
<tr>
<td>Paik 2004 (212)</td>
<td>Clinical Validation</td>
<td>For ER, LBA vs. Oncotype DX, in multivariate analysis, fmol/mg levels not predictive. RS predictive for likelihood of distant recurrence HR=2.81 (1.70-4.64), p&lt;0.001</td>
<td>Study re: prognostic for distant recurrence with TAM, ER+, node negative</td>
</tr>
<tr>
<td>Esteva 2005 (213)</td>
<td>Assay Development</td>
<td>For ER, IHC vs. RS, k=0.81 +/- 0.05. For PR, IHC vs. RS, k=0.48. Not prognostic of distant recurrence-free survival for women without treatment</td>
<td>Evaluating as prognostic with, node negative, no TAM or chemotherapy</td>
</tr>
<tr>
<td>Cobleigh 2005(214)</td>
<td>Assay Development</td>
<td>For ER, RT-PCR (86 genes) vs. IHC, k=0.83. For PR, IHC vs. RS, k=0.40. Univariate Cox proportional hazard model of DRFS - ER (% cells) HR=0.993 (0.986-1.000); ER (+ vs. -) HR=0.59 (0.34-1.0), p=0.07, PR (% cells) HR=0.988b (0.979-0.998). PR (+/-) HR=0.817 (0.476-1.403), p=0.46.</td>
<td>Study was a basis of 21-gene assay</td>
</tr>
<tr>
<td>Gianni 2005(220)</td>
<td>Clinical Validations</td>
<td>After adjusting for tumour size and grade, the RS was associated with risk of breast cancer death in ER-positive, tamoxifen-treated (p=.003) and -untreated (p=.03) patients. RS associated with risk of recurrence for tamoxifen-treated and untreated patients (both p&lt;0.0001). RS associated with BrCa mortality among women with ER+ BrCa without TAM treatment and also among women with ER- BrCa.</td>
<td>Predictive for response to neoadjuvant chemotherapy</td>
</tr>
<tr>
<td>Habel 6015(215)</td>
<td>Clinical Validation</td>
<td>Comparison of RS and 70-gene profile: 81% agreement. Comparison of RS and 70-gene profile for ER+ dataset: 76.9% agreement. Intermediate vs. low RS: relapse-free survival HR=1.81 (0.70-4.68), p=0.22. Overall survival HR=1.81 (0.39-8.27), p=0.45. High vs. low RS: Relapse-free survival HR=4.27 (2.05-8.92), &lt;0.001. Overall survival HR=6.14 (1.84-20.4), p=0.003.</td>
<td>Comparison with other gene expression assays</td>
</tr>
<tr>
<td>Fan 2006(224)</td>
<td>Clinical validation</td>
<td>For ER expression, IHC vs. rt-PCR: positive correlation. Non-significant correlation between RS and partial/complete response to primary chemotherapy. Correlation between RS and pCR = .67.</td>
<td>Prognostic-chemotherapy benefit</td>
</tr>
<tr>
<td>Paik 2006(218)</td>
<td>Clinical validation</td>
<td>Likelihood ratio tests of the interaction of chemotherapy treatment with the clinical variables and the gene expression variables: Quantitative ER &gt;/= 50 HR=1.96 (0.73-5.30), p=0.183. Quantitative PR &gt;/= 50 HR=1.87 (0.70-4.97), p=0.214.</td>
<td>Prognostic/predictive-chemotherapy benefit</td>
</tr>
<tr>
<td>Mina 2007(221)</td>
<td>Assay development</td>
<td>For ER expression, IHC vs. rt-PCR: positive correlation. Correlation between RS and pCR = .67.</td>
<td>Prognostic-chemotherapy benefit</td>
</tr>
<tr>
<td>Cronin 2007(216)</td>
<td>Platform technology</td>
<td>In amplification efficiency, assay linearity, quantitative bias and precision, and assay reproducibility Oncotype specifications can be reported within a SD of 2 RS units (100 unit scale)</td>
<td>Publication is primarily to show outcomes of another test</td>
</tr>
<tr>
<td>Loi 2007(222)</td>
<td>Concordance, other gene assay</td>
<td>Correlation between Genomic Grade Index and RS classifications: r=0.7; 95% CI, 0.63-0.76, p&lt;0.0001. The correlation of classifications for tamoxifen-treated data set p&lt;0.0001.</td>
<td>Gene meta analysis</td>
</tr>
<tr>
<td>Smith 2008(223)</td>
<td>Concordance, other gene</td>
<td>Top 1% upregulated genes from ER+ tumours, only 1 overlapped with Oncotype DS, and for ER-, 2 genes overlapped. In top 5% upregulated genes, 4 overlapped with Oncotype DS for</td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>Assay Parameter*</td>
<td>Main Conclusions</td>
<td>Role/evaluation</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Wolf 2008(217)</td>
<td>Clinical validation</td>
<td>For PR, Oncotype vs. IHC, negative PR staining correlated with increased risk of recurrence after tamoxifen treatment in categories defined by Oncotype. PR staining: low-risk 11%, intermediate-risk 25%, high-risk 51%, p&lt;.0001. For ER, p=0.5.</td>
<td>Prognostic/predictive</td>
</tr>
<tr>
<td>Chang 2008(219)</td>
<td>Predictive</td>
<td>For ER, IHC and rt-PCR: k coefficients and 95% CIs k=1.00(1.00,100). For PR, IHC and rt-PCR, k=0.57(0.37-0.77). RS and cCR for ER gene group (n=72) OR 0.73 95% CI(0.52-1.00), p=0.049. CR was more likely with a high RS (p=0.008).</td>
<td>Neoadjuvant chemotherapy</td>
</tr>
</tbody>
</table>

**Abbreviations:** BrCa, Breast Cancer; cCR=clinical complete response; CI(s), confidence interval(s); DRFS, distant recurrence free survival; ER, estrogen receptor; HR, hazard ratio; OR, odds ratio; pCR, pathologic complete response; PR, progesterone receptor; RS, recurrence score; rt-PCR, reverse-transcriptase polymerase chain reaction; SD, standard deviation; TAM, tamoxifen; vs., versus.

*Categories that Genomic Health used to group publications.
DISCUSSION

Immunohistochemistry as an assay to test for hormone receptors in breast cancer is currently in widespread use across the developed world. In the literature reviewed, IHC was shown to obtain results that were at least as reliable as the reference standards to which it was compared. Hundreds of studies have examined various ways to better the performance of IHC, increase its discriminatory power, and standardize results within and across laboratories. Within IHC testing itself, this review attempted to arrange studies into categories for ease of organization. Preanalytic, analytic, and postanalytic elements (reporting) and proficiency testing or quality assurance are all areas of abundant and ongoing research.

In terms of test performance, it appears that IHC can be “accurately” performed in many ways. An important message, emphasized many times by researchers in the field, is that initial and ongoing validation, internal and external quality assurance, and the continued training of pathologists is essential to reliable HR testing results. For each of the separate elements examined (i.e., preanalytic and analytic variables), it is possible to draw some conclusions regarding practices that seem to offer the best results.

Core biopsies and, when fixed in NBF, FNAs may be considered a reliable alternative to the testing of surgical specimens. Core biopsy may on occasion be more accurate, likely due to the shorter time to fixative after tissue procurement that is sometimes observed, when compared to a surgical specimen. In clinical practice, the assessment of HR on core biopsies is limited to cases for which neoadjuvant treatment is considered or when there is no residual invasive tumour on subsequent excision. FNA may be especially valuable in metastatic disease if a core biopsy is not a possibility. The literature reviewed showed high concordance between HR testing results when using any of these tissue specimens. Tissue microarrays are a reliable method for assessing receptor status, and lend themselves especially well to proficiency testing programs.

Several approaches have been tried to preserve the antigenicity of tissue in unstained slides. From the literature reviewed, it appears that the following steps may be of use in attempting to preserve antigenicity: protection from light, heat, and high temperatures, along with paraffin coating.

In studies specifically examining the role of fixatives, NBF seemed to have the most consistent performance. While other fixatives have been tested and are used in laboratories, the performance of these fixatives against NBF needs to be properly validated. In terms of time to fixation, the data show that tissue should be placed in fixative and then submitted to initiate the fixation process as soon as possible after removal. Fixation time in many of the studies reviewed was variable; while it appears that a fixation time of 24 hours may give the best results, a few studies have shown that a minimum of six hours may be reasonable, which may be more practical for day-to-day use. Finally, studies have shown that, with proper antigen retrieval, tissue that has been fixed for several days will continue to immunoreact for ER and PR.

In studies comparing the use of one antibody against another for IHC testing, the results show that many of the commercially available antibodies are highly concordant with each other. Therefore, the choice of a particular antibody should probably depend on the full assay or kit being used and the experience and preference of the laboratory doing the testing, as long as rigorous validation is performed before any changes are made.

In general, automated staining appears to improve the reliability of testing results, compared to manual staining. While evidence is emerging that image analysis is probably comparable to manual estimation, strong evidence is lacking that manual estimation should be wholly replaced by image analysis.
Finally, the percentage of positively HR-stained nuclear cells appears to be reliable and comparable to other scoring methods, as well as providing valuable predictive and prognostic information. However, surveys have shown that laboratories use many different cutoff levels to determine the positivity for ER. Further, when reporting the results of comparative studies in the literature, investigators likewise have used varying cutoff levels. One of the most common cutoff levels reported is 10%, but there is evidence to suggest that cutoff levels as low as 1% predict a benefit following endocrine therapy.

CONFLICT OF INTEREST
The authors declared that there were no conflicts of interest.

JOURNAL REFERENCE
The following systematic review has been published by *Applied Immunohistochemistry & Molecular Morphology* (Copyright © 2011 by Lippincott Williams & Wilkins; [http://journals.lww.com/appliedimmunohist](http://journals.lww.com/appliedimmunohist):)


ACKNOWLEDGEMENTS
The authors would like to thank Emily Vella for taking the lead in drafting this systematic review. They would like to thank M. Elizabeth H. Hammond and Antonio C. Wolff for their significant contributions as reviewers, as well as Patricia Hurley for her involvement in data extraction. The authors would also like to thank R. Bryan Rumble, Denise Kam, and Lava Bahirathan for their contributions toward data calculations and auditing.

**Funding**
The PEBC is a provincial initiative of Cancer Care Ontario supported by the Ontario Ministry of Health and Long-Term Care through Cancer Care Ontario. All work produced by the PEBC is editorially independent from its funding source.

**Copyright**
This report is copyrighted by Cancer Care Ontario; the report and the illustrations herein may not be reproduced without the express written permission of Cancer Care Ontario. Cancer Care Ontario reserves the right at any time, and at its sole discretion, to change or revoke this authorization.

**Disclaimer**
Care has been taken in the preparation of the information contained in this report. Nonetheless, any person seeking to apply or consult the report is expected to use independent medical judgment in the context of individual clinical circumstances or seek out the supervision of a qualified clinician. Cancer Care Ontario makes no representation or guarantees of any kind whatsoever regarding the report content or use or application and disclaims any responsibility for its application or use in any way.

**Contact Information**
For further information about this report, please contact:

**Dr. Wedad Hanna**, Professor, Department of Laboratory Medicine and Pathobiology
Faculty of Medicine, University of Toronto
Fax: 416-480-4271  Email: Wedad.Hanna@sunnybrook.ca

For information about the PEBC and the most current version of all reports, please visit the CCO website at [http://www.cancercare.on.ca/](http://www.cancercare.on.ca/) or contact the PEBC office at:

Phone: 905-527-4322, ext. 42842  Fax: 905-526-6775  E-mail: ccopgi@mcmaster.ca

EVIDENTIARY BASE - page 66
REFERENCES


Appendix 1. Members of the Guideline on Hormone Receptor Testing in Breast Cancer Working Group

Systematic Review and Guideline Development Working Group Members (Sections 1, 2, 3)

Chair:
Sharon Nofech-Mozes MD
Department of Anatomic Pathology
Sunnybrook Health Sciences Centre
Assistant Professor, Department of Laboratory Medicine & Pathobiology
Faculty of Medicine, University of Toronto, Canada

Sukhinder Dhesy-Thind MD, FRCPC
Medical Oncologist, Juravinski Cancer Centre
Associate Professor, Department of Oncology
McMaster University, Hamilton, ON, Canada

Wedad M. Hanna MD
Department of Anatomic Pathology
Sunnybrook Health Sciences Centre
Professor, Department of Laboratory Medicine & Pathobiology
Faculty of Medicine, University of Toronto, Canada

Emily Vella PhD
Research Coordinator
Program in Evidence-based Care, Cancer Care Ontario, Hamilton, ON, Canada

Additional Working Group Members for the Systematic Review (Section 2)

Karen L. Hagerty, MD
Director, Reimbursement Policy
American Society of Clinical Oncology

Pamela B. Mangu, MA
Senior Practice Guideline Specialist
American Society of Clinical Oncology

Sarah Temin, MSPH
Guideline Resource Specialist
American Society of Clinical Oncology
Appendix 2. Literature search strategy.

Database: Ovid MEDLINE® <1950 to May Week 4 2008> Search Strategy:

1. exp Breast Neoplasms/ (155716)
2. (breathe or mammar$) and (cancer? or carcinoma? or neoplasm? or tumor? or malignan$).tw. (164695)
3. 1 or 2 (200070)
4. exp Receptors, Estrogen/ or exp Receptors, Progesterone/ (30928)
5. ((?estrogen or progesterone or hormone) and receptor?).mp. (109585)
6. (“PgR” or “ER/PR” or “ER/PgR” or ER positive or “ER+” or PR positive or “PR+” or PgR positive or “PgR+”).mp. [mp=title, original title, abstract, name of substance word, subject heading word] (50162)
7. Tumor Markers, Biological/ (48073)
8. 4 or 5 or 6 or 7 (190772)
9. 4 or 5 or 6 (145421)
10. 3 and 8 (28749)
11. 10 not (comment or letter or editorial or note or erratum or short survey or news or newspaper article or patient education handout).pt. (27789)
12. limit 11 to (english language and humans and yr=”1980 - 2008”) (22927)
13. limit 12 to yr=”1990 - 2008” (19525)
14. 3 and 9 (23541)
15. 14 not (comment or letter or editorial or note or erratum or short survey or news or newspaper article or patient education handout).pt. (22906)
16. limit 15 to (english language and humans and yr=”1980 - 2008”) (18729)
17. limit 16 to yr=”1990 - 2008” (15498)
18. exp “laboratory techniques and procedures”/ or exp laboratories, hospital/ or exp laboratories/ (1461337)
19. exp biopsy, needle/ or exp biopsy, fine-needle/ (41153)
20. exp Neoplasm Staging/ (79673)
21. exp Immunohistochemistry/ (398787)
22. exp “Staining and Labeling”/ (131989)
23. exp Reference Standards/ (23410)
24. exp Image Processing, Computer-Assisted/ (97843)
25. exp Analytic Sample Preparation Methods/ (84)
26. exp Histocytological Preparation Techniques/ (180936)
27. exp Specimen Handling/ (267572)
28. gene expression.mp. or exp Gene Expression/ (531683)
29. exp Gene Expression Profiling/ (33224)
30. rtPCR.mp. or exp Reverse Transcriptase Polymerase Chain Reaction/ (78615)
31. or/18-30 (2208508)
32. 12 and 31 (10972)
33. 13 and 31 (10031)
34. 16 and 31 (8373)
35. 17 and 31 (7508)
36. exp “Reproducibility of Results”/ (160522)
37. exp “Sensitivity and Specificity”/ (261650)
38. exp “Predictive Value of Tests”/ (85294)
39. reliability.mp. (59293)
40. validity.mp. (63336)
41. sensitivity.mp. (527300)
42. specificity.mp. (601348)
43. predictive value.mp. (106759)
44. exp diagnostic errors/ or exp false negative reactions/ or exp false positive reactions/ or exp observer variation/ (72872)
45. false positive?.mp. (41522)
46. false negative?.mp. (27390)
47 or/36-46 (1184377)
48 12 and 47 (3836)
49 13 and 47 (3445)
50 16 and 47 (2457)
51 17 and 47 (2114)
52 exp Algorithms/ (89109)
53 exp Decision Support Techniques/ (54341)
54 exp Computational Biology/ (18069)
55 algorithm?.mp. (113092)
56 or/52-55 (170753)
57 12 and 56 (316)
58 13 and 56 (311)
59 16 and 56 (176)
60 17 and 56 (173)
61 exp Quality Control/ (30788)
62 exp quality assurance, health care/ or exp benchmarking/ or exp medical audit/ or exp total quality management/ or exp quality indicators, health care/ (184975)
63 quality control.mp. (40407)
64 quality assurance.mp. (43893)
65 exp “outcome and process assessment (health care)”/ or exp “outcome assessment (health care)”/ or exp “process assessment (health care)”/ (400688)
66 or/61-65 (609147)
67 12 and 66 (1541)
68 13 and 66 (1497)
69 16 and 66 (1291)
70 17 and 66 (1247)
71 exp Guideline/ or exp Practice Guideline/ (17521)
72 guideline?.mp. (155514)
73 Evaluation Studies as Topic/ (119429)
74 program evaluation.mp. or exp Program Evaluation/ (37293)
75 exp validation studies as topic/ (76)
76 or/71-75 (306773)
77 12 and 76 (357)
78 13 and 76 (320)
79 16 and 76 (262)
80 17 and 76 (225)
81 exp Neoplasm Recurrence, Local/ (60803)
82 disease-free survival.mp. (30877)
83 exp disease-free survival/ or exp treatment outcome/ (368949)
84 survival rate?.mp. or exp Survival Rate/ (129489)
85 or/81-84 (515749)
86 12 and 85 (4246)
87 13 and 85 (3886)
88 16 and 85 (3523)
89 17 and 85 (3188)
90 (prognos? Or diagnos?).mp. [mp=title, original title, abstract, name of substance word, subject heading word] (42561)
91 exp diagnosis/ or exp prognosis/ (4569143)
92 90 or 91 (4584394)
93 12 and 92 (11532)
94 13 and 92 (10138)
95 16 and 92 (8682)
96 17 and 92 (7382)
97 32 or 48 or 57 or 67 or 77 or 86 or 93 (16056)
98 33 or 49 or 58 or 68 or 78 or 87 or 94 (14215)
Database: EMBASE <1980 to 2008 Week 23>

Search Strategy:

2  ((breast or mammar$) and (cancer? Or carcinoma? Or neoplasm? Or tumor? or malignan$)).tw. (137452)
2  exp Breast Tumor/ (149074)
3  1 or 2 (171805)
4  ((?estrogen or progesterone or hormone) and receptor?).tw. (75844)
5  exp progesterone receptor/ or exp progesterone receptor a/ or exp progesterone receptor b/ or exp estrogen receptor/ or exp estradiol receptor/ or exp estrogen receptor alpha/ or exp estrogen receptor beta/ (33474)
6  (PgR or “ER/PR” or “ER/PgR” or ER positive or “ER+” or PR positive or “PR+” or PgR positive or “PgR+”).tw. (41008)
7  Tumor Marker/ (18840)
8  4 or 5 or 6 or 7 (130728)
9  4 or 5 or 6 (112642)
10  3 and 8 (24501)
11  10 not (comment or letter or editorial or note or erratum or short survey or news or newspaper article or patient education handout).pt. (23002)
12  limit 11 to (human and english language and yr="1990 – 2008") (16436)
13  3 and 9 (22171)
14  13 not (comment or letter or editorial or note or erratum or short survey or news or newspaper article or patient education handout).pt. (20846)
15  limit 14 to (human and english language and yr="1990 – 2008") (14909)
16  4 and 3 (19320)
17  exp clinical laboratory/ or exp laboratory/ (19320)
18  exp hospital laboratory/ (875)
19  exp “diagnosis, measurement and analysis”/ or exp clinical assessment tool/ or exp rating scale/ or exp scoring system/ or exp sensitivity analysis/ or exp hormone determination/ or exp breast biopsy/ or exp tumor biopsy/ or exp needle biopsy/ or exp diagnostic error/ or exp false negative result/ or exp false positive result/ or exp abnormal laboratory result/ or exp laboratory test/ or exp cancer grading/ or exp cancer staging/ (5184517)
19  exp immunohistochemistry/ (186254)
20  exp STAINING/ (41082)
21  exp antibody labeling/ (5906)
22  exp standard/ or exp cancer survival/ or exp disease free survival/ or exp survival rate/ (174856)
23  exp Image Processing/ (11737)
24  exp Laboratory Diagnosis/ (82193)
25  exp Gene Expression Profiling/ (12081)
26  rtPCR.mp. or exp Reverse Transcription Polymerase Chain Reaction/ (85669)
27  gene expression.mp. or exp Gene Expression/ (501622)
28  exp ALGORITHM/ or algorithm?.mp. (63666)
29  exp decision support system/ (1352)
30  exp HORMONE SENSITIVITY/ or exp “SENSITIVITY AND SPECIFICITY”/ or sensitivity.mp. or exp RECEPTOR SENSITIVITY/ (401131)
31  specificity.mp. or exp ANTIBODY SPECIFICITY/ (253395)
32  exp Diagnostic Value/ or exp Prediction/ or exp Prognosis/ or predictive value.mp. (380259)
33  reliability.mp. or exp INTRARATER RELIABILITY/ or exp TEST RETEST RELIABILITY/ or exp INTERRATER RELIABILITY/ or exp RELIABILITY/ (82376)
34  exp REPRODUCIBILITY/ (33558)
35  exp Observer Variation/ (6608)
36  false positive?.mp. (25648)
37  false negative?.mp. (15914)
38  exp Treatment Outcome/ (437816)
<table>
<thead>
<tr>
<th></th>
<th>Survival rate.mp. (73026)</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>exp Cancer Recurrence/ (45969)</td>
</tr>
<tr>
<td>41</td>
<td>prognosis.mp. (225894)</td>
</tr>
<tr>
<td>42</td>
<td>disease-free survival.mp. (15871)</td>
</tr>
<tr>
<td>43</td>
<td>exp practice guideline/ or exp clinical protocol/ or exp consensus development/ or exp good clinical practice/ or exp quality control/ or exp good laboratory practice/ or exp instrument validation/ or exp total quality management/ or exp validation process/ (246512)</td>
</tr>
<tr>
<td>44</td>
<td>exp Medical Audit/ (11529)</td>
</tr>
<tr>
<td>45</td>
<td>exp health care quality/ or exp biomedical technology assessment/ or exp clinical effectiveness/ or exp clinical indicator/ or exp performance measurement system/ or exp medical error/ or exp professional standard/ or exp advisory committee/ or exp consensus/ or exp “institutional review”/ or exp “professional standards review organization”/ or exp quality circle/ or exp accreditation/ or exp certification/ or exp recertification/ or exp disease free interval/ or exp outcome assessment/ or exp outcomes research/ (725037)</td>
</tr>
<tr>
<td>46</td>
<td>guideline?.mp. (140909)</td>
</tr>
<tr>
<td>47</td>
<td>program evaluation.mp. (962)</td>
</tr>
<tr>
<td>48</td>
<td>exp validation study/ (4995)</td>
</tr>
<tr>
<td>49</td>
<td>exp evaluation/ (52258)</td>
</tr>
<tr>
<td>50</td>
<td>quality assurance.mp. (10213)</td>
</tr>
<tr>
<td>51</td>
<td>quality control.mp. (63245)</td>
</tr>
<tr>
<td>52</td>
<td>exp INTERNAL VALIDITY/ or exp EXTERNAL VALIDITY/ or exp VALIDITY/ or exp PREDICTIVE VALIDITY/ or exp validity.mp. (56332)</td>
</tr>
<tr>
<td>53</td>
<td>or/16-52 (6139877)</td>
</tr>
<tr>
<td>54</td>
<td>12 and 53 (13768)</td>
</tr>
<tr>
<td>55</td>
<td>15 and 53 (12356)</td>
</tr>
</tbody>
</table>

Database: EBM Reviews - Cochrane Database of Systematic Reviews <2nd Quarter 2008>

Search Strategy:

<table>
<thead>
<tr>
<th></th>
<th>((breast or mammar$) and (cancer? or carcinoma? or neoplasm? or tumo?r? or malignan$)).mp. [mp=title, abstract, full text, keywords, caption text] (236)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>((?estrogen or progesterone or hormone) and receptor?).mp. [mp=title, abstract, full text, keywords, caption text] (141)</td>
</tr>
<tr>
<td>3</td>
<td>(PgR or “ER/PR” or “ER/PgR” or ER positive or “ER+” or PR positive or “PR+” or PgR positive or “PgR+”).mp. [mp=title, abstract, full text, keywords, caption text] (167)</td>
</tr>
<tr>
<td>4</td>
<td>tumo?r? marker?.mp. [mp=title, abstract, full text, keywords, caption text] (6)</td>
</tr>
<tr>
<td>5</td>
<td>2 or 3 or 4 (296)</td>
</tr>
<tr>
<td>6</td>
<td>2 or 3 (291)</td>
</tr>
<tr>
<td>7</td>
<td>1 and 5 (66)</td>
</tr>
<tr>
<td>8</td>
<td>1 and 6 (62)</td>
</tr>
</tbody>
</table>
Appendix 3. International guideline developers and key organizations.

National Guideline Clearing House

International Guideline Developers:
NICE (UK)
SIGN (UK)
ASCO (US)
NCCN (US)
Agency for Healthcare Research and Quality (US)
European Society for Medical Oncology
National Health and Medical Research Council (Aus)
New Zealand Guidelines Group

Cancer agencies:
National Cancer Institute (US)
International Breast Cancer Research Foundation
The Cancer Council Australia

Canadian provincial cancer agencies:
BC Cancer Agency
Alberta Cancer Board
Saskatchewan Cancer Agency
Cancer Care Manitoba
Cancer Care Nova Scotia

Organizations:
College of American Pathologists
Clinical and Laboratory Standards Institute
Canadian Association of Pathologists
Canadian Medical Association
Standards Council of Canada
Quality Management Program - Laboratory Services

Abbreviations: Aus, Australia; UK, United Kingdom; US, United States.
Appendix 4. References comparing the performance of IHC against DCC, EIA, ELISA or flow cytometry without patient outcomes.


Leers MP, Hoop JG, van BM, van RN, Pannebakker M, Nap M. Determination of threshold values for determining the size of the fraction of steroid hormone receptor-positive tumor...


Wilbur DC, Willis J, Mooney RA, Fallon MA, Moynes R, Sant’Agnese PA. Estrogen and progesterone receptor detection in archival formalin-fixed, paraffin-embedded tissue from breast carcinoma: a comparison of immunohistochemistry with the dextran-coated...


Guideline on Hormone Receptor Testing in Breast Cancer:  
EBS Development Methods and External Review Process

S. Nofech-Mozes, E. Vella, S. Dhesy-Thind, and W. Hanna

A Quality Initiative of the  
Program in Evidence-Based Care (PEBC), Cancer Care Ontario (CCO)

Report Date: April 8, 2011

THE PROGRAM IN EVIDENCE-BASED CARE

The Program in Evidence-based Care (PEBC) is an initiative of the Ontario provincial cancer system, Cancer Care Ontario (CCO) (1). The PEBC mandate is to improve the lives of Ontarians affected by cancer, through the development, dissemination, implementation, and evaluation of evidence-based products designed to facilitate clinical, planning, and policy decisions about cancer care.

The PEBC supports a network of disease-specific panels, termed Disease Site Groups (DSGs), as well as other groups or panels called together for a specific topic, all mandated to develop the PEBC products. These panels are comprised of clinicians, other health care providers and decision makers, methodologists, and community representatives from across the province.

The PEBC is well known for producing evidence-based guidelines, known as Evidence-based Series (EBS) reports, using the methods of the Practice Guidelines Development Cycle (1,2). The EBS report consists of an evidentiary base (typically a systematic review), an interpretation of and consensus agreement on that evidence by our Groups or Panels, the resulting recommendations, and an external review by Ontario clinicians and other stakeholders in the province for whom the topic is relevant. The PEBC has a formal standardized process to ensure the currency of each document, through the periodic review and evaluation of the scientific literature and, where appropriate, the integration of that literature with the original guideline information.
The Evidence-Based Series
Each EBS is comprised of three sections:

- **Section 1: Guideline Recommendations.** Contains the clinical recommendations derived from a systematic review of the clinical and scientific literature and its interpretation by the Group or Panel involved and a formalized external review in Ontario by review participants.
- **Section 2: Evidentiary Base.** Presents the comprehensive evidentiary/systematic review of the clinical and scientific research on the topic and the conclusions reached by the Group or Panel.
- **Section 3: EBS Development Methods and External Review Process.** Summarizes the evidence-based series development process and the results of the formal external review of the draft version of **Section 1: Guideline Recommendations** and **Section 2: Evidentiary Base.**

DEVELOPMENT OF THIS EVIDENCE-BASED SERIES
Development and Internal Review
This EBS was developed jointly by the CCO PEBC and ASCO and was reviewed by the Expert Panel, which consists of members from CCO’s Pathology and Laboratory Medicine Program. The series is a convenient and up-to-date source of the best available evidence on HR testing in breast cancer, developed through review of the evidentiary base, evidence synthesis, and input from external review participants in Ontario.

Expert Panel
Prior to the submission of this EBS draft report for external review, the report was reviewed and approved by the Expert Panel, which consists of six members of CCO’s Pathology and Laboratory Medicine Program. The Expert Panel raised the following key issues:

- I’m not sure that I’m clear on the recommendations on ER PR testing for non-invasive cancers. It doesn’t sound like it’s ready for routine use though some centres are doing this. I (and others) don’t understand the presumed underlying pathophysiologic basis for how testing a non-invasive cancer would help prevent breast cancer morbidity and mortality. The eligibility criteria of the systematic review should be described in more detail.
- I think that it would be useful to understand why we need to include so much technical information in our pathology reports to clinicians about the clone used, results of quality control, etc. I prefer having comprehensive laboratory policies and procedures as well as written (or electronic) worksheets and log sheets of all of this information, so that if there is ever any doubt about a specific result, these could be reviewed.
- Target population: is it worth stating why male breast cancers were not included and if testing recommended at all for ER PR?
- Fixation time: should there be a statement regarding time in formalin whilst in tissue processor if that can be added to fixation time or not?

**Modifications in Response to Expert Panel Feedback**
- The sentence in Section 1, “This guideline should be used for DCIS cases tested for HR as well as for invasive breast cancer.” was changed to “If DCIS is tested for HR status, this guideline should be used.” This was to reflect the fact that not all cases of DCIS
are routinely tested for HR, but if they are, then our guideline should be followed. In addition, the inclusion criteria were described in more detail.

- In Section 1, asterisks were placed next to reporting elements that were required and the following statement was added, “* Asterisked items should be included in the final report that goes to the physician. Other information would be available in the laboratory records.”
- In Section 1, the target population was changed from “Adult female patients with in situ or invasive breast cancer” to “Tissue from adult female patients with in situ or invasive breast cancer was the focus, since most of the literature includes tissue from female patients. However, the recommendations in this guideline would also apply to tissue from male breast cancer patients.”
- In Section 1, the following recommendation on fixation time was changed from “Tissue can be fixed for approximately 8 to 72 hrs, with proper antigen retrieval.” to “With proper antigen retrieval, tissue can be fixed for approximately 8 to 72 hrs. This includes time in the tissue processor.”

**Internal Review: PEBC Director**

Prior to the External Review of this EBS draft report, it was submitted for Internal Review to the Director of the PEBC, Dr. Melissa Brouwers, a researcher with expertise in methodological issues. After that review and approval, the key issues raised by the Director included the following:

- The research questions of the systematic review should be explicitly stated.
- The eligibility criteria of the systematic review should be described in more detail.
- A more extensive summary of the outcomes assessed should be provided in the systematic review.
- The study designs and quality should be described in more detail.

**Modifications in Response to Feedback from the PEBC Director**

- The four research questions that were used for the systematic review were explicitly stated.
- The Inclusion and Exclusion Criteria were explained in more detail. Studies with any clinical outcome were included, which was explicitly stated.
- A summary of the outcomes assessed was included. This included any clinical outcome, especially the correlation between HR status and benefit from endocrine therapy, as well as a definition of concordance and the lack of a gold standard.
- The study designs and quality were described in more detail. This description included factors that affected the quality of the studies, as well as the limitation of using standard operating procedures as the reference standard without clinical outcomes.
- Clinical outcomes were described in more detail in the Outcomes section.
- For the second research question, a description of the reference standard (formalin-fixed paraffin-embedded surgical sections), as well as how the data were organized, was included, for example, the concordance values were summarized as minimum and median values across studies.

**External Review by Ontario Clinicians and Other Experts**

The PEBC external review process is two-pronged and includes a targeted peer review that is intended to obtain direct feedback on the draft report from a small number of
specified content experts and a professional consultation that is intended to facilitate dissemination of the final guidance report to Ontario practitioners.

Following the review and discussion of Section 1: Recommendations and Section 2: Evidentiary Base of this EBS and the review and approval of the report by the PEBC Report Approval Panel, the authors circulated Sections 1 and 2 to external review participants for review and feedback. Box 1 summarizes the draft recommendations and supporting evidence developed by the authors.

**BOX 1:** DRAFT RECOMMENDATIONS (approved for external review December 15 2010)

**PURPOSE**

The overall purpose of this guideline is to improve the quality and accuracy of hormone receptor (HR) testing and its utility as a prognostic and predictive marker for in situ and invasive breast cancer.

**QUESTIONS**

5. **Clinical Validity of Immunohistochemistry (IHC)**
   Can IHC reliably determine the levels of expression of the HR pathway and potentially correlate with the clinical outcome compared to other assays (dextran-coated charcoal [DCC] or ligand-binding assay [LBA], enzyme immunoassay [EIA], enzyme-linked immunosorbent assay [ELISA], and flow cytometry)?

6. **Optimizing IHC**
   How should HR testing be performed optimally to assess true HR status? This includes evaluation of the effect of the following:
   - Pre-analytic variables (i.e., variables of testing involving the collection, fixation, and storage of samples)
   - Analytic variables (i.e., variables associated with the method of testing itself)
   - Thresholds to define results
   - Post-analytic variables (i.e., variables associated with handling of the results, such as reporting)

7. **Quality Assurance of IHC**
   What parameters should be used to assess the proficiency of an individual laboratory performing HR status testing?

8. **Clinical Validity of the Oncotype DX Assay**
   Can Oncotype DX reliably determine the levels of expression of the HR pathway?

**TARGET POPULATION**

Tissue from adult female patients with in situ or invasive breast cancer was the focus, since most of the literature includes tissue from female patients. However, the recommendations in this guideline would also apply to tissue from male breast cancer patients.
INTENDED USERS
Pathologists and breast cancer oncologists.

CONTEXT
This guideline is not intended to advise clinicians when to perform HR testing for in situ or invasive breast cancer but to improve the quality and accuracy of HR testing. Limited data, based on the retrospective analysis of the National Surgical Adjuvant Breast and Bowel Project (NSABP) B24 study suggest that estrogen receptor (ER) expression is an important predictor of the response to tamoxifen in patients with ER-positive ductal carcinoma in situ (DCIS) (1). The results for ER-negative DCIS were inconclusive (1). Although HR testing for DCIS is not mandatory, some oncologists are seeking the HR status in DCIS when considering the benefit of adjuvant endocrine treatment according to the Cancer Care Ontario treatment guidelines for DCIS (2). If DCIS is tested for HR status, this guideline should be used.

RECOMMENDATIONS AND KEY EVIDENCE

5. Clinical Validity of IHC
Can IHC reliably determine the levels of expression of the HR pathway and potentially correlate with the clinical outcome compared to other assays (DCC/LBA, EIA, ELISA, and flow cytometry)?

- IHC should be used instead of DCC/LBA, EIA, ELISA, or flow cytometry.
  In 20 out of 22 studies that compared IHC to another test and included patient outcomes, ER and/or PR using IHC was found to predict patient response to endocrine therapy and/or provide prognostic data such as overall survival, disease-free survival/interval, progression-free survival, metastasis-free interval, and recurrence/relapse-free survival at least as well as or better than the reference assays of DCC or EIA or ELISA (3-24). IHC has certain advantages over older forms of HR testing. These include the requirement of smaller amounts of tissue; the ability to conduct testing on formalin-fixed, paraffin-embedded (FFPE) tissue (allowing for histological examination); and the storage and retrieval of archived slides for retrospective analysis.

6. Optimizing IHC
How should hormone receptor testing be performed optimally to assess the true HR status?

Preanalytical Variables
Core Biopsies

- Core biopsies may be used to assess ER and progesterone-receptor (PR) status in the primary tumour prior to neoadjuvant therapy or in the case of metastatic disease. However, as they are derived from only a small sample of a larger tumour and in view of the heterogeneity in tumour HR expression, it is preferable to test the tumour in the surgical excision specimen. Core biopsies may also be preferentially assessed when the surgical specimen is inadequate, due to inadequate fixation or lack of an invasive component.
  - Eighteen comparative studies found concordance values greater than 83% (median 95%) for ER and greater than 69% (median 88.5%) for PR from core biopsies against standard surgical specimens (25-42).
Variability between cores and whole tumours may be dependent on the size of the tumour or be affected by treatment (30,39).

**Tissue Microarray (TMA)**
- TMA should not be used for diagnostic purposes. The test may be used for quality assurance and as a research tool. At least two cores should be removed from each tumour.
- Seven comparative studies found concordance values greater than 95% (median 97%) for ER and greater than 85% (median 93.25%) for PR using TMAs against standard whole sections (43-49). Six retrospective studies found prognostic value of HR using TMAs (7,48,50-53). Three studies found good agreement between cores (43,46,52). However, three other studies suggested there may be heterogeneity in the amount of tumour tissue between cores and that cores may be lost (44,47,54). The working group recommended against the use of TMAs for diagnostic purposes for several reasons, including the heterogeneity in the cores, the requirement of specialized skill and reading, the possibility for errors with low-positive tumours, and the delay in HR testing caused by waiting for TMAs to be constructed from multiple patients.

**Fine Needle Aspiration (FNA)**
- FNA may be used to assess ER or PR status using IHC, especially in metastatic disease where a core biopsy may not be a possibility.
- FNAs should be fixed in neutral buffered formalin (NBF), using cell blocks from pellets.
- Use at least 100 nuclei when scoring HR from FNA samples, when available.
  - Sixteen comparative studies found concordance values greater than 86% (median 94.5%) for ER and greater than 73% (median 80.2%) for PR from FNA against the standard of IHC on paraffin sections (55-70). Two studies showed a high degree of reliability using FNAs (56,71).
  - The working group chose formalin fixation to maintain consistency with fixative recommendations for paraffin sections (please see below).
  - One study suggested at least 100 nuclei should be used as a criteria to reject technically suboptimal slides for scoring FNAs (68).

**Frozen Sections**
- Paraffin sections should be used instead of frozen sections. However, frozen sections can be used to validate the quality of fixation of paraffin sections.
  - The concordance between frozen and paraffin sections among 21 studies was over 72% (median 91.4%) for ER and over 84% (median 90.8%) for PR (9,21,65,72-89). One study found that ER assessed in paraffin, frozen samples or using DCC were similar in predicting response to endocrine therapy (72), whereas another study found that ER assessed in paraffin sections was a better predictor of endocrine response than ER assessed in frozen samples (21).
  - It is common practice to use paraffin sections, because histological and HR assessment can be determined on the same block. In addition,
tumours that are small and/or non-palpable may be difficult to sample for frozen section analysis.

**Fixation Type**

- Ten percent NBF should be used as a fixative for optimal results.
- Other fixatives may be used if they have been validated against NBF.
- Ethanol should not be used as a fixative.
- Two studies, using paraffin sections or TMAAs, suggest that 10% neutral buffered formalin as a fixative achieves the best results and that ethanol should be avoided as a fixative (90-91). Furthermore, the working group felt that NBF is commonly used in many laboratories. If laboratories use other fixatives, then their results should be validated against tissue fixed in NBF. (See the internal quality assurance [QA] section below.)

**Fixation Time**

- Place tissue into fixative as soon as possible following removal (preferably immediately and not more than 1 hour afterwards).
- Avoid freezing and cryosectioning before fixing the sample.
- Fix for 24 hours for optimal results.
- With proper antigen retrieval, tissue can be fixed for approximately 8 to 72 hours. This includes the time in the tissue processor.
- Under-fixation is more critical than over-fixation; fixation for less than 8 hours precludes the specimen from being used for HR testing.
- Two comparative studies found the best results were achieved when tissue was fixed for 24 hours in NBF, immediately after surgery (90-91). Poorer results were found when tissue was frozen and cryosectioned before fixing or when there was a delay in onset of fixation of 12 hours (90). Another study found that paraffin sections from tissue that had been fixed 1 to 4 hours after removal had poorer concordance with frozen tissue than paraffin sections from tissue fixed within a few minutes after removal (73). The working group chose to use a conservative estimate of a 1-hour threshold for cold ischemic time to fixation.
- Two comparative studies found that tissues fixed for as little as 4 hours in one study and 6 to 8 hours in another study could achieve similar results to tissue fixed for longer periods of time (73,92-93). Again, the working group chose the most conservative time of 8 hours as a minimum requirement for the duration of fixation.
- Four comparative studies showed that with proper antigen retrieval, tissue fixed for several days will continue to immunoreact for HR (77,86,90,94). The longest duration of fixation used was 48 hours for PR, compared to frozen sections, and 96 hours for ER, compared to tissue fixed for 24 hours. The working group chose 72 hours for both ER and PR as maximum requirements for the duration of fixation in order to be consistent with HER2 recommendations (95). In addition, longer fixation does not exclude the specimen for HR testing.

**Tissue Processing**

- Surgical specimens received in the pathology laboratory should be
oriented, carefully inked for surgical margin assessment, and carefully sliced at 5-10 mm intervals before being placed into formalin. If the tissue samples are to be further sectioned and placed into tissue cassettes at a later time, gauze pads or paper towels should be placed in between tissue slices to assist with the penetration of the formalin. If a gross tumour is easily identifiable, normal breast tissue from around the tumour can be placed in the same cassette and immediately fixed at the time of the initial gross evaluation. This will ensure good tissue fixation and that the normal breast tissue, acting as an internal positive control, will be fixed in the same manner as the tumour. For samples obtained remotely from the grossing laboratory, pathologists, in collaboration with the personnel at the remote locations, should ensure the sample is promptly placed in fixative prior to its refrigeration and transport. Although this is less optimal than the rapid gross examination of fresh samples by the pathologist, it is preferable to uncut, fixed or unfixed tissue stored in the refrigerator.

- The block samples from the specimen should be 3-5 mm in thickness. Similarly, the appropriate thickness of the sections from the paraffin block should be 3-5 µm.
  - No published studies were found to inform these recommendations. Therefore, these recommendations were developed through consensus with the working group, based on common practices in IHC laboratories.

Storage of Slides

- Use freshly cut sections for optimal results, especially with TMAs.
- Storing at low temperatures (4°C) or in a nitrogen chamber, protection from light and paraffin coating may be used to preserve antigenicity in slides.
  - Two comparative studies using TMAs showed better HR detection was found with freshly cut sections compared with slides stored for longer periods of time (96-97). As well, one study found that longer storage decreased HR associations with tumour-specific survival (98). Several studies suggested possible methods to preserve antigenicity (protection from light and heat and paraffin coating) (96,99-100); however, the working group decided that, for diagnostic purposes, freshly cut sections should be recommended.

Analytical Variables

Controls

- Controls should include positive and negative breast cancer cases, plus a low positive case if possible.
- The control tissue should be fixed and processed in the same manner as the patient samples.
- Controls should be run with each HR IHC batch run.
- Internal controls of normal breast elements must be evaluated. In the majority of breast cancer resection specimens, 5-10% of luminal epithelial cells will express ER in normal breast elements. If normal breast elements are completely ER negative, a false negative result should be considered, and another block should be tested or ER expression may be measured by another method.
• Other types of tissue such as the endometrium or myometrium could be used for controls. The variation in the level of expression during the menstrual cycle in premenopausal women would give a broad level of expression for ER.
• The controls should use tissue. Cell lines may be used as controls but not in isolation, because they are not practical since they do not have stroma, are not subjected to the same processing methods, and do not have normal tissue that serves as a positive control.
  – No articles were found to inform these recommendations. Therefore, these recommendations were developed through consensus with the working group based on common practices in their laboratories.

Antigen Retrieval
• Antigen retrieval may be used to recover immunoreactivity from overfixed tissue; however the method (e.g., wet autoclave, pressure cooker, steamer, microwave, enzymatic digestion, water bath), reagents (e.g., citrate buffer, ethylenediaminetetraacetic acid) and duration must be optimized for each antibody.
• Stringent compliance to validated standard operating procedures developed in assay validation must be adhered to, and quality control (QC) documentation must be in place.
• Any modifications from package inserts must be validated.
  – Eleven comparative studies were highly variable in their choice of antibody and antigen retrieval method (77,89,101-109). Since no specific recommendations for antigen retrieval could be derived from the evidence, the working group recommended compliance with validated standard operating procedures used in the laboratory, either following package inserts or validating against a reference standard (see internal QA section below).

Reagents
• Stringent compliance to validated standard operating procedures developed in assay validation must be adhered to, and QC documentation must be in place.
• Any modifications from package inserts must be validated.
  – Due to the lack of evidence in this area, these recommendations were developed through consensus with the working group based on common practices and to be consistent with recommendations developed for HER2 testing (95).

Antibodies
• Currently used, or commercially available, validated antibodies such as ER antibodies 1D5, 6F11, and SP1 and PR antibodies 1A6, 1294, and 312 may be used; however, any modification to the product insert must be validated.
• Any change in the type of antibody used, or the use of home-brewed antibodies, must be validated against a validated antibody.
  – The antibodies in the above recommendation correlated with patient response to endocrine therapy, using 1% or 10% thresholds, and were
validated in large studies with samples greater than 1000 (7,15,19,22,110-111). Although three studies did find prognostic variability across ER antibodies, the studies were all retrospective in design (7,76,112).

**Antibody Detection**

- Antibody incubation and detection must be optimized for the specific antibodies used and must be validated with every batch of antibody.
  - Although there were comparative studies that identified differences between antibody detection systems for various antibodies (90,109,113-115), the working group did not want to specifically endorse any type of detection system as each protocol needs to be validated or the package inserts need to be followed, regardless of the detection system used (see internal QA section below).

**Automated versus Manual Staining**

- Correctly operated and validated automated staining protocols and equipment should be used for all steps in the process.
- Maintenance and service records should be regularly updated and filed in the laboratory.
  - There were two comparative studies that found a higher percentage of ER positivity using immunostainers compared to manual staining (116-117). One study found automated staining was more reliable than manual staining but did not find a statistically higher percentage of ER positivity with immunostainers (118). The working group decided that automated staining, if available, is recommended, but validated methods should be used.

**Image Analysis**

- Validated image analysis systems may be used, particularly for low positive cases, to reduce subjectivity and improve interobserver reliability.
- Pathologists must supervise all image analyses.
  - IHC using image analysis in 29 comparative studies had good reliability and was found to be highly concordant and correlated with manual estimation, but there were no prospective trials with patient outcomes to convince the working group that image analysis should replace manual estimation (6,11,13,50-52,73,119-140).

**Thresholds to Define Results**

- There are three categories:
  - Positive: ≥10% staining for ER or PR
  - Low positive: 1% to 9% staining for ER or PR
  - Negative: < 1% staining for ER and PR
- A statement of intensity is not a mandatory element in reporting but could be valuable if the tumour is low positive with weak intensity.
  - Different scoring methods were shown to be comparable, but the percent of cells stained positive is the simplest method and is correlated with patient outcomes (5,7,10-11,18,21,50,53,110-111,135,141-142). Moreover, six studies found that the proportion of positive staining was at least as good a
predictive or prognostic indicator, or better, than the intensity of stained cells or a combination score of the proportion and intensity of stained cells (12-13,16,20,142-143).

- Eight studies suggest that ER values as low as 10% and possibly as low as 1% can have predictive or prognostic value (9,12,15-16,20,23,111,142). Furthermore, other studies have shown the percentage of PR positivity, to as low as 1%, provides additional predictive or prognostic value (5,19,22,53,111,144), independent of the ER values (20,142). However, all of these studies are retrospective and mainly evaluate the prognostic value of HR in patients treated with endocrine therapy. They do not directly assess the predictive ability of HR at different cutoffs by examining the interaction between patients that receive endocrine therapy versus those patients that do not receive endocrine therapy at various cut points. Since most patients receive endocrine therapy, these studies would require large sample sizes. Therefore, in order to be maximally sensitive, and based on the retrospective evidence, the working group decided to use 1% positivity for ER or PR as a threshold for defining low-positive samples. However, since there is high interlaboratory variability in assessing low-positive samples resulting in high false negative rates (91,145) and the nature of the evidence to support a 1% threshold is retrospective, the working group felt that clinical judgment should be used when assessing patients with low-positive HR tumours for endocrine therapy, especially when chemotherapy may not be a tolerable option for some patients. The working group felt there was sufficient evidence to exclude staining intensity in the evaluation of HR positive and negative status. However, intensity could be reported with low-positive samples where clinical judgment may be necessary.

Post-Analytical Variables

After the standard background patient information (see below), results should be stated clearly and bolded, while the additional information of the quality of the test should follow in synoptic format.

- The following items should be reported. Asterisked items should be included in the final report that goes to the physician. Other information would be available in the laboratory records.
  - *Standard background patient information
  - *Specimen identification (case and block number)
  - *Specimen site and type
  - Specimen fixative type (if not 10% NBF)
  - Time to fixation (if available)
  - Duration of fixation (if available)
  - *Antibody clone and vendor
  - Method used (test and vendor)
  - Image analysis method (if used)
  - Adequate controls
  - Adequacy of sample for evaluation
  - *Results:
    - Percentage of positively stained cells for ER and PR
  - *Interpretation:
    - Low Positive 1-9 % for ER or PR
Positive ≥ 10% for ER or PR
Negative < 1% for ER and PR
Not interpretable

*Sample exclusion criteria to perform or interpret a HR IHC assay include the following:
- Tissue fixed using other than 10% NBF unless validated
- Excisional and needle biopsies fixed for less than 8 hrs, samples where fixation is delayed beyond 1 hour
- Core needle biopsies with
  - Edge or retraction artifact involving entire core
  - Crush artifact (thin-gauge vacuum-extraction needle samples)
  - Tissues where controls exhibit unexpected results
- If assay is negative, then:
  - Look at histology and grade; some cancer types like classic lobular and tubular carcinoma or low grade tumours are only rarely ER negative.
  - For cases where there is incongruence between HR testing and histology and/or grade, repeat assay or have another laboratory repeat assay
- Decalcified specimens are often negative. When possible, tease out fragments of the tumour so that no decalcification is required.
- In samples with only DCIS, the type of DCIS should be mentioned and scored.
- In tumours that contain both the in situ and invasive components, only the invasive component should be tested.

Reporting elements not supported by evidence described in the answers to question 2, were developed through consensus with the working group, based on commonly used practices. These reporting elements are also in keeping with the recommendations for HER2 (95).

### 7. Quality Assurance of IHC

What parameters should be used to assess the proficiency of an individual laboratory performing HR-status testing?

**Caseload**

- Appropriate training is recommended for pathologists who report on HR status. The appropriate training may be part of a residency or fellowship program, mentorship with an experienced pathologist, or a formal didactic course. When possible, cases should be reviewed with an experienced pathologist, and at least a 90% concordance should be achieved; any discordance should be assessed on a dual head microscope.
- Staff should be encouraged to show each other borderline cases; experienced pathologists should be consulted in cases with low-positive staining or weak intensity.
- These activities could be tracked in pathology computer systems.
- The number of tests performed by each pathologist should be considered to ensure competency.
- Test volume should be addressed in conjunction with the laboratory’s adherence to strict QC and QA practices.
- Technologists should undergo appropriate training.
**Internal QA**

- Initial test validation should take place together with ongoing QC and equipment maintenance.
- Initial and ongoing education of laboratory personnel, training, and competency assessment should also be implemented.
- Use of standardized operating procedures, including routine use of control materials, should be enforced, and modified procedures should be revalidated.
- Ongoing competency assessment and education of pathologists should take place.
- When validating a new antibody, at least 50 samples should be tested. However, if the laboratory has little experience with performing HR testing, a minimum of 100 sample tests is advisable.
- An assay accuracy of 90% for ER-positive or PR-positive tumours and a 95% concordance rate for ER-negative or PR-negative tumours should be achieved.
- Validation documentation must be kept.
- Any modification to the procedure requires additional validation to ensure accurate performance.
- Periodic trend analysis should be performed; approximately 70% of samples should be ER positive and approximately 50% of samples should be PR positive; this may vary depending on the referred patient population.

**External QA**

**HR-specific external QA**

- Participation in an external proficiency testing for the analytical and postanalytical program (such as the External Quality Assessment program organized by the Quality Management Program - Laboratory Services in Ontario or the College of American Pathologists) with at least two testing events (mailings) annually is mandatory.
- Satisfactory performance requires at least 90% correct responses in graded challenges of several levels of HR expression, as well as a negative control.
- Unsatisfactory performance will require a laboratory to respond according to accreditation agency program requirements.
- Unsatisfactory performance results in suspension of laboratory testing for ER and/or PR.

**IHC external QA**

- Each laboratory should be accredited to perform IHC and follow standard operating procedures (according to the Ontario Laboratory Accreditation requirements).
- Onsite inspection for IHC should take place every other year, with an annual requirement for self-inspection.
- Review of laboratory validation, procedure, QA results and processes, results, and reports for IHC should be put into place.
All recommendations addressing question 3 were developed, through consensus with the working group, to be consistent with HER2 recommendations (95). The percentages of ER and PR positivity for periodic trend analysis were derived from the mean frequency observed across 71 laboratories, using a 10% threshold in the United Kingdom’s National External Quality Scheme.

8. **Clinical Validity of Oncotype DX**

Can Oncotype DX reliably determine levels of expression of HR pathway?

- **Oncotype DX can accompany IHC results but should not replace them.**
  - Fourteen studies investigating Oncotype DX have evaluated the predictive validity of the recurrence score in ER-positive breast cancer, and none were found that correlated the expression of the HR-related genes to clinical outcome (146-159). The working group decided there was insufficient evidence to suggest that HR assessed with Oncotype DX is a better predictive or prognostic indicator than HR assessed using IHC.

**FUTURE RESEARCH**
Future research that would provide valuable information for these recommendations would include studies that validated the score derived from the HR-related genes assessed by Oncotype DX, prospective trials that included patient outcomes comparing image analysis with manual estimation, and studies comparing 1% versus 10% thresholds as predictive HR markers.

**Methods**

**Targeted Peer Review:** During the guideline development process, six targeted peer reviewers from Ontario and Alberta, considered to be clinical and/or methodological experts on the topic, were identified by the authors. Several weeks prior to completion of the draft report, the nominees were contacted by email and asked to serve as reviewers. All six potential reviewers agreed, and the draft report and a questionnaire were sent via email for their review. The questionnaire consisted of items evaluating the methods, results, and interpretive summary used to inform the draft recommendations and whether the draft recommendations should be approved as a guideline. Written comments were invited. The questionnaire and draft document were sent out on December 15, 2010. Follow-up reminders were sent at two weeks (email) and at four weeks (telephone call). The authors reviewed the results of the survey.

**Professional Consultation:** Feedback was obtained through a brief online survey of health care professionals who are the intended users of the guideline. Members of the Canadian Association of Pathologists and members of the Quality Management Program Laboratory Services were contacted by email to inform them of the survey. Participants were asked to rate the overall quality of the guideline (Section 1) and whether they would use and/or recommend it. Written comments were invited. Participants were contacted by email and directed to the survey Web site where they were provided with access to the survey, the guideline recommendations (Section 1), and the evidentiary base (Section 2). The notification email was sent on December 15, 2010. The consultation period ended on January 26, 2011. The authors reviewed the results of the survey.
**Results**

*Targeted Peer Review:* Six responses were received from six reviewers. Key results of the feedback survey are summarized in Table 1.

Table 1. Responses to nine items on the targeted peer reviewer questionnaire.

<table>
<thead>
<tr>
<th>Question</th>
<th>Reviewer Ratings (N=6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lowest Quality (1)</td>
</tr>
<tr>
<td>1. Rate the guideline development methods.</td>
<td>0</td>
</tr>
<tr>
<td>2. Rate the guideline presentation.</td>
<td>0</td>
</tr>
<tr>
<td>3. Rate the guideline recommendations.</td>
<td>0</td>
</tr>
<tr>
<td>4. Rate the completeness of reporting.</td>
<td>0</td>
</tr>
<tr>
<td>5. Does this document provide sufficient information to inform your</td>
<td>0</td>
</tr>
<tr>
<td>decisions? If not, what areas are missing?</td>
<td></td>
</tr>
<tr>
<td>6. Rate the overall quality of the guideline report.</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>7. I would make use of this guideline in my professional decisions.</td>
<td>0</td>
</tr>
<tr>
<td>8. I would recommend this guideline for use in practice.</td>
<td>0</td>
</tr>
<tr>
<td>9. What are the barriers or enablers to the implementation of this</td>
<td></td>
</tr>
<tr>
<td>guideline report?</td>
<td></td>
</tr>
<tr>
<td>One barrier is that the process to monitor the competency of</td>
<td></td>
</tr>
<tr>
<td>pathologists is difficult to enforce.</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Summary of written comments by targeted peer reviewers and modifications/actions taken.

<table>
<thead>
<tr>
<th>Summary of Written Comments</th>
<th>Modifications/Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Although it is gratifying that the recommendation notes that “longer fixation (e.g. &gt;72</td>
<td>See action for #2.</td>
</tr>
<tr>
<td>hours) does not exclude the specimen for HR testing” I am unclear as to the evidence</td>
<td></td>
</tr>
<tr>
<td>behind the current guideline of 72 hours, which is difficult to accommodate with long (e.g.</td>
<td></td>
</tr>
<tr>
<td>3-4 day) weekends. The kinetic studies of formaldehyde binding in tissue (Helander KG.</td>
<td></td>
</tr>
<tr>
<td>Biotechnic &amp; Histochemistry 1994;69(3):177-179) indicate that the cross-linking induced by</td>
<td></td>
</tr>
<tr>
<td>formaldehyde is completely reversible. Therefore extended fixation times is not a theoretical</td>
<td></td>
</tr>
<tr>
<td>obstacle to reliable ER testing. This is supported by the testimony of Bryan Hewlett (</td>
<td></td>
</tr>
<tr>
<td>Commission of Inquiry on Hormone Receptor Testing, Exhibit P-3359, page 24) when he</td>
<td></td>
</tr>
<tr>
<td>indicated that they have tested ER, HER2 and CD117 for up to 90 days of fixation time,</td>
<td></td>
</tr>
<tr>
<td>finding NO significant loss in immunoreactivity after 24 hours.</td>
<td></td>
</tr>
</tbody>
</table>
2. Fixation time: I would disagree that fixation time shorter than 8 hours should preclude specimen for ER testing. Some decrease in testing was observed with shorter fixation time. Knowing that most positive cases are clearly positive even with shorter fixation time, it would be unreasonable to “forbid” the use of such tissue for ER testing. It will suffice to state that other fixatives and other fixation conditions may be used if properly validated. Therefore, similarly to the statement for 72h fixation, I recommend to state at the end of paragraph: “Shorter fixation time does not exclude the specimen for HR testing if proper validation of such protocols is well documented.” Regarding 72h fixation as an upper limit, I suggest that the last sentence should also contain a statement on proper validation of such protocols.

3. Sample exclusion criteria: I would recommend that it is stated that “Samples may be excluded from testing based on …. “. However, I would allow attempting to test samples even if they are fixed shorter than 8h and even if fixation if delayed for more than 1h as some of the samples will be the only samples available for that patient and in many occasions, the results will still be acceptable as internal controls will show positivity and the tumors will show significant positivity. If results are negative and internal controls (if available) are negative, it should be stated that the testing was attempted, but false-negative results cannot be excluded due to….. (whatever suboptimal conditions occurred).

4. Antigen retrieval: I strongly disagree that any modifications from package inserts must be validated. Package inserts are irrelevant for any IHC product that is not FDA approved kit. If FDA approved kits are used, than no modifications are allowed. If FDA kit is not used, it is expected that the laboratory will test different conditions and identify the one that is optimally working and could be validated externally. Often, package inserts are completely discrepant with optimal conditions. Your statement that modifications from package inserts must be validated are giving completely wrong impression that package inserts should be followed, when this is not true.

5. I agree that the preference for testing should be given to tumour in the surgical excision specimen. However, if this approach is used, than additional recommendation is necessary, that is: “If tumour in the surgical excision specimen is HR-negative, the IHC testing should be repeated in the core biopsy if it is available.” This recommendation is necessary because it is not always possible to know if negative result is due to inadequate fixation even when internal controls are demonstrated as expression of HR in benign tissue is greatly variable and cannot be used as internal reference for calibration of sensitivity.

The recommendation “Underfixation is more critical than overfixation; fixation for less than 8 hours precludes the specimen from being used for HR testing.” was changed to “Underfixation is more critical than overfixation, however, shorter or longer fixation times do not exclude the specimen for HR testing if proper validation of such protocols is well documented.”

The recommendation “fixation for less than 8 hours precludes the specimen from being used for HR testing” was removed.

The recommendation for sample exclusion criteria under reporting elements “Excisional and needle biopsies fixed for less than 8 hrs, samples where fixation is delayed beyond 1 hour” was removed.

The recommendation for sample exclusion criteria under reporting elements “When controls were inappropriate” was added.

The recommendations “Any modifications from package inserts must be validated.” was changed to “Any modifications must be validated.”

Added “If the surgical specimen is adequately fixed and internal and external controls are concordant it is not mandatory to repeat testing on negative tumours.”
6. Fine needle Aspiration (FNA): Although it is not stated, it appears that immunocytochemistry (immunoassay using cytological smears) is not recommended. If this is the case, this should be strongly emphasized as on occasion cell pellets may not be available. Immunocytochemistry is a valid approach to HR demonstration, but only if protocols are validated. There are several good protocols (using various fixatives) that can be used. Cytological smears can also be fixed in 10% neutral formalin. I would personally recommend the following: “Cytological smears are not recommended for demonstration of HR status by immunocytochemistry. However, if immunoassays are internally validated, such protocols may be used if cell pellet is not available and no other tissue sampling is planned before the treatment has initiated.”

| Included “If not fixed in formalin, the protocol should be validated and have appropriate controls” |

7. Fixation type: Some argue that “10% NBF” is insufficient description of what type of formalin should be used as exact chemical formulations vary from one supplier to another. It is not clear if this variation has an impact on demonstration of HR. At least one should consider recommending that “when laboratory changes supplier of the 10% NBF, exact chemical composition should be checked and if different, the IHC protocol for HR should be validated for the new type of 10% NBF”.

| The recommendation “The formulation of NBF should be confirmed with each lot and if different should be validated.” was added. |

8. Storage of slides: Since TMAs are not recommended, there is no point in giving recommendations for handling TMA specimens. Mentioning TMAs distracts greatly from the main important message that sections should not be cut longer than one to couple of days prior to testing. Also, none of the studies that suggest that preservation of epitopes may be good by using special storing conditions were sufficiently documented and my personal experience is suggesting otherwise, especially for storing slides at 4°C with which I have particularly bad experience. All of the mentioned methods, as they are part of pre-analytical phase, need to be properly validated for use which is practically impossible.

| Under storage of slides, removed “especially with TMAs” and also removed “Storing at low temperatures (4°C) or in a nitrogen chamber, protection from light and paraffin coating may be used to preserve antigenicity in slides.” |

9. Controls do not need to include either positive or negative breast cancer cases. Breast cancer does not need to be included at all. It will suffice to include any tissues with predictable levels of expression to include high and low expressors. This is not always possible with tumors and if tumors are included, they have to be properly validated before inclusion with external laboratory. It is much better to recommend benign tissues that express ER and PR like endometrium, endocervix, and for weak expression one may consider tumors, but I recommend lung tumors that are often positive for ER and when positive, they are often weakly positive. As to variation of expression in these benign tissues, this argument is equally true for tumors, too. Prostate shows weak expression of PR in the stroma of BPH tissue which is often available in abundance.

| The recommendation “Internal controls of normal breast elements must be evaluated.” was changed to “Internal controls of normal breast elements must be evaluated when present on the section.” |

| The recommendation “For progesterone receptors, hyperplastic benign prostatic tissue could be used for controls.” was added. |

10. Controls: Specify which controls you are talking about (internal, external, positive, and negative as they all have their specific requirements and problems)

| The Working Group inserted which controls they were referring to in their recommendation. |
11. External positive controls should be run on each slide in an automated laboratory. If manual staining is done, this is also preferred way of how to use controls, but is less critical. One external control per run is not recommended.

The recommendation “However, on slide external controls for automated processes are preferred.” was added.

12. Manual staining may be superior if done properly. Manual staining should not be discouraged as evidence that has published results comparing the two could not be universally applied, but only refers to their internal methods of manual staining. Manual staining is often superior to automated staining if incubations are done overnight at 4°C which allows for more biologically appropriate dilution of the primary Ab and gives necessary time for Ab-Ag reaction to occur as it takes about 12 hours for such reaction. With automated staining we use oversaturated conditions to allow rapid reaction (about 30 min or less) for Ab-Ag binding, which always has a potential to cause false positive and false negative reactions. This is much less likely to occur with manual staining and longer incubation with less concentrated Ab and it often results with brilliant signal-to-noise results. Therefore, it is biologically incorrect to recommend automated staining, but it should rather recommend any protocols that are executed properly.

Added “validated manual staining may be used.”

13. I agree with the cut off points, but do not agree with recommendations that intensity does not need be reported. FDA-approved image analysis uses both percentage of positive cells and measure of intensity (1 to 3+). While actual H-scores are not required some descriptive estimate of intensity will not preclude, but has a potential to help proper stratification of the patients for therapies. Also, it does not take extra time from the pathologists to do so.

The recommendation “A statement of intensity is not a mandatory element in reporting but could be valuable if the tumour is low positive with weak intensity.” was changed to “A statement of intensity can be reported as weak, moderate or strong; however this is more important if the tumour is low positive with weak intensity.” Also, the explanation “The Working Group felt there was sufficient evidence to exclude staining intensity in the evaluation of HR positive and negative status. However, intensity could be reported with low-positive samples where clinical judgment may be necessary.” was changed to “Although the Working Group suggested that staining intensity could be included, they felt there was sufficient evidence to not make it a mandatory requirement in the evaluation of HR positive and negative status. The Working Group felt that reporting intensity was more important with low-positive samples where clinical judgment
14. Decalcification is not a problem if EDTA is used for decalcification. This is often not the case. I would suggest that this recommendation is given for decalcification of any bone tissue with suspected metastatic tumor. General statement that decalcified specimens are often negative is not acceptable to be included in guidelines.

The recommendation “Decalcified specimens are often negative.” was changed to “Controlled decalcified specimens with EDTA can be used.”

15. Caseload: concordance is often stated to be taken into account when QA of IHC is addressed. Concordance defined as at least 90% is meaningless if there is no definition of what parameters need to be agreed upon and how they are measured are also recommended. Statistically, concordance is difficult to calculate and in this case, we are not talking about concordance, but about agreement between observers. Therefore, kappa-value of 0.80 which defines near perfect agreement is much better measure. As this value is calculated, statistical power should also be defined, meaning that this recommendation needs to say exactly how many specimens need to be evaluated. It takes 80 specimens to have scientifically well acceptable power and it takes a minimum of 40 samples to have minimal power for diagnostic purposes.

The Working Group chose to align their recommendations for caseload with those developed by the ASCO/CAP guidelines by Fitgibbons et al 2010.

16. Under Internal QA, Terminology: “Any modification to the procedure” is not acceptable as it is not clear if this is addressing pre-analytical, analytical, or post-analytical component. If all of these components are addressed, it should be stated clearly as assumptions are often made that such statement is referring to analytical component only.

The recommendation “Any modification to the procedure requires additional validation to ensure accurate performance.” was changed to “Any modification to preanalytical, analytical, or postanalytical procedures requires additional validation to ensure accurate performance.”

17. Please state regarding “periodic trend analysis” that this refers to institution and it should be done by internal audit. I also strongly recommend the same for each pathologist as it may be very useful to be able to compare to other colleagues in the department and more widely.

The recommendation “Periodic trend analysis should be performed” was changed to “At least semiannual periodic trend analysis should be performed by internal audit for each institution and preferably for each pathologist”

18. Validation of new Ab or new protocols could be done with 50 to 100 samples. However, what will determine how valid this validation is, is the design of the validation samples. Validation samples should include about 25% strongly positive cases, at least 25% weakly positive cases, and 30 to 50% negative cases. If this is not the cases, all calculations will be irrelevant.

The recommendation “Adequate validation should be ensured, preferably by using approximately 30% cases that are unequivocally positive, approximately 30% cases that are low positive and approximately 40% cases that are negative.” was added.

19. I recommend considering an article Applied Immunohistochemistry & Molecular Morphology. 17(5):375-382, October 2009 to be considered for inclusion for consideration when EQA is discussed.

This article by Terry et al 2009 is outside the literature search dates of the systematic review; however, the Canadian EQA program mentioned in the article (Canadian
Professional Consultation: Thirty-five out of 53 (66%) responses were received. Key results of the feedback survey are summarized in Table 2.

Table 3. Responses to four items on the professional consultation survey.

<table>
<thead>
<tr>
<th>General Questions: Overall Guideline Assessment</th>
<th>Number (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lowest Quality (1)</td>
</tr>
<tr>
<td>1. Rate the overall quality of the guideline report.</td>
<td>0</td>
</tr>
<tr>
<td>2. I would make use of this guideline in my professional decisions.</td>
<td>Strongly Disagree (1)</td>
</tr>
<tr>
<td>3. I would recommend this guideline for use in practice.</td>
<td>0</td>
</tr>
</tbody>
</table>

4. What are the barriers or enablers to the implementation of this guideline report?
This guideline report would implement a standardized approach for the handling and testing of surgical specimens throughout the province and would further advance the necessity for standardized reporting. One barrier includes the difficulty in the adherence of these guidelines by pathologists and medical laboratory technologists in respect to their scope and practice. Another barrier remains the problem of off-site specimens arriving in a timely fashion to the laboratory. There is also a need to have “buy-in” from general surgeons to ensure that the preanalytic components are adequately addressed. Perhaps there may be minor financial barriers due to professional and technical hours required to implement these recommendations. Also in smaller institutions consultation fees with external reference laboratories may incur additional costs. A yearly onsite inspection would be difficult and expensive. Another of the potential barriers, breast surgery continues to be community based, where pathologists may not have the volume to maintain competency. As well, the lack of dissemination of information to all pathologists who routinely arrange or perform this testing is a barrier.

Table 4. Summary of written comments by professional consultants and modifications/actions taken.

<table>
<thead>
<tr>
<th>Summary of Written Comments</th>
<th>Modifications/Actions</th>
</tr>
</thead>
</table>
| 1. I did not make a direct comparison between these guidelines and those published by the ASCO/CAP group last April. However, it would be useful to know if there are any differences and to read an explanation thereof. | Under core biopsies, this statement was inserted: “These recommendations differ from the ASCO/CAP recommendations where core biopsies are preferred. The potential lack of normal ducts and lobules and tumour heterogeneity are reasons the
<table>
<thead>
<tr>
<th></th>
<th>Working Group preferred core biopsies. ASCO/CAP preferred core biopsies because cores may be exposed to more uniform and consistent tissue fixation. In the case of inadequate fixation of surgical specimens, the Working Group preferred core biopsies as well.”</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2.</strong></td>
<td>The guideline should state that only nuclear staining should be counted and scored. Occasionally, there is strong cytoplasmic staining but no nuclear staining (some apocrine carcinomas, for example). This needs to be ignored. Pathologists need to know this.</td>
</tr>
<tr>
<td><strong>Under thresholds</strong></td>
<td>“There are three categories for nuclear, not cytoplasmic, staining:” was added.</td>
</tr>
<tr>
<td><strong>3.</strong></td>
<td>Include medical laboratory technologists under intended users.</td>
</tr>
<tr>
<td><strong>The recommendation under intended users</strong></td>
<td>“Pathologists and breast cancer oncologists” was changed to “Any personnel involved in HR testing and interpretation and any health care provider involved in the management of breast cancer patients.”</td>
</tr>
<tr>
<td><strong>4.</strong></td>
<td>Intended users - the term “breast cancer oncologists” is vague. Specify as radiation oncologists, medical oncologists, surgical oncologists. Also add general surgeons. After all, a significant proportion of breast surgery is performed in community hospitals.</td>
</tr>
<tr>
<td><strong>See above #3</strong></td>
<td></td>
</tr>
<tr>
<td><strong>5.</strong></td>
<td>No mention, or maybe this is not applicable or I missed it, of a protocol for dealing with or significance of testing multicentric and multifocal invasive carcinoma.</td>
</tr>
<tr>
<td><strong>Under context</strong></td>
<td>“In cases of multifocal invasive carcinoma with different histotypes and/or grades separate HR testing should be conducted.” was added.</td>
</tr>
<tr>
<td><strong>6.</strong></td>
<td>There are no guidelines provided for retesting of breast cancers when they recur, either locally in the breast or axilla, or distant metastases. For example, should retesting be always performed after recurrence after treatment with tamoxifen or not?</td>
</tr>
<tr>
<td><strong>The target population was changed to</strong></td>
<td>“Tissue from adult female patients with primary or metastatic invasive or in situ breast cancer.”</td>
</tr>
<tr>
<td><strong>7.</strong></td>
<td>Another reason to NOT routinely perform ER PR on core biopsy specimens is that normal ducts and lobules are frequently not present in these core samples and thus it is hard to validate negative results.</td>
</tr>
<tr>
<td><strong>Changed</strong></td>
<td>“However, as they are derived from only a small sample of a larger tumour and in view of the heterogeneity in tumour HR expression, it is preferable to test the tumour in the surgical excision specimen.” to “However, as they are derived from only a small sample of a larger tumour where normal ducts and lobules are frequently not present and in view of the heterogeneity in tumour HR expression, it is preferable to test the tumour in the surgical excision specimen.”</td>
</tr>
<tr>
<td><strong>8.</strong></td>
<td>Under tissue processing, a very brief passing reference is made to “samples obtained remotely from the grossing laboratory.” This is insufficient</td>
</tr>
<tr>
<td><strong>The recommendation</strong></td>
<td>“If the laboratory is remote from the site of the surgery then...”</td>
</tr>
</tbody>
</table>
guidance for small community hospital general surgeons where an itinerant pathologist is not available. Should the surgeon ink the specimen and make some cuts into it? Sometimes, we receive mastectomy specimens for cancer that have been resected 48 hours earlier and were not adequately fixed in the first place, because of infrequent courier pickups from the community hospital to the academic hospital, 100 km away. The surgeons need to assume some of the responsibility, as part of a team approach to quality. In some cases, they are willing and able. But if this guideline were stronger in its wording and more descriptive in its guideline, it would put some pressure on those reluctant or unwilling surgeons to cooperate. The surgeon should ink and slice the sample and place it in formalin.” was added.

<table>
<thead>
<tr>
<th>9.</th>
<th>The Nordic.qc group recommends uterine muscle as the primary positive external control material, not positive and negative breast cancers, which can be variable in expression, and difficult to externally validate. Did the consensus group review their website? The bullet that refers to endometrium is not sufficiently emphasized as control material.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The recommendation “Other types of tissue such as the endometrium or myometrium could be used for controls.” was changed to “Other types of tissue such as the endometrium, myometrium, or uterine cervix could be used for controls.”</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10.</th>
<th>Antigen retrieval: All examples of retrieval methods and reagents should be removed from this section as they are technically specific to each laboratory providing this service. Blanket statements to optimization and validation should be included.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The recommendation “Antigen retrieval may be used to recover immunoreactivity from overfixed tissue; however the method (e.g., wet autoclave, pressure cooker, steamer, microwave, enzymatic digestion, water bath), reagents (e.g., citrate buffer, ethylenediaminetetraacetic acid) and duration must be optimized for each antibody.” was changed to “Antigen retrieval may be used to recover immunoreactivity from tissue; however the method, reagents and duration must be optimized for each antibody.”</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>11.</th>
<th>Antibodies: Remove the examples. Home-brewed antibodies cannot be used. All antibodies used for ER and PR testing must be Health Canada approved and have a Class II medical device license.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The recommendation “Currently used, or commercially available, validated antibodies such as ER antibodies 1D5, 6F11, and SP1 and PR antibodies 1A6, 1294, and 312 may be used; however, any modification to the product insert must be validated.” was changed to “Currently approved antibodies by Health Canada that have a Class II medical device license may be used. The Working Group chose to recommend antibodies approved by Health Canada based on the suggestion from external reviewers. Also, the Quality Management Program – Laboratory</td>
<td></td>
</tr>
</tbody>
</table>

DEVELOPMENT & REVIEW - page 22
<table>
<thead>
<tr>
<th>Services in Ontario recognizes all tests approved and licensed by Health Canada.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The asterisk in front of “Antibody clone and vendor” was removed.</td>
</tr>
<tr>
<td>The recommendation “Technologists should undergo appropriate training.” was changed to “Medical laboratory technologists who perform Class II IHC testing should undergo appropriate training.”</td>
</tr>
<tr>
<td>The recommendation “The evidence to guide exact caseloads per laboratory or pathologist is not available. However, laboratories could use the recommendations for HER2, but this is not mandatory. These should be carried out by experienced pathologists at laboratories participating in IHC external QA accreditation programs.” was added.</td>
</tr>
<tr>
<td>Unsatisfactory performance results will be addressed according to the regulations of the accreditation program. The recommendation “Satisfactory performance requires at least 90% correct responses in graded challenges of several levels of HR expression, as well as a negative control.” was removed. This would depend on the accreditation agency. The recommendation “Unsatisfactory performance will require a laboratory to respond according to accreditation agency program requirements.” was removed.</td>
</tr>
<tr>
<td>If a site is to select one external QA program, QMPLS-EQA is mandatory. I am not sure why CAP is included especially as the actual slides are not reviewed by CAP. These two organizations work in completely different ways.</td>
</tr>
<tr>
<td>The Working Group did not see the need to restrict the external QA process to a Canadian program.</td>
</tr>
<tr>
<td>The Working Group felt it should be determined in each lab: director of immunohistochemistry or director of surgical pathology or a designated breast pathologist.</td>
</tr>
</tbody>
</table>
Conclusion
This EBS report reflects the integration of feedback obtained through the external review process with final approval given by the authors and the Report Approval Panel of the PEBC. Updates of the report will be conducted as new evidence informing the question of interest emerges.

Funding
The PEBC is a provincial initiative of Cancer Care Ontario supported by the Ontario Ministry of Health and Long-Term Care through Cancer Care Ontario. All work produced by the PEBC is editorially independent from its funding source.

Copyright
This report is copyrighted by Cancer Care Ontario; the report and the illustrations herein may not be reproduced without the express written permission of Cancer Care Ontario. Cancer Care Ontario reserves the right at any time, and at its sole discretion, to change or revoke this authorization.

Disclaimer
Care has been taken in the preparation of the information contained in this report. Nonetheless, any person seeking to apply or consult the report is expected to use independent medical judgment in the context of individual clinical circumstances or seek out the supervision of a qualified clinician. Cancer Care Ontario makes no representation or guarantees of any kind whatsoever regarding the report content or use or application and disclaims any responsibility for its application or use in any way.

Contact Information
For further information about this report, please contact:

Dr. Wedad Hanna, Professor, Department of Laboratory Medicine and Pathobiology
Faculty of Medicine, University of Toronto
Fax: 416-480-4271   Email: Wedad.Hanna@sunnybrook.ca

For information about the PEBC and the most current version of all reports, please visit the CCO website at http://www.cancercare.on.ca/ or contact the PEBC office at:
Phone: 905-527-4322, ext. 42842   Fax: 905-526-6775   E-mail: ccopgi@mcmaster.ca
REFERENCES


