Assessing HER2 testing quality in breast cancer: variables that influence HER2 positivity rate from a large, multicenter, observational study in Germany

Josef Rüschoff¹, Annette Lebeau², Hans Kreipe³, Peter Sinn⁴, Claus Dieter Gerharz⁵, Winfried Koch⁶, Stefanie Morris⁷, Johannes Ammann⁷, Michael Untch⁸, and on behalf of the Nicht-interventionelle Untersuchung (NIU) HER2 Study Group*

¹Institut für Pathologie Nordhessen and Targos Molecular Pathology GmbH, Kassel, Germany; ²Institut für Pathologie, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany; ³Gemeinschaftspraxis für Pathologie, Lübeck, Germany; ⁴Institut für Pathologie, Medizinische Hochschule Hannover, Hannover, Germany; ⁵Sektion Gynäkopathologie, Pathologische Institut, Heidelberg, Germany; ⁶Institute of Pathology, Evangelical Bethesda Clinical Center, Duisburg, Germany; ⁷Biostatistical Data Services Koch, Schwetzingen, Germany; ⁸Roche Pharma AG, Grenzach-Wyhlen, Germany and ⁸Clinic for Gynecology, Gynecologic Oncology, and Obstetrics, Breast Cancer Center, Berlin, Germany

Despite > 10 years of routine human epidermal growth factor receptor 2 (HER2) testing in breast cancer, testing quality is still an issue. Guidelines recommend assessing HER2 positivity rates as a quality indicator; however, the extent to which patient- or tumor-related factors influence HER2 positivity is still unknown. The present study analyzed these influences to identify pathology centers with HER2 positivity rates unexplained by patient- or tumor-related factors. This observational, prospective study monitored routine HER2 testing at 57 institutes of pathology in Germany (January 2013–August 2014). Data collected included HER2 test result, patient- and tumor-related factors, sample source, and method of sample retrieval. Factors influencing HER2 positivity rates were identified by multiple logistic regression. Individual center effects were assessed in an extended multiple logistic regression model by their statistical significance after adjusting for the combined effect of patient- or tumor-related covariates and multiple testing. Analyses included 15 332 invasive breast cancer samples. Histologic grade showed the strongest influence on HER2 positivity, followed by hormone receptor status, histologic subtype, age, and nodal status (all $P < 0.0001$). The overall HER2 positivity rate across centers was 14.4% (range 7.1–27.3%). A statistically significant center effect on the HER2 positivity rate was identified for three centers ($P < 0.05$), with a trend toward a center effect for a further three ($P < 0.2$). This study, the first of its kind, highlights that assessing HER2 testing quality with HER2 positivity rates should include standardized assessment of patient- or tumor-related characteristics to identify centers with HER2 testing quality issues more effectively. As treatment options for HER2-positive breast cancer continue to evolve, identifying the right patients is key. *Modern Pathology* (2017) 30, 217–226; doi:10.1038/modpathol.2016.164; published online 21 October 2016

Overexpression of the human epidermal growth factor receptor 2 (HER2) protein or amplification of

the *HER2* gene occurs in ~15–20% of all breast cancers.¹ HER2-positive breast cancer has a worse prognosis than HER2-negative disease, having an increased risk of recurrence and a more aggressive disease course.^{2–4} HER2 is a prognostic biomarker predictive for the response to HER2-targeted therapies including the recombinant monoclonal antibodies trastuzumab and pertuzumab, and the antibody–drug conjugate ado-trastuzumab emtansine, all of which specifically target HER2 and are effective treatments for HER2-positive breast

Correspondence: Professor J Rüschoff, MD, Molecular Pathology GmbH, Institut für Pathologie Nordhessen, Germaniastr. 7, 34119 Kassel, Germany.
E-mail: rueschoff@patho-nordhessen.de

*The investigators in the NIU HER2 Study Group are listed in the Supplementary Appendix.
Received 19 April 2016; revised 8 August 2016; accepted 8 August 2016; published online 21 October 2016

cancer.^{5–13} Pertuzumab, in combination with trastuzumab plus docetaxel, for example, demonstrated statistically significant and clinically relevant improvements in progression-free and overall survival as first-line treatment of patients with HER2-positive metastatic breast cancer (compared with placebo, trastuzumab, and docetaxel).^{9,13}

A prerequisite for receiving HER2-targeted therapy is the identification of HER2-overexpression either in the primary tumor or in metastatic lesions. Reliable, high-quality HER2 testing for clinical use is of paramount importance for the correct identification of patients who would benefit from HER2-targeted therapies.^{1,14–18} False-negative HER2 assessments could result in denial of an effective treatment, while false-positive assessments may lead to inappropriate administration of a potentially harmful, costly, and ineffective HER2-targeted therapy, or exclusion from treatment with other targeted therapies in the HER2-negative setting.^{1,8,11,19,20} Despite >10 years of routine HER2 testing, studies have demonstrated variability in HER2 positivity assessments between local pathology laboratories and central testing centers, posing a challenge for clinicians.^{15,21–25} In some of these studies it was assumed that variations in HER2 positivity at individual centers equated to problems with testing quality.^{15,22,26,27} To overcome this issue, monitoring of HER2 positivity rates has been suggested as a means for HER2 testing quality control in international guidelines.^{1,28} Test accuracy, however, is likely to be only one of several factors which can influence overall positivity rates. The extent to which other potential factors, such as patient- and tumor-related characteristics, influence HER2 positivity rates has not yet been investigated systematically because relevant data, eg, age, tumor size, tumor grade, lymph node status, hormone receptor status, sample origin, etc, were not collected in the previous studies.

In the present study, routine HER2 test results from patients with histologically confirmed breast cancer, along with patient- and tumor-related characteristics, were collected from 57 institutes of pathology in Germany. The extent to which these characteristics influenced HER2 positivity rates was assessed at individual institutes. Institutes where HER2 positivity rates could not be explained by patient- or tumor-related characteristics were also identified.

Materials and methods

Patient Eligibility Criteria and Parameters

Eligible tumor samples were collected from randomly selected patients with histologically confirmed breast cancer (any stage). HER2 testing was performed routinely, within daily practice, and documentation of HER2 test result, patient- and

tumor-related characteristics, sample origin, and sample retrieval method was required for each sample (Supplementary Table S1). Retrospective documentation or testing results obtained before the study started were not included.

Study Design

This noninterventional, prospective monitoring study of routine HER2 testing was performed by pathologists at 57 institutes of pathology ('centers') in Germany. Each center provided up to 750 breast cancer cases and documented samples via an electronic case report form. Information on scoring algorithm, fixative, antibodies, and test platforms was recorded, although it was not mandatory to do so.

This study was conducted in accordance with the Declaration of Helsinki. Test results and patient information were anonymous. The study was approved by the local ethical committee of the principal investigator and was available for submission to the local ethical committees of the participating centers.

Evaluations

Evaluations included HER2 positivity rate (defined as immunohistochemistry 3+, or immunohistochemistry 2+/*in situ* hybridization-positive, or *in situ* hybridization-positive) pooled between all centers, HER2 positivity rate at individual centers, and the variability in HER2 positivity rate between centers. Patient- or tumor-related characteristics and their association with HER2 positivity were assessed.

Statistical Methods

The main objectives of the statistical analyses were: (1) to identify patient- or tumor-related characteristics that influence the HER2 positivity rate; (2) to develop a statistical model to predict HER2 positivity for individual patients and centers; (3) to compare the documented HER2 positivity rates of centers with those predicted from the statistical model; and (4) to identify centers for which the documented rate deviated significantly from the predicted rate.

The planned sample size was 15 992 from 160 centers, representative of a population of 180 000 patients with breast cancer in Germany and assuming a HER2 positivity rate of 16.7%²² (maximal variation $\pm 10\%$). The tolerable error rate was considered to be 5%, and 95% confidence intervals were calculated.

The statistical influence of patient- or tumor-related characteristics on HER2 positivity was assessed using multiple logistic regression, which fits the probabilities of the two response levels of the dependent variable HER2 status (positive/negative)

using a logistic function of the patient- or tumor-related characteristics. Missing values for each independent variable were categorized as 'unknown', and the level of 'unknown' was modeled as a separate level for each independent variable (covariate); thus, no samples were excluded due to missing covariate data. Sensitivity analyses are described in Supplementary Material S1.

Likelihood ratio tests were used to assess the effects of covariates on HER2 positivity. All *P*-values were adjusted for the influence of all other independent variables. The importance of the influential variables was quantified by an effect measure. A receiver operating characteristic curve was used to investigate sensitivity and specificity; the area under the curve was used as a measure for the predictive strength of the model.

A prediction profiler was applied to visualize the relationship between HER2 positivity rate and the combined influence of covariates. The dependence of HER2 positivity rate from the levels of one covariate was standardized at a certain level of all other covariates and tested in the developed model.

Center effects, in addition to covariates, were assessed with an extended multiple logistic regression analysis that included centers as an additional covariate (Supplementary Material S2). Twenty-six samples from centers with < 10 test results were pooled into one center, termed 'Center P'.

Statistical analyses were performed using SAS JMP version 11.2.

Results

Patient- and Tumor-related Characteristics and Study Centers

Data were collected from 16 528 breast cancer samples (January 2013–August 2014). The final analysis set included 15 332 invasive breast cancer samples, with exclusions explained in Figure 1. Overall, 14.4% of cases were HER2-positive. Of these, 79.1% were positive based on immunohistochemistry 3+ scores, 14.2% were positive based on immunohistochemistry 2+ and *in situ* hybridization-positive scores, and 5.0% were positive based only on an *in situ* hybridization-positive score (Supplementary Table S2). The overall rates of HER2 positivity compared with the individual variables and combinations thereof across all centers are shown in Figure 2.

The Statistical Influence of Patient-Related Variables on HER2 Positivity

In the multiple logistic regression model, histologic grade, hormone receptor status, histologic subtype, age, and nodal status all had a statistically significant influence only on HER2 positivity rate ($P < 0.0001$; Table 1). Overall, the ranking of covariate importance was the same for all three measures used (*P*-value, main effect, and total effect). No significant

Figure 1 Flow diagram of main analyses. Additional sensitivity analyses were conducted and are reported in the Supplementary material.

influence was detected for the variables of tumor status (size and/or extent of tumor), metastasis status, sample origin (primary tumor, metastasis status [primary, secondary], and local recurrence), or method of sample retrieval (biopsy vs resection); these variables were omitted from the final model.

For most covariates, the proportion of data classified as 'missing' or 'not evaluable' was relatively low (histologic grade, 7.8%; hormone receptor status, 0.2%; histologic subtype, 1.2%; and age, 0%); however, nodal status was documented only for 34.6% of the samples.

Prediction of HER2 Positivity Rate Based on Covariates

The relative importance of an individual covariate in predicting HER2 positivity is illustrated by the magnitude of change in its prediction profiler trace (Figure 3). For an example patient (histologic grading: G2, hormone receptor status: estrogen receptor-negative/progesterone receptor-negative, histologic subtype: ductal, age: 63 years, nodal status: [y]pN1), our model predicted the likelihood for HER2 positivity as 19.4% (95% confidence interval: 16.3–22.8%).

Assessment of Center Effects in the Context of Covariates

For every center in the study, a HER2 positivity rate was predicted based on the patient- and tumor-related characteristics of its documented samples (Figure 4). After adjustment for the influence of covariates, 13 centers were identified as having a significant 'center effect' on HER2 positivity rate ($P < 0.05$; Figure 4 and Supplementary Table S3).

Figure 2 Rates of HER2 positivity with 95% confidence intervals by individual covariates. The overall rates of HER2 positivity are shown against (a) histologic grading, (b) hormone receptor status, (c) histologic subtype, (d) age, (e) nodal status, (f) tumor status, (g) metastasis status, (h) sample origin, and (i) method of sample retrieval (biopsy vs resection). ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; PgR, progesterone receptor.

Following Bonferroni-Holm correction for multiple testing, a statistically significant center effect on HER2 positivity rate was identified for three centers (Centers 4, 28, and 32; $P < 0.05$), with a trend toward a center effect for a further three (Centers 1, 17, and 35; $P < 0.2$; Figure 4 and Supplementary Table S3).

Assessment of the relative importance of center effect in predicting HER2 positivity identified a main effect of 0.159 and a total effect of 0.177, ranking the still-unexplained center effect third behind the covariates of histologic grade and hormone receptor status and higher than histologic subtype, age, and nodal status (Supplementary Table S4).

Correlation of Center Effect with Center Properties

Overall, there was a trend toward an increasing center effect in 'private practices', compared with 'other hospitals' and particularly when compared with centers described as 'university hospitals' (Supplementary Material S3.1). There was no correlation between a significant center effect and the number of HER2 breast diagnostic tests performed annually by each center, whether centers performed consecutive or selected documentation of samples, whether centers carried out reference HER2 testing, or the degree of HER2 testing automation (Supplementary Material S3.2–5).

Figure 2 Continued.

Predicted HER2 Positivity Rates Compared with 99% Confidence Intervals of Documented Positivity Rates of Centers

To visualize their relationship, covariate-predicted HER2 positivity rates were sorted in increasing order and plotted together with documented HER2 positivity rates of centers and their 99% confidence intervals (Figure 5). The predicted HER2 positivity rates for centers reflect the distribution of histologic grading, hormone receptor status, age, histologic subtype, and nodal status across centers. The overall predicted HER2 positivity rate across all centers was 14.4% (95% confidence intervals: 13.8–14.9%; predicted range: 11.1–19.1%); for most centers the predicted rate was close to the documented rate and within the 99% confidence intervals limits. Eight centers, however, had predicted HER2 positivity rates outside the 99% confidence intervals limits; among these were all three centers which showed the statistically significant center effect and the three centers with a trend toward a center effect.

Discussion

In addition to trastuzumab,^{29–31} routine treatment for patients with HER2-positive breast cancer includes

HER2-targeted therapies such as pertuzumab and ado-trastuzumab emtansine. These drugs define a new standard of care, but accurate HER2 testing is critical to limit false-positive/-negative results and to select patients who might benefit from these treatments.¹ Proficiency testing, along with recording and monitoring HER2 positivity rates by pathology institutes as a means of quality control, has been beneficial in identifying centers that may have testing quality issues.^{1,22} However, monitoring HER2 testing accuracy by HER2 positivity rate alone does not account for patient- or tumor-related factors that influence overall HER2 positivity at pathology centers.²⁸ This study aimed to identify patient- or tumor-related characteristics that influence HER2 positivity rate and develop a statistical model to predict HER2 positivity and identify centers with deviating HER2 positivity rates.

The data granularity of this large and unique dataset allowed us to perform in-depth analysis of patient cohorts at an institutional level. In our multiple testing model, five patient- or tumor-related covariates were identified as having a statistically significant influence on the HER2 positivity rate; these were histologic grade, hormone receptor status, histologic subtype, age, and nodal status, in that order. The use of these covariates in a statistical model allowed prediction of HER2

Table 1 Likelihood ratio test results of covariates and their importance in predicting HER2 positivity in the multiple logistic regression model

Variable	Likelihood ratio test of effect of independent variables on HER2 positivity ^a (P-value)	Covariate importance in predicting HER2 positivity	
		Main effect ^b	Total effect ^b
Histologic grading	< 0.0001	0.396	0.444
Hormone receptor status	< 0.0001	0.285	0.310
Histologic subtype (lobular vs ductal)	< 0.0001	0.142	0.157
Age	< 0.0001	0.088	0.095
Nodal status	< 0.0001	0.050	0.053

^aAll P-values are adjusted for the influence of all other independent variables. ^bThe main effect reflects the relative contribution of that factor alone. The total effect reflects the relative contribution of that factor both alone and in combination with other factors.

Figure 3 Prediction profile with 95% confidence intervals for each covariate and predicted HER2 positivity rate for selected levels (dashed red lines). The dependence of the HER2 positivity rate from the levels of one covariate were standardized at a certain level of all other covariates (patient example: histologic grading: G2, hormone receptor status: estrogen receptor-negative/progesterone receptor-negative, histologic subtype: ductal, age: 63.5 years, nodal status: [y]pN1) and tested in the developed model. The predicted mean positivity rate was 19.4% (horizontal red dashed line). ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; HR, hormone receptor; PgR, progesterone receptor.

positivity rates for each center, based on the individual patient cohort, and comparison of documented rates with predicted rates. In contrast, previous studies of HER2 positivity rates assessed individual centers in relation to the overall rate of a standard population.²² For 40 centers (83.3%), the documented HER2 positivity rate was close to the predicted rate or the predicted HER2 positivity rate fell within the 99% confidence intervals. Eight centers (16.7%) were identified where the predicted HER2 positivity rate was outside the 99% confidence intervals limits, with six of these centers having a statistically significant center effect or a trend toward a center effect.

Closer examination of centers identified as having significant center effects identified several which would not have been detected using a more traditional method (Figure 5); eg, Center 4, which had a documented HER2 positivity rate of 18.1 and 99% confidence intervals overlapping the overall positivity rate of 14.4%. In addition, Center 28 underlines the importance of considering covariates; this center

had the second-highest documented HER2 positivity rate (25.7%), but a markedly different predicted HER2 positivity rate. Centers 4 and 28 had above-average HER2 positivity rates, but below-average predicted HER2 positivity rates, mainly due to a relatively high proportion of histologic grade G1 tumors. The opposite was true for Center 32, mainly because of the relatively high percentage of histologic grade G3 tumors, and lower than average documented HER2 positivity rates and 99% confidence intervals.

For Centers 4, 28, and 32, the significant center effect identified by the model-based approach was greater than that obtained with the traditional confidence intervals-based method, as the predicted HER2 positivity rate was further from the documented HER2 positivity rate than it was from the overall HER2 positivity rate. In contrast, for Center 2 our model-based approach indicated no significant center effect, while a critical deviation in HER2 positivity rate was indicated by the traditional confidence intervals-based method.

Figure 4 Documented and predicted HER2 positivity rates. Documented HER2 positivity rates for individual centers are shown, with two-sided 99% confidence intervals included as the lower and upper limits. The documented mean positivity rate across all centers was 14.4%. The predicted HER2 positivity rate for each center is shown as a black diamond. Centers with a statistically significant center effect are highlighted in dark gray and those with a trend toward a center effect in light gray. HER2, human epidermal growth factor receptor 2.

While centers may vary in the extent to which distribution of histologic grades influences their predicted HER2 positivity rate, differences in grading accuracy, as reported in other studies,^{32–34} should be considered. To explore the robustness of our model, we simulated random grading errors (eg, G2 randomly changed to G1 or G3) or systematic increases or decreases in grading (eg, a defined percentage of G1 tumors were increased to G2 or G3) (data not shown). The results of these simulations indicated that our model was robust enough to tolerate random or systematic grading errors in up to every fourth case, before an influence on identified center effects was observed. Overall, we are confident that errors in histologic grade did not play a relevant role in our study due to the robustness and high predictive power of our statistical model, both across and within study centers. Nevertheless, differences in grading practices cannot be excluded and centers with an identified center effect should verify whether cases were graded according to standard assessment rules.

This study indicates that methods for assessing HER2 positivity rates based on the overall calculated 95% or 99% confidence intervals alone²² were insufficient to assess variations in HER2 positivity rate or accurately draw conclusions on HER2 testing quality. This was particularly true when covariate levels of a center deviated from those expected in a standard population (Figure 5). Although the confidence intervals-based method may be easier to implement routinely, the model-based approach described here

has a stronger statistical justification and offers greater reassurance for pathology institutes.

The overall mean HER2 positivity rate in this study was 14.4%. When anti-HER2 therapies were initially introduced, studies of HER2 testing were performed in cohorts with metastatic disease or in high-risk patients, and HER2 positivity rates were 20–30%.^{2,3,14,35} The data presented here are in accordance with the recently published US Surveillance, Epidemiology, and End Results (SEER) registry data, and with data from the UK and Canada that showed HER2 positivity rates of 14.5–15.0% in large patient cohorts.^{28,36,37} Therefore, our study reflects published and realistic expectations of HER2 positivity rates in pathology institutes today.

This study is not without limitations. Data collected were sample-oriented, not patient-oriented. Therefore, there may be multiple samples per patient. As the study achieved target enrollment more quickly than initially anticipated, it is also possible there may have been some documentation bias at centers that grouped samples according to particular criteria. Thus, sensitivity analyses were conducted to minimize any potential bias. Although it might seem that the unexpectedly low overall relative weight of influence attributed to nodal status may have been due to lack of documentation (this being reported only for 34.6% of samples), additional sensitivity analyses demonstrated that assessment of center effects was not significantly influenced by this lack of documentation (Supplementary Material S1). Although sample origin was excluded from the main model as it did

Figure 5 Predicted HER2 positivity rate for individual centers in relation to 99% confidence intervals of their documented HER2 positivity rate. HER2 positivity rates for individual centers are represented by gray circles and their two-sided 99% confidence intervals represented by gray crosses (lower limit) and gray diamonds (upper limit). The mean positivity rate (14.4%) across all centers is shown as a horizontal black line. The centers were ordered in the graph based on increasing predicted HER2 positivity rate (gray crosses along the broken gray line). Centers where the predicted value was outside the 99% confidence intervals are highlighted. Those centers which showed a statistically significant center effect after multiplicity correction according to Bonferroni-Holm are marked by a dark gray rectangle; centers with a Bonferroni-Holm-adjusted P -value < 0.2 (but > 0.05) are marked by a light gray rectangle. HER2, human epidermal growth factor receptor 2.

not significantly influence HER2 positivity, it is possible that metastatic site may have had an effect on sample origin. As the majority of samples analyzed were collected from primary tumors and, due to the limited information available on metastatic site from the relatively low number of distant metastases, we were unable to consider the impact of metastatic site within our model.

Despite these limitations, and in contrast to previously reported studies,^{15,22,25} the data collected here were unique as they were provided by 57 centers from routine diagnostics and not from a highly selective clinical study cohort, a single center, or a small grouping of centers. The statistical model developed was novel as it enables prediction of HER2 positivity based on five covariates. It has the potential to indicate whether measured differences in HER2 positivity rates were specifically due to different patient populations, or tended to relate to testing quality issues. Variability in the rate of HER2 positivity has been a cause of discussion and

concern and this model-based approach provides centers with the means to account for differences. Most importantly, this is the first time that a statistical modeling approach has been used to assess variability in HER2 positivity rates and, although validation is required, it has the potential to highlight center effects. A table providing predicted positivity rates for a comprehensive set of 2000 combinations of covariate factor levels has been generated (Supplementary Table S5).

To ensure high-quality HER2 testing in breast cancer, centers with identified center effects should take further steps to examine the underlying reasons. Testing variation can occur during the preanalytic, analytic, and postanalytic stages as a result of factors such as fixation, antigen retrieval, reagent used, antibodies used, immunohistochemical protocol, and interpretation of results.²⁸ Participation in proficiency testing and collaboration with a reference laboratory might help centers recognize variability in their own procedures.

In conclusion, this study is the first of its kind, reporting on the multifactorial parameters from the patient population and tumor characteristics that may impact routine HER2 positivity rates. The results highlight that routine assessment of HER2 testing quality based on the HER2 positivity rate should include patient- and tumor-related characteristics, assessed in a standardized fashion, to identify institutes with potential HER2 testing quality issues more effectively.

Acknowledgments

We would like to thank the investigators and centers of the Nicht-interventionelle Untersuchung (NIU) HER2 Study Group who provided information on samples, along with clinical and diagnostic data for the analyses (listed in the Supplementary Material). In addition, we thank Dennis Fortun, Ines Reuter, and Michael Selg for operative support. Support for third-party writing assistance for this manuscript, furnished by Islay Steele, PhD, of Health Interactions, was provided by Roche Pharma AG, Grenzach-Wyhlen. This work was supported by Roche Pharma AG, Grenzach-Wyhlen.

Disclosure/conflict of interest

AL has received honoraria from Roche Pharma and Novartis, has performed a consulting or advisory role for Genomic Health Inc., Roche Pharma and Menarini Diagnostics, and has received institute research funding from Sysmex Europe and BioNTech Diagnostics. SM is an employee of Roche. JA is an employee of Roche Pharma AG, Germany and holds shares in Roche Pharmaceuticals. The remaining authors declare no conflict of interest.

References

- 1 Wolff AC, Hammond ME, Hicks DG, *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.
- 2 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;235:177–182.
- 3 Slamon DJ, Godolphin W, Jones LA, *et al*. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* 1989;244:707–712.
- 4 Hayes DF, Thor AD. C-erbB-2 in breast cancer: Development of a clinically useful marker. *Semin Oncol* 2001;29:231–245.
- 5 Cobleigh MA, Vogel CL, Tripathy D, *et al*. Multinational study of the efficacy and safety of humanized anti-HER2 monoclonal antibody in women who have HER2-overexpressing metastatic breast cancer that has progressed after chemotherapy for metastatic disease. *J Clin Oncol* 1999;17:2639–2648.
- 6 Vogel CL, Cobleigh MA, Tripathy D, *et al*. Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *J Clin Oncol* 2002;20:719–726.
- 7 Slamon DJ, Leyland-Jones B, Shak S, *et al*. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 2001;344:783–792.
- 8 Perez EA, Suman VJ, Davidson NE, *et al*. Cardiac safety analysis of doxorubicin and cyclophosphamide followed by paclitaxel with or without trastuzumab in the North Central Cancer Treatment Group N9831 adjuvant breast cancer trial. *J Clin Oncol* 2008;26:1231–1238.
- 9 Baselga J, Cortes J, Kim SB, *et al*. Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. *N Engl J Med* 2012;366:109–119.
- 10 Gianni L, Pienkowski T, Im YH, *et al*. Efficacy and safety of neoadjuvant pertuzumab and trastuzumab in women with locally advanced, inflammatory, or early HER2-positive breast cancer (NeoSphere): a randomised multicentre, open-label, phase 2 trial. *Lancet Oncol* 2012;13:25–32.
- 11 Romond EH, Jeong JH, Rastogi P, *et al*. Seven-year follow-up assessment of cardiac function in NSABP B-31, a randomized trial comparing doxorubicin and cyclophosphamide followed by paclitaxel (ACP) with ACP plus trastuzumab as adjuvant therapy for patients with node-positive, human epidermal growth factor receptor 2-positive breast cancer. *J Clin Oncol* 2012;30:3792–3799.
- 12 Verma S, Miles D, Gianni L, *et al*. Trastuzumab emtansine for HER2-positive advanced breast cancer. *N Engl J Med* 2012;367:1783–1791.
- 13 Swain SM, Baselga J, Kim SB, *et al*. Pertuzumab, trastuzumab, and docetaxel in HER2-positive metastatic breast cancer. *N Engl J Med* 2015;372:724–734.
- 14 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;19:354–363.
- 15 Paik S, Bryant J, Tan-Chiu E, *et al*. Real-world performance of HER2 testing—National Surgical Adjuvant Breast and Bowel Project experience. *J Natl Cancer Inst* 2002;94:852–854.
- 16 Press MF, Sauter G, Bernstein L, *et al*. Diagnostic evaluation of HER-2 as a molecular target: An assessment of accuracy and reproducibility of laboratory testing in large, prospective, randomized clinical trials. *Clin Cancer Res* 2005;11:6598–6607.
- 17 Bartlett JM, Ibrahim M, Jasani B, *et al*. External quality assurance of HER2 fluorescence in situ hybridisation testing: Results of a UK NEQAS pilot scheme. *J Clin Pathol* 2007;60:816–819.
- 18 Lebeau A, Turzynski A, Braun S, *et al*. Reliability of human epidermal growth factor receptor 2 immunohistochemistry in breast core needle biopsies. *J Clin Oncol* 2010;28:3264–3270.
- 19 Russell SD, Blackwell KL, Lawrence J, *et al*. Independent adjudication of symptomatic heart failure with the use of doxorubicin and cyclophosphamide followed by trastuzumab adjuvant therapy: A combined review of cardiac data from the National Surgical

- Adjuvant Breast and Bowel Project B-31 and the North Central Cancer Treatment Group N9831 clinical trials. *J Clin Oncol* 2010;28:3416–3421.
- 20 Bowles EJ, Wellman R, Feigelson HS, *et al*. Risk of heart failure in breast cancer patients after anthracycline and trastuzumab treatment: A retrospective cohort study. *J Natl Cancer Inst* 2012;104:1293–1305.
 - 21 Noske A, Loibl S, Darb-Esfahani S, *et al*. Comparison of different approaches for assessment of HER2 expression on protein and mRNA level: prediction of chemotherapy response in the neoadjuvant GeparTrio trial (NCT00544765). *Breast Cancer Res Treat* 2011;126:109–117.
 - 22 Choritz H, Büsche G, Kreipe H, *et al*. Quality assessment of HER2 testing by monitoring of positivity rates. *Virchows Arch* 2011;459:283–289.
 - 23 Denkert C, Huober J, Loibl S, *et al*. HER2 and ESR1 mRNA expression levels and response to neoadjuvant trastuzumab plus chemotherapy in patients with primary breast cancer. *Breast Cancer Res* 2013;15:R11.
 - 24 Avril S, von Minckwitz G, Gündisch S, *et al*. Assessment of phosphorylated HER2 protein as predictive biomarker to stratify anti-HER2 treatment in HER2 non-amplified patients – A translational study in the GeparQuattro and GeparQuinto trials. Presented at the 37th San Antonio Breast Cancer Symposium, San Antonio, TX, USA, held on 9–13 December 2014.
 - 25 Bianchi S, Caini S, Paglierani M. Accuracy and Reproducibility of HER2 Status in Breast Cancer Using Immunohistochemistry: a Quality Control Study in Tuscany Evaluating the Impact of Updated 2013 ASCO/CAP Recommendations. *Pathol Oncol Res* 2015;21:477–485.
 - 26 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;94:855–857.
 - 27 Perez EA, Suman VJ, Davidson NE, *et al*. HER2 testing by local, central, and reference laboratories in specimens from the North Central Cancer Treatment Group N9831 intergroup adjuvant trial. *J Clin Oncol* 2006;24:3032–3038.
 - 28 Rakha EA, Pinder SE, Bartlett JM, *et al*. Updated UK Recommendations for HER2 assessment in breast cancer. *J Clin Pathol* 2015;68:93–99.
 - 29 Cardoso F, Costa A, Norton L, *et al*. ESO-ESMO 2nd international consensus guidelines for advanced breast cancer (ABC2). *Ann Oncol* 2014;25:1871–1888.
 - 30 National Comprehensive Cancer Network: NCCN Clinical Practice Guidelines in Oncology: Breast Cancer, V1 2015.
 - 31 Giordano SH, Temin S, Kirshner JJ, *et al*. Systemic therapy for patients with advanced human epidermal growth factor receptor 2-positive breast cancer: American Society of Clinical Oncology clinical practice guideline. *J Clin Oncol* 2014;32:2078–2099.
 - 32 Ellis IO, Coleman D, Wells C, *et al*. Impact of a national external quality assessment scheme for breast pathology in the UK. *J Clin Pathol* 2006;59:138–145.
 - 33 Fanshawe TR, Lynch AG, Ellis IO, *et al*. Assessing agreement between multiple raters with missing rating information, applied to breast cancer tumour grading. *PLoS One* 2008;3:e2925.
 - 34 Rakha EA, Reis-Filho JS, Baehner F, *et al*. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res* 2010;12:207.
 - 35 Owens MA, Horten BC, Da Silva MM. HER2 amplification ratios by fluorescence in situ hybridization and correlation with immunohistochemistry in a cohort of 6556 breast cancer tissues. *Clin Breast Cancer* 2004;5:63–69.
 - 36 Bilous M, Morey AL, Armes JE, *et al*. Assessing HER2 amplification in breast cancer: findings from the Australian In Situ Hybridization Program. *Breast Cancer Res Treat* 2012;134:617–624.
 - 37 Howlader N, Altekruse SF, Li CI, *et al*. US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. *J Natl Cancer Inst* 2014;106. pii: dju055.

Supplementary Information accompanies the paper on Modern Pathology website (<http://www.nature.com/modpathol>)