Pathol. Oncol. Res. (2008) 14:253–259
DOI 10.1007/s12253-008-9079-z

ORIGINAL PAPER

Practice of HER-2 Immunohistochemistry in Breast Carcinoma in Austria

A. Reiner-Concin · P. Regitnig · H. P. Dinges · G. Höfler · S. Lax · E. Müller-Holzner · P. Obrist · M. Rudas

Received: 30 March 2008 / Accepted: 18 June 2008 / Published online: 28 August 2008

Abstract Practice and accuracy of immunohistochemistry is known to vary highly. Reliability of HER-2 immunohistochemistry is critical because of its role in patient selection for therapeutical options in breast cancer. Therefore reliability of HER-2 immunohistochemistry in pathology laboratories in Austria was assessed. Ten tissue specimens of invasive ductal breast carcinomas and three cell line samples were tested. Presence/absence of gene amplification was determined by FISH to be used as a gold standard. Laboratories were asked to stain and assess slides using their routine immunohistochemical staining protocol. Overall the study consisted of 311 tests on tissue specimens and 142 on cell lines. In all cases manual scoring was performed. Participation was voluntary and was 94%. Overall sensitivity was 90.5% and specificity 99.2%. Overscoring including true false positive results were found in 6.7% and 6.3% in tissue specimens and cell lines, respectively. False negative determinations were obtained in 1.9% and 2.8% of tissue specimens and cell lines, respectively. HercepTest™ showed slightly higher reliability in comparison with individualized staining methods. By manual scoring inaccurate scoring affected 12.3% of test results and 62% of the laboratories. In conclusion participation rate and accuracy of HER-immunohistochemistry was high all over the country. Manually performed scoring demonstrated some limitations.

Keywords Breast cancer · HER-2 · Immunohistochemistry · Quality assurance · Reproducibility

Abbreviations
IHC immunohistochemistry
FISH fluorescence in situ hybridization

Introduction

Immunohistochemical testing of HER-2 in invasive breast cancer has become a routine diagnostic procedure. It is recommended by the ASCO/CAP guidelines in every invasive breast cancer [1]. Testing is usually performed in a two tired

A. Reiner-Concin
Department of Pathology, Donauspital,
Vienna, Austria

P. Regitnig · G. Höfler
Department of Pathology, Medical University of Graz,
Graz, Austria

H. P. Dinges
Department of Pathology, Hospital Klagenfurt,
Klagenfurt, Austria

S. Lax
Department of Pathology, LKH-West Graz,
Graz, Austria

E. Müller-Holzner
Department of Obstetrics and Gynecology,
Medical University of Innsbruck,
Innsbruck, Austria

P. Obrist
Pathohistological Laboratory,
Landeck, Austria

M. Rudas
Department of Pathology, Medical University of Vienna,
Vienna, Austria

A. Reiner-Concin (✉)
Pathologisch-Bakteriologisches Institut,
Donauspital, Langobardenstrasse 122,
1220 Vienna, Austria
e-mail: angelika.reiner@wienkav.at

system. Immunohistochemistry (IHC) being performed in the first step followed by in situ hybridisation in equivocal cases. Equivocal results are considered IHC staining score 2+ or staining artifacts. The second test level most frequently consists of fluorescence in situ hybridisation (FISH) assessing gene amplification or non-amplification at the DNA level.

HER-2 is accepted as a predictive factor for targeted therapeutics, e.g. the humanised anti-HER-2 monoclonal antibody trastuzumab and with respect to hormonal and chemotherapy under certain circumstances [2–5]. Therefore accurate testing especially at the first level by IHC has become a critical issue. However, some limitations are well known for IHC. It is generally known to be difficult to standardize due to methodological problems from tissue fixation including duration of fixation and used fixative, pretreatment and antibodies used. With regard to sensitivity of HER-2 antibodies significant differences between 6% and 82% are reported [6]. Microscopic interpretation in IHC for HER-2 is recommended to be performed semiquantitatively. But semiquantitative interpretation of IHC particularly when performed manually is susceptible to subjectivity and variability [7]. Inter- and intraobserver variability in IHC in general are well known problems and also are reported for HercepTest™ [8]. These problems with HER-2 IHC are an important issue for low volume laboratories where approximately 20% of HER-2 immunohistochemical assays proved to be false on retesting in a central high volume laboratory [9]. Mainly false positive HER-2 reporting was affected. But false negative HER-2 assessment was also reported in slide circulations [10]. Both false positive and false negative results have to be avoided because due to such test results patients would be withheld from therapy and wrong expectations would be raised. On the other hand unnecessary side effects would be taken irresponsibly.

Austria has a low population of approximately 8 million inhabitants and thus a limited number of primary breast cancers. Approximately 5,000 primary breast carcinomas are diagnosed per year. Histologic diagnosis is performed in 34 pathology laboratories over the country and most of the laboratories perform IHC for HER-2. Therefore there exist only a few high volume laboratories and many laboratories are only low volume performers with respect to HER-2 IHC. Therefore we wanted to assess the performance and reliability of HER-2 IHC all over the country and encourage pathologists to participate in slide circulations for quality assurance.

Materials and Methods

Tissue Specimens

All tissues were fixed in 7% phosphate buffered neutral formalin between 24 and 48 h. Paraffin-embedded tissue specimens were selected from the surgical pathology archive at the Donauspital, Vienna, Austria and the Department of Pathology at the Medical University Graz, Austria. The tissues consisted of primary invasive ductal breast cancer specimens. From the paraffin blocks core biopsies were punched out by a core biopsy punch instrument usually taken for skin biopsies at a diameter of 5 mm and blocked into macrotissue arrays. Two macrotissue arrays containing five tissue specimens each were built. Altogether the trial consisted of 10 tissue specimens of invasive breast cancer with known HER-2 amplification status using PathVysion (Abbott Molecular, Illinois formerly Vysis) and three cell lines as controls as provided with the HercepTest™ (Dako, Glostrup, Denmark). Two of the selected breast carcinomas were proven to be high level amplified and eight were not amplified for the HER-2 gene. All cases showed CEP 17 diploidy. In addition to tissue specimens on each slide a section containing three cell lines (MDA-231, MDA-175 and SK-BR-3) as provided with the HercepTest™ was mounted. Sections of the cell lines were provided by Dako. Unstained slides were sent to participating laboratories to be stained using the in-house laboratory protocol.

Scoring

Scoring of HER-2 of the stained slides was performed according to the HercepTest™ which is now also recommended by ASCO/CAP guidelines. Briefly score 0 and 1+ were considered negative and defined by no staining or weak incomplete membrane staining of tumor cells. Score 2+ was considered equivocal for HER-2 and defined by weak or moderate complete membrane staining in at least 10% of tumor cells. Score 3+ was considered positive for HER-2 and characterized by uniform intense membrane staining in more than 10% of tumor cells. This study was carried out before the ASCO/CAP guidelines were published and the new threshold of 30% of positive tumor cells was introduced.

Participants were asked to report all the results including single parameters and the HER-2 score on standardized forms. Single parameters asked for were staining intensity (low, intermediate, high or absent), percentage of positive stained tumor cells graded stepwise by 10 percentiles and membrane staining being either complete, incomplete or absent. A second form was requested to be filled out concerning the immunohistochemical staining protocol used.

Participation

The study was designed for participation of 34 histology laboratories in departments of pathology in public hospitals in Austria. Participation was voluntary. 94% (32/34) of pathology laboratories participated. All laboratories were

coded for data entry guaranteeing anonymity. The study center was situated in the Department of Pathology in the Donauspital, Vienna, Austria.

Statistics

Excel 2003 (Microsoft, Redmond, USA) was used for data input and score calculation. STATA 6.0 (Stata Corporation, Texas, USA) was used for statistical analysis of the results. Sensitivity and specificity were calculated for positive and negative test results as the main factors for treatment decisions. Multirater kappa was chosen as a measure of interobserver and interlaboratory variability. Multirater kappa is widely used and hence comparable in medical studies dealing with interobserver variability. Kappa values were calculated according to Fleiss [11] for each group separately and for all groups (overall kappa). Overall kappa is the summary of the agreement across all observers, adjusted for the level of agreement that would be expected to occur solely by chance. Kappa was interpreted according to Landis and Koch [12]. Nevertheless multirater kappa harbours some disadvantages when used for three or more categories, as the number of possibilities that theoretically can be chosen by observers is not included in the calculation and therefore different estimation systems in other studies are difficult to compare.

Results

Ninety-four percent (32/34) of pathology laboratories in public hospitals in Austria participated. Thus the participation rate was very high and the results can be assumed to be representative for the accuracy of HER-2 IHC in Austria.

Overall, 453 HER-2 immunohistochemical test results were achieved. 311 of the tests were performed on tissue specimens and 142 on cell lines. Correct results according to FISH and defined by cell lines were found in 97.4% of tissue specimens and 95% of cell lines. Detailed results for tissue specimens and cell lines are demonstrated in Tables 1 and 2.

Table 2 Results of HER-2 IHC: Cell lines, $n=142$

Case	Defined score	Score 0	Score 1+	Score 2+	Score 3+
1f	0	20	1	1	1
2f	0	20	1	1	2
1g	1+	8	13	2	0
2g	1+	7	16	2	0
1h	3+	2	0	0	21
2h	3+	2	0	1	21
Total		59	31	7	45

In six of the tissue specimens with HER-2 gene amplification laboratories reported score 1+ resulting in a false negative report. Both breast cancer samples with HER-2 amplification were affected by these false negative assessments. The false negative assessments happened in six different laboratories and are demonstrated in Fig. 1. Only two of the six laboratories were able to report a score 3+ in one of the two HER-2 amplified tissue sample. Four of the laboratories (13% of participating laboratories) clearly reported results of too low staining intensity in both amplified tissue samples.

In cell line samples score 0 assessments were reported four times in the HER-2 amplified cell line (SK-BR-3) resulting in false negative assessment. Two of these false negative assessments occurred in one laboratory which also performed one of the false negative assessments in tissue samples. The other two false negative results occurred in two different laboratories which were able to provide correct results for the tissue specimens. Overall, 1.9% and 2.8% of the IHC assays resulted in false negative determinations in tissue specimens and cell lines, respectively.

Taking into account score 3+ and 2+ IHC results in samples without HER-2 amplification together overscoring was found in 6.7% and 6.3% in tissue specimens and cell lines, respectively. These were in detail two false positive assessments showing score 3+ in tissue samples without HER-2 amplification. In cell lines without HER-2 amplifi-

Table 1 Results of HER-2 IHC: Tissue specimens, $n=311$

Case	Amplification	Score 0	Score 1+	Score 2+	Score 3+
1a	No	16	12	2	0
1b	No	11	16	4	1
1d	No	27	5	0	0
1e	No	30	1	0	0
2a	No	27	0	0	0
2b	No	9	13	10	0
2c	No	18	14	0	0
2e	No	11	17	3	1
1c	Amplified	0	2	3	26
2d	Amplified	0	4	9	19
Total		149	84	31	47

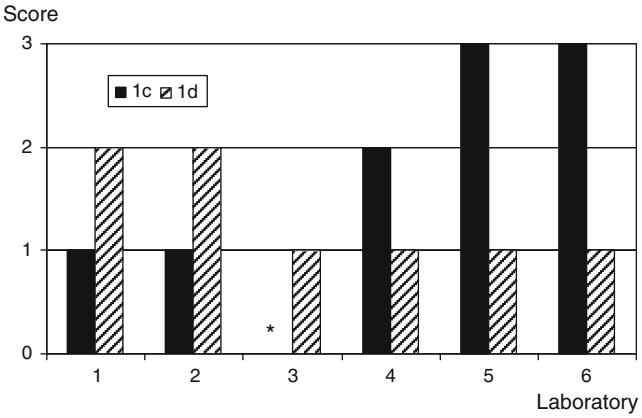

Fig. 1 False negative HER-2 assessment for individual laboratories in both HER- 2 gene amplified tissue samples. *1c*, *1d* tissue specimens. *Asterisk* Tissue lost during staining

cation three false positive results (score 3+) were found (Table 2).

The results for the score 2+ category were somewhat heterogeneous. In total 31 score 2+ results occurred in tissue specimens (Table 1). Of these 19 occurred in HER-2 non amplified carcinomas and thus could be interpreted as overscored and 12 occurred in HER-2 amplified carcinomas representing underscoring. In cell lines a total of 7 score 2+ assessments was found (Table 2). 6 of them were found in cell lines without HER-2 amplification and could be interpreted as overscoring. Only one assessment of score 2+ was found in the cell line with HER-2 amplification demonstrating underscoring.

With respect to recommendations of guidelines where retesting of score 2+ by in situ hybridisation is recommended overall sensitivity was 90.5% and specificity 99.2%. Positive and negative predictive values were 96.6% and 97.6%, respectively.

With respect to variability of microscopic interpretation scores reported by laboratories and scores determined by recalculation of reported single parameters according to guidelines (staining intensity and membrane staining) were compared. 56 of 453 (12.3%) scores presented with discordant scoring results. These discrepancies affected 21 (62%) laboratories. By individual laboratory on average 2.6 discrepant results were found. The maximum number of discrepant reporting was 6 (Fig. 2). Errors were due as well to incorrect discrimination between complete and incomplete membrane staining and due to estimation of percentage of stained cells. No significant single factor contributing to this finding could be identified (data not shown). In Fig. 3 reported and recalculated scores are demonstrated for tissue specimens and cell lines by individual cases. No significant trends with respect to differences can be demonstrated. There were only two specimens of tissue samples and one cell line where no differences were found.

Analysing immunohistochemical staining protocols 27 laboratories gave sufficient information in order to compare results with respect to laboratory protocols. Thirteen laboratories performed the HercepTest^{TM} strictly according to the manufacturer's protocol. Fourteen laboratories reported a variety of individualized immunohistochemical staining protocols. Overall, HercepTest^{TM} showed a higher κ value (κ=0.46) in comparison with individualized staining methods (κ=0.35). Concerning the single categories at the extremes of the spectrum represented by score 0 and score 3+ kappa was in the very good to good range for HercepTest^{TM}. In comparison kappa values in individualized staining protocols were lower in these scores (Table 3).

Discussion

HER-2 is a critical predictive marker for response to several therapeutic options in breast cancer patients. Thus meanwhile clinical demand on HER-2 is a routine question. Therefore testing is recommended by the ASCO/CAP

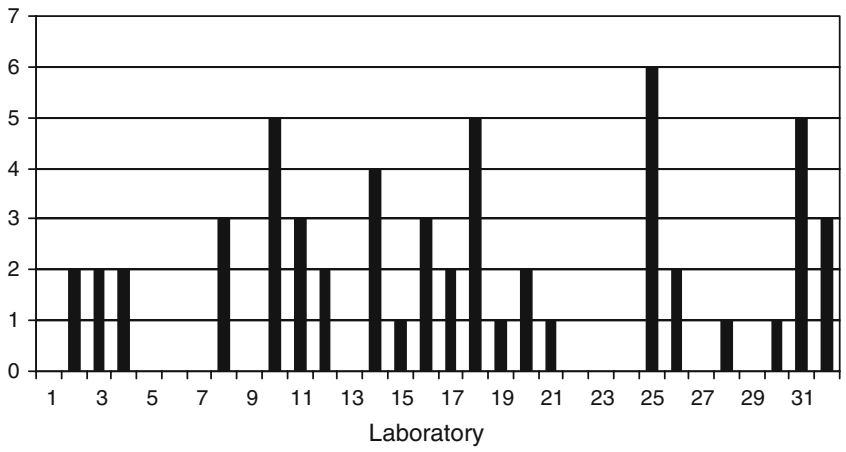

Fig. 2 Number of discrepancies by laboratory between scoring reported by laboratories and recalculated score by reported single parameters

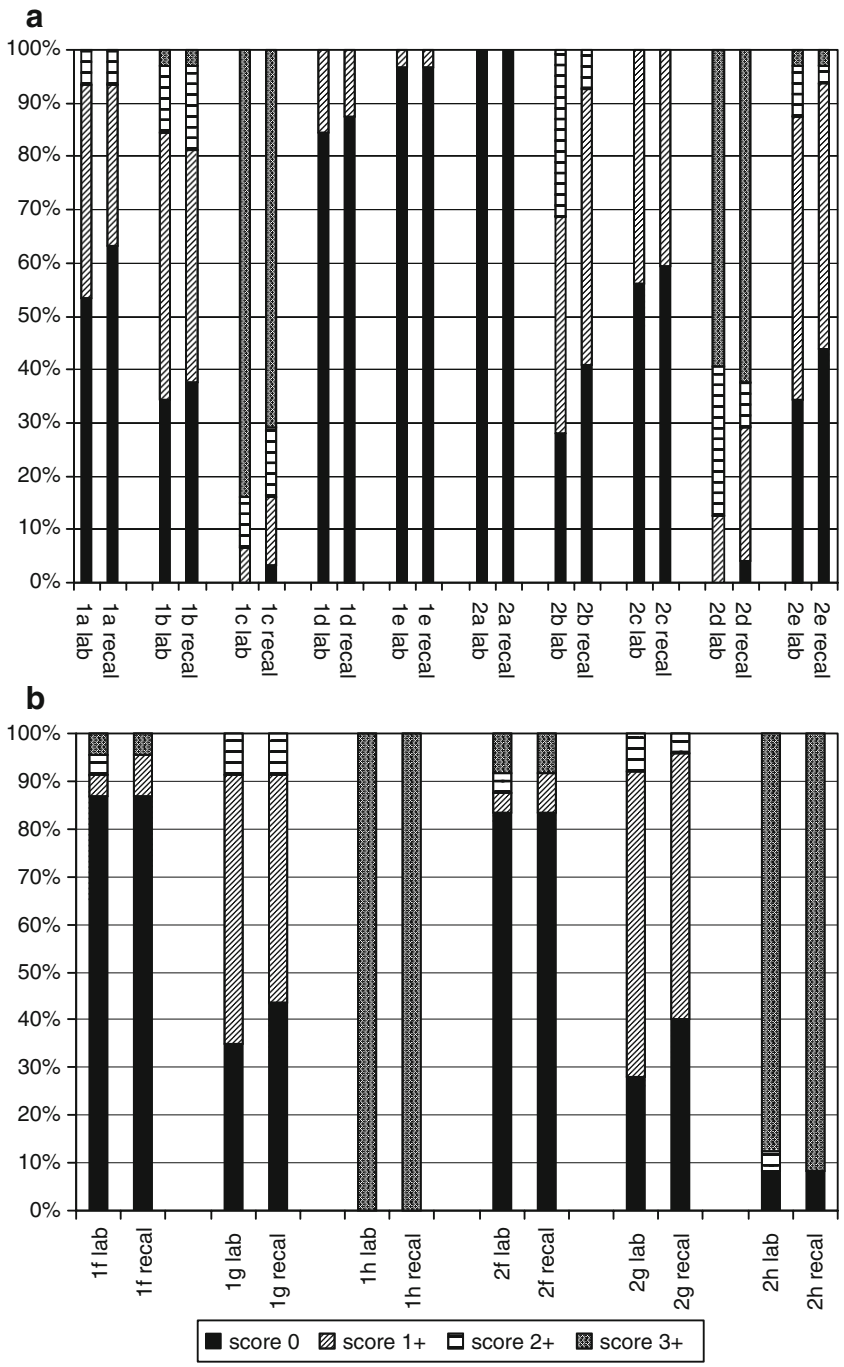

Fig. 3 Comparison between scores reported by laboratories and recalculated by reported single parameters. **a** Tissue specimens, **b** cell lines

guidelines in every invasive breast cancer [1]. It usually consists of testing by IHC at the first level followed by in situ hybridisation performed either by fluorescence or chromogenic in situ hybridisation at the second level in certain cases. At the immunohistochemical level a positive result is defined by score 3+. A negative result is defined as score 0 and 1+ respectively. Equivocal results defined as score 2+ require further testing. It is known from the literature that approximately 20% of current HER-2 testing may be inaccurate [9] and correlation between FISH which

Table 3 Comparison of IHC by staining protocols demonstrated by kappa-statistics

	Herceptest κ	Individual protocol κ
Score 0	0.54	0.35
Score 1+	0.31	0.19
Score 2+	0.15	0.12
Score 3+	0.71	0.65

is currently widely accepted as the gold standard and particularly score 2+ in IHC may be as low as 25% [13]. Due to incorrect testing not all patients may receive appropriate therapies. Therefore urgent need for quality assurance exists. It was the goal of our study to assess the performance and reliability of HER-2 IHC testing in breast cancer in Austrian pathology laboratories. In addition by this initiative laboratories should be convinced of the usefulness of participation in slide circulations and encouraged to participate in such programs.

Our findings demonstrate that the overall accuracy of HER-2 assessment in IHC all over the country was high. This finding guaranteeing patient safety with respect to selection of therapy is important because comparing Austria as a small country with larger countries relatively low numbers of primary breast cancers occur. Therefore only a small number of pathology laboratories would meet the minimum numbers to test recommended for quality assured testing [14]. However, daily practice of HER-2 testing i.e. IHC does not take place in only a small number of high volume laboratories but is rather organized as part of routine histologic procedures in almost every histopathology laboratory over the country.

In a few laboratories problems with too low sensitivity of immunohistochemical staining were identified. The majority of these laboratories did not achieve appropriate results in more than one specimens. Too low sensitivities may be due to several technical aspects. It may be due to insufficiency and variation in tissue fixation. This can be ruled out in our study because all the tissue specimens derived from two centers using fixation protocols according to the ASCO/CAP guidelines and the majority of laboratories were able to achieve appropriate results. It may also be due to low sensitivity of antibodies used or insufficient antigen retrieval applied. Both explanations are possible in our study because false negative results determined by individualized staining methods occurred slightly more frequently than by HercepTest™. Another explanation is given by reports in the literature reporting prolonged storage of sections cut from paraffin blocks at room temperature may result in antigen loss leading to decreased sensitivity [15]. This limitation is not true for our study because sections were cut immediately before mailing to participants. According to the recommendation of the manufacturer of the HercepTest™ participants were instructed performing stainings within four weeks. Thus time limits in our slide circulation were kept within the limit of 36 weeks which was described as the maximum shelf life of cut sections for HER-2 IHC [16]. In addition one has to keep in mind that all participating laboratories had the same requirements with respect to time and most performed testing accurately.

In the literature high rates of inaccuracy up to 20% are reported and these are mostly due to false positive assessments [17–19]. As described in these studies false positive testing in IHC was associated also with HercepTest™ where standardized staining procedures can be assumed. This was described particularly for cases scoring 2+. Regarding amplification status score 2+ can be assumed as overscoring. With respect to this view the majority of our score 2+ results could be interpreted as overscoring in IHC. They occurred in tissue samples and cell lines, in individualized staining procedures and also HercepTest™. Overscoring with score 2+ will never be eliminated completely since it is inherent to the method but the goal should be to reduce it to a minimum even when followed by retesting by in situ hybridisation and thus being of no clinical relevance.

As could be demonstrated by kappa statistics, differences between HercepTest™ and individualized staining procedures were demonstrated. Overall kappa was in the moderate range for the HercepTest™ while it was in the slight range for individualized staining procedures. This is in accordance with the literature where HercepTest™ was shown to be more reliable in comparison with individualized methods [10, 20]. In addition higher kappa values at the extremes at score 0 and 3+ were found for the HercepTest™. Particularly for score 3+ was kappa in the substantial range while it was only fair in score 1+ and 2+. This is in accordance to the literature suggesting that IHC scoring is highly predictive for gene amplification status only at the extreme ends of the scores represented by score 0 and 3+. For these scores in the literature also interobserver agreement was found to be highly satisfactory. On the other hand IHC is neither reliable for prediction of gene status in score 1+ and 2+ nor is interobserver agreement satisfactory in these scores [21].

Several guidelines and several organisations running quality assurance programs suggest inclusion of cell lines as standards in the testing process [13, 16]. It was demonstrated during a two year study period in an international slide circulation that the number of laboratories achieving appropriate results in cell lines was significantly improved [16]. Surprisingly in our study there were slightly more inappropriate results in cell lines compared with inappropriate results in tissue specimens. There is no clear explanation for this result. But possibly participants have had more difficulties in applying scoring in cell lines than in tissue specimens. This could be in agreement with the finding of Rhodes et al. finding that scoring by image analysis on cell lines was improved in comparison to manual scoring [16]. In our study all the scoring was performed manually.

As discussed above most inappropriate assessments occurred in score 1+ and 2+ specimens in our study and all scoring was performed manually. Thus some of the inappropriate results could be explained not only by

technical reasons in staining procedures but also by subjectivity of microscopic interpretation of staining itself. Participants were instructed of using scoring defined by HercepTest™. In addition results needed to be given in defined reporting forms thus minimizing variability of reporting. As is demonstrated by this approach stringent application even of defined scoring criteria is difficult and may result in variability. This was proven by recalculation of scores in the study where a substantial proportion of laboratories presented with discordant scores. A solution to this problem is suggested in the literature with application of image analysis. As demonstrated by comparison of results derived from analysis by an automated cellular imaging system with manually established results could be improved in particular for score 2+ cases [7]. However, image analysis is not feasible in daily practice yet. Therefore to our opinion training of scoring at the individual level is very important.

In conclusion our study demonstrated high accuracy of HER-2 testing at the first level by IHC all over the country. In a few cases false negative assessments could have prevented patients from therapies and those laboratories were informed for the need of improvement of their IHC procedure. Manual scoring is limited by subjectivity and this is of special importance in the intermediate categories of the score. As long as image analysis is not available readily generous testing at the second level by in situ hybridisation seems to be prudent. This approach may become more realistic with the current development of in situ hybridisation methods at the light microscopic level as for instance the newly developed silver in situ hybridisation [22].

Acknowledgment The project was supported by a grant of the Austrian Society of Senology.

References

1. Wolff AC, Hammond ME, Schwartz JN et al (2007) 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 131:18–43
2. Gennari A, Sormani MP, Pronzato P et al (2008) HER2 status and efficacy of adjuvant anthracyclines in early breast cancer: a pooled analysis of randomized trials. J Natl Cancer Inst 100:14–20
3. Yamauchi H, Stearns V, Hayes DF (2001) When is a tumor marker ready for prime time? A case study of c-erbB-2 as a predictive factor in breast cancer. J Clin Oncol 19:2334–2356
4. Konecny G, Pauletti G, Pegram M et al (2003) Quantitative association between HER-2/neu and steroid hormone receptors in hormone receptor-positive primary breast cancer. J Natl Cancer Inst 95:142–153
5. Ménard S, Valagussa P, Pilotti S et al (2001) Response to cyclophosphamide, methotrexate, and fluorouracil in lymph node-positive breast cancer according to HER2 overexpression and other tumor biologic variables. J Clin Oncol 19:329–335
6. Press MF, Hung G, Godolphin W et al (1994) Sensitivity of HER-2/neu antibodies in archival tissue samples: potential source of error in immunohistochemical studies of oncogene expression. Cancer Res 54:2771–2777
7. Wang S, Saboorian MH, Frenkel EP et al (2001) Assessment of HER-2/neu status in breast cancer. Automated Cellular Imaging System (ACIS)-assisted quantitation of immunohistochemical assay achieves high accuracy in comparison with fluorescence in situ hybridization assay as the standard. Am J Clin Pathol 116:495–503
8. Italian Network for Quality Assurance of Tumor Biomarkers (INQAT) Group; SIAPEC-IAP (2005) Interobserver reproducibility of immunohistochemical HER-2/neu assessment in human breast cancer: an update from INQAT round III. Int J Biol Markers 20:189–194
9. Paik S, Bryant J, Tan-Chiu E et al (2002) Real-world performance of HER2 testing—National Surgical Adjuvant Breast and Bowel Project experience. J Natl Cancer Inst 94:852–854
10. Rhodes A, Jasani B, Anderson E et al (2002) Evaluation of HER-2/neu immunohistochemical assay sensitivity and scoring on formalin-fixed and paraffin-processed cell lines and breast tumors: a comparative study involving results from laboratories in 21 countries. Am J Clin Pathol 118:408–417
11. Fleiss JG (1981) Statistical methods for rates and proportions. Wiley, New York, NY, pp 225–232
12. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174
13. Bilous M, Dowsett M, Hanna W et al (2003) Current perspectives on HER2 testing: a review of national testing guidelines. Mod Pathol 16:173–182
14. Ellis IO, Bartlett J, Dowsett M et al (2004) Best Practice No 176: Updated recommendations for HER2 testing in the UK. J Clin Pathol 57:233–237
15. Bertheau P, Cazals-Hatem D, Meignin V et al (1998) Variability of immunohistochemical reactivity on stored paraffin slides. J Clin Pathol 51:370–374
16. Rhodes A, Borthwick D, Sykes R et al (2004) The use of cell line standards to reduce HER-2/neu assay variation in multiple European cancer centers and the potential of automated image analysis to provide for more accurate cut points for predicting clinical response to trastuzumab. Am J Clin Pathol 122:51–60
17. Jacobs TW, Gown AM, Yaziji H et al (1999) Comparison of fluorescence in situ hybridization and immunohistochemistry for the evaluation of HER-2/neu in breast cancer. J Clin Oncol 17:1974–1982
18. Lebeau A, Deimling D, Kaltz C et al (2001) Her-2/neu analysis in archival tissue samples of human breast cancer: comparison of immunohistochemistry and fluorescence in situ hybridization. J Clin Oncol 19:354–63
19. Tubbs RR, Pettay JD, Roche PC et al (2001) Discrepancies in clinical laboratory testing of eligibility for trastuzumab therapy: apparent immunohistochemical false-positives do not get the message. J Clin Oncol 19:2714–2721
20. Hoang MP, Sahin AA, Ordòñez NG et al (2000) HER-2/neu gene amplification compared with HER-2/neu protein overexpression and interobserver reproducibility in invasive breast carcinoma. Am J Clin Pathol 113:852–859
21. Thomson TA, Hayes MM, Spinelli JJ et al (2001) HER-2/neu in breast cancer: interobserver variability and performance of immunohistochemistry with 4 antibodies compared with fluorescent in situ hybridization. Mod Pathol 14:1079–1086
22. Dietel M, Ellis IO, Höfler H et al (2007) Comparison of automated silver enhanced in situ hybridisation (SISH) and fluorescence ISH (FISH) for the validation of HER2 gene status in breast carcinoma according to the guidelines of the American Society of Clinical Oncology and the College of American Pathologists. Virchows Arch 51:19–25