Optimizing breast cancer survival models based on conventional biomarkers and stromal parameters

Dekker, T.J.A.; Dekker T.J.A.

Citation

Cover Page

The handle http://hdl.handle.net/1887/55957 holds various files of this Leiden University dissertation

Author: Dekker T.J.A.
Title: Optimizing breast cancer survival models based on conventional biomarkers and stromal parameters
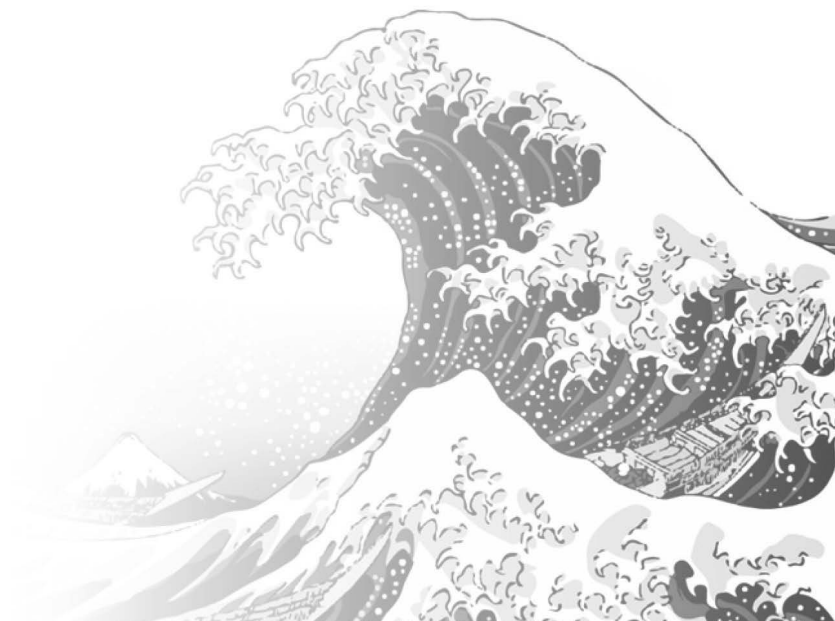**Date**: 2017-09-26

4

Quality assessment of estrogen receptor and progesterone receptor testing in breast cancer using a tissue microarray-based approach

T.J.A. Dekker, S. ter Borg, G.K.J. Hooijer, S.L. Meijer, J. Wesseling, J.E. Boers,
E. Schuuring, J. Bart, J. van Gorp, P. Bult, S.A. Riemersma, H.F.B.M. Sleddens,
W.E. Mesker, J.R. Kroep, V.T.H.B.M. Smit, M.J. van de Vijver

Introduction

Testing estrogen receptor (ER) expression is mandatory for all breast carcinomas as this biomarker predicts response to estrogen-modulating therapy [1]. Adequate testing of ER expression via immunohistochemistry is considered the gold standard for selecting patients for neoadjuvant and adjuvant hormonal therapies [2]. The progesterone receptor (PR) has been assessed as a prognostic factor [3] and as a potential predictive marker [4, 5]. Initial studies on the quality of hormone receptor (HR) testing have shown cause for concern with a low percentage of laboratories showing acceptable performance [6]. An American Society of Clinical Oncology (ASCO) and College of American Pathologists (CAP) panel addressed the need for improving ER and PR testing and published a set of guidelines concerning this matter [7]. Recommendations were also made to lower the positivity threshold from 10% to 1%. Unfortunately, a significant (although decreasing) number of laboratories still fail to achieve sufficient testing quality in the NordiQC and/or NEQAS ER and PR assessment runs.

This current study was designed to evaluate a tissue microarray (TMA)-based method for assessing ER and PR testing quality. This method allows pathology laboratories to evaluate the reproducibility of IHC testing results by retesting a high number of ER and PR assays on TMAs. By comparing the original result to the retested assay on TMAs, discordances between local report and retested tumors can be easily assessed at large scale. Additionally, the effect of the recommended threshold change of 10 to 1% positive cells on testing reproducibility was investigated.

Methods

Tissues

Formalin-fixed paraffin-embedded (FFPE) tumors blocks were collected for TMA construction from 9 laboratories in the Netherlands; the Academic Medical Center (AMC, Amsterdam), Netherlands Cancer Institute/Antoni van Leeuwenhoek (NKI/AVL, Amsterdam), Diakonessenhuis (Utrecht), Isala (Zwolle), Leiden University Medical Center (LUMC, Leiden), University Medical Center Groningen (UMCG, Groningen), Eramus Medical Center (EMC, Rotterdam), Radboud University Medical Center (Radboud UMCNijmegen) and Laboratory Pathology Eastern Netherlands (LabPON) (table S1). The tissue blocks contained invasive breast carcinomas that were previously tested for ER, PR and/or HER2 expression by immunohistochemistry as part of routine pathological diagnostics. HER2 testing quality for a subset of the included tumors was investigated in a previous publication [8]. According to Dutch law, these tissue

Quality assessment of estrogen receptor and progesterone receptor testing in breast cancer
using a tissue microarray-based approach
Breast Cancer Research and Treatment 2015; 152(2): 247-252

blocks can be freely used for research purposes after anonymization, provided these are handled according to national ethical guidelines ('Code for Proper Secondary Use of Human Tissue', Dutch Federation of Medical Scientific Societies). TMA sections were stained with SP1 (for ER) and 1E2 (for PR) antibodies using the Benchmark XT autostainer (Ventana Medical Systems, Tucson, AZ, United States).

Comparison of ER and PR test results

The TMA cores were scored by determining the percentage of nuclear staining, invasive tumor cells (staining intensity was not accounted) in increments of 10%. ER and PR results from the original tests were retrieved from the local pathology reports. These ER and PR scores were compared to the results that were obtained from the TMA cores. For discordant cases, whole tissue sections were sectioned and stained for ER and PR. This was done to rule out that discordant results were due to sampling errors introduced by the use of TMAs. If the results between the local pathology reports were concordant with the whole slide, the final result was considered concordant. If the result was still discordant with the original pathology report, this tumor was considered as truly discordant and the reason for the discordancy was then investigated. For this purpose, the original slides used for the local ER and PR diagnosis were centrally reviewed. If the revision of the original testing slide by the central revision panel revealed discordance with the local observer, the reason of the discordant result was considered to be observer inaccuracy. If the original testing slide showed positive nuclear staining in revision, but this positive IHC result could not be reproduced on both TMA and subsequent whole-sized slides despite appropriate positive controls, the reason for the discordant result was a false positive IHC procedure. In case of the opposite result (negative local IHC result with ER-positive results on TMA and whole-sized slides), the reason of discordance was considered to be inaccurate IHC leading to false-negative results. The workflow of the study is summarized in figure S1.

Adjustment from 10% to 1% threshold for HR positivity

Since all these materials were originally tested prior to the recommended threshold of 1% for hormone receptor positivity, we then investigated the influence of the change of this threshold from 10% to 1% positive cells as is recommended by the ASCO/CAP guidelines. For all discordant cases, we investigated whether this discordancy would still exist after changing this scoring methodology.

Results

ER concordance

A number of 1736 invasive breast carcinomas that were tested for ER in 9 different pathology laboratories were included in the study. Of these, 163 tumors were omitted from the analysis when the original ER result could not be retrieved, TMA cores were lost during the staining procedure or due to the absence of invasive breast cancer on the TMA cores. A further 4 tumors were excluded because material was not available for subsequent retesting after an initial discordant result was found between the TMA and the original testing result. The subsequent analysis was performed on the remaining cohort of 1569 breast tumors (figure 1). When comparing the local testing result with the TMA result, 52 tumors were considered to be discordant. For these tumors, the whole-sized sections were stained for ER in order to assess the reason for discordance. If the whole-slide result was concordant with the original ER testing result, the discordance was decided to be due to sampling error due to use of a TMA and the final results were thus concordant (*N*=36). If the discordance remained, this was considered a true discordant result (*N*=16). Of the 16 discordant cases, 12 were false-positive and 4 were false-negative (figure 1 and table 1). Overall concordance was 99.0% and the sensitivity and specificity for all ER tests performed by the combined 9 centers showed a sensitivity of 99.7% (range 98.7-100.0%) and specificity of 95.4% (range 83.3-100.0%). Positive predictive value (PPV) and negative predictive value (NPV) for all centers combined were 99.1% (range 97.4-100.0%) and 98.4% (range 90.9-100%).

The next step was to investigate whether the discordant results were due to observer inaccuracy or inaccurate IHC procedures. To assess the possibility of observer error, the original slides were revised when available (*N*=15). In 12 tumors, discordance between the local observer and the revision panel was present, which can be considered to be observer inaccuracy. Three discordant cases were due to inaccurate IHC procedures. Two showed ER positive staining in the local testing center (which was also verified with slide revision), while no positive test result was obtained if the staining was repeated (example shown in figure 2). The opposite was true for the third discordant case. The reason for the discordant result could not be ascertained for the sole remaining tumor, since the unavailability of the original slide leaves it impossible to determine whether the discordance was due to inaccurate scoring or IHC procedure (table 1).

Quality assessment of estrogen receptor and progesterone receptor testing in breast cancer
using a tissue microarray-based approach
Breast Cancer Research and Treatment 2015; 152(2): 247-252

PR concordance

A number of 1518 PR tested cases were provided by 8 laboratories that performed PR testing. A number of 171 cases were excluded from the final analysis. This left a number of 1347 PR-tested tumors available for the comparison with the TMA results (figure S2). A total number of 150 tumors were discordant between the original PR testing result and the TMA, and for all these cases, the PR test was performed centrally on a whole tissue block. True discordant results were seen in 80 cases, which led to an overall concordance of 94.1%. Of these 80 discordant cases, 32 tumors were deemed false-positive and 48 tumors were considered false-negative (table S2 and figure S2). Overall sensitivity and specificity for PR testing was slightly lower than for ER testing, with overall sensitivity of 94.8% and overall specificity of 92.6%. Sensitivity and specificity values of individual laboratories ranged from 87.1-97.8% and 85.7-97.0%, respectively. PPV and NPV overall were 96.4% (range 92.6-98.7%) and 89.3% (range 80.0-96.6%). With the aid of the revision of the local PR-test (available for 59 of the 80 tumors) and the whole tissue retesting, the reason for discordant results was investigated. Observer inaccuracy was detected in 20 cases, and the IHC test was irreproducible in 39 cases (table S2).

Consequence of threshold adjustment

All discordant cases were again reviewed to determine whether adjusting the original or retested ER or PR result, based on the 2010 ASCO/CAP guidelines, would influence the discordant result. For some cases, this required the availability of data regarding the number of HR-positive cells (if any) observed during the original, local HR testing. This is important in the case of a tumor that was determined to be negative at local testing according to the 10% cut-off, since such tumors might either be completely negative or have some positive staining but less than 10% overall. For some cases this information was unavailable in the pathology report (*N*=8). Regardless, out of 96 initially discordant results, applying the recommended 1% cut-off lead to a concordant result for 36 tumors (further described in table 3).

Discussion

Our study assessed the reproducibility of immunohistochemical ER and PR testing performed in 9 testing laboratories in the Netherlands. For this purpose, TMAs were used to facilitate retesting relatively high numbers of previously tested tumors and thus provide an accurate assessment of the reproducibility of these IHC tests. We compared the original ER and PR result from the pathology archives with the result

that was detected on TMA. For discordant results, whole tissue sections were tested to rule out the possibility of sampling error. If a tumor tested negative at a local center, but showed positive HR expression on both TMA and whole slide examination, this tumor is likely to indeed have HR expression. If a tumor shows positive HR expression at the local center, but both TMA and whole-sized stainings are unable to replicate this staining (despite appropriate internal and external controls), it is hard to say whether the first positive result was truly false-positive. Careful examination of the slide with knowledge of expected staining patterns might however be helpful (figure 2). Unfortunately, no gold standard exists that could have been used to determine which assessment is correct which remains a weakness of this study design. Response to hormonal therapy should be the gold standard in these cases, but this is also dependent on other known and unknown variables and information regarding hormonal response is not always available. Viale et al. showed that the group of tumors that were locally ER-positive while centrally ER-negative tended to follow the overall survival patterns of ER-negative tumors (namely early relapse with following plateau, whereas ER-positive tumors follow a slower rate of relapse) [9]. These observations speak in favour of centrally performed HR tests in general, but this cannot be applied to each individual. Other studies have used RT-PCR as an additional method for determining HR status in addition to local and central IHC, but these assays are neither free from reproducibility issues themselves nor have these been shown to correlate more closely to response to hormonal therapy [10].

Fortunately, concordance between local and retested HR results was high for both ER (99.0%) and PR (94.1%) in this current study. Remarkably, irreproducible test results obtained for ER were only rarely due to errors in the IHC procedure whereas the ratio of IHC procedure error to observer error was more balanced in the PR tested group. This might be due to the quality of the antibodies, as traditionally more emphasis has been placed on ER testing quality.

A 2010 report by an ASCO/CAP panel has suggested lowering the threshold of positivity from 10% HR positive cells to 1%. These guidelines were established along a similar methodology as an earlier report concerning HER2 testing which recommended increasing the positivity threshold to 30% positive cells [11]. The ER/PR guideline adjustments were not designed to improve testing accuracy, but were based on the observation that even patients with low percentage HR cells (1-10%) still respond to tamoxifen. This is despite the observation that most tumors with 1-10% HR+ cells share more common biologic features with ER- tumors [12]. Regardless, this change might also have consequences for HR testing reproducibility in this rare [13] group of tumors, which was investigated in this study. We found that a substantial number

Quality assessment of estrogen receptor and progesterone receptor testing in breast cancer
using a tissue microarray-based approach
Breast Cancer Research and Treatment 2015; 152(2): 247-252

of these cases that were discordant between local and TMA testing were concordant when following the 2010 ASCO/CAP guidelines, suggesting that adherence to the 2010 guidelines improves the reproducibility of HR testing results.

Central assessment of ER and PR status of tumors that were included into the Breast International Group (BIG) 1-98 trial showed that locally-tested ER-negative tumors tend to show ER positivity in a relatively high number of cases (69.5%)[9]. Discordance was even more pronounced for PR testing [9]. Retesting of HR-tested tumors included in the Eastern Cooperative Oncology Group (ECOG) study E2197, showed a concordance of 90% and 84%, respectively between locally tested and centrally tested ER and PR results [10]. Central review of local HR testing performed in the Adjuvant Lapatinib and/or Trastuzumab Treatment Optimisation (ALTTO) showed that local ER-positives could not be reproduced for 4.3% of cases. Even more worrisome was the poor reproducibility of 21.6% of ER-negative results which displayed positive staining when retesting of the original result was performed [14]. All of these studies indicate (i) a relatively poor reproducibility of ER-negative test results, (ii) average reproducibility of ER testing below 95%, (iii) an even lower reproducibility for PR testing. A 2014 report by Viale et al. published the concordance from the ER and PR testing performed locally for the first 800 participants of the MINDACT trial with central IHC retesting [15]. Concordance for ER and PR IHC tests was determined as 97.6% and 89.6% respectively. These last results and ours indicate an improving trend in ER and PR testing reproducibility. The relatively high reproducibility in our study might be explained by the routine use of autostainers among all participating laboratories. Also, the participating centers in this study were all accredited laboratories in the Netherlands, leaving the question whether these results apply to all individual centers.

Continuous improvement of local IHC methods and validation of these is of essential importance to provide and maintain optimal care for breast cancer patients. Participation in such quality control schemes should be considered as mandatory for every individual HR-testing laboratory. The tissue microarray approach described in this study can provide important feedback regarding testing reproducibility.

Acknowledgments

Funding

Conflict of interest

Quality assessment of estrogen receptor and progesterone receptor testing in breast cancer
using a tissue microarray-based approach
Breast Cancer Research and Treatment 2015; 152(2): 247-252

References

1. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials: Lancet 2005;365:1687-1717

2. Harvey JM et al.: Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. J Clin Oncol 1999;17:1474-1481

3. Reiner A et al.: Immunocytochemical localization of estrogen and progesterone receptor and prognosis in human primary breast cancer. Cancer Res 1990;50:7057-7061

4. Bardou VJ et al.: Progesterone receptor status significantly improves outcome prediction over estrogen receptor status alone for adjuvant endocrine therapy in two large breast cancer databases. J Clin Oncol 2003;21:1973-1979

5. Davies C et al.: Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. Lancet 2011;378:771-784

6. Rhodes A et al.: Study of interlaboratory reliability and reproducibility of estrogen and progesterone receptor assays in Europe. Documentation of poor reliability and identification of insufficient microwave antigen retrieval time as a major contributory element of unreliable assays. Am J Clin Pathol 2001;115:44-58

7. Hammond ME et al.: American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. J Clin Oncol 2010;28:2784-2795

8. Dekker TJ et al.: Determining sensitivity and specificity of HER2 testing in breast cancer using a tissue micro-array approach. Breast Cancer Res 2012;14:R93

9. Viale G et al.: Prognostic and predictive value of centrally reviewed expression of estrogen and progesterone receptors in a randomized trial comparing letrozole and tamoxifen adjuvant therapy for postmenopausal early breast cancer: BIG 1-98. J Clin Oncol 2007;25:3846-3852

10. Badve SS et al.: Estrogen- and progesterone-receptor status in ECOG 2197: comparison of immunohistochemistry by local and central laboratories and quantitative reverse transcription polymerase chain reaction by central laboratory. J Clin Oncol 2008;26:2473-2481

11. Wolff AC et al.: Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. J Clin Oncol 2013;31:3997-4013

12. Iwamoto T et al.: Estrogen receptor (ER) mRNA and ER-related gene expression in breast cancers that are 1% to 10% ER-positive by immunohistochemistry. J Clin Oncol 2012;30:729-734

4

13. Collins LC et al.: Bimodal frequency distribution of estrogen receptor immunohistochemical staining results in breast cancer: an analysis of 825 cases. Am J Clin Pathol 2005;123:16-20

14. Gelber RD, Gelber S: Facilitating consensus by examining patterns of treatment effects. Breast 2009;18 Suppl 3:S2-S8

15. Viale G et al.: High concordance of protein (by IHC), gene (by FISH; HER2 only), and microarray readout (by TargetPrint) of ER, PgR, and HER2: results from the EORTC 10041/BIG 03-04 MINDACT trial. Ann Oncol 2014;25:816-823

Quality assessment of estrogen receptor and progesterone receptor testing in breast cancer
using a tissue microarray-based approach
Breast Cancer Research and Treatment 2015; 152(2): 247-252

Table 1. Discordant ER results.

N	Local ER testing result	ER result after revision of original slide	TMA and whole slide ER result	Conclusion	Reason for discordance
1	Negative	Negative	Positive	False negative	IHC error
3	Negative	Positive	Positive	False negative	Observer error
2	Positive	Positive	Negative	False-positive	IHC error
9	Positive	Negative	Negative	False positive	Observer error
1	Positive	Unknown	Negative	False positive	Unknown

Table 2. Discordant results reevaluated according to 2011 ASCO/CAP guidelines.

N	Percentage of HR positive cells in local result	Threshold at 1%	Threshold at 10% (reported in pathology report)	Percentage of HR positive cells at retesting	Threshold at 1%	Threshold at 10%	Discordant at 1% threshold?	Discordant at 10% threshold?
12	< 10% but ≥ 1%	Positive	Negative	≥10%	Positive	Positive	Yes	No
24	≥ 10%	Positive	Positive	< 10% but ≥ 1%	Positive	Negative	Yes	No
8	Not reported	Unknown	Negative	≥10%	positive	positive	Yes	Unknown
29	≥ 10%	Positive	Positive	0%	Negative	Negative	Yes	Yes
23	0%	Negative	Negative	≥ 10%	Positive	Positive	Yes	Yes

4

Figure 1. Concordance for ER testing results.

Quality assessment of estrogen receptor and progesterone receptor testing in breast cancer
using a tissue microarray-based approach
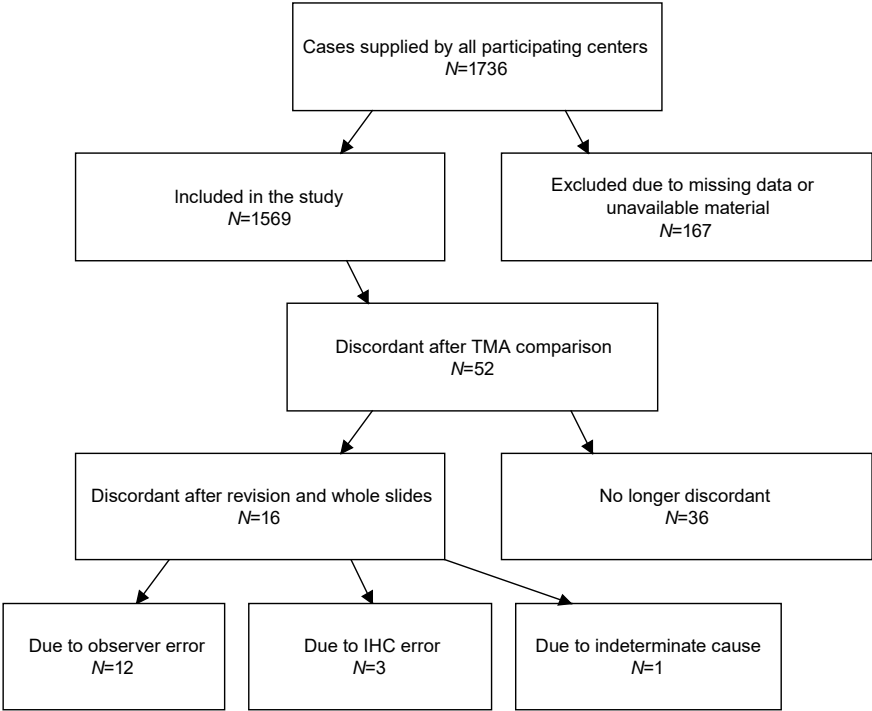Breast Cancer Research and Treatment 2015; 152(2): 247-252

Figure 2. A case where the local result was determined as ER-positive, while this staining was not reproduced on the TMA core and whole slide testing. (A) The local slide which showed both nuclear and smudgy, weaker cytoplasmic staining in the tumor cells as well as associated fibroblasts. A nearby duct is strongly positive. (B) The TMA test showing no staining in tumor cells. (C) Whole size test which verified the ER-negative staining of the TMA, while the normal duct shows an appropriate positive control.
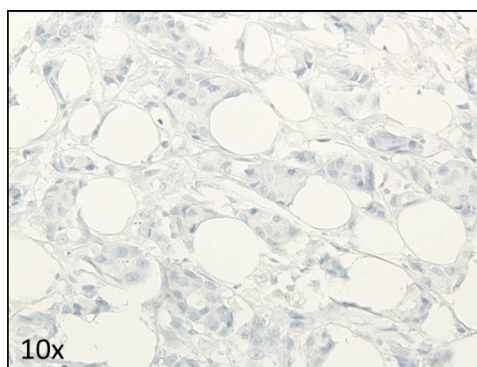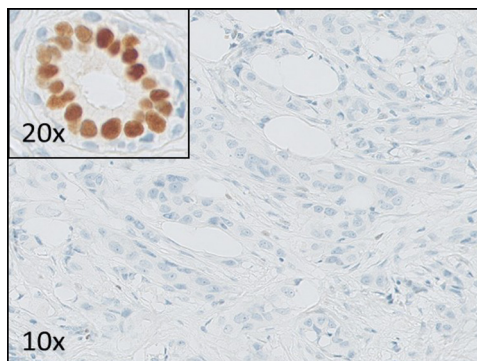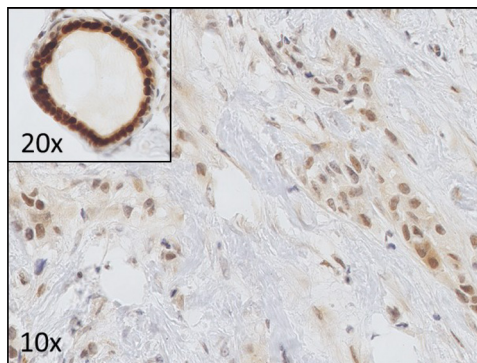
Table S1. Evaluable cases provided by local testing centers and corresponding original testing results.

Center	Antibody clone for ER and PR	ER-tested cases	ER+	ER-	PR-tested cases	PR+	PR-
1	SP1, 1E2	147	127	20	150	110	40
2	1D5, 1A6	296	264	32	297	198	99
3	1D5, PgR1294	179	143	36	176	113	163
4	SP1	208	186	22	Does not perform PR testing	Does not perform PR testing	Does not perform PR testing
5	EP1, PgR 636	99	83	16	95	67	28
6	SP1, 1E2	184	138	46	183	110	73
7	SP1, PgR 636	153	115	38	153	95	58
8	SP1, PgR 636	88	77	11	79	59	20
9	SP1/EP1, PgR 636	215	185	30	214	158	56
Total		1569	1318	251	1347	910	437

Table S2. Discordant PR results.

N	Local PR testing result	PR result after revision of original slide	TMA and whole slide ER result	Conclusion	Reason for discordance
27	Negative	Negative	Positive	False negative	IHC error
12	Negative	Positive	Positive	False negative	Observer error
12	Positive	Positive	Negative	False-positive	IHC error
8	Positive	Negative	Negative	False positive	Observer error
9	Negative	Unknown	Positive	False negative	Unknown
12	Positive	Unknown	Negative	False positive	Unknown

Quality assessment of estrogen receptor and progesterone receptor testing in breast cancer
using a tissue microarray-based approach
Breast Cancer Research and Treatment 2015; 152(2): 247-252

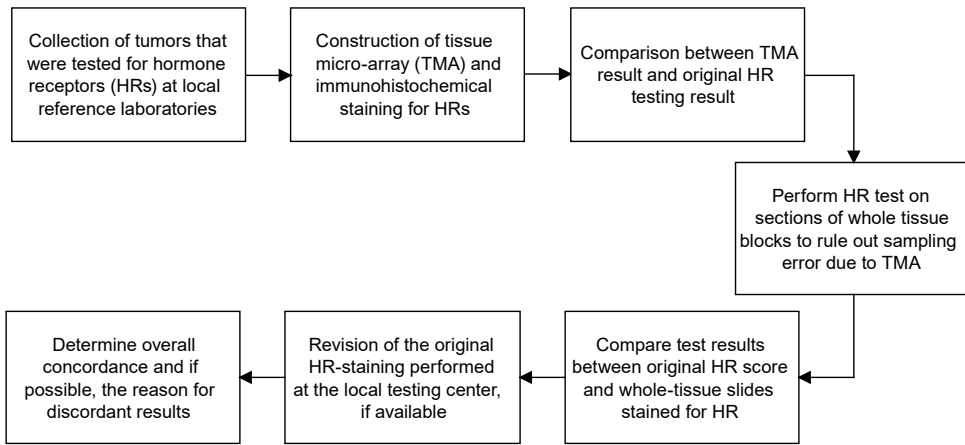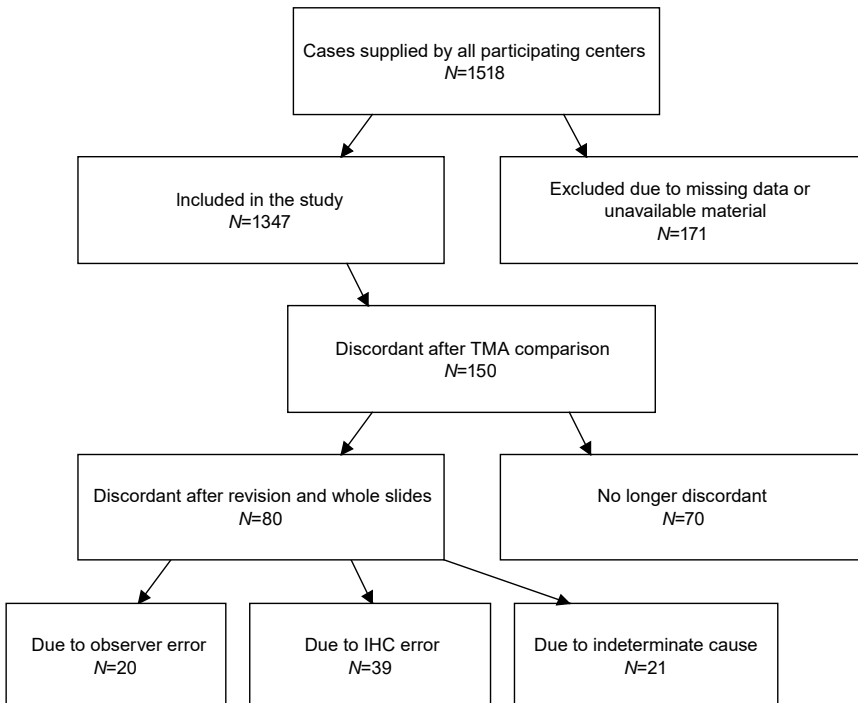**Figure S1.** Study workflow. HR = hormone receptors, TMA = tissue micro-array.

Figure S2. Concordance for PR testing results.