EUROPEAN JOURNAL OF CANCER 42 (2006) 2946–2953

available at www.sciencedirect.com

journal homepage: www.ejconline.com

High inter-observer agreement in immunohistochemical evaluation of HER-2/*neu* expression in breast cancer: A multicentre GEFPICS study ☆

Magali Lacroix-Triki[a,*], *Simone Mathoulin-Pelissier*[b], *Jean-Pierre Ghnassia*[c], *Gaetan Macgrogan*[d], *Anne Vincent-Salomon*[e], *Véronique Brouste*[b], *Marie-Christine Mathieu*[f], *Pascal Roger*[g], *Frédéric Bibeau*[h], *Jocelyne Jacquemier*[i], *Frédérique Penault-Llorca*[j], *Laurent Arnould*[k]

[a]*Department of Pathology, Institut Claudius Regaud, 20–24 rue du pont Saint Pierre, 31052 Toulouse Cedex, France*
[b]*Department of Biostatistics, Institut Bergonié, Bordeaux, France*
[c]*Centre Paul Strauss, Strasbourg, France*
[d]*Institut Bergonié, Bordeaux, France*
[e]*Institut Curie, Paris, France*
[f]*Institut G. Roussy, Villejuif, France*
[g]*CHRU Montpellier, France*
[h]*Centre Val d'Aurelle, Montpellier, France*
[i]*Institut P. Calmette, Marseille, France*
[j]*Centre J. Perrin, Clermont-Ferrand, France*
[k]*Centre G.F. Leclerc, Dijon, France*

ARTICLE INFO

Article history:
Received 26 January 2006
Received in revised form
31 March 2006
Accepted 26 June 2006
Available online 20 September 2006

Keywords:
Breast cancer
HER-2/*neu*
Immunohistochemistry
Inter-observer reproducibility
Quality assurance
Ring study

ABSTRACT

The accurate determination of HER-2 in invasive breast cancer has become a critical issue, particularly in the context of the results of recent trastuzumab (Herceptin®) adjuvant trials. This multicentre study evaluated inter-observer reproducibility in interpretation of HER-2 immunostains performed in different laboratories according to their in-house technique. A total of 74 HER-2 immunostains were evaluated by 16 pathologists and by a central review committee. As determined by central review, the HER-2 score was 0 in 33 cases (44%), 1+ in 10 cases (13%), 2+ in 9 cases (12%) and 3+ in 23 cases (31%). The overall kappa value was good (kappa = 0.75). Agreement was excellent for the 0/1+ group (kappa = 0.85) and for the 3+ group (kappa = 0.82). As expected, the score 2+ group showed poor agreement (kappa = 0.38). A quality assurance process showed that ring studies and adherence to national guidelines greatly improve inter-observer reproducibility.

☆ GEFPICS: Groupe d'Etude des Facteurs Pronostiques par Immunohistochimie dans le Cancer du Sein (Group for the Evaluation of Immunohistochemical Prognostic Factors in Breast Cancer). Members of the group: the authors and M. Antoine, A. Balaton, M.-C. Baranzelli, V. Becette, J.P. Bellocq, M.-C. Blanc-Fournier, Y. Denoux, F. Ettore, M. Frassard, V. Fridman, B. Lannes, C. Migeon, C. Rigaud, C. Sagan, B. Sigal-Zafrani, I. Treilleux, V. Verrièle.

* *Corresponding author:* Tel.: +335 61 42 42 31; fax: +335 61 42 46 02.
E-mail address: lacroix.magali@claudiusregaud.fr (M. Lacroix-Triki).

0959-8049/$ - see front matter
doi:10.1016/j.ejca.2006.06.020

1. Introduction

Accurate determination of HER-2 status has become of major clinical importance in breast cancer. HER-2 over-expression, which is observed in 10–30% of breast carcinomas, is associated with poor clinical outcome, as it is correlated with shorter disease-free survival.[1] Moreover, HER-2 over-expression predicts for response to anti-HER-2 therapy with the recombinant humanised anti-p185$^{HER-2/neu}$ antibody trastuzumab (Herceptin®). Promising results of three trastuzumab adjuvant trials were presented at the American Society of Clinical Oncology 2005 annual meeting.[2,3] These three trials: the National Surgical Adjuvant Breast and Bowel Project B-31 (NSABP B-31); the North Central Cancer Treatment Group N9831 (NCCTG N9831); and the HERceptin® Adjuvant (HERA) trial, showed a 52% risk reduction for recurrence with adjuvant trastuzumab compared with controls.[2,3] Furthermore, accumulating evidence suggests that HER-2 over-expression may also predict sensitivity to anthracycline-based chemotherapy [4–6] and may help in the choice of endocrine therapy.[7]

Currently, no single assay is globally accepted as the gold standard for HER-2 determination, but among the three validated techniques, i.e. immunohistochemistry (IHC), fluorescence *in situ* hybridisation (FISH) and chromogenic *in situ* hybridisation (CISH), IHC is the most commonly used technique and the only one to study the treatment's target, i.e. the HER-2 protein. Indeed, IHC is a reliable, easy to perform and accessible technique, which is far less expensive and time-consuming than FISH. Moreover, several studies have demonstrated a high concordance between results for IHC and FISH.[8–12] However, the choice of technique for HER-2 determination remains a matter of debate. Firstly, the IHC technique was criticised because of a lack of inter-laboratory reproducibility due to the variability in fixation, tissue processing, IHC protocol, anti-HER-2 antibodies and scoring system used in each laboratory, i.e. the so-called 'real world'. Secondly, HER-2–IHC was considered to be a subjective test for the assessment of staining intensity and percentage of labelled tumoural cells, and hence prone to inter-observer variability.[13] In a clinical laboratory assay study of HER-2 testing, the College of American Pathologists thus emphasised the importance of adhering to standardised protocols for IHC-HER-2, claiming that the initial validation of assays against a gold standard is mandatory for accuracy.[14]

In this context, the French multicentre group Groupe d'Etude des Facteurs Pronostiques par Immunohistochimie dans le Cancer du Sein (GEFPICS) performed a study comparing the HER-2 status of 119 breast invasive carcinomas, determined by IHC in 12 different laboratories before and after calibration by reference to FISH on the corresponding frozen tissue sections.[15] This study showed that a high accuracy of IHC could be obtained for the determination of HER-2 status in all laboratories using their in-house IHC technique, provided that a calibration process was performed. Therefore, after examining the question of inter-laboratory variability, we aimed to study inter-observer reproducibility, particularly as there are few studies addressing this problem.

Inter-observer reproducibility was determined among 19 pathologists from 19 different institutions, who were asked to interpret the IHC–HER-2 stains of 75 breast invasive carcinomas.

2. Material and methods

2.1. Selection of cases

The cases for this study were derived from a previous French multicentre GEFPICS study, which concerned the calibration of IHC for assessment of HER-2 in breast cancer. Details of this study have been reported elsewhere.[15] Briefly, 12 different French laboratories selected 119 invasive breast carcinomas. Initial criteria for tumour selection were the availability of frozen tissue and previously immunohistochemically determined HER-2 status. Tumours were then secondary selected so as to balance the series with negative and positive cases in a range close to the distribution observed in clinical practice. At the end of this study, the 12 different laboratories had calibrated their HER-2 immunohistochemical technique, by reference to FISH performed on frozen sections. Among this series, 75 cases were used for the present study.

2.2. IHC procedure and FISH analysis

For each case, an IHC–HER-2 stain was performed by the laboratory of origin. Therefore, 4 different fixatives were used in this series (Table 1). Nineteen tumours were fixed in alcohol–formalin–acetic (ethanoic) acid (ethanoic acid 5%/ethanol 100° 75%/water 18%/commercial formalin (methanal) 2%), 43 in neutral buffered formalin, 9 in Hollande–Bouin's fixative and 4 in Bouin's fixative. The monoclonal antibody CB11 (Novocastra, Newcastle, United Kingdom (UK)) was used in 39 cases, the polyclonal antibody A0485 (Dako, Glostrup, Denmark) in 36 cases (Table 1). Details of the IHC procedure for each antibody are described in Table 2. Briefly, 11 out of 12 laboratories used a heat-induced antigen retrieval technique, mainly with a citrate pH 6 buffer. Primary antibody dilution and incubation time varied between the 12 laboratories. All the detection procedures relied on avidin-biotin-based systems, except in one centre which used a dextran polymer enhancing system (Envision®, Dako) for CB11 detection.

The FISH status was known in 72 cases, as it was performed in the previously described study.[15] These FISH experiments were carried out on frozen samples using the

Table 1 – Fixative and antibodies used for the HER-2 immunostains

Fixative	Cases (*n*)	Anti-HER-2	
		CB11[a]	A0485[b]
Acetic (ethanoic) acid /formalin/alcohol	19	15	4
Neutral buffered formalin	43	24	19
Hollande–Bouin's fixative	9	0	9
Bouin's fixative	4	0	4
Total	75	39	36

a Antibody CB11 (Novocastra, Newcastle, United Kingdom (UK)).
b Antibody A0485 (Dako, Glostrup, Denmark).

Table 2 – Immunohistochemistry technical procedures

Antibody	CB11 (6 laboratories)	A0485 (6 laboratories)
Antigen retrieval	Yes (5/6 laboratories)	Yes
Buffer	Citrate pH6 or EDTA pH7[a]	Citrate pH6
Method	Microwave, steamer or waterbath[b]	Steamer or waterbath[c]
Antibody dilution[d]	1:400–1:800	1:1200–1:2000
Incubation time[d] (min)	25–60	20–90

a One laboratory used EDTA pH7.
b Microwave in 2 laboratories, steamer in 2, and waterbath in 1.
c Steamer in 3 laboratories, waterbath in 3.
d Variation interval between the different laboratories.

PathVysion HER-2 probe kit (Vysis, Downer's Grove, IL, United States of America (USA)). Details of the technique used have been provided elsewhere.[15] Among the 72 cases, 31 were found to have an amplification of the HER-2 gene.

2.3. Study design

All of the 75 IHC–HER-2 stains were evaluated by 19 pathologists from 19 different institutions, in a 1-d session in one location. These pathologists recorded their evaluation on a worksheet (Appendix A). Briefly, the following criteria were used for IHC–HER-2 stain interpretation: percentage of stained invasive tumour cells and intensity of the staining, i.e. weak, moderate or strong intensity. Cytoplasmic staining without membranous staining was scored as negative. Negativity of normal glands was the prerequisite for interpreting the cases, according to the recommendations of the College of American Pathologists.[14] The final IHC score was given according to the HercepTest® for Immunoenzymatic Staining (Dako Corp., Carpinteria, CA, USA) scoring system, i.e. 0, absence of membranous staining or <10% stained cells; 1+, >10% stained cells with a weak and incomplete staining; 2+, >10% stained cells with a weak or moderate complete staining; 3+, >10% stained cells with a strong and complete staining. A case was considered to be HER-2 over-expressed if it scored 2+ (weak over-expression) or 3+ (strong over-expression) using the HercepTest® reporting system.

In addition to this individual IHC–HER-2 evaluation, a central review committee, composed of 5 of the 19 pathologists participating in the study (L.A., M.-C.M., P.R., G.M.G. and A.V.-S.), simultaneously interpreted the 75 slides on a multihead microscope in order to achieve consensus, following the previously defined recommendations.

2.4. Data analysis and statistical methods

Pathologists independently evaluated all of the immunohistochemical slides in order to evaluate the inter-rater reliability. The inter-observer reliability was measured with the kappa index for multiple readers.[16] Kappa is an index of agreement over and above that expected by chance.[17] The kappa index was calculated for the HercepTest® score in three (0/1+; 2+; 3+) and two categories (0/1+ and 2+/3+). This index was also calculated for percentage of stained cells using two cut-off points: 10%, as used in the HercepTest® reporting system, and 60%, as suggested by the GEFPICS's previous study.[15] All kappa indexes were presented with 95% confidence intervals (95% CI). Finally, all slides for which at least 4 pathologists of 16 were discordant (conflicting slides) were analysed.

3. Results

3.1. Characteristics of patients and tumours (Table 3)

The 75 cases of invasive breast carcinoma were distributed as follows: 62 (82.6%) invasive ductal carcinomas, 6 invasive lobular carcinomas, 3 apocrine carcinomas and 3 invasive micropapillary carcinomas. The majority of the tumours corresponded to histological grade III (62.7%), with a mean size of 32 mm. As determined by the central review committee, the HER-2 IHC score was 0 in 33 cases (44%), 1+ in 10 cases (13%), 2+ in 9 cases (12%) and 3+ in 23 cases (31%). The hormonal status of the tumours was known in 49 cases, showing oestrogen receptor (ER) and progesterone receptor (PR) expression in 63% and 47% of these cases, respectively. Forty-five of the 75 patients (60%) had histologically proven axillary lymph node metastases.

Table 3 – Characteristics of patients and tumours (75 cases)

Characteristics	
Age (years)	
Median	55 (range 32–96)
NA	1
Tumour size (mm)	
Median	28 (range 9–100)
NA	2
Histological type	
Ductal	62 (82.6%)
Lobular	6 (8%)
Others	6 (8%)
NA	1 (1.4%)
Histological grade	
I	8 (10.6%)
II	17 (22.7%)
III	47 (62.7%)
NA	3 (4%)
Axillary node status	
+	45 (60%)
–	26 (34.7%)
NA	4 (5.3%)

NA, not available.

3.2. Data analysis and quality of the slides

Among the 19 participating pathologists, 3 provided worksheets that were ineligible for statistical analysis due to missing data. For the same reason, 1 case out of the 75 slides examined was also excluded. Therefore, the statistical analysis was performed on 74 cases reviewed by 16 pathologists, i.e. on a total of 1184 immunostained slide interpretations.

Concerning the quality of the slides, a great majority (74%) of the slide interpretations was reported as satisfactory by the 16 reviewers, with only 5% reported as unsatisfactory. On average, 3 slides per pathologist were considered inadequate for interpretation. Among the 74 slides analysed, 4 were reported as unsatisfactory by at least 4 of the 16 pathologists. The reasons given for such an inadequacy included normal tissue staining, particularly in columnar cell change (2 cases), air bubble (1 case), non-specific background staining and loss of tissue adherence (1 case).

Table 4 – Kappa values (95% confidence interval) between the 16 pathologists for HER-2–IHC interpretation of the 74 invasive breast carcinoma cases

	Kappa values	95% CI
Herceptest® score		
3 categories		
global	0.75	0.59–0.91
0/1+	0.85	
2+	0.38	
3+	0.82	
2 categories[a]	0.85	0.83–0.87
Stained cells percentage		
10%	0.79	0.77–0.81
60%	0.75	0.73–0.77

IHC, immunohistochemistry; CI, confidence interval.
a Categories: 0/1+ and 2+/3+.

3.3. Agreement for the HercepTest® score (Table 4)

Using the HercepTest® reporting system, the overall kappa value was good (0.75, 95% CI 0.59–0.91) for the three main categories, i.e. 0 and 1+ versus 2+ versus 3+. Considering each category, the agreement was excellent for the 0/1+ group (specific kappa value = 0.85) and for the 3+ group (specific kappa value = 0.82). As expected, the intermediate category, i.e. score 2+ showed a poor agreement, with a specific kappa value of 0.38. Of interest, considering the 0/1+ group, the agreement was good for the score 0 (specific kappa value = 0.77), but mild for the score 1+ (specific kappa value = 0.44). Among the 160 slide interpretations in the 1+ group (i.e. 10 cases by 16 pathologists), 84 were scored as 2+, 73 as 1+, and 3 as 0. Thus, up-scoring in the 2+ category was responsible for the disagreement in the 1+ group. The overall agreement was slightly improved (overall kappa value = 0.85, 95% CI 0.83–0.87) when considering two main categories in the HercepTest® reporting system, i.e. 0/1+ versus 2+/3+. As expected, the kappa values were the same when considering the final conclusion, i.e. absence of over-expression, weak over-expression or strong over-expression, reported for each case (Table 4).

3.4. Agreement for the percentage of stained cells (Table 4)

As suggested by the previous GEFPICS study,[15] we used two cut-off values for stained cells, i.e. 10% (HercepTest® reporting system) and 60% to analyse the inter-observer agreement. Considering these two cut-off values, the 16 pathologists showed a good homogeneity in the distribution of the 74 cases (Fig. 1). Moreover, the overall agreement between the 16 pathologists was good, as shown by the kappa value of 0.79 (95% CI 0.77–0.81) and 0.75 (95% CI 0.73–0.77) obtained for the cut-off values of 10% and 60% stained cells, respectively.

3.5. Conflicting slides

Among the 74 cases analysed, 7 slides (9.4%) were reported as conflicting because of a disagreement over interpretation between at least 4 of the 16 pathologists. Characteristics and

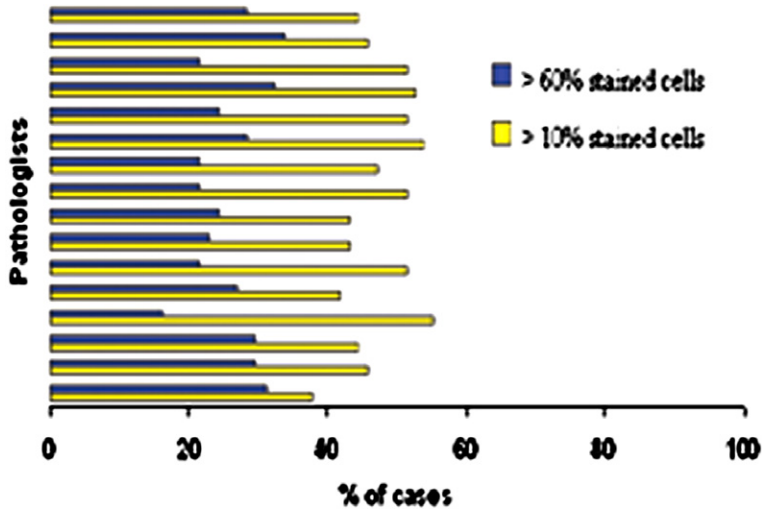

Fig. 1 – Distribution of the 74 cases by the 16 pathologists, using two different values of HER-2 positive tumour cells as cut-off values.

Table 5 – Characteristics and IHC-HER-2 interpretation results of the 7 conflicting slides

Slide number	Fixative	Antibody	Distribution of scores of the 16 pathologists (Herceptest® scoring system)					Central review	FISH
			0	1+	2+	3+	nd		
E4	Formol	A0485	2	10	3	1	0	1+	NA
I3	Formol	A0485	0	12	4	0	0	0	NA
K1	Holland	A0485	0	4	10	2	0	2+	A
K4	Formol	CB11	1	9	4	0	2	0	NA
R3	AFA	CB11	1	4	9	2	0	2+	A
O2	AFA	CB11	1	6	8	1	0	1+	A
R2	AFA	A0485	0	8	5	2	1	1+	NA

IHC, immunohistochemistry; FISH: fluorescent *in situ* hybridisation; nd, not determined; AFA, acetic (ethanoic) acid/formalin/alcohol; NA, non-amplified; A, amplified.

HER-2 interpretation results of these 7 cases are detailed in Table 5. These 7 slides were provided by 5 different laboratories. All of these cases were invasive ductal carcinomas. A disagreement was observed between 4 pathologists in 4 cases (slide numbers E4, I3, K1 and K4), between 5 pathologists in 1 case (slide number R3), and between 7 pathologists in 2 cases (slide numbers O2 and R2) (Table 5). Of note, 1 conflicting slide (number K4) was also reported as unsatisfactory by 4 pathologists, because of a normal tissue staining. Among these 7 cases, 2 cases showed a 2+ score, 3 cases showed a 1+ score and the remaining 2 cases showed a score of 0, as determined by the central review committee. In most of the cases, the majority of the pathologists was in agreement with the final interpretation of the central review committee. In 2 cases (slide numbers I3 and O2), the immunostaining was heterogeneous, thus partly explaining the disagreement (Fig. 2). Interestingly, only 1 of these conflicting cases showed a discrepancy between the IHC–HER-2 stain and the FISH result, i.e. was reported immunohistochemically as without over-expression while the FISH showed an amplification (slide number O2). This case presented with an unusual architectural papillary pattern (Fig. 2).

4. Discussion

Assessing the HER-2 status has now became an integral part of the optimal management of breast cancer patients, not only to identify a group of patients with worse prognosis, eligible for trastuzumab targeted therapy, but also to help in the choice of endocrine therapy. Indeed, recent studies have shown that patients presenting with an invasive breast cancer showing HER-2 over-expression and a positive oestrogen receptor status should preferentially receive aromatase inhibitors as endocrine adjuvant therapy.[7] Moreover, the results of three adjuvant trials (NSABP B–31, NCCTG N9831 and HERA trials) show the benefit to patients with HER-2-positive breast

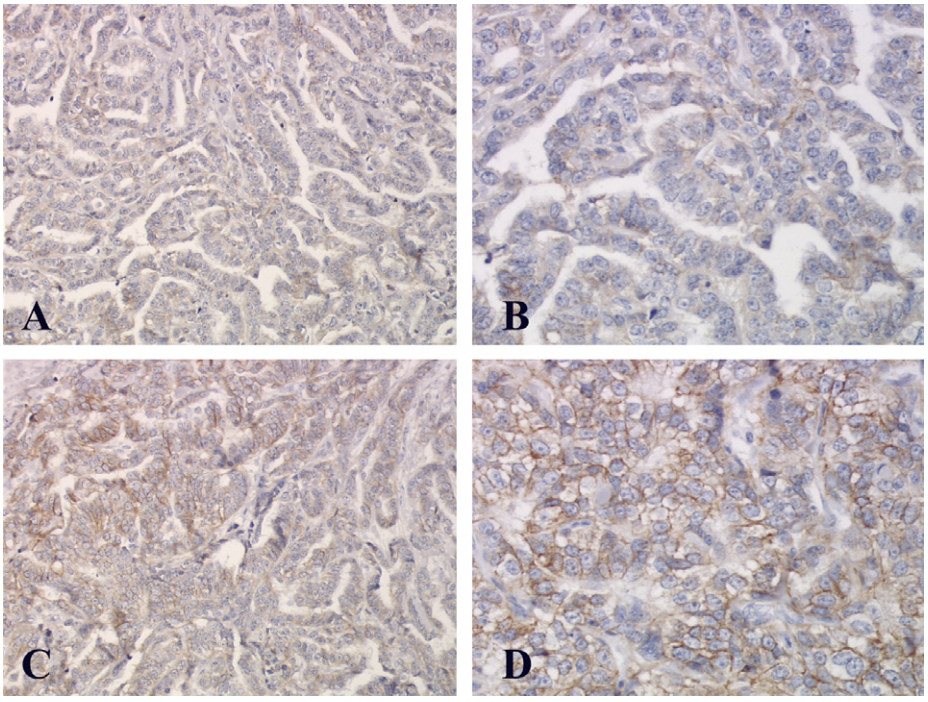

Fig. 2 – HER-2 immunostain of a discordant case (slide number O2) showing heterogeneous incomplete membrane staining (score 1+). (A, B) Area of focal and weak staining (A × 100, B × 200). (C, D) Area of incomplete moderate staining (C × 100, D × 200).

cancer of adjuvant trastuzumab therapy in reducing the risk of recurrence.[2,3] These recent data are positive additive arguments for the early and systematic determination of HER-2 status in patients with invasive breast cancer. Besides, in order best to determine which patients might benefit from trastuzumab therapy, three validated and US Food and Drug Administration (FDA)-approved techniques are now performed for HER-2 testing, i.e. IHC, FISH and CISH.

Whereas FISH and CISH remain expensive techniques performed by experienced laboratories, IHC–HER-2 is now widely used among pathology laboratories. Among the existing IHC–HER-2 scoring systems, the HercepTest® reporting system is the most commonly used. According to this system, 0 and 1+ scores are considered as non-HER-2 over-expressing cases, while a 2+ score corresponds to weak HER-2 over-expression and a 3+ score to strong over-expression. Present recommendations state that a 2+ score should be confirmed by FISH or CISH.[18–20] Hence, there are sufficient arguments supporting IHC as a valid technique for screening patients for trastuzumab treatment, i.e. low cost, convenience and biological relevance. It is now of great importance to fulfil the question of inter-laboratory and inter-observer reproducibility.

Numerous authors have emphasised the need to standardise the IHC–HER-2 technique and quality.[9,14,21,22] In a subset study of the NSABP B–31 trial, Paik and colleagues found a HER-2 discordance rate of 18% between local small-volume laboratories and a reference laboratory, supporting the idea of a central testing facility for HER-2.[22] Nevertheless, the unsolved question of inter-laboratory agreement with special regards to different methods of tissue fixation and processing is partly addressed by our previous GEFPICS multicentre study.[15] Indeed, we have shown a high accuracy of HER-2–IHC in all laboratories using their in-house IHC technique, after calibration against FISH.[15] Thus, our data and others clearly highlight the importance of quality control measures, such as periodic testing for concordance with FISH, to ensure accuracy and inter-laboratory agreement.[15,22]

Concerning the inter-observer reproducibility of IHC–HER-2 assessment, few data are available in the literature. The published studies report varying kappa values for inter-observer reproducibility, from 0.44 to 0.98.[23–27] Such a range of kappa value is probably explained by the differences in study methods, with special regard to the distribution of the HER-2 expression levels in the population analysed. In a study of inter- and intra-observer reproducibility, Santinelli and colleagues evaluated the agreement between three expert pathologists in IHC–HER-2 assessment of 108 breast cancer patients, including 65 HER-2 over-expressing cases.[24] With the HercepTest® they reached excellent inter-observer (kappa = 0.911) and intra-observer reproducibility (kappa = 0.863–0.926), but the distribution of the 108 cases in 0/1+/2+ and 3+ score is not clearly specified from the outset.[24] This HER-2 score distribution of the study population is of great importance in the overall inter-observer agreement, as several authors have shown that the majority of the inter-observer discordances occurred in the 2+ HER-2 over-expressing cases with a clear decrease in the kappa value for this 'grey-zone' category.[23,25,27] Indeed, the problems raised by the 2+ score are obvious in a study by Bilous and colleagues, obtaining a moderate overall agreement (kappa = 0.44) between 14 expert pathologists in the interpretation of only 5 IHC–HER-2 slides.[27] Among these 5 circulating slides, 2 HER-2 amplified cases as determined by FISH were evaluated as 2+ by half of the pathologists and as 3+ by the others. [27] In the present study, with a balanced cohort distribution close to the one observed in 'real-life' practice (44% score 0, 13% score 1+, 12% score 2+, 31% score 3+), we reach good overall inter-observer agreement (kappa = 0.75, 95% CI 0.59–0.91), with an expected decrease for the 2+ cases (kappa = 0.38). Of note, the percentage of 2+ cases in our study, which impact on the overall inter-observer agreement, is slightly higher than usually observed in clinical practice. The surprising mild agreement in the 1+ category (kappa = 0.44) was mainly due to an upscoring of the cases in the 2+ group. Nevertheless, this upgrading, which implies a secondary FISH determination in accordance with our clinical practice, is not damaging for the patients. Indeed, among the 10 cases scoring 1+ in this study, 9 showed no HER-2 amplification by FISH and would therefore not have received trastuzumab. Our results of inter-observer reproducibility between 16 pathologists are in accordance with those obtained by Thomson and colleagues between 3 pathologists (kappa = 0.64–0.79) and with a close population distribution (69% score 0, 6% score 1+, 5% score 2+, 20% score 3+).[26]

Among the 7 conflicting slides, 2 cases consisted of HER-2 2+ score with a low level of gene amplification, as determined by FISH. Only 1 case remained definitely equivocal as even the central review committee could not be in accordance with the FISH result (slide number O2). This case was the only one showing HER-2–FISH amplification among the 10 cases 1+ in this study. Such a discordance is extremely rare in our experience. It should be emphasised that in this case, as well as in 5 other cases, tumoural heterogeneity was mainly responsible for our disagreement. Only 1 slide was discordant because of a poor quality as revealed by normal gland staining. The factors contributing to inter-observer variability are exhaustively listed by Thomson and colleagues and Bilous and colleagues.[19,26] The most frequently invoked factors are heterogeneity of staining, edge and stroma retraction artefacts, non-specific staining of stroma and epithelium, weak staining of more than 30% of tumoural cells and extensive incomplete membrane staining. The presence of such factors on HER-2 immunostains should urge pathologists further to determine HER-2 status by FISH.

In the context of the striking results of the three trastuzumab trials NSABP B-31, NCCTG N9831 and HERA,[2,3] the need to improve our inter-observer and inter-laboratory reproducibility in IHC–HER-2 has now become critical. Recent years have seen the onset of comprehensive quality recommendations and national guidelines, which have certainly initiated an improvement in IHC–HER-2 assessment.[14,18,19,28] Among other things, these exhaustive guidelines recommend scoring the percentage and intensity of cells showing a complete membrane staining, to assess staining in the invasive component, to reject a test in case of normal gland staining, to be aware of retraction artefacts, and to exclude cytoplasmic staining.[19] Furthermore, as it is now crucial for laboratories to join external technical quality assurance programs, it is also highly recommended they join ring studies to promote training and experience.[18,19]

In conclusion, good inter-observer agreement was obtained in a ring study, performed on cases that were initially calibrated against FISH to improve the IHC–HER-2 technique. This study is in keeping with the quality assurance process, which is more than ever crucial in the context of the results of trials of adjuvant trastuzumab in invasive breast cancer. The standardisation of IHC–HER-2 protocols, associated with calibration against FISH and periodic external controls, allow high inter-laboratory reproducibility. This study shows that inter-observer reproducibility is improved by adherence to the published national guidelines.

Conflict of interest statement

None declared.

Appendix A. IHC–HER-2 stain worksheet

1. Pathologist' name
2. Case number
3. Slide quality
   - ☐ Satisfactory
   - ☐ Partly satisfactory
   - ☐ Unsatisfactory
4. Percentage of stained cells with complete membranous staining
5. Percentage of stained cells with incomplete membranous staining
6. Staining intensity
   - ☐ Absent
   - ☐ Weak
   - ☐ Moderate
   - ☐ Strong
7. Cytoplasmic staining
   - ☐ Present
   - ☐ Absent
8. Normal glands staining
   - ☐ Absent
   - ☐ Cytoplasmic staining
   - ☐ Membranous staining
   - ☐ Cytoplasmic and membranous staining
9. Score (Herceptest® reporting system, Dako)
   - ☐ 0
   - ☐ 1+
   - ☐ 2+
   - ☐ 3+
10. Conclusion
    - ☐ Absence of over-expression
    - ☐ Weak over-expression
    - ☐ Strong over-expression
    - ☐ Indeterminate case
11. Comment

REFERENCES

1. Slamon DJ, Clark GM, Wong SG, et al. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/*neu* oncogene. *Science* 1987;**235**:177–82.
2. Romond EH, Perez EA, Bryant J, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med* 2005;**353**:1673–84.
3. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med* 2005;**353**:1659–72.
4. Di Leo A, Larsimont D, Gancberg D, et al. HER-2 and topoisomerase II alpha as predictive markers in a population of node-positive breast cancer patients randomly treated with adjuvant CMF or epirubicin plus cyclophosphamide. *Ann Oncol* 2001;**12**:1081–9.
5. Paik S, Bryant J, Tan-Chiu E, et al. HER2 and choice of adjuvant chemotherapy for invasive breast cancer. National Surgical Adjuvant Breast and Bowel Project Protocol B-15. *J Natl Cancer Inst* 2000;**92**:1991–8.
6. Penault-Llorca F, Cayre A, Bouchet Mishellany F, et al. Induction chemotherapy for breast carcinoma: predictive markers and relation with outcome. *Int J Oncol* 2003;**22**:1319–25.
7. Ellis MJ, Coop A, Singh B, et al. Letrozole is more effective neoadjuvant endocrine therapy than tamoxifen for ErbB-1 and/or ErbB-2-positive, estrogen receptor-positive primary breast cancer. Evidence from a phase III randomized trial. *J Clin Oncol* 2001;**19**:3808–16.
8. Jacobs TW, Gown AM, Yaziji H, et al. Comparison of fluorescence in situ hybridization and immunohistochemistry for the evaluation of HER-2/neu in breast cancer. *J Clin Oncol* 1999;**17**:1974–82.
9. Jacobs TW, Gown AM, Yaziji H, et al. HER-2/neu protein expression in breast cancer evaluated by immunohistochemistry. A study of interlaboratory agreement. *Am J Clin Pathol* 2000;**113**:251–8.
10. Couturier J, Vincent-Salomon A, Nicolas A, et al. Strong correlation between results of fluorescent in situ hybridization and immunohistochemistry for the assessment of the ERBB2 (HER-2/neu) gene status in breast carcinoma. *Mod Pathol* 2000;**13**:1238–43.
11. Jimenez RE, Wallis T, Tabasczka P, et al. Determination of HER-2/neu status in breast carcinoma: comparative analysis of immunohistochemistry and fluorescent in situ hybridization. *Mod Pathol* 2000;**13**:37–45.
12. Lebeau A, Deimling D, Kaltz C, et al. HER-2/neu analysis in archival tissue samples of human breast cancer: comparison of immunohistochemistry and fluorescence in situ hybridization. *J Clin Oncol* 2001;**19**:354–63.
13. Kay EW, Walsh CJ, Cassidy M, et al. C-erbB-2 immunostaining: problems with interpretation. *J Clin Pathol* 1994;**47**:816–22.
14. Cell Markers and Cytogenetics Committees. College of American Pathologists. Clinical laboratory assays for HER-2/neu amplification and overexpression: quality assurance, standardization, and proficiency testing. *Arch Pathol Lab Med* 2002;**126**:803–8.
15. Vincent-Salomon A, MacGrogan G, Couturier J, et al. Calibration of immunohistochemistry for assessment of HER2 in breast cancer: results of the French multicentre GEFPICS study. *Histopathology* 2003;**42**:337–47.
16. Fleiss JL. *Statistical methods for rates and proportions*. New York: John Wiley and Sons; 1981.
17. McGinn T, Wyer PC, Newman TB, et al. Tips for learners of evidence-based medicine: 3. Measures of observer variability (kappa statistic). *CMAJ* 2004;**171**:1369–73.
18. Di Leo A, Dowsett M, Horten B, et al. Current status of HER2 testing. *Oncology* 2002;**63**:25–32.
19. Bilous M, Dowsett M, Hanna W, et al. Current perspectives on HER2 testing: a review of national testing guidelines. *Mod Pathol* 2003;**16**:173–82.

20. Namer M, Gligorov J, Luporsi E, et al. Saint-Paul-de-Vence. 'Cancers du sein' Recommandations pour la pratique clinique. *Oncologie* 2005;**7**:342–79.
21. Roche PC, Suman VJ, Jenkins RB, et al. Concordance between local and central laboratory HER2 testing in the breast intergroup trial N9831. *J Natl Cancer Inst* 2002;**94**:855–7.
22. Paik S, Bryant J, Tan-Chiu E, et al. Real-world performance of HER-2 testing – National Surgical Adjuvant Breast and Bowel Project experience. *J Natl Cancer Inst* 2002;**94**:852–4.
23. Tsuda H, Sasano H, Akiyama F, et al. Evaluation of interobserver agreement in scoring immunohistochemical results of HER-2/neu (c-erbB-2) expression detected by HercepTest, Nichirei polyclonal antibody, CB11 and TAB250 in breast carcinoma. *Pathol Int* 2002;**52**:126–34.
24. Santinelli A, Baccarini M, Colanzi P, et al. Immunohistochemical evaluation of HER-2/neu expression in infiltrating breast carcinoma: a study of reproducibility. *Anal Quant Cytol Histol* 2002;**24**:54–62.
25. Hoang MP, Sahin AA, Ordonez NG, et al. HER-2/neu gene amplification compared with HER-2/neu protein overexpression and interobserver reproducibility in invasive breast carcinoma. *Am J Clin Pathol* 2000;**113**:852–9.
26. Thomson TA, Hayes MM, Spinelli JJ, et al. HER-2/neu in breast cancer: interobserver variability and performance of immunohistochemistry with 4 antibodies compared with fluorescent in situ hybridization. *Mod Pathol* 2001;**14**:1079–86.
27. Bilous M, Ades C, Armes J, et al. Predicting the HER-2 status of breast cancer from basic histopathology data: an analysis of 1500 breast cancers as part of the HER2000 International Study. *Breast* 2003;**12**:92–8.
28. Penault-Llorca F, Balaton A, Sabourin JC, et al. Evaluation immunohistochimique du statut HER-2 dans les carcinomes mammaires infiltrants : mise au point du protocole technique et de la lecture des résultats – recommandations. *Ann Pathol* 2002;**22**:150–7.