ORIGINAL ARTICLE

Quality assessment of HER2 testing by monitoring of positivity rates

**Harald Choritz · Guntram Büsche · Hans Kreipe ·
On behalf of the Study Group HER2 Monitor**

Abstract Interlaboratory variation in human epidermal growth factor receptor 2 (HER2) testing provides a challenge for targeted therapy in breast and gastric cancer. Assessment of positivity rates among laboratories could help monitor their performance and define reference values for positivity rates to be expected in a geographic region. Pathologists regularly determined the number of HER2-positive cases (HER2 3+, HER2 2+/amplified or amplified) in their laboratory, and figures were continuously entered into a central website. The overall positivity rate of each participant was calculated and compared with the average rates of all other institutes ($n=42$). A total of 18,081 test results on breast cancer and 982 on gastric cancer were entered into the system. Positivity rates for HER2 in breast cancer ranged from 7.6% to 31.6%. Statistically, the results from six institutions qualified as outliers ($p<0.000005$). From the remaining institutions encompassing 10,916 assessments, the mean proportion of positive cases was 16.7±3.2% (99% confidence interval 16.6–16.8). The results from six institutions were in between the 95% and 99.5% confidence intervals. For gastric cancer, there was one outlier and the mean positivity rate was 23.2±5.7%. The proportion of HER2-positive breast cancer cases is considerably lower than could have been expected from published studies. By assessing the positivity rates and comparing them with that of all breast or gastric cancers in a given population, pathologists will be alerted to a potential systematic error in their laboratory assay, causative for over- or underestimation of cancer cases suited for anti-HER2 therapy.

Introduction

Targeted therapy against human epidermal growth factor receptor 2 (HER2)-overexpressing tumours provides a major breakthrough in cancer therapy [1]. The identification of cancer patients who are suited for anti-HER2 therapy depends on the analysis of cancer tissue by immunohistochemistry (IH) or in situ hybridisation (ISH), which usually is performed by pathology departments. Central retesting within the framework of therapy trials has revealed considerable interlaboratory variation [2–4]. Testing inaccuracy was identified as a major issue with either assay, IH and ISH [5]. Proficiency testing by round robin tests was launched in several countries as a potential remedy [6–10]. Although useful and indispensable, proficiency testing surveys render only an incomplete and ephemeral assessment of testing performance and do not necessarily reflect the lasting reliability. Furthermore, they rely on artificial systems such as tissue microarrays or cell lines [9, 10]. Usually, they do not cover the whole process and omit decisive steps such as tissue fixation and processing [11].

From regular proficiency tests, it became obvious that inaccurate results were not haphazardly distributed but followed a systematic trait [7, 8]. Participating pathologists who were unsuccessful in most instances failed either because of systematic false-positive or false-negative staining [8]. In a central review of 1,459 cases from Germany in an international therapy trial and tested locally as HER2 3+, only 1,167 could be confirmed by central testing (80%) (results not published). The 1,459 cases were derived from 116 different centres, from which a small number (6%) were responsible for 23% of discrepant cases

H. Choritz · G. Büsche · H. Kreipe (✉)
Institute of Pathology, Medizinische Hochschule Hannover,
Carl Neuberg Str. 1,
30625 Hannover, Germany
e-mail: kreipe.hans@mh-hannover.de

with an average discordance rate of 50% (unpublished data). These observations led us to the conclusion that surveillance of positivity rates in HER2 testing may help identify laboratories with insufficient testing assays and a high yield of false-positive or negative results. Consequently, pathologists were offered the opportunity to compare their positivity rates with those of others in Germany, Austria and Switzerland. Because German guidelines require that every case of invasive breast cancer is tested for HER2, there are more than 40,000 HER2 tests of breast cancer in Germany per year [12]. From published results, it is difficult to calculate the proportion of positive cases to be expected when optimal testing circumstances are present. Initial studies suggested overexpression in as many as 30% of cases [13]. Larger series recently revealed lower positivity rates with either ISH or IH ranging from 18% to 22.7% [14, 15]. Therefore, a second aim of the study was to obtain an estimate of the positivity rate which has to be expected among a population of breast and gastric cancers in central Europe.

Material and methods

In 2010, all pathology departments in Germany, Austria and Switzerland were offered the opportunity by the German Society of Pathology and the Association of German Pathologists to enter their positivity rates for HER2 testing on breast and gastric cancer into a central web page. Institutes willing to participate received an access code to guarantee confidentiality. The individual figures could be entered on a weekly or monthly basis. The figures which were entered comprised the number of cases being HER2 0; HER2 1+; HER2 2+; HER2 2+; and amplified in ISH, HER2 3+. For those laboratories which only perform ISH, the numbers for cases not amplified or amplified or falling into the equivocal grey zone [5] were entered.

The average positivity rates of all other institutes corrected for the number of cases entered into the survey were compared to the individual result. A system of traffic lights was indicated to institutes whether they lay outside the 95% confidence interval (yellow) or the 99.5% confidence interval (red).

Statistics

Differences between institutions were analysed by using the χ^2 test. Before determining the rate of HER2-positive cases, the data were checked for outliers. The data from every institution were compared to the pooled data from the other institutions, and the data from those institutions differing highly significantly from the other institutions ($p<0.0005$) were excluded in a stepwise manner. The exclusion procedure was stopped when none of the remaining institutions differed highly significantly ($p<0.0005$) from the pooled data of the other laboratories that had not been excluded. Institutions excluded by this procedure were regarded as outliers and therefore not taken into consideration when determining the rate of HER2-positive cases.

On the basis of the rate of HER2-positive cases determined by the results from the laboratories not excluded as outliers, 95% and 99.5% confidence intervals were calculated for the number of HER2-positive cases applying the binomial distribution for $n\leq500$ and approximating the binomial distribution by the normal distribution for $n>500$.

Results

Within 1 year, 42 institutes of pathology (9 in academic institutions, 17 in community hospitals and 16 in private praxis) entered the results of their HER2 testing in breast cancer into the system. Test results on 18,081 breast cancers were communicated. The average number of cases per institute was 430.5 ranging from 4 to 2,733 cases. Seven institutes entered results of fewer than 50 cases of breast cancer. With regard to gastric cancer, 3 institutions communicated more than 50 assessments during the period under study. Positivity rates for HER2 in breast cancer ranged from 7.6% to 31.6%. The average positivity rate of all 42 institutes corrected for the number of cases was 14.61±4.55%. In order to exclude regional differences, the data were screened for potential association with postal codes, which turned out not to be the case (data not shown). Statistically, the results from six institutions were considered to be outliers ($p<0.000005$). Therefore, the results of these institutes were not included when the expected rate of HER2-positive cases per institute and the number of assays were determined. Of the remaining 10,916 assessments, the mean proportion of positive cases was 16.7% (99% confidence interval 16.6–16.8). Six institutions were outside of the 99.5% confidence interval (Fig. 1). The number of HER2 assessments performed by the institutes outside the 99.5% confidence interval ranged from 189 to 3,287 cases. There were two institutes outside the 99.5% confidence interval which had entered more than 2,500 cases. Two institutes assessed HER2 exclusively by in situ hybridisation and did not rely on immunohistochemistry. One of these institutions had performed 491 assessments and proved to be outside the 99.5% confidence interval with 8.35% unequivocally amplified cases.

Of the remaining 36 participating institutes, 6 institutions were in between the 95% and 99.5% confidence interval ($p<0.0005$) (Fig. 1). All of these had commu-
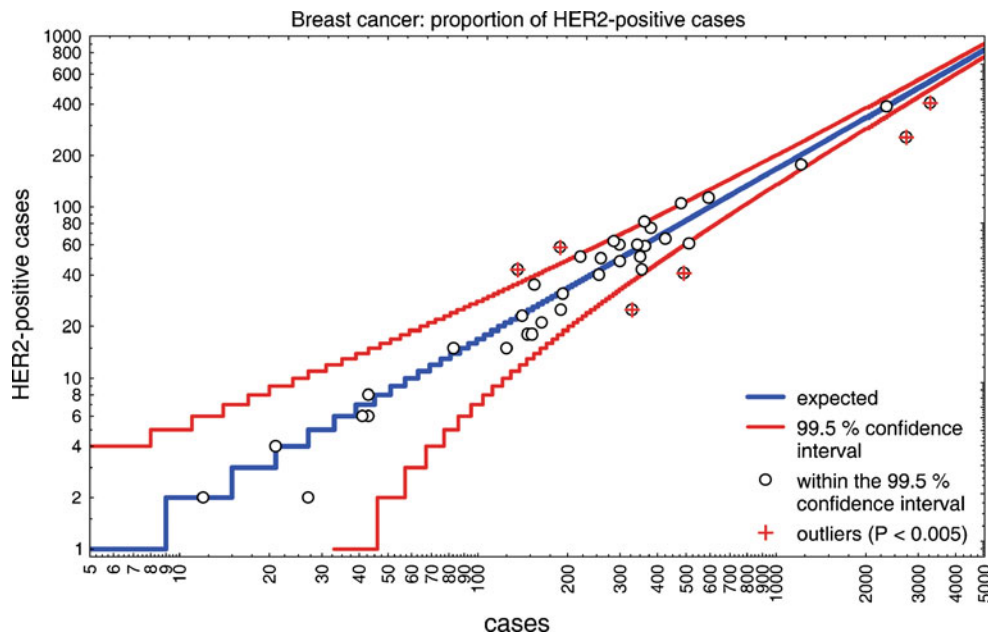
Fig. 1 The number of HER2-positive breast cancer cases (HER2 3+, HER2 2+/amplified, amplified) per institute of pathology in relation to the number of cases investigated was plotted on a logarithmic scale. The 99.5% confidence interval is indicated by *red lines*. In institutions with a low number of assessments, the confidence interval is broader. The expected rate calculated from the mean value of 36 institutions within the 99.5% confidence interval is demonstrated by a *blue line*. There are six institutions outside the 99.5% confidence interval (indicated by *red crosses*). Four of these potentially underestimate HER2 and two have a higher positivity rate than could have been expected. Institutes with a positivity rate within the 99.5% confidence interval are represented by *white circles*

nicated between 153 and 567 HER2 assays in breast cancer cases. No correlation to the type of institute (academic, community hospital or private praxis) could be observed.

The proportion of cases tested immunohistochemically as HER2 2+ ranged from 0% to 60.1% of all assessments (mean $16.5 \pm 15.5\%$). With regard to the 36 reference institutes within the 99.5% confidence interval, the mean percentage of HER2 2+ cases was $18.7 \pm 14.0\%$ (Table 1). Of the HER2 cases which were further analysed by in situ hybridisation, $17.9 \pm 17.0\%$ were amplified (range 0.0–75.0%) (Table 1). Two of the six institutes outside the 99.5% confidence interval rendered a HER2 2+ assessment on 2.8% and 7.6% of cases, respectively. Institutes, which had lower numbers of HER2-positive cases, also revealed a low percentage of 2+ assessments. There was a highly significant correlation between low HER2 positivity rates and low proportion of cases within the 2+ category ($p < 0.000005$).

With regard to gastric cancer, 15 institutes of pathology took part and entered 982 results of their assays. The average positivity rate was $24.11 \pm 7.35\%$. After correction for one outlier, the mean positivity rate was $23.2 \pm 5.7\%$ (Table 1). Because the number of cases per institute was rather small, there was a broad range of positivity rates which fell into the 99.5% confidence interval (Fig. 2). Of the 15 participating institutes, only one institute was outside

the 99.5% confidence interval and none further outside the 95% interval (Fig. 2). The percentage of cases tested as HER2 2+ was $28.7 \pm 12.7\%$ (range 0.0–71.4%). Of these, $30.5 \pm 12.1\%$ were amplified by in situ hybridisation (range 0–52.2%) (Table 1).

Discussion

HER2 testing provides the prototype of a new field in pathology, which has been termed predictive pathology. The results of clinical trials demonstrated a significant benefit of HER2-targeted therapy for early and late stages of breast cancer [1, 16] as well as recently also for gastric cancer [17]. Interlaboratory variation in HER2 testing became obvious from trials with central re-testing [3, 4]. Although regular participation in proficiency testing significantly improved the performance of individual institutes [8], there are doubts that the current quality assurance methods are sufficient to reduce testing variation. In order to improve the reliability of testing, several efforts have been undertaken. Guideline recommendations have been published which set standards for thresholds between positive and negative HER2 test results and define algorithms [5, 18]. Furthermore, regular and predominantly tissue microarray-based proficiency tests are organised in Europe and USA [6, 9, 10].

Table 1 HER2 positivity rates in breast and gastric cancer

	Breast cancer	Gastric cancer
Number of HER2 assay results	18,081	982
Participating institutes of pathology	42	15
Average HER2 assay number per institute±standard deviation	430±705	65±118
Mean HER2 positivity rate[a]±standard deviation	16.7±3.2%	23.2±5.7%
Range[a]	11.8–23.1%	17.6–28.1%
Outliers	6 (14.3%)	1 (6.7%)
HER2 2+ (mean rate[a]±standard deviation)	18.7±14.0%	28.7±12.7%
Range[a]	0.0–54.0%	13.9–54.6%
HER2 gene amplification rate among HER2 2+ cases	17.9±17.0%	30.5±12.1%
Range[a]	0.0–75.0%	8.3–52.2%

[a] Institutes with outlying positivity rates were not included in the calculation of the mean value and range; institutes with ≤20 gastric cancer cases and <35 breast cancer cases were merged and considered as a single group

Proficiency tests take place once or twice a year and do not reflect the permanent accuracy of HER2 assessment in routine practice. An auxiliary instrument to compensate for this limitation and to assure quality of HER2 testing is presented here. By monitoring positivity rates in HER2 testing, institutes of pathology were identified, which lay outside the 99.5% confidence interval of expected results. The exact frequency of HER2-overexpressing or amplified cancers was not known and had to be determined in order to define a reference value. The positivity rates reported in the literature range from 18% to 30% [13–15]. On the basis of 10,916 assessments in 36 institutes of pathology, a mean positivity rate of 16.7% was determined (Table 1). Because HER2 testing is performed on every breast cancer in Germany, there is no selection bias in this study as might be the case in therapy trials. Six institutes were outside of the 99.5% confidence interval (Fig. 1). These outliers were informed that a systematic error in the methodology of HER2 assessment in their laboratory might cause over- or underestimation of HER2 in cancer. Of the six institutes
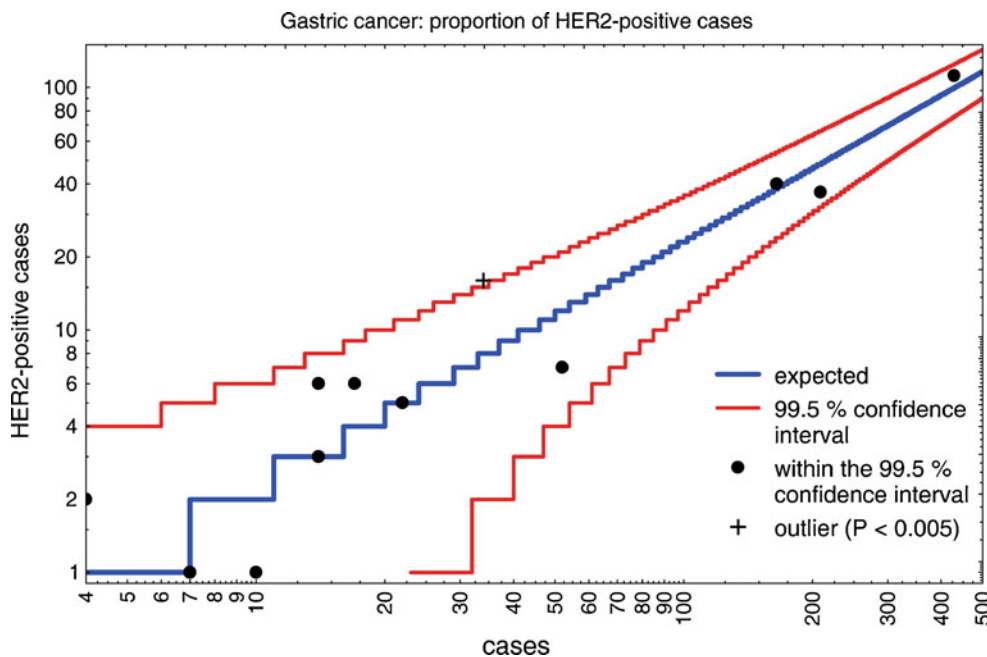
Fig. 2 The number of HER2-positive gastric cancer cases (HER2 3+, HER2 2+/amplified) per institute of pathology in relation to the number of cases investigated was plotted on a logarithmic scale. The 99.5% confidence interval is indicated by *red lines*. In institutions with a low number of assessments, the confidence interval is broader. The expected rate calculated from the mean value of 14 institutions within the 99.5% confidence interval is demonstrated by a *blue line*. There is one institution outside the 99.5% confidence interval (indicated by a *black cross*). Institutes with a positivity rate within the 99.5% confidence interval are represented by *black points*. The confidence interval might narrow over time when more assessments are available for consideration

which were outside the 99.5% confidence interval, five had participated in round robin tests on HER2 assessment offered in Germany. Three of the institutes with low positivity rates had received the information that the sensitivity of their detection method might be too low in at least one of the annual quality assurance trials. Interestingly, a high frequency of assessments did not protect from potential systematic errors. Two institutes which revealed positivity rates outside the 99.5% confidence interval had entered more than 2,500 cases (Fig. 1). It remains to be determined by further studies whether the traffic light system is efficient in improving HER2 assessments in underperforming institutes.

Diversity of positivity rates was highest when the HER2 2+ category was considered (Table 1). This finding indicates that the HER2 2+ category might be limited by subjectivity and poor reproducibility [19] (Fig. 3). In a recent meta-analysis on 17 studies encompassing 8,410 patients, the mean proportion of the HER2 2+ category was 23.2% with a broad range from 2.0% to 87.5% [20]. Only a slight enrichment for amplified cases was found (26.5% vs. 21.1%) [20]. When compared with ISH results in this study, there was no significant enrichment of amplified cases in the HER2 2+ group (Table 1).

Institutes which rely completely on ISH instead of IH to assess HER2 positivity were too few to allow for comparison (Fig. 3). Whether ISH or IH is more reliable and reproducible is a matter of debate [19]. In this study, one of the two institutes which exclusively performed ISH was outside the 99.5% confidence interval.

In order to keep the entering of data into the HER2 monitor as simple as possible and not to reduce the compliance of participants, no detailed information on methods or composition of cases was requested from the participants. It cannot be excluded that an abnormal proportion of low-grade cancers or other specific conditions may be responsible for an aberrant positivity rate. Therefore, a positivity rate outside the 99.5% confidence interval does not necessarily imply that the HER2 assessment method in use is inadequate. Such a finding should, however, urge pathologists to consider this possibility. The primary aim of monitoring HER2 positivity rates is to alert institutes of pathology to potential systematic errors which require further measures to assure quality of testing. As a consequence of abnormal positivity rates, tests in use could be validated or participation in proficiency tests could take place with higher frequency. Only if methodological problems have been excluded should secondary influences such as abnormal composition of the set of samples in which HER2 has been assessed be taken into consideration. Two institutes with a high number of tests and a low positivity rate outside the 99.5% confidence interval also documented extremely low HER2 2+ rates. Unlike the 3+ category, the 2+ category is not related to histological grade. Therefore, it appears highly unlikely that in these two institutes, which together performed more than 6,000 HER2 assessments in breast cancer, a selection bias towards grade 1 and 2 cases might be responsible for the low total positivity rate.

Most therapy trials on targeted HER2 therapy require central retesting of samples which were locally assessed as HER2 positive. As a consequence, central retesting in trials alerts pathologists to false-positive but not to false-negative assessments. This inherent tendency might explain why there are twice as many institutes which potentially underestimate HER2 positivity than institutes with potential overestimation ($n=4$; 99.5% confidence interval). Thus, without eliminating outliers, the mean rate of HER2-

Fig. 3 A high degree of variability between institutes was observed with regard to the HER2 2+ category. In particular, the differentiation between HER2 2+ and HER2 1+ might be handled differently. This case illustrates the borderline between the 2+ category (intraductal carcinoma, *double arrow*) and the 1+ category (invasive carcinoma, *single arrow*) (immunohistochemistry with 4B5 anti-HER2 monoclonal antibody, ×200)
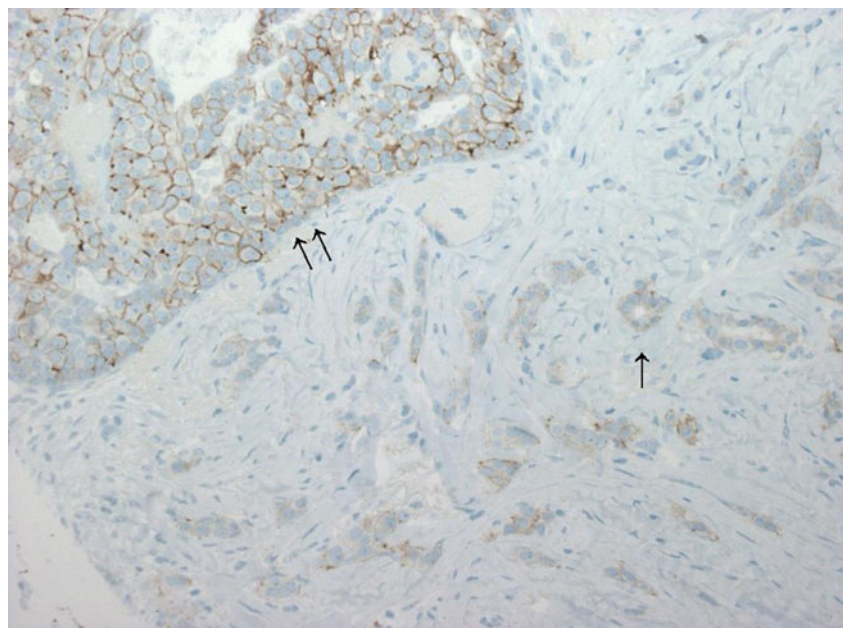
positive cases was lower (14.61±4.55% in 18,221 breast cancers). An almost identical rate was found by questionnaires on 4,940 breast cancer samples in Sweden [21] and slightly higher in Australia (17.1%, 6,512 cases) [22]. In contrast to the HER2 monitor presented here, in both studies, outliers had not been eliminated from the calculation of the expected positivity rate.

There are several measures which institutes of pathology can take to assure quality of HER2 testing. Besides on-slide controls [23], participation in proficiency tests and adherence to guidelines [5–10, 18, 24], a further instrument is proposed here. Monitoring of positivity rates and comparison with an expected value will help identify potential errors in HER2 assessment, which lead to systematic over- or underestimation of HER2 in cancer.

Conflicts of interest None.

References

1. Piccart-Gebhart MJ, Procter M, Leyland-Jones B et al (2005) Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. N Engl J Med 353:1659–1672

2. Paik S, Bryant J, Tan-Chiu E, Romond E, Hiller W et al (2002) Real-world performance of HER2 testing—National Surgical Adjuvant Breast and Bowel Project experience. J Natl Cancer Inst 94:852–854

3. Perez EA, Suman VJ, Davidson NE et al (2006) HER2 testing by local, central, and reference laboratories in specimens from the North Central Cancer Treatment Group N9831 intergroup adjuvant trial. J Clin Oncol 24:3032–3038

4. Roche PC, Suman VJ, Jenkins RB et al (2002) Concordance between local and central laboratory HER2 testing in the breast intergroup trial N9831. J Natl Cancer Inst 94:855–857

5. 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. J Clin Oncol 25:118–145

6. 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

7. Rüdiger T, Höfler H, Kreipe HH et al (2002) Quality assurance in immunohistochemistry: results of an interlaboratory trial involving 172 pathologists. Am J Surg Pathol 26:873–882

8. von Wasielewski R, Hasselmann S, Rüschoff J et al (2008) Proficiency testing of immunohistochemical biomarker assays in breast cancer. Virchows Arch 453:537–543

9. von Wasielewski R, Mengel M, Wiese B et al (2002) Tissue array technology for testing interlaboratory and interobserver reproducibility of immunohistochemical estrogen receptor analysis in a large multicenter trial. Am J Clin Pathol 118:675–682

10. Fitzgibbons PL, Murphy DA, Dorfman DM et al (2006) Interlaboratory comparison of immunohistochemical testing for HER2: results of the 2004 and 2005 College of American Pathologists HER2 Immunohistochemistry Tissue Microarray Survey. Arch Pathol Lab Med 130:1440–1445

11. Tong LC, Nelson N, Tsourigiannis J et al (2011) The effect of prolonged fixation on the immunohistochemical evaluation of estrogen receptor, progesterone receptor, and HER2 expression in invasive breast cancer: a prospective study. Am J Surg Pathol 35:545–552

12. Wöckel A, Kreienberg R (2008) First revision of the German S3 guideline 'diagnosis, therapy, and follow-up of breast cancer'. Breast Care (Basel) 3:82–86

13. Slamon DJ, Clark GM, Wong SG et al (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science 235:177–182

14. Yaziji H, Goldstein LC, Barry TS et al (2004) HER-2 testing in breast cancer using parallel tissue-based methods. JAMA 291:1972–1977

15. Owens MA, Horten BC, Da Silva MM (2004) HER2 amplification ratios by fluorescence in situ hybridization and correlation with immunohistochemistry in a cohort of 6556 breast cancer tissues. Clin Breast Cancer 5:63–69

16. Dinh P, de Azambuja E, Piccart-Gebhart MJ (2007) Trastuzumab for early breast cancer: current status and future directions. Clin Adv Hematol Oncol 5:707–717

17. Bang YJ, Van Cutsem E, Feyereislova A et al (2010) Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. Lancet 376:687–697

18. Rüschoff J, Dietel M, Baretton G et al (2010) HER2 diagnostics in gastric cancer-guideline validation and development of standardized immunohistochemical testing. Virchows Arch 457:299–307

19. Sauter G, Lee J, Bartlett JM (2009) Guidelines for human epidermal growth factor receptor 2 testing: biologic and methodologic considerations. J Clin Oncol 27:1323–1333

20. Dendukuri N, Khetani K, McIsaac M et al (2007) Testing for HER2-positive breast cancer: a systematic review and cost-effectiveness analysis. CMAJ 176:1429–1434

21. Rydén L, Haglund M, Bendahl PO et al (2009) Reproducibility of human epidermal growth factor receptor 2 analysis in primary breast cancer: a national survey performed at pathology departments in Sweden. Acta Oncol 48:860–866

22. Francis GD, Dimech M, Giles L et al (2007) Frequency and reliability of oestrogen receptor, progesterone receptor and HER2 in breast carcinoma determined by immunohistochemistry in Australasia: results of the RCPA Quality Assurance Program. J Clin Pathol 60:1277–1283

23. Mengel M, Hebel K, Kreipe H et al (2005) Standardized on-slide control for quality assurance in the immunohistochemical assessment of therapeutic target molecules in breast cancer. Breast J 11:34–40

24. von Wasielewski R, Krusche CA, Rüschoff J et al (2008) Implementation of external quality assurance trials for immunohistochemically determined breast cancer biomarkers in Germany. Breast Care (Basel) 3:128–133