Acta Oncologica, 2015; 54: 1040–1048

informa
healthcare

ORIGINAL ARTICLE

Highly reproducible results of breast cancer biomarkers when analysed in accordance with national guidelines – a Swedish survey with central re-assessment

MARIA EKHOLM[1,2], DORTHE GRABAU[1,3], PÄR-OLA BENDAHL[1], JONAS BERGH[4], GÖRAN ELMBERGER[5], HANS OLSSON[6], LEILA RUSSO[7], GIUSEPPE VIALE[7] & MÅRTEN FERNÖ[1]

[1]*Department of Clinical Sciences Lund, Division of Oncology and Pathology, Lund University, Lund, Sweden,* [2]*Department of Oncology, Ryhov County Hospital, Jönköping, Sweden,* [3]*Department of Pathology, Skåne University Hospital, Lund, Sweden,* [4]*Department of Oncology-Pathology, Radiumhemmet, Cancer Center Karolinska, Karolinska Institutet and University Hospital, Stockholm, Sweden,* [5]*Department of Laboratory Medicine, Pathology, Faculty of Medicine and Health, Örebro University, Örebro, Sweden,* [6]*Molecular and Immunological Pathology, Department of Clinical and Experimental Medicine, Faculty of Health Sciences, Linköping University, Linköping, Sweden, and* [7]*Department of Pathology, European Institute of Oncology, University of Milan, Milan, Italy*

ABSTRACT

Background. Biomarkers are crucial for decisions regarding adjuvant therapy in primary breast cancer, and their correct assessment is therefore of the utmost importance.

Aims. To investigate the concordance between Swedish pathology departments and a reference laboratory, for routine analysis of oestrogen receptor (ER), progesterone receptor (PR), Ki67, and human epidermal growth factor receptor 2 (HER2), alone, and in combination (St Gallen subtypes).

Methods. This survey included 27 of the 28 pathology laboratories in Sweden, covering 98% of cases of primary breast cancer surgery in Sweden. Paraffin-embedded tumour blocks (n = 270) were collected and sent to the central reference laboratory, together with the originally stained slides, for re-analysis. The primary evaluations were previously performed according to national Swedish guidelines, without any knowledge of the subsequent central assessment.

Results. The agreement for ER, PR, and Ki67 was 99% [kappa value (κ) = 0.95], 95% (κ = 0.85), and 85% (κ = 0.70), respectively. The agreement for HER2 (0/1 + vs. 2+/3+) was 85% (κ = 0.64), but when equivocal tumours were further analysed with in situ hybridisation, only one discrepancy was observed. Discrepancies between results for ER and PR seem to be explained by analytical differences, whereas the interpretation of staining seems to be more critical for Ki67 and HER2 immunohistochemistry. The agreement between the results from the Swedish laboratories and the reference laboratory, based on the St Gallen subtypes, was 88% (κ = 0.81).

Conclusions. When applying national guidelines, highly reproducible results were obtained in routine assessment of breast cancer biomarkers, and the results of this study confirm the clinical utility of these markers for decisions regarding the treatment of primary breast cancer.

Correct assessments of prognostic and predictive markers are essential in the choice of adjuvant therapy in primary breast cancer. False negative oestrogen receptor (ER) or human epidermal growth factor receptor 2 (HER2) results may lead to useful targeted therapy being withheld, while false positive results may lead to the use of unnecessary therapies that may be very expensive, or subject the patient to unnecessary side effects. The proliferation marker Ki67, progesterone receptor (PR), and HER2, are important in distinguishing patients with ER-positive tumours in whom adjuvant chemotherapy can be

Correspondence: M. Fernö, Department of Clinical Sciences Lund, Division of Oncology and Pathology, Building 406, Medicon Village, Scheelevägen 8, 223 63 Lund, Sweden. Tel: + 46 46 2752550. Fax: + 46 46 147327. E-mail: marten.ferno@med.lu.se

(Received 20 February 2015; accepted 26 March 2015)

ISSN 0284-186X print/ISSN 1651-226X online
DOI: 10.3109/0284186X.2015.1037012

omitted due to the low risk of recurrence [1–4]. A false low value of Ki67 may result in patients with a high risk of recurrence not being given chemotherapy, while a false high value of Ki67 may result in overtreatment of patients with a good prognosis.

Immunohistochemical (IHC) techniques are still the gold standard for assessing ER, PR, and Ki67 in routine breast cancer pathology. HER2 is also analysed by IHC, but for equivocal cases further analysis with in situ hybridisation (ISH) is needed in order to obtain the final HER2 status (positive vs. negative). These biomarkers are all included in the clinicopathological surrogate definition of the intrinsic subtypes according to the St Gallen International Expert Consensus of 2013: 'Luminal A-like', 'Luminal B-like/HER2-negative', 'Luminal B-like/HER2-positive', 'HER2-positive/non-luminal', and 'Triple-negative' (Table I) [1]. IHC techniques are generally available, and at a lower cost than commercial multi-gene assays. When combined into an immunohistochemical score (IHC4), these markers have also provided similar prognostic information as Oncotype DX [4]. Furthermore, IHC techniques can be applied to both routinely processed and formalin-fixed paraffin-embedded tissue samples and have the advantage of ensuring that only invasive tumour cells are assessed when evaluating small tumours [5]. However, moderate interobserver and interlaboratory reproducibility have been reported, especially for Ki67 [6–10]. Standardised and reproducible analyses, based on generally accepted guidelines are, therefore, of the utmost importance.

Table I. St Gallen 2013 clinicopathological surrogate definition of the intrinsic subtypes of breast cancer.

Intrinsic subtype	Clinicopathological surrogate definition	Characteristics
Luminal A	'Luminal A-like'	ER-positive and high PR HER2-negative low Ki67
Luminal B	'Luminal B-like/HER2-negative'	ER-positive HER2-negative At least one of: • high Ki67 • negative or low PR
	'Luminal B-like/HER2-positive'	ER-positive HER2-positive Any Ki67 Any PR
ErbB2 overexpression	'HER2-positive/non-luminal'	ER- and PR-negative HER2-positive
Basal-like	'Triple-negative'	ER- and PR-negative HER2-negative

ER and PR (positive vs. negative)
- LA > 1%
- RA and IEO ≥ 1%

PR (high vs. low): > 20%
Ki67 (high vs. low): > 20%

Validation of an assay includes standardisation of pre-analytical procedures (e.g. sample collection, fixation, storage, etc.), analytical procedures (e.g. method of antigen retrieval, antibody, antibody detection and staining method), as well as interpretative methods (e.g. scoring method and thresholds/cut-offs) [11,12]. As IHC techniques are evolving, continuous validation is required to ensure consistent reporting [5,12]. The importance of quality assurance is highlighted in guidelines and recommendations, and several organisations perform external proficiency testing, e.g. the United Kingdom National External Quality Assessment Scheme for Immunocytochemistry (UK NEQAS), the College of American Pathologists (CAP), and the organisation Nordic Immunohistochemical Quality Control (NordiQC), as well as other smaller national organisations [6,8,11–16].

This study was undertaken by Swedish Quality Assurance breast cancer (SweQA), as part of the yearly quality assurance program for breast cancer biomarkers in Sweden. Paraffin-embedded tumour tissue blocks from breast carcinomas were collected together with the originally stained sections, according to a predefined protocol, and sent to the European Institute of Oncology (IEO), Milan, Italy, for re-assessment and re-analysis. The aim was to investigate the pairwise agreement between Swedish pathology departments and the central laboratory regarding the routine biomarkers of primary breast cancer (ER, PR, Ki67, and HER2), and to investigate whether possible discrepancies could be explained by analytical or interpretative procedures. To investigate the possible clinical consequences of central testing, the tumours were re-classified according to the St Gallen clinicopathological surrogate definition of the intrinsic subtypes [1].

Methods

Participants and case selection

Written invitations to participate in this study were sent by the organisation External Quality Assessment for Clinical Laboratory Investigations (Equalis) to all 28 pathology departments in Sweden carrying out histopathological assessments after breast cancer surgery [17]. Twenty-seven of these 28 laboratories, covering 98% of all cases of primary breast cancer surgery in Sweden, accepted the invitation to participate in the study [18]. Seven were university clinics and 20 were general hospitals. To achieve a reasonably random selection of tumours, the laboratories were instructed to identify the first breast carcinoma diagnosed after the 15th day of each month during 2012, except for July and December, giving a total of 270 tumours. The

Swedish laboratories were then asked to collect the formalin-fixed paraffin-embedded tumour tissue blocks analysed in the clinical setting, whenever possible, together with the originally stained slides. These were then sent to IEO by Equalis. The study design and the analyses performed are illustrated in Figure 1.

Analyses

The samples had been immunohistochemically stained, assessed, and scored for ER, PR, Ki67, and HER2 at each laboratory in clinical routine, using Swedish guidelines for the assessment of breast carcinomas (Supplementary Table I, available online at http://informahealthcare.com/doi/abs/10.3109/0284186X.2015.1037012) [19,20]. These results are referred to below as local assessments (LA).

The originally stained sections were re-evaluated at IEO by an experienced pathologist (LR) with regard to ER, PR, Ki67, and HER2, according to their local interpretative guidelines. These results are hereafter referred to as the reviewed assessments (RA).

New 4-µm sections were cut from the collected formalin-fixed paraffin-embedded tumour blocks, stained, and scored with regard to ER, PR, Ki67, and HER2, at IEO by the same pathologist (LR) using their local guidelines (Supplementary Table I available online at http://informahealthcare.com/doi/abs/10.3109/0284186X.2015.1037012). These results are hereafter referred to as the IEO.

The following comparisons were made:

- LA vs. RA – for interpretative comparisons,
- LA vs. IEO – for analytic and interpretative comparisons, and
- RA vs. IEO – for analytic comparisons.

The routines for the LA can briefly be described as follows; ER and PR are assessed as the percentage of positive cells without taking intensity into consideration. The Ki67 index is obtained by counting 200 nuclei in hotspots disregarding the intensity. All tumours locally assessed as HER2/IHC 2+/3+ (except two 3 + tumours) were further analysed with ISH as part of the clinical routine. To define the final HER2 status (positive vs. negative), tumours with IHC results from all three assessments were included,

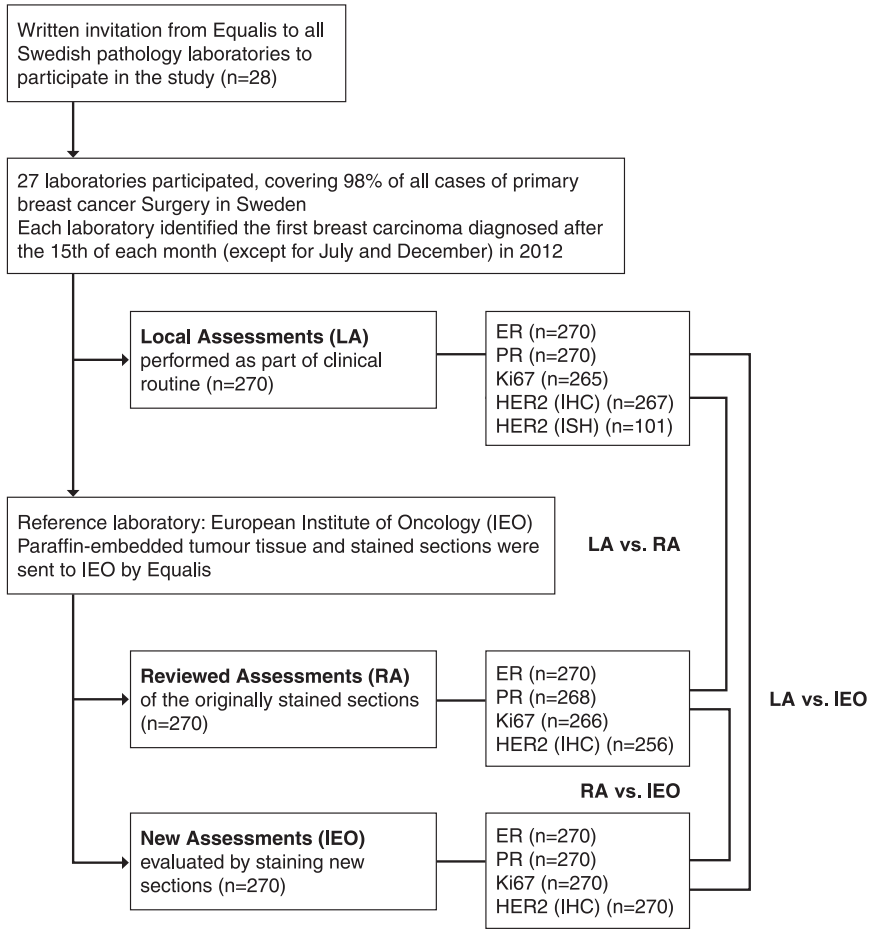

Figure 1. CONSORT-diagram illustrating the study design and the different analyses.

and they were considered HER2 positive when assessed as 3 + and/or amplified by ISH performed in Sweden. Our group has previously shown that the reproducibility of HER2/ISH is very good (κ = 0.92–0.96), and HER2 was therefore not centrally re-analysed by this method [21].

Cut-off values

ER, PR, Ki67, and HER2 were used as dichotomous variables. High values of Ki67 and PR were defined as > 20% positive nuclei in all three assessments. The cut-off used for ER- and PR-positivity was ≥ 1% in the RA and IEO analyses, but > 1% in the LA, due to the way in which the percentage of positive cells is recorded in Sweden [19]. This affected ER in five tumours and PR in two tumours.

According to current Swedish guidelines, each pathology department should use laboratory-specific cut-offs for Ki67, defined as the 67th percentile of the first 100 breast carcinomas each year [19]. The cut-off values for the individual laboratories are listed in Supplementary Table I available online at http://informahealthcare.com/doi/abs/10.3109/0284186X.2015.1037012, and these were used in a separate analysis to investigate the effect of laboratory-specific cut-offs for Ki67.

St Gallen 2013 clinicopathological subtype classification

The clinicopathological surrogate definition of the intrinsic subtypes according to St Gallen 2013 was used (Table I) [1]. Only tumours with complete data for all four tumour markers by all three assessments were included. Six tumours, assessed as ER negative and PR positive, by any of the three assessments, were excluded since these tumours cannot be classified according to St Gallen 2013.

Statistical methods

Agreement (%) and Cohen's kappa (κ) were used to compare the results for all three comparisons, including only cases with complete data for all three assessments. The following definition is frequently used to define the chance-corrected measure of agreement: κ ≤ 0.40 poor-fair agreement; 0.41–0.60 moderate agreement; 0.61–0.80 substantial agreement; and > 0.80 almost perfect agreement [22]. The reader should keep in mind that this measure is prevalence dependent and that values for variables with different number of categories are not comparable. The Wilcoxon-matched pairs signed-rank test was used to compare PR expression across different antibodies. Data management and statistical analyses were carried out using Stata 13.1 (StataCorp LP, College Station, TX, USA).

Ethics

This study was undertaken as part of the yearly Swedish quality assurance programme, performed by SweQA. All samples were anonymised and, according to Swedish law, there is no need for ethical approval for quality assurance studies.

Results

ER

The overall agreement for ER (positive vs. negative) in the three comparisons was 99% (κ = 0.96) for LA vs. RA, 99% (κ = 0.95) for LA vs. IEO, and 98% (κ = 0.91) for RA vs. IEO (n = 270) (Table II). Three of the five tumours showing discrepancies were assessed as positive according to LA and RA, but negative according to IEO (Table III).

PR

The overall agreement for PR (positive vs. negative) in the three comparisons was 98% (κ = 0.94) for LA vs. RA, 95% (κ = 0.85) for LA vs. IEO, and 97% (κ = 0.91) for RA vs. IEO (n = 268) (Table II). Seven of the 13 tumours showing discrepancies were assessed as positive according to LA and RA, but negative according to IEO (Table III). These seven tumours were from departments using the Ventana 1E2 ready-to-use (RTU) antibody kit; four of them were ER negative (ER = 0) and two were borderline ER negative (ER ≤ 1) according to all three assessments. To further investigate the influence of the specific antibodies, we compared the PR levels of the tumours analysed with the Ventana 1E2 RTU kit (RA), with the PR levels of the tumours obtained using the PharmDx kit (clone PgR 1294) (IEO) (n = 158). In total 120 tumours were assessed as having the same levels, 25 tumours had higher levels, and 13 tumours had lower levels of PR, according to the Ventana 1E2 RTU kit. This shift towards higher PR values was statistically significant (p = 0.03; Wilcoxon-matched pairs signed-rank test).

Ki67

The overall agreement for Ki67 (low vs. high) was 89% (κ = 0.77) for LA vs. RA, 85% (κ = 0.70) for LA vs. IEO, and 94% (κ = 0.89) for RA vs. IEO (n = 265) (Table II and Figure 2). Twelve out of 42 tumours showing discrepancies were categorised in

Table II. Results of agreement analyses for ER, PR, Ki67, HER2, and St Gallen subtype classification 2013, for tumours with complete data for all three assessments.

	All (n)	LA	RA	IEO	LA vs. RA	LA vs. IEO	RA vs. IEO
ER (positive vs. negative)	270	88% positive	89% positive	87% positive			
Agreement (%)					99	99	98
κ					0.96	0.95	0.91
PR (positive vs. negative)	268	80% positive	81% positive	79% positive			
Agreement (%)					98	95	97
κ					0.94	0.85	0.91
Ki67 (high vs. low)	265	47% high	42% high	46% high			
Agreement (%)					89	85	94
κ					0.77	0.70	0.89
HER2 IHC (0/1 + vs. 2+/3+)	256	35% 2 + or 3+	27% 2 + or 3+	27% 2 + or 3+			
Agreement (%)					86	85	96
κ					0.66	0.64	0.91
HER2 (final result*)	248	12% positive	12% positive	12% positive			
Agreement (%)					99.6	99.6	100
κ					0.98	0.98	1.00
St Gallen subtype classification (five groups)	233						
'Luminal A-like'		40%	47%	43%			
'Luminal B-like/HER2-negative'		41%	35%	38%			
'Luminal B-like/HER2-positive'		9%	9%	9%			
'HER2-positive/non-luminal'		3%	3%	3%			
'Triple-negative'		7%	8%	8%			
Agreement (%)					91	88	94
κ					0.86	0.81	0.91

IEO, New assessment at European Institute of Oncology; LA, local assessment; RA, reviewed assessment.
[a]HER2 3 + for all three assessments and/or amplification according to in situ hybridisation (see Methods).

the same group according to both LA and RA, but differently according to the IEO analysis, and 27 tumours were categorised in the same group according to RA and the IEO analysis, but differently according to LA. The remaining three discrepant tumours were categorised identically according to LA and IEO, but differently according to RA. In 38 of the 42 tumours showing a discrepancy, at least one of the assessments was located near the cut-off (15–25%).

When only including tumours from the 24 Swedish laboratories reporting laboratory-specific cut-off for Ki67 (Supplementary Table I available online at http://informahealthcare.com/doi/abs/10.3109/0284186X.2015.1037012), the agreement between LA and IEO was reduced from 86% (κ = 0.72) (cut-off >20%) to 80% (κ = 0.57) (laboratory-specific cut-off). The corresponding figures for LA vs. RA were 90% and 83% (κ = 0.80 and 0.64), respectively.

HER2

The overall agreement for HER2 (IHC) (0/1 + vs. 2+/3+) was 86% (κ = 0.66) for LA vs. RA, 85% (κ = 0.64) for LA vs. IEO, and 96% (κ = 0.91) for RA vs. IEO (n = 256) (Table II). Of the 60 equivocal tumours (2+), four (7%) were amplified. In four of the six cases assessed as 1 + according to LA, but 2 + according to RA or IEO, no results of ISH were available, and these were therefore excluded from the agreement analyses concerning the final HER2 status. The agreement for final HER2 status (positive vs. negative) varied between 99.6% and 100% for the three comparisons, with corresponding κ between 0.98 and 1.00 (n = 248) (Table II).

The St Gallen 2013 clinicopathological surrogate definition of the intrinsic subtypes

The overall agreement in the comparisons of the five different subtypes for tumours with complete data for all four tumours markers was 91% (κ = 0.86) for LA vs. RA, 88% (κ = 0.81) for LA vs. IEO, and 94% (κ = 0.91) for RA vs. IEO (n = 233) (Tables II and IV).

Seventeen tumours were considered 'Luminal A-like' according IEO, but 'Luminal B-like/HER2-negative' according to LA, whereas 10 tumours showed the opposite pattern. One tumour was assessed as 'Triple-negative' according to IEO, but 'Luminal B-like/HER2-negative' according to LA, and one tumour was considered 'Luminal B-like/HER2-positive' according to LA, instead of 'Luminal A-like' according to IEO (Table IV).

Table III. Percentages of ER, PR, and Ki67, and HER2 status in tumours showing discrepancies in ER and PR (positive vs. negative), respectively.

		ER (%)			PR (%)			Ki67 (%)			HER2 (IHC)			
Lab	Antibody (ER)	LA	RA	IEO	LA	RA	IEO	LA	RA	IEO	LA	RA	IEO	HER2 (ISH)
4	6F11	40	40	0	0	0	0	80	70	78	0	–	0	*n.a.*
20	6F11	100	99	0	0	0	0	–	–	45	0	0	0	*n.a.*
29	SP1	50	30	0	0	0	0	90	95	90	1+	1+	1+	N
19	SP1	1	1	0	10	50	0	5	7	10	2+	2+	2+	N
29	6F11	0	1	0	0	3	2	40	31	32	0	0	0	N
		ER (%)			PR (%)			Ki67 (%)			HER2 (IHC)			
Lab	Antibody (PR)	LA	RA	IEO	LA	RA	IEO	LA	RA	IEO	LA	RA	IEO	HER2 (ISH)
26	1E2	0	0	0	50	50	0	80	72	82	1+	2+	2+	N
19	1E2	1	1	0	10	50	0	5	7	10	2+	2+	2+	N
19	1E2	0	0	0	100	5	0	40	40	35	3+	3+	3+	N
8	1E2	0	0	0	20	20	0	10	19	22	3+	3+	3+	A
8	1E2	1	<1	1	5	1	<1	76	66	85	0	1+	1+	*n.a.*
8	1E2	0	0	0	60	60	<1	60	54	71	3+	3+	3+	A
12	1E2	20	50	50	80	50	0	5	15	15	0	0	0	*n.a.*
8	1E2	100	99	99	0	0	2	6	11	11	1+	1+	1+	*n.a.*
18	1E2	80	90	90	0	1	1	15	20	22	2+	2+	2+	N
29	16,SAN27	0	1	0	0	3	2	40	31	32	0	0	0	N
5	636	100	99	99	0	20	20	12	6	7	1+	0	0	*n.a.*
5	636	95	99	99	0	5	5	10	7	4	1+	1+	1+	*n.a.*
9	636	100	99	99	3	<1	<1	15	15	14	2+	2+	2+	N

ER and PR (positive vs. negative)
- LA > 1%
- RA and IEO ≥ 1%

PR (high vs. low): > 20%
Ki67 (high vs. low): > 20%
Abbreviations: LA, Local Assessment; RA, Reviewed Assessment; IEO, New assessment at European Institute of Oncology; IHC, Immunohistochemistry; ISH, in situ hybridization; A, Amplified; N, Normal; n.a., Not analysed, –: missing.

Discussion

When the results of the primary routine analyses of breast cancer biomarkers in Sweden were re-assessed and re-analysed at the IEO reference laboratory in Milan, very good agreement was found for ER, PR, and HER2 ($\kappa \geq 0.85$), and good agreement was seen for Ki67 ($\kappa \geq 0.70$). To the best of our knowledge, no one has previously reported a national comparison of test results for all four biomarkers included in the St Gallen 2013 subtype classification [1]. In quality assurance trials, the participants are often aware that their assessments will be evaluated, and therefore may assess the samples more carefully than is usually the case in the routine clinical setting. A particular

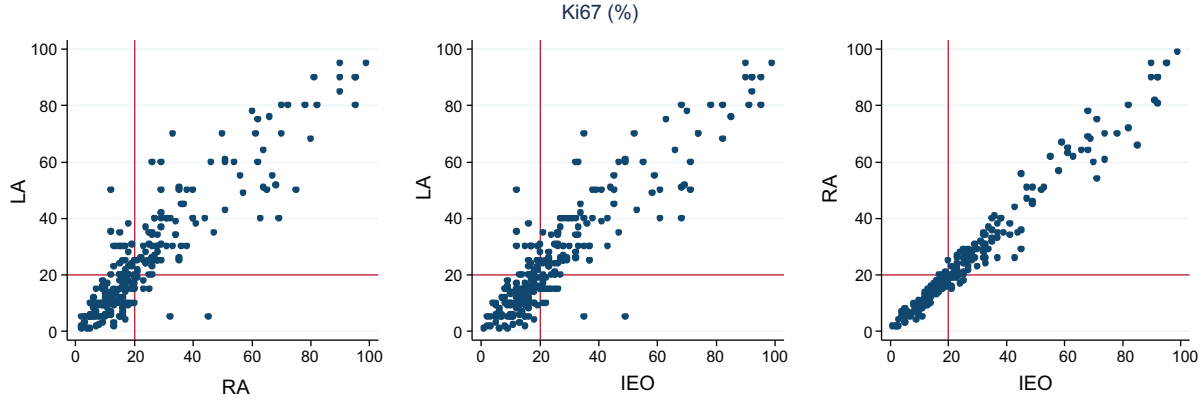

Figure 2. Comparisons of the Ki67 content (percentage) for the different assessments (LA vs. RA, LA vs. IEO, and RA vs. IEO).

Table IV. Comparison of subtype classification (St Gallen 2013), according to LA and IEO, for tumours with complete data for all three assessments.

	IEO					
	'Luminal A-like'	'Luminal B-like/ HER2-negative'	'Luminal B-like/ HER2-positive'	'HER2-positive/ non-luminal'	'Triple-negative'	Total (%)
LA						
'Luminal A-like'	83[a]	10	0	0	0	93 (40)
'Luminal B-like/HER2-negative'	17	78	0	0	1	96 (41)
'Luminal B-like/HER2-positive'	1	0	20	0	0	21 (9)
'HER2-positive/non-luminal'	0	0	0	6	0	6 (3)
'Triple-negative'	0	0	0	0	17	17 (7)
Total (%)	101 (43)	88 (38)	20 (9)	6 (3)	18 (8)	233 (100)

IEO, new assessment at European Institute of Oncology; LA, local assessment.
[a]number. Dark grey: No discrepancies; Light grey: Discrepancies.

strength of this study is, therefore, that the local assessments of the biomarkers ER, PR, Ki67, and HER2 were performed as part of the clinical routine, and the pathologists thus had no knowledge of the subsequent central testing, and the risk of bias must therefore be considered negligible. When comparing the results with the Swedish population-based figures from 2012 (cut-off: $>10\%$ for ER- and PR-positivity), a representative selection of tumours was found in the present study, where the percentage of ER-positivity was 87% compared to 87% in the population-based data [18]. The corresponding figures for PR and HER2 were 73% vs. 75% and 11% vs. 13%, respectively [18]. Although the present study is based on only 270 cases of breast cancer, the representativeness of the samples included indicates that the results could be generalised to the current overall quality of IHC assessment of these biomarkers in Sweden.

Regarding the tumours showing discrepancies in ER and PR, the results from LA and RA were more often in agreement than the RA and IEO results. Regitnig and co-workers also found greater variability in ER and PR when the participating laboratories performed both staining and assessment, compared to assessment of pre-stained sections [8]. Seven tumours, all analysed using the Ventana 1E2 RTU kit, were found to be PR-positive according to LA and RA, but PR-negative according to IEO, where the PharmDX kit (clone PgR 1294) was used. Similar observations concerning the Ventana 1E2 RTU kit have previously been described in both 2012 and 2014 assessment runs for PR performed by NordiQC [23]. Based on these findings, we performed a not predefined analysis demonstrating significantly higher PR values when the tumours were analysed by the Ventana 1E2 RTU kit (RA) compared to the PharmDX kit (clone PgR 1294) used at IEO ($p = 0.03$). Since all tumours were assessed by the same pathologist (LR), the results indicate a difference concerning the sensitivity of these antibodies. Thus, the Ventana1E2 RTU kit seems to give false PR positivity for some tumours, at worst resulting in adjuvant chemotherapy being withheld from patients with 'Luminal B-like/HER2-negative' tumours.

The discrepancy in HER2 classification was mainly due to differences in interpretation between 0/1+ and 2+. Since equivocal cases are analysed with ISH, which is a more robust and reproducible method than IHC analysis, the discrepancies in IHC assessments would have little influence on the final HER2 status, as was confirmed in the present study.

As there are concerns about the reproducibility of Ki67 assessments, an agreement of 85–94%, with corresponding values of κ of 0.70–0.89, can be considered very good. The interpretation of staining seems to be the most critical part of Ki67 analysis, since the concordance between the RA and IEO results was clearly better than between LA and RA, and LA and IEO (Figure 2), despite the fact that new tumour tissue sections were cut and stained for the IEO assays. In 38 of the 42 tumours showing a discrepancy, at least one of the assessments was located near the cut-off (15–25%). The laboratory-specific cut-offs for Ki67 used in Sweden may offer an advantage, in automatically adjusting for methodological changes. However, when these laboratory-specific cut-offs were applied in the local assessments the agreement was reduced from 90% to 83% ($\kappa = 0.80$–0.64) for LA vs. RA, and from 86% to 80% ($\kappa = 0.72$–0.57) for LA vs. IEO. Although laboratory-specific cut-off for Ki67 have been advocated by others, our results do not support this approach [1,9].

In order to illustrate the possible clinical consequences of discrepancies in results on the choice of adjuvant therapy, we used the subtype classification and the recommendations concerning medical adjuvant therapy from St Gallen 2013 [1]. As only

ER, PR, Ki67, and HER2, were considered in this study, we were not able to consider other well established prognostic factors, such as the TNM classification, histological grade, or the age of the patient at diagnosis. The agreement between the LA and the IEO was 88%, with a corresponding κ value of 0.81. As the recommendations for medical adjuvant therapy differ between the five subtypes, 12% of the patients would potentially have been treated differently. Most of the tumours showing discrepancies (27 of 29) were considered to be luminal A-like according to the local Swedish assessments, but luminal B-like according to central assessment at the reference laboratory, or vice versa. This could to a great extent be explained by the above mentioned problems of Ki67 values close to the cut-off.

In conclusion, the present national survey shows that adherence to national guidelines regarding the assessment of the breast cancer biomarkers ER, PR, Ki67, and HER2 ensures highly reproducible results, and this study confirms the clinical utility of these markers for decisions regarding adjuvant therapy in primary breast cancer. Participation in external quality assurance programmes should be mandatory for departments performing breast cancer pathology, as this provides information concerning the current performance of the laboratories, as well as providing opportunities for training and improvement.

Acknowledgements

This work was supported by funding from the Swedish Breast Cancer Association (BRO), Futurum – the Academy of Health and Care, Jönköping County Council, and the Foundation for Clinical Cancer Research in Jönköping. We are indebted to the participating pathology departments in Sweden, Equalis, and the SweQA Breast Cancer Working Group; Sara Baker, Tomasz Bujny, Maria Ekholm, Göran Elmberger, Margareta Eriksson, Mårten Fernö, Dorthe Grabau, Elisabeth Hagberg, Thomas Hatschek, Suzanne Johansson, Anikó Kovács, Barbro Linderholm, Hans Olsson, Malin Westerlund, Cecilia Winge, and Fredrik Wärnberg.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

[1] Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thurlimann B, et al. 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;24:2206–23.

[2] Prat A, Cheang MC, Martin M, Parker JS, Carrasco E, Caballero R, et al. Prognostic significance of progesterone receptor-positive tumor cells within immunohistochemically defined luminal A breast cancer. J Clin Oncol 2013;31:203–9.

[3] Cheang MC, Chia SK, Voduc D, Gao D, Leung S, Snider J, et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. J Natl Cancer Inst 2009; 101:736–50.

[4] Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E, et al. Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. J Clin Oncol 2011;29:4273–8.

[5] Fitzgibbons PL, Murphy DA, Hammond ME, Allred DC, Valenstein PN. Recommendations for validating estrogen and progesterone receptor immunohistochemistry assays. Arch Pathol Lab Med 2010;134:930–5.

[6] Rhodes A, Jasani B, Balaton AJ, Miller KD. 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.

[7] Mengel M, von Wasielewski R, Wiese B, Rudiger T, Muller-Hermelink HK, Kreipe H. Inter-laboratory and inter-observer reproducibility of immunohistochemical assessment of the Ki-67 labelling index in a large multi-centre trial. J Pathol 2002;198:292–9.

[8] Regitnig P, Reiner A, Dinges HP, Hofler G, Muller-Holzner E, Lax SF, et al. Quality assurance for detection of estrogen and progesterone receptors by immunohistochemistry in Austrian pathology laboratories. Virch Arch 2002;441: 328–34.

[9] Polley MY, Leung SC, McShane LM, Gao D, Hugh JC, Mastropasqua MG, et al. An international Ki67 reproducibility study. J Natl Cancer Inst 2013;105:1897–906.

[10] Varga Z, Diebold J, Dommann-Scherrer C, Frick H, Kaup D, Noske A, et al. How reliable is Ki-67 immunohistochemistry in grade 2 breast carcinomas? A QA study of the Swiss Working Group of Breast- and Gynecopathologists. PLoS One 2012;7:e37379.

[11] Dowsett M, Nielsen TO, A'Hern R, Bartlett J, Coombes RC, Cuzick J, et al. Assessment of Ki67 in breast cancer: Recommendations from the International Ki67 in Breast Cancer working group. J Natl Cancer Inst 2011;103:1656–64.

[12] Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, 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;28:2784–95.

[13] Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, 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.

[14] Wasielewski R, Hasselmann S, Ruschoff J, Fisseler-Eckhoff A, Kreipe H. Proficiency testing of immunohistochemical biomarker assays in breast cancer. Virch Arch 2008;453: 537–43.

[15] Parker RL, Huntsman DG, Lesack DW, Cupples JB, Grant DR, Akbari M, et al. Assessment of interlaboratory variation in the immunohistochemical determination of estrogen receptor status using a breast cancer tissue microarray. Am J Clin Pathol 2002;117:723–8.

[16] Wells CA, Sloane JP, Coleman D, Munt C, Amendoeira I, Apostolikas N, et al. Consistency of staining and reporting of oestrogen receptor immunocytochemistry within the European Union – an inter-laboratory study. Virch Arch 2004;445: 119–28.
[17] External quality assessment for clinical laboratory investigations (Equalis). [cited 2015 Mar 23]. Available from: http://www.equalis.se/en/start.aspx.
[18] Report from the national Swedish breast cancer register, year 2012. [cited 2015 Mar 23]. Available from: http://www.cancercentrum.se/sv/INCA/kvalitetsregister/Brostcancer/Brostcancer11/rapporter/.
[19] Swedish guidelines concerning breast pathology. [cited 2015 Mar 23] Available from: http://www.svfp.se/files/docs/kvast/brostpatologi/Brostcancerdokument_godkant_maj_2014.pdf.
[20] Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. Arch Pathol Lab Med 2007;131:18–43.
[21] Ryden L, Haglund M, Bendahl PO, Hatschek T, Kolaric A, Kovacs A, et al. 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;48:860–6.
[22] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33: 159–74.
[23] NordiQC Assessment Run for Progesterone receptor (PR); B14 2012 and B18 2014. [cited 2015 Mar 23]. Available from: http://www.nordiqc.org/Epitopes/PGR/PGR.htm.

Supplementary material available online

Supplementary Table I available online at http://informahealthcare.com/doi/abs/10.3109/0284186X.2015.1037012