



Article

Variability in Breast Cancer Biomarker Assessment and the Effect on Oncological Treatment Decisions: A Nationwide 5-Year Population-Based Study

Balazs Acs^{1,2,†}, Irma Fredriksson^{3,4,†} , Caroline Rönnlund^{1,2}, Catharina Hagerling⁵, Anna Ehinger^{5,6}, Anikó Kovács⁷, Rasmus Røge^{8,9} , Jonas Bergh^{1,10} and Johan Hartman^{1,10,*}

- ¹ Department of Oncology and Pathology, Karolinska Institutet, 17176 Stockholm, Sweden; balazs.acs@ki.se (B.A.); caroline.ronnlund@ki.se (C.R.); jonas.bergh@ki.se (J.B.)
 - ² Department of Clinical Pathology and Cytology, Karolinska University Laboratory, 11883 Stockholm, Sweden
 - ³ Department of Breast, Endocrine Tumors and Sarcoma, Karolinska University Hospital, 17176 Stockholm, Sweden; irma.fredriksson@ki.se
 - ⁴ Department of Molecular Medicine and Surgery, Karolinska Institutet, 17176 Stockholm, Sweden
 - ⁵ Division of Clinical Genetics, Department of Laboratory Medicine, Lund University, 22185 Lund, Sweden; catharina.hagerling@med.lu.se (C.H.); anna.ehinger@med.lu.se (A.E.)
 - ⁶ Division of Oncology and Pathology, Department of Clinical Sciences Lund, Lund University, 22184 Lund, Sweden
 - ⁷ Department of Clinical Pathology, Sahlgrenska University Hospital, 41345 Gothenburg, Sweden; aniko.kovacs@vgregion.se
 - ⁸ Department of Clinical Medicine, Aalborg University, 9000 Aalborg, Denmark; rr@rn.dk
 - ⁹ NordiQC, Institute of Pathology, Aalborg University Hospital, 9000 Aalborg, Denmark
 - ¹⁰ Breast Center, Cancer Theme, Karolinska University Hospital and Karolinska Comprehensive Cancer Center, Gävlegatan 55, 17164 Solna, Sweden
- * Correspondence: johan.hartman@ki.se; Tel.: +46-73-976-02-42
† These authors contributed equally to this work.



Citation: Acs, B.; Fredriksson, I.; Rönnlund, C.; Hagerling, C.; Ehinger, A.; Kovács, A.; Røge, R.; Bergh, J.; Hartman, J. Variability in Breast Cancer Biomarker Assessment and the Effect on Oncological Treatment Decisions: A Nationwide 5-Year Population-Based Study. *Cancers* **2021**, *13*, 1166. <https://doi.org/10.3390/cancers13051166>

Academic Editor: Rupert Bartsch

Received: 1 January 2021

Accepted: 26 February 2021

Published: 9 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Simple Summary: Biomarkers that define breast cancer treatment recommendations include estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth-factor receptor 2 (HER2); histological grade; and in many countries, the Ki67 proliferation index. However, the subjective nature and degree of variability in breast cancer biomarker assessment might result in under- or overtreatment. We demonstrated that limited variability exists in ER, PR, and HER2 positivity rates among 29 departments in Sweden, including 43,261 patients. However, even a few outlier labs affect endocrine and anti-HER2 treatment rates in a clinically relevant proportion, indicating a need for improvement. Despite international guidelines, standardized protocols, and external quality control procedures, very high variability was found in Ki67 scoring and histological grading, indicating a need for new methods. Monitoring rates of biomarker expression and treatments among departments should be mandatory in order to detect variability issues affecting the clinical management of breast cancer.

Abstract: We compared estrogen receptor (ER), progesterone receptor (PR), human epidermal growth-factor receptor 2 (HER2), Ki67, and grade scores among the pathology departments in Sweden. We investigated how ER and HER2 positivity rates affect the distribution of endocrine and HER2-targeted treatments among oncology departments. All breast cancer patients diagnosed between 2013 and 2018 in Sweden were identified in the National Quality Register for Breast Cancer. Cases with data on ER, PR, HER2, Ki67, grade, and treatment were selected (43,261 cases from 29 departments following the guidelines for biomarker testing). The ER positivity rates ranged from 84.2% to 97.6% with 6/29 labs out of the overall confidence intervals (CIs), while PR rates varied between 64.8% and 86.6% with 7/29 labs out of the CIs. HER2 positivity rates ranged from 9.4% to 16.3%, with 3/29 labs out of the overall CIs. Median Ki67 varied between 15% and 30%, where 19/29 labs showed significant intra-laboratory variability. The proportion of grade-II cases varied between 42.9% and 57.1%, and 13/29 labs were outside of the CI. Adjusting for patient characteristics, the proportion of endocrine and anti-HER2 treatments followed the rate of ER and HER2 positivity, illustrating the clinical effect

of inter- and intra-laboratory variability. There was limited variability among departments in ER, PR, and HER2 testing. However, even a few outlier pathology labs affected endocrine and HER2-targeted treatment rates in a clinically relevant proportion, suggesting the need for improvement. High variability was found in grading and Ki67 assessment, illustrating the need for the adoption of new technologies in practice.

Keywords: positivity rate; biomarker; breast cancer; endocrine treatment; HER2-targeted treatment; variability

1. Introduction

In the era of precision medicine, patient management of invasive breast cancer is largely based on high-quality tumor biomarker testing. Indeed, breast cancer care has been a pioneer in the use of molecular features of the tumor to guide clinical decision making. Beside stage and tumor size, the leading biomarkers that define breast cancer treatment recommendations are estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth-factor receptor 2 (HER2); histological grade; and in many countries, the Ki67 proliferation index [1]. Furthermore, ER and HER2 status are strongly predictive for anti-estrogen therapies [2,3] and HER2-targeted therapies [4,5]. All these biomarkers are established by the pathological analysis of tumor tissue, which, according to international guidelines, is mandatory in the analysis of all diagnosed primary breast cancer cases [1]. In clinical practice, immunohistochemical (IHC) detection is the gold standard for assessing ER, PR, and Ki67 in routine breast cancer pathology [6,7]. HER2 is analyzed by IHC, and in equivocal cases by in situ hybridization (ISH) [8]. IHC and ISH methods have several advantages, including wide applicability and availability with a low cost, ensuring the assessment of only invasive tumor cells, which may be challenging—especially in small tumors [9]. However, several studies have demonstrated moderate inter-observer and inter-laboratory reproducibility in breast cancer biomarker assessment [10,11]. Therefore, tests with proven analytical and clinical validity as well as clinical utility [12] are critical, as false-negative tests result in withholding effective treatment, while false-positive tests lead to overtreatment with costly and ineffective therapy, simultaneously resulting in unwanted direct and long-term side effects [13,14]. Parameters that predominantly influence the IHC and ISH results include pre-analytical (type of biopsy, tissue handling), analytical (IHC and ISH protocol), interpretive, and scoring steps [15]. Standardized procedures should be used in all steps of biomarker testing. International and national guidelines for the analysis and interpretation steps have been in place for several years [16]. In addition, to ensure compliance with validated protocols for testing, it is recommended that laboratories participate in regular internal and external quality control procedures like UK-NEQAS-ICC [17] or NordiQC [18]. The quality control work may also include comparisons between local and central laboratories.

Although external quality control has an important role in maintaining analytical validity, it cannot be said to represent daily routine, where time-constrained limitation always applies, and several pathologists perform the diagnostic evaluation. In quality-control sets it is often the most experienced pathologist with the highest skills in the specific field who performs the assessment. Furthermore, it is very difficult to compare labs using different antibodies, detection systems, and protocols. For this reason, the monitoring of positivity rates has been proposed as a tool to identify laboratories with insufficient assays and a high yield of false-positive or false-negative results [19,20]. A comparison of department performances in breast cancer biomarker testing is necessary to ensure that daily diagnostic practice meets clinical requirements.

In this study we compared the 5-year distribution of ER, PR, and HER2; Ki67 proliferation index; and histological grade evaluation scores from daily clinical practice among

pathology departments in Sweden. We also investigated how ER and HER2 positivity rates affect the use of ER- and HER2-targeted treatments among oncology departments.

2. Materials and Methods

2.1. Data Source and Study Population

We retrieved patient data from NKBC, the Swedish National Quality Register for Breast Cancer (<https://statistik.incanet.se/brostcancer/>), which contains detailed clinical data on patient and tumor characteristics, treatment, and follow-up for all newly diagnosed primary cases of in situ and invasive breast cancer diagnosed in Sweden since 2008 [21]. The registry database is considered 99% complete [21]. The reported pathology data contains biomarker assessment data from the surgical specimen and, if preoperative systemic treatment, from pre-treatment core needle biopsy [21].

For this study, all invasive breast cancer cases diagnosed between 2013 and 2018 in Sweden were identified, and cases with available data on ER, PR, HER2, Ki67, and histological grade were selected ($n = 43,959$). Only pathology departments with more than 450 diagnosed breast cancer cases in the given period of time were included in the study, resulting in 43,261 valid cases from 29 investigated departments across Sweden. In cases with primary surgery, biomarker status was analyzed on the surgical specimen ($n = 38,076$), while in cases with neoadjuvant treatment, biomarker status was performed on the pre-treatment core biopsy ($n = 5185$). All studied pathology departments follow the national guidelines of biomarker evaluation [22], and participate in external quality control, passing breast cancer biomarker runs (Table S1). The distribution of breast cancer phenotypes across the regions of Sweden is considered homogenous [21]. For each included case we extracted registry data on patient characteristics (age, date of diagnosis, and date of surgery) and tumor characteristics (type of tissue specimen, tumor size, histologic grade, lymph node status, ER, PR, and HER2 status) as well as detailed treatment data. Treatment data on neoadjuvant and adjuvant endocrine and anti-HER2 therapies were handled as binary variables (yes or no).

2.2. Biomarker Assessment

According to the Swedish guidelines [22], ER and PR status are considered positive when $\geq 10\%$ of tumor cells show ER- and PR-specific staining in tumor nuclei detected by IHC. Both ER and PR status were taken into account as binary variables, either positive ($\geq 10\%$) or negative ($< 10\%$). HER2 status was considered either positive or negative, detected by IHC and/or ISH as recommended by the Swedish guidelines; IHC was performed first, followed by amplification testing in case of a 2+ IHC score. Between 2013–2018, the Swedish guidelines were following the 2013 ASCO/CAP guidelines in reporting HER2 status.

The Ki67 proliferation index was analyzed as a continuous variable and defined as the percentage of tumor cells with positive nuclear staining counted in a hot-spot with a minimum of 200 tumor cells (in accordance with Swedish guidelines [22]). Histological grade was performed on hematoxylin and eosin (HE)-stained slides according to the Nottingham grade scoring system [23]. As intra-tumoral heterogeneity critically affects histological grade scoring on core biopsy specimens, these were excluded from the study.

2.3. Quality Assurance of Pathology Data Reported to the NKBC

In order to ensure the validity of NKBC data, 8 pathology departments from 3 different regions were selected to compare NKBC data with original pathology reports. Two included breast cancer cases per month at each site were randomly selected for validation ($n = 1076$). During part of the studied period, Ki67 was reported as either high or low (two categories), while more recently this was changed to high, intermediate, or low (three categories). Therefore, we decided to take into account the reported percentage values only (Ki67 proliferation index). Validity in biomarker reporting was 95–98.5% and considered acceptable (relative percentage error rate: 1.5–5%, Table S2).

2.4. Questionnaire on Analytical Procedures

Since the labs use different antibodies, detection systems, protocols, etc., information on the analytical factors that might affect the inter-laboratory variation of the biomarkers was investigated with a questionnaire sent out covering antibody clone, IHC platform, and protocol used during the studied period. We could not detect any specific analytical factors causing inter- and intra-laboratory variability (Table S3).

2.5. Statistical Analysis

Separate analysis was performed for ER, PR, HER2, Ki67, histological grade status, endocrine treatment, HER2-targeted treatment, age, tumor size, and lymph node status as outcome measures. Kruskal–Wallis test with Dunn’s post-hoc analysis and χ^2 tests were used to compare the departments with one another, resulting in a total of 406 comparisons per marker. To investigate interannual variability for each department (intra-laboratory variability), χ^2 test was used. We applied Bonferroni correction in multiple comparisons. To interpret the results, error bars and bar charts with 95% CI levels were used, where the biomarkers’ rates per department were plotted with the overall national proportion represented with 95% confidence limits as target. In all statistical analyses, the level of significance was set at $p < 0.05$. For statistical analysis, SPSS 25 software was used (IBM, Armonk, NY, USA).

3. Results

3.1. Patient and Tumor Characteristics

In total, 43,261 patients were included, with biomarker assessment performed at 29 different laboratories. Patient and tumor characteristics for all included cases are listed in Table 1. The mean age was 63.5 years and the median tumor size was 15 mm. Three out of four patients were node-negative (72%) based on sentinel-node biopsy and/or axillary dissection due to the mammography screening introduced over 20 years. More than half of the tumors were categorized as grade II (51.4%), 87.9% as ER-positive, and 72.8% as PR-positive. HER2 was considered positive in 13.1%. The median Ki67 proliferation index was 22%. Endocrine treatment was given to 73.8% of the patients, and HER2-targeted therapy in 11.6%.

Table 1. Patient and tumor characteristics and treatment in 43,261 cases with biomarker assessment performed at 29 pathology departments.

| Variable | Category | Statistics | Results |
|--|---------------|--------------------|---------|
| Patients | | Total <i>n</i> , % | 43,959 |
| | | Valid <i>n</i> , % | 43,261 |
| Age (years) | | Mean \pm SD | 63.5 |
| | Missing data | % | 0.1 |
| Tumor size (mm) | | Median, IQR | 15 |
| | Missing data | % | 1.1 |
| Lymph node status | Node-negative | %, CI | 72.0 |
| | Node-positive | %, CI | 28.0 |
| | Missing data | % | 2.3 |
| Histological grade (<i>n</i> = 38,076) | Grade I | %, CI | 21.4 |
| | Grade II | %, CI | 51.4 |
| | Grade III | %, CI | 27.2 |
| | Missing data | % | 1.3 |

Table 1. Cont.

| Variable | Category | Statistics | Results |
|-----------------------|--------------|-------------|-------------------|
| ER | Positive | %, CI | 87.9 86.7–89.2 |
| | Negative | %, CI | 12.1 10.8–13.2 |
| | Missing data | % | 4.4 |
| PR | Positive | %, CI | 72.8 70.9–74.8 |
| | Negative | %, CI | 27.2 25.2–29.1 |
| | Missing data | % | 5.7 - |
| HER2 | Positive | %, CI | 13.1 12.4–13.8 |
| | Negative | %, CI | 86.9 86.2–87.6 |
| | Missing data | % | 2.3 - |
| Ki67 | | Median, IQR | 22.0 26 |
| | Missing data | % | 2.0 - |
| Endocrine treatment | Received | %, CI | 73.8 71.6–75.9 |
| | Not received | %, CI | 26.2 24.1–28.4 |
| | Missing data | % | 0.7 - |
| HER2-targeted therapy | Received | %, CI | 11.6 10.9–12.3 |
| | Not received | %, CI | 88.4 87.7–89.1 |
| | Missing data | % | 0.7 - |

CI = 95 % confidence interval; IQR = interquartile range; SD = Standard deviation.

3.2. Variability of ER Positivity Rates and Endocrine Therapy

The distribution of ER-positive cases ranged from 84.2% to 97.6% among the departments. All 29 labs were compared to each other, encompassing 406 comparisons in total (inter-lab variability). We found that 88 out of 406 lab comparisons showed significantly different ER positivity rates, and 5 labs had a positivity rate outside of the overall CI (Figure 1A). Four labs overvalued, while one lab undervalued the overall CI of ER positivity rates. Considering intra-laboratory variation in ER scoring, 7 out of 29 labs showed interannual variation in ER positivity rates (Figure 2A, Figure S1).

We compared the rate of endocrine therapy among the oncology departments, and found that 179 out of 406 department comparisons were significantly different. Furthermore, 13/29 departments had a rate of endocrine therapy outside of the CI (Figure 1A). However, we also found statistically significantly different patient characteristics among departments with regard to age, tumor size, and lymph node status at diagnosis (Figure S2). Furthermore, there were regional differences in offering endocrine treatment for tumors 10 mm or smaller. Therefore, we applied a case selection including only patients aged 70 or less and with tumor sizes greater than 10 mm. Adjusting for differences in age and tumor size among departments, we found that the rate of endocrine therapy strictly followed that of ER positivity (6 out of 406 department comparisons were statistically significant) (Figure 1B).

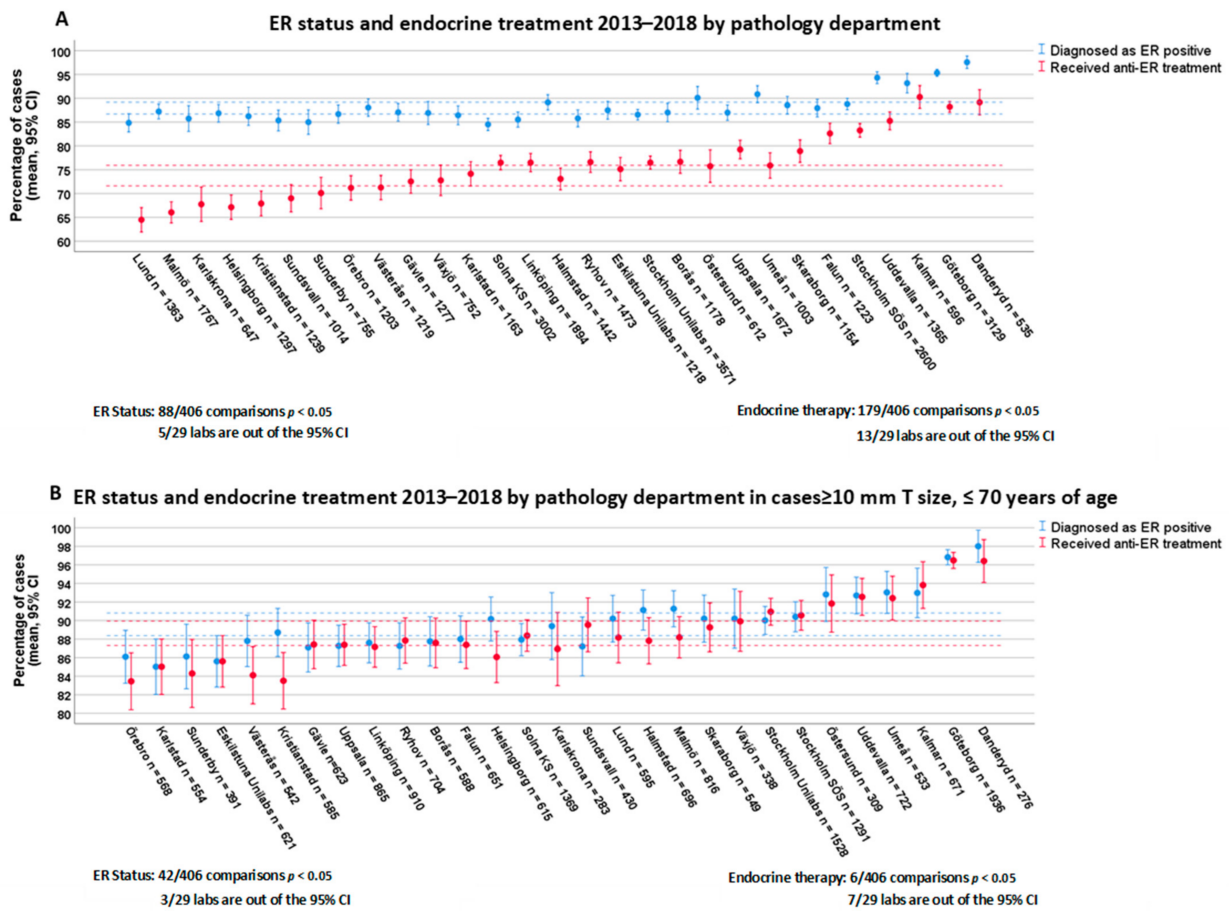


Figure 1. Endocrine receptor (ER) status and endocrine treatment between 2013 and 2018 by department. Dashed lines represent the CI for overall positivity (blue) and overall treatment rate (red). (A) includes all the patients while (B) shows results with cases ≥ 10 mm T size, ≤ 70 years of age.

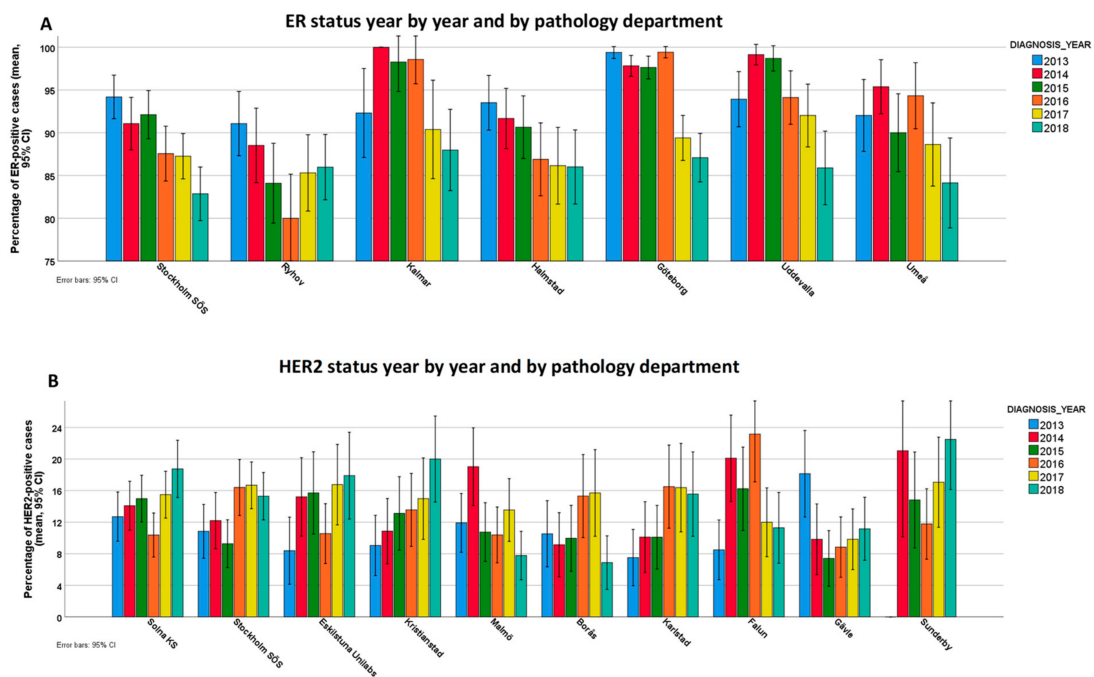


Figure 2. Intra-lab variability for ER (A) and human epidermal growth-factor receptor 2 (HER2) (B) status among pathology departments. Only laboratories showing statistically significant differences are shown. Please refer to Figure S1 to see all the labs.

3.3. Variability of PR Positivity Rates

Positivity rates for PR among the departments varied between 64.8% and 86.6%. Overall, 119 out of 406 lab comparisons were significant, and 7 labs showed positivity rates outside of the CI (Figure S3). Regarding intra-laboratory variability, 10 out of 29 labs showed significantly different PR positivity rates between 2013 and 2018 (Figure S3).

3.4. Variability of HER2 Positivity Rates and Anti-HER2 Treatment

The distribution of HER2-positive cases among departments ranged from 9.4% to 16.3%. Regarding inter-lab variability, 14 out of 406 lab comparisons were significantly different, and two labs overvalued, while one lab undervalued the overall CI of HER2 positivity rates (Figure 3A). Considering intra-laboratory variation in HER2 scoring, 10 out of 29 labs showed significant interannual variation (Figure 2B, Figure S1).

A comparison of the rate of HER2-targeted treatments among the oncology departments showed that 19 out of 406 department comparisons were significantly different. Furthermore, five departments had treatment rates outside of the CI (Figure 3A). Adjusting for differences in age and tumor size among departments, we found that HER2-targeted therapy strictly followed that of HER2 positivity, with only one oncology department outside of the overall CI and 3 out of 406 comparisons statistically significantly different (Figure 3B).

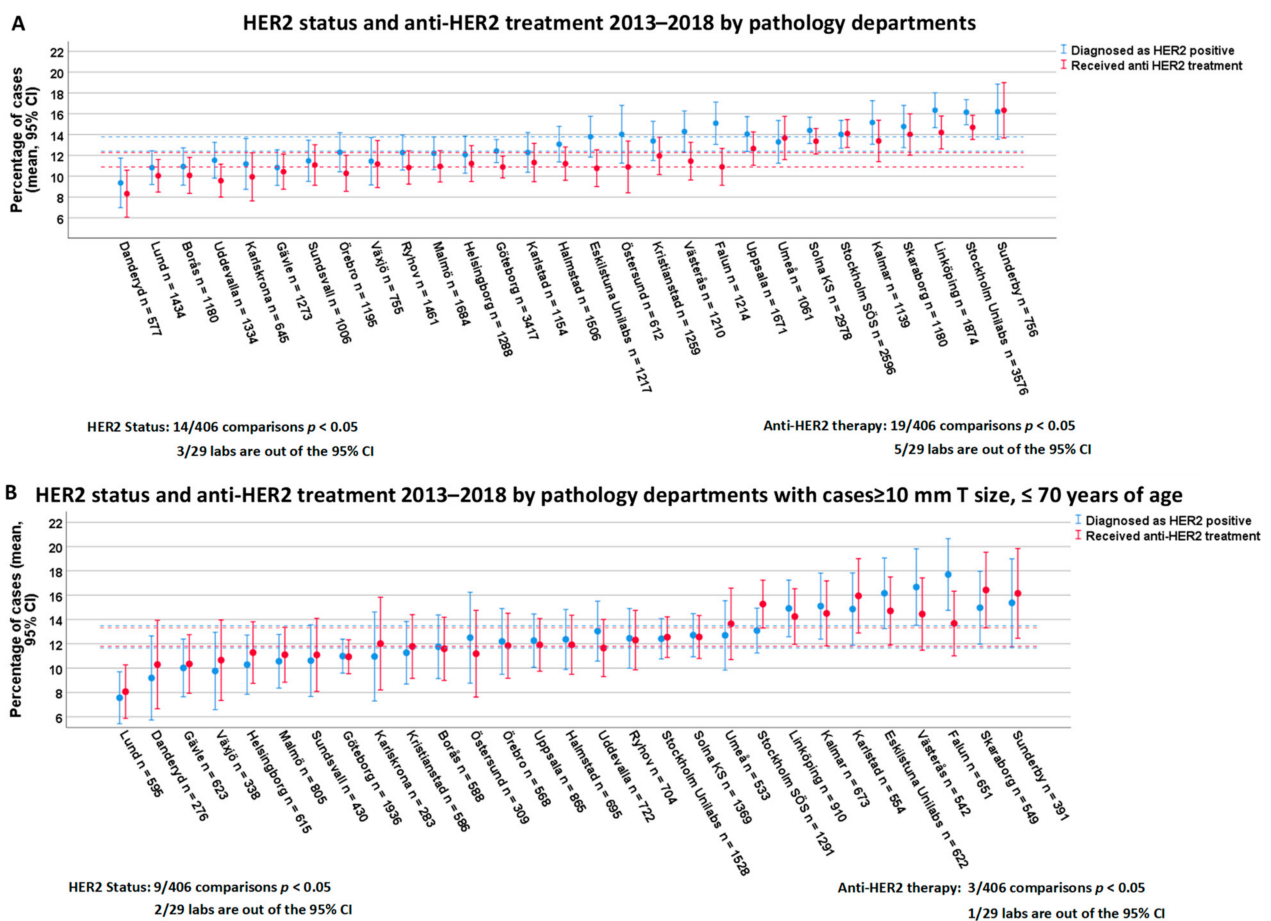


Figure 3. HER2 status and anti-HER2 treatment between 2013 and 2018 by pathology department. Dashed lines represent the CI for overall positivity (blue) and overall treatment rate (red). (A) includes all the patients while (B) shows results with cases ≥ 10 mm T size, ≤ 70 years of age.

3.5. Variability of Ki67 Proliferation Index

The median Ki67 proliferation index varied between 15% and 30% among the departments. Overall, 257 out of 406 lab comparisons were significant (Figure 4A). Regarding intra-laboratory variability, 19 out of 29 labs showed significant interannual variation (Figure 4B).

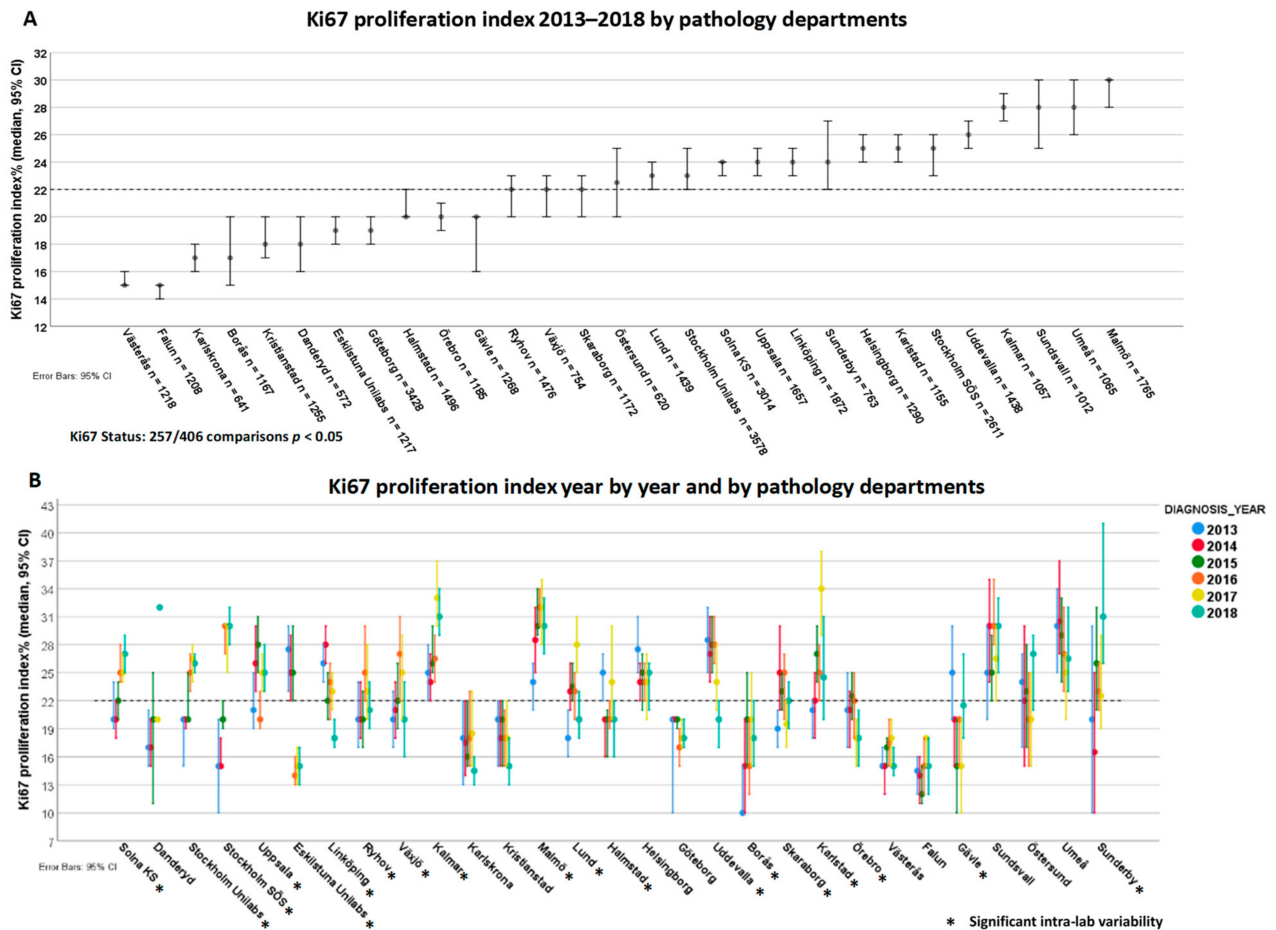


Figure 4. Inter- (A) and intra-laboratory variability (B) for Ki67 among 29 pathology departments.

3.6. Variability of Histological Grade

The proportion of histological grade II cases among the departments ranged from 42.9% to 57.1%. Of 406 lab comparisons, 104 were significantly different, and 13 labs had a distribution rate outside of the overall CI (Figure 5A). When considering histological grade as a binarized variable, similar inter-lab variability was seen (93 out of 406 lab comparisons with $p < 0.05$). Considering intra-laboratory variation, 15 out of 29 labs showed significant interannual variation (Figure 5B).

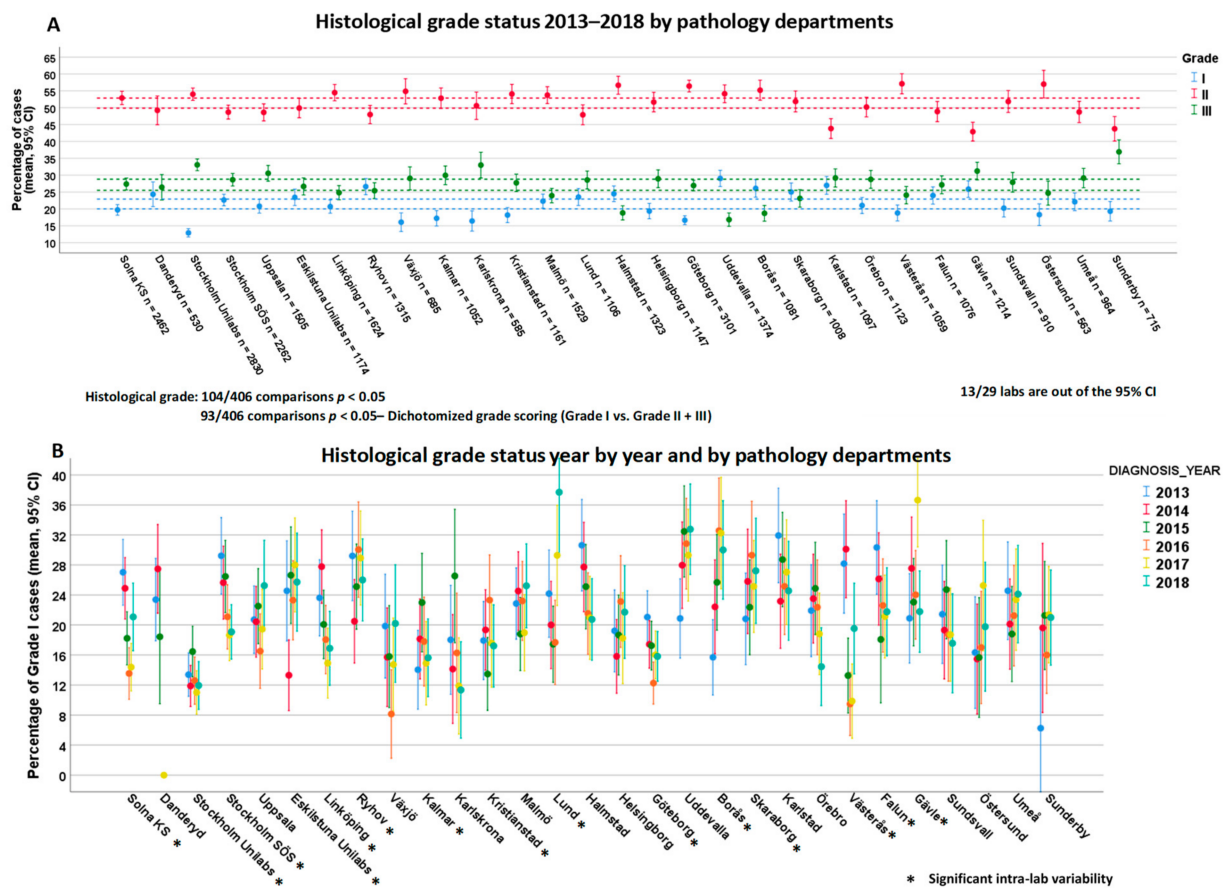


Figure 5. Inter- (A) and intra-laboratory variability (B) for histological grade status among 29 pathology departments.

4. Discussion

We investigated inter- and intra-laboratory variability in the assessment of ER, PR, HER2, Ki67, and histological grade in a non-selective population-based nationwide cohort of 43,261 invasive breast cancer patients diagnosed between 2013 and 2018, using real-world data from the Swedish National Quality Register for Breast Cancer (NKBC). The patients involved in this study represent 98.4% of breast cancer cases occurring between 2013 and 2018 in Sweden. We demonstrated that there was limited variability in ER, PR, and HER2 positivity rates among pathology departments. Overall positivity rates were 87.9% for ER, 72.8% for PR, and 13.1% for HER2, with a median Ki67 proliferation index of 22% and 51.4% of tumors deemed grade II, which are in line with international findings [19,20,24–30]. Although the proportion HER2 positivity ranged from 9.4% to 16.3% (thus varying from 1 positive out of 6 tested patients to 1 positive out of 11 patients tested), the confidence intervals were broadly overlapping for the vast majority of the labs. The same observation applies for positivity rates of ER and PR among the pathology departments. A recent investigation involving 33,046 patients performed in the Netherlands using a similar study design showed comparable variability among pathology departments in ER, PR, and HER2 positivity rates [25]. Other studies have shown significant variation among pathology laboratories, with HER2 positivity rates varying from 7.6% to 31.6% [19,20] and with significant outliers even after adjusting for population characteristics [20]. Moreover, our results demonstrate that when adjusting for clinicopathological factors, the distribution of endocrine therapy and HER2-targeted treatments strictly followed the differences of ER and HER2 positivity rates among pathology departments, leading to the same relative difference in drug prescription rates among oncology departments.

To the best of our knowledge, this is the first study to investigate the association between positivity rates of ER and HER2 among pathology laboratories and the distribution of endocrine therapy and HER2-targeted treatments among oncology departments.

Although the number of outlying departments is limited (ER: 6/29; PR: 7/29; HER2: 3/29), the slightly higher rate of intra-laboratory variability shows that there is still room for improvement—especially because such variation will be directly translated into treatment decisions. Stage, grade, and biomarker assessment are pillars for the multidisciplinary recommendations regarding surgery, chemotherapy, endocrine, and targeted therapies. Valid analyses with low variability could potentially improve patient outcomes.

Two previous studies investigated reproducibility in routine breast cancer bio-marker assessment at Swedish pathology departments. In the first study, a tissue microarray of 11 breast cancers to stain and evaluate for HER2 status was sent out twice to 24 departments [31]. The authors demonstrated a very good reproducibility in this round robin study, and showed a good analytical variability. However, the study did not reflect daily diagnostic routine [31]. In the second study, 10 cases from each of 27 participating departments were systematically collected and sent to a reference lab for re-staining and re-scoring of ER, PR, HER2, and Ki67 [32]. Although the authors demonstrated a very good agreement and observed discrepancies seemed to be explained by analytical differences, the study design did not enable comparison of individual departments [32]. In the present study, limited variability regarding ER, PR, and HER testing was found, which is in accordance with previous studies. On the other hand, we also demonstrated unacceptable inter- and intra-laboratory variability in histological grade and Ki67 assessment, urging imminent efforts to further standardize testing procedures. In a large study of 33,043 non-selected breast cancer patients, van Doijeweert et al. also demonstrated a substantial variability in histological grading among pathology departments [29]. It has long been acknowledged that both Ki67 and histological grade provide useful prognostic information [30,33,34]. However, they are subject to reproducibility issues which diminish clinical utility [11,35,36]. According to the Swedish guidelines, all ER-positive breast cancer patients are categorized into luminal-A-like and luminal-B-like subtypes based on tumor histological grade, Ki67, and PR status. All grade I and grade III cases are categorized as luminal-A-like and luminal-B-like, respectively. Grade II cases with low and high Ki67 values (based on local cut-offs) are divided into luminal-A-like and luminal-B-like, respectively. Grade II cases with intermediate Ki67 score are further measured by PR status in order to being divided into luminal-A-like (PR expression $\geq 20\%$) and luminal-B-like (PR expression $< 20\%$) subtypes [37]. As both histological grade and Ki67 status play important roles in the prognostication of breast cancer patients, the high variability found between pathology departments regarding these markers has serious clinical implications. Although significant efforts have been made to standardize pathologist-read scoring for these markers [38], introducing new methods to this field such as digital-image analysis and molecular methods would hold the promise of increasing reproducibility [39–43]. Besides, we believe that Ki67 IHC and its evaluation should undergo rigorous external quality-control procedures (similarly to ER, PR, and HER2), which is not currently the case. Furthermore, the new International Ki67 in Breast Cancer Working Group recommendations have been established, providing specific and detailed guidelines for each step of Ki67 assessment with the levels of evidence for its technical validity and clinical utility as a biomarker [44]. Following these guidelines in clinical practice has the potential to reach an acceptable reproducibility for Ki67 assessment.

There are several potential limitations to this study. First of all, as the proportion of neoadjuvant treatment differs among oncology departments, the proportion of cases analyzed based on pretreatment biopsy instead of surgical specimen differ. Tumor heterogeneity might affect the concordance in biomarker expression rates between core biopsy sample and resection specimen. However, several studies have demonstrated good agreement in biomarkers between core biopsy sample and resection specimen, especially for ER [45–47]. However, this does not apply to histological grade, as only a small tumor area is presented in the biopsy specimen. For this reason, we excluded all grade scores

reported based on core biopsy specimens only. Additionally, although the distribution of breast cancer molecular phenotypes seems to be homogenous between regions in Sweden, some differences in case-mix may remain despite adjustments for age and stage among different departments. Furthermore, the study design did not enable us to control the pre-analytical and analytical factors, potentially affecting assessments. For this reason, a questionnaire regarding used antibodies, IHC platforms, and protocols was sent out, and results from the NordiQC were retrieved to make sure that all participating departments had previously passed analytical validity testing. As IHC and ISH staining needs to be run overnight in practice, we hypothesized that the definitive biomarkers status might have been added to the pathology report afterward, and thus be missing in the report to NKBC. However, we found low rates of missing data, and it equally affected receptor-positive and receptor-negative tumors.

5. Conclusions

We demonstrated in a nation-wide population-based cohort study that there is limited variability in ER, PR, and HER2 testing among laboratories. However, as even a few outlier pathology labs affected endocrine and anti-HER2 treatment rates in a clinically relevant proportion, our study results indicate a need for further improvement. Despite international and national guidelines for analysis and interpretation, standardized protocols, and previous internal and external quality control procedures, very high variability was found in Ki67 scoring and histological grading, which suggests a need for the use of new methods. Monitoring and comparing rates of biomarker expression and treatments among departments should be mandatory in order to detect variability issues affecting the clinical management of breast cancer.

Supplementary Materials: The following are available online at <https://www.mdpi.com/2072-6694/13/5/1166/s1>, Figure S1: Intra-laboratory variability for ER (A) and HER2 (B) status among 29 pathology departments, Figure S2: Distribution of Age, T size, and LN status, Figure S3: Inter- (A) and intra-laboratory variability (B) for PR status among 29 pathology departments, Table S1: NordiQC—Sweden—Mamma module—extended data, Table S2: Quality assurance of NKBC data, Table S3: Questionnaire on analytical procedures.

Author Contributions: Conceptualization, B.A., I.F. and J.H.; methodology, B.A., I.F. and J.H.; software, B.A.; validation, B.A., C.R., C.H., A.E., A.K., and R.R.; formal analysis, B.A. and I.F.; investigation, B.A.; resources, I.F. and J.H.; data curation, B.A., I.F. and J.H.; writing—original draft preparation, B.A., I.F., J.B. and J.H.; writing—review and editing, all authors; visualization, B.A.; supervision, I.F. and J.H.; project administration, B.A.; funding acquisition, J.H. All authors have read and agreed to the published version of the manuscript.

Funding: Our research is supported by grants from the Swedish Cancer fund, Stockholm County Council, the Swedish Research Council, Swedish Breast Cancer Association, and the Stockholm Cancer Society.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethical Committee of Karolinska Institutet and the Swedish National Quality Register for Breast Cancer (#2019-01908; 15 April 2019).

Informed Consent Statement: Patient consent was waived because all cancer patients are admitted to the Swedish National Quality Register for Breast Cancer upon diagnosis. All data were retrieved and handled in compliance with the General Data Protection Regulation act.

Data Availability Statement: Data are available on the website of Swedish National Quality Register for Breast Cancer (NKBC) and can be accessed via this link: <https://statistik.incanet.se/brostcancer/>.

Conflicts of Interest: J.H. was former member of the advisory board at Visiopharm A/S. J.H. has obtained speaker's honoraria or advisory board remunerations from Roche, Novartis, AstraZeneca, Eli Lilly, Pfizer and MSD. J.H. is co-founder and shareholder of Stratipath AB. J.H. has received institutional research grants from Cepheid and Novartis. Other authors have no conflicts of interest

to disclose. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Cardoso, F.; Kyriakides, S.; Ohno, S.; Penault-Llorca, F.; Poortmans, P.; Rubio, I.T.; Zackrisson, S.; Senkus, E. Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **2019**, *30*, 1194–1220. [[CrossRef](#)]
- McGuire, W.L. Current status of estrogen receptors in human breast cancer. *Cancer* **1975**, *36*, 638–644. [[CrossRef](#)]
- Davies, C.; Godwin, J.; Gray, R.; Clarke, M.; Cutter, D.; Darby, S.; McGale, P.; Pan, H.C.; Taylor, C.; Wang, Y.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. [[PubMed](#)]
- Hayes, D.F. Steady progress against HER2-positive breast cancer. *N. Engl. J. Med.* **2011**, *365*, 1336–1338. [[CrossRef](#)]
- Fehrenbacher, L.; Cecchini, R.S.; Geyer, C.E., Jr.; Rastogi, P.; Costantino, J.P.; Atkins, J.N.; Crown, J.P.; Polikoff, J.; Boileau, J.F.; Provencher, L.; et al. NSABP B-47/NRG oncology phase III randomized trial comparing adjuvant chemotherapy with or without trastuzumab in high-risk invasive breast cancer negative for HER2 by FISH and with IHC 1+ or 2. *J. Clin. Oncol.* **2020**, *38*, 444–453. [[CrossRef](#)] [[PubMed](#)]
- Allison, K.H.; Hammond, M.E.H.; Dowsett, M.; McKernin, S.E.; Carey, L.A.; Fitzgibbons, P.L.; Hayes, D.F.; Lakhani, S.R.; Chavez-MacGregor, M.; Perlmutter, J.; et al. Estrogen and progesterone receptor testing in breast cancer: ASCO/CAP guideline update. *J. Clin. Oncol.* **2020**, *38*, 1346–1366. [[CrossRef](#)] [[PubMed](#)]
- Senkus, E.; Kyriakides, S.; Ohno, S.; Penault-Llorca, F.; Poortmans, P.; Rutgers, E.; Zackrisson, S.; Cardoso, F. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **2015**, *26*, v8–v30. [[CrossRef](#)]
- Wolff, A.C.; Hammond, M.E.H.; Allison, K.H.; Harvey, B.E.; Mangu, P.B.; Bartlett, J.M.S.; Bilous, M.; Ellis, I.O.; Fitzgibbons, P.; Hanna, W.; et al. Human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. *J. Clin. Oncol.* **2018**, *36*, 2105–2122. [[CrossRef](#)] [[PubMed](#)]
- Fitzgibbons, P.L.; Murphy, D.A.; Hammond, M.E.; Allred, D.C.; Valenstein, P.N. Recommendations for validating estrogen and progesterone receptor immunohistochemistry assays. *Arch. Pathol. Lab. Med.* **2010**, *134*, 930–935. [[CrossRef](#)]
- Rhodes, A.; Jasani, B.; Balaton, A.J.; Miller, K.D. Immunohistochemical demonstration of oestrogen and progesterone receptors: Correlation of standards achieved on in house tumours with that achieved on external quality assessment material in over 150 laboratories from 26 countries. *J. Clin. Pathol.* **2000**, *53*, 292–301. [[CrossRef](#)]
- Polley, M.Y.; Leung, S.C.; McShane, L.M.; Gao, D.; Hugh, J.C.; Mastropasqua, M.G.; Viale, G.; Zabaglo, L.A.; Penault-Llorca, F.; Bartlett, J.M.; et al. An international Ki67 reproducibility study. *J. Natl. Cancer Inst.* **2013**, *105*, 1897–1906. [[CrossRef](#)] [[PubMed](#)]
- Hayes, D.F. Biomarker validation and testing. *Mol. Oncol.* **2015**, *9*, 960–966. [[CrossRef](#)] [[PubMed](#)]
- Romond, E.H.; Jeong, J.H.; Rastogi, P.; Swain, S.M.; Geyer, C.E., Jr.; Ewer, M.S.; Rathi, V.; Fehrenbacher, L.; Brufsky, A.; Azar, C.A.; et al. Seven-year follow-up assessment of cardiac function in NSABP B-31, a randomized trial comparing doxorubicin and cyclophosphamide followed by paclitaxel (ACP) with ACP plus trastuzumab as adjuvant therapy for patients with node-positive, human epidermal growth factor receptor 2-positive breast cancer. *J. Clin. Oncol.* **2012**, *30*, 3792–3799.
- Russell, S.D.; Blackwell, K.L.; Lawrence, J.; Pippen, J.E., Jr.; Roe, M.T.; Wood, F.; Paton, V.; Holmgren, E.; Mahaffey, K.W. Independent adjudication of symptomatic heart failure with the use of doxorubicin and cyclophosphamide followed by trastuzumab adjuvant therapy: A combined review of cardiac data from the National Surgical Adjuvant breast and Bowel Project B-31 and the North Central Cancer Treatment Group N9831 clinical trials. *J. Clin. Oncol.* **2010**, *28*, 3416–3421. [[PubMed](#)]
- Dowsett, M.; Nielsen, T.O.; A'Hern, R.; Bartlett, J.; Coombes, R.C.; Cuzick, J.; Ellis, M.; Henry, N.L.; Hugh, J.C.; Lively, T.; et al. Assessment of Ki67 in breast cancer: Recommendations from the International Ki67 in Breast Cancer working group. *J. Natl. Cancer Inst.* **2011**, *103*, 1656–1664. [[CrossRef](#)]
- Bilous, M.; Dowsett, M.; Hanna, W.; Isola, J.; Lebeau, A.; Moreno, A.; Penault-Llorca, F.; Ruschoff, J.; Tomicic, G.; van de Vijver, M. Current perspectives on HER2 testing: A review of national testing guidelines. *Mod. Pathol.* **2003**, *16*, 173–182. [[CrossRef](#)]
- Ellis, I.O.; Bartlett, J.; Dowsett, M.; Humphreys, S.; Jasani, B.; Miller, K.; Pinder, S.E.; Rhodes, A.; Walker, R. Best practice no 176: Updated recommendations for HER2 testing in the UK. *J. Clin. Pathol.* **2004**, *57*, 233–237. [[CrossRef](#)]
- Vyberg, M.; Nielsen, S. Proficiency testing in immunohistochemistry—Experiences from Nordic Immunohistochemical Quality Control (NordiQC). *Virchows Arch.* **2016**, *468*, 19–29. [[CrossRef](#)]
- Choritz, H.; Büsche, G.; Kreipe, H. Quality assessment of HER2 testing by monitoring of positivity rates. *Virchows Arch.* **2011**, *459*, 283–289. [[CrossRef](#)]
- Rüschoff, J.; Lebeau, A.; Kreipe, H.; Sinn, P.; Gerharz, C.D.; Koch, W.; Morris, S.; Ammann, J.; Untch, M. Assessing HER2 testing quality in breast cancer: Variables that influence HER2 positivity rate from a large, multicenter, observational study in Germany. *Mod. Pathol.* **2017**, *30*, 217–226. [[CrossRef](#)]
- National Quality Registry for Breast Cancer (NKBC). Available online: <https://statistik.incanet.se/brostcancer/> (accessed on 26 February 2021).
- Swedish Society of Pathology—Quality and Standardization Committee (KVASt): Breast Cancer Guideline. 2020. Available online: <http://www.svfp.se/foreningar/uploads/L15178/kvast/brostpatologi/Brostcancer2020korr.pdf> (accessed on 26 February 2021).

23. Elston, C.W.; Ellis, I.O. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: Experience from a large study with long-term follow-up. *Histopathology* **1991**, *19*, 403–410. [[CrossRef](#)]
24. Parise, C.A.; Caggiano, V. breast cancer survival defined by the ER/PR/HER2 subtypes and a surrogate classification according to tumor grade and immunohistochemical biomarkers. *J. Cancer Epidemiol.* **2014**, *2014*, 469251. [[CrossRef](#)] [[PubMed](#)]
25. van Doonijeweert, C.; Deckers, I.A.G.; Baas, I.O.; van der Wall, E.; van Diest, P.J. Hormone- and HER2-receptor assessment in 33,046 breast cancer patients: A nationwide comparison of positivity rates between pathology laboratories in the Netherlands. *Breast Cancer Res. Treat.* **2019**, *175*, 487–497. [[CrossRef](#)] [[PubMed](#)]
26. Owens, M.A.; Horten, B.C.; Da Silva, M.M. HER2 amplification ratios by fluorescence in situ hybridization and correlation with immunohistochemistry in a cohort of 6556 breast cancer tissues. *Clin. Breast Cancer.* **2004**, *5*, 63–69. [[CrossRef](#)]
27. Wolff, A.C.; Hammond, M.E.; Hicks, D.G.; Dowsett, M.; McShane, L.M.; Allison, K.H.; Allred, D.C.; Bartlett, J.M.; Bilous, M.; Fitzgibbons, P.; 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. [[CrossRef](#)] [[PubMed](#)]
28. Goldhirsch, A.; Winer, E.P.; Coates, A.S.; Gelber, R.D.; Piccart-Gebhart, M.; Thurlimann, B.; Senn, H.J. Personalizing the treatment of women with early breast cancer: Highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann. Oncol.* **2013**, *24*, 2206–2223. [[CrossRef](#)]
29. van Doonijeweert, C.; van Diest, P.J.; Willems, S.M.; Kuijpers, C.; van der Wall, E.; Overbeek, L.I.H.; Deckers, I.A.G. Significant inter- and intra-laboratory variation in grading of invasive breast cancer: A nationwide study of 33,043 patients in the Netherlands. *Int. J. Cancer.* **2020**, *146*, 769–780. [[CrossRef](#)]
30. Schwartz, A.M.; Henson, D.E.; Chen, D.; Rajamarthandan, S. Histologic grade remains a prognostic factor for breast cancer regardless of the number of positive lymph nodes and tumor size: A study of 161 708 cases of breast cancer from the SEER Program. *Arch. Pathol. Lab. Med.* **2014**, *138*, 1048–1052. [[CrossRef](#)] [[PubMed](#)]
31. Rydén, L.; Haglund, M.; Bendahl, P.O.; Hatschek, T.; Kolaric, A.; Kovács, A.; Olsson, A.; Olsson, H.; Strand, C.; Fernö, M. Reproducibility of human epidermal growth factor receptor 2 analysis in primary breast cancer: A national survey performed at pathology departments in Sweden. *Acta Oncol.* **2009**, *48*, 860–866. [[CrossRef](#)]
32. Ekholm, M.; Grabau, D.; Bendahl, P.O.; Bergh, J.; Elmberger, G.; Olsson, H.; Russo, L.; Viale, G.; Fernö, M. Highly reproducible results of breast cancer biomarkers when analysed in accordance with national guidelines—A Swedish survey with central re-assessment. *Acta Oncol.* **2015**, *54*, 1040–1048. [[CrossRef](#)]
33. Stuart-Harris, R.; Caldas, C.; Pinder, S.E.; Pharoah, P. Proliferation markers and survival in early breast cancer: A systematic review and meta-analysis of 85 studies in 32,825 patients. *Breast* **2008**, *17*, 323–334. [[CrossRef](#)]
34. Criscitiello, C.; Disalvatore, D.; De Laurentiis, M.; Gelao, L.; Fumagalli, L.; Locatelli, M.; Bagnardi, V.; Rotmensz, N.; Esposito, A.; Minchella, I.; et al. High Ki-67 score is indicative of a greater benefit from adjuvant chemotherapy when added to endocrine therapy in luminal B HER2 negative and node-positive breast cancer. *Breast* **2014**, *23*, 69–75. [[CrossRef](#)]
35. Meyer, J.S.; Alvarez, C.; Milikowski, C.; Olson, N.; Russo, I.; Russo, J.; Glass, A.; Zehnbauer, B.A.; Lister, K.; Parwaresch, R. Breast carcinoma malignancy grading by Bloom-Richardson system vs proliferation index: Reproducibility of grade and advantages of proliferation index. *Mod. Pathol.* **2005**, *18*, 1067–1078. [[CrossRef](#)] [[PubMed](#)]
36. Frierson, H.F., Jr.; Wolber, R.A.; Berean, K.W.; Franquemont, D.W.; Gaffey, M.J.; Boyd, J.C.; Wilbur, D.C. Interobserver reproducibility of the Nottingham modification of the Bloom and Richardson histologic grading scheme for infiltrating ductal carcinoma. *Am. J. Clin. Pathol.* **1995**, *103*, 195–198. [[CrossRef](#)] [[PubMed](#)]
37. Swedish National Breast Cancer Care Program. Available online: <https://kunskapsbanken.cancercentrum.se/diagnoser/brostcancer/vardprogram/kategorisering-av-tumoren/> (accessed on 26 February 2021).
38. Polley, M.Y.; Leung, S.C.; Gao, D.; Mastropasqua, M.G.; Zabaglo, L.A.; Bartlett, J.M.; McShane, L.M.; Enos, R.A.; Badve, S.S.; Bane, A.L.; et al. An international study to increase concordance in Ki67 scoring. *Mod. Pathol.* **2015**, *28*, 778–786. [[CrossRef](#)] [[PubMed](#)]
39. Rimm, D.L.; Leung, S.C.Y.; McShane, L.M.; Bai, Y.; Bane, A.L.; Bartlett, J.M.S.; Bayani, J.; Chang, M.C.; Dean, M.; Denkert, C.; et al. An international multicenter study to evaluate reproducibility of automated scoring for assessment of Ki67 in breast cancer. *Mod. Pathol.* **2019**, *32*, 59–69. [[CrossRef](#)]
40. Stalhammar, G.; Fuentes Martinez, N.; Lippert, M.; Tobin, N.P.; Molholm, I.; Kis, L.; Rosin, G.; Rantalainen, M.; Pedersen, L.; Bergh, J.; et al. Digital image analysis outperforms manual biomarker assessment in breast cancer. *Mod. Pathol.* **2016**, *29*, 318–329. [[CrossRef](#)]
41. Acs, B.; Pelekanou, V.; Bai, Y.; Martinez-Morilla, S.; Toki, M.; Leung, S.C.Y.; Nielsen, T.O.; Rimm, D.L. Ki67 reproducibility using digital image analysis: An inter-platform and inter-operator study. *Lab. Invest.* **2019**, *99*, 107–117. [[CrossRef](#)]
42. Sotiropoulos, C.; Wirapati, P.; Loi, S.; Harris, A.; Fox, S.; Smeds, J.; Nordgren, H.; Farmer, P.; Praz, V.; Haibe-Kains, B.; et al. Gene expression profiling in breast cancer: Understanding the molecular basis of histologic grade to improve prognosis. *J. Natl. Cancer Inst.* **2006**, *98*, 262–272. [[CrossRef](#)]
43. Wang, M.; Klevebring, D.; Lindberg, J.; Czene, K.; Grönberg, H.; Rantalainen, M. Determining breast cancer histological grade from RNA-sequencing data. *Breast Cancer Res.* **2016**, *18*, 48. [[CrossRef](#)]
44. Nielsen, T.O.; Leung, S.C.Y.; Rimm, D.L.; Dodson, A.; Acs, B.; Badve, S.; Denkert, C.; Ellis, M.J.; Fineberg, S.; Flowers, M.; et al. Assessment of Ki67 in breast cancer: Updated recommendations from the international Ki67 in breast cancer working group. *J. Natl. Cancer Inst.* **2020**. [[CrossRef](#)] [[PubMed](#)]

45. Arnedos, M.; Nerurkar, A.; Osin, P.; A'Hern, R.; Smith, I.E.; Dowsett, M. Discordance between core needle biopsy (CNB) and excisional biopsy (EB) for estrogen receptor (ER), progesterone receptor (PgR) and HER2 status in early breast cancer (EBC). *Ann. Oncol.* **2009**, *20*, 1948–1952. [[CrossRef](#)] [[PubMed](#)]
46. Dekker, T.J.; Smit, V.T.; Hooijer, G.K.; Van de Vijver, M.J.; Mesker, W.E.; Tollenaar, R.A.; Nortier, J.W.; Kroep, J.R. Reliability of core needle biopsy for determining ER and HER2 status in breast cancer. *Ann. Oncol.* **2013**, *24*, 931–937. [[CrossRef](#)] [[PubMed](#)]
47. Robertson, S.; Rönnlund, C.; de Boniface, J.; Hartman, J. Re-testing of predictive biomarkers on surgical breast cancer specimens is clinically relevant. *Breast Cancer Res. Treat.* **2019**, *174*, 795–805. [[CrossRef](#)] [[PubMed](#)]