



HER2 testing in gastric cancer diagnosis: insights on variables influencing HER2-positivity from a large, multicenter, observational study in Germany

Gustavo Baretton¹ · Hans H. Kreipe² · Peter Schirmacher³ · Timo Gaiser⁴ · Ralf Hofheinz⁵ · Karl-Heinz Berghäuser^{6,7} · Winfried Koch⁸ · Claudia Künzel⁹ · Stefanie Morris⁹ · Josef Rüschoff¹⁰ · On behalf of the Nicht-interventionelle Untersuchung (NIU) HER2 Study Group

Received: 21 August 2018 / Revised: 18 January 2019 / Accepted: 4 February 2019 / Published online: 2 March 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

HER2 testing in metastatic gastric or gastroesophageal junction cancer (mGC/mGEJC) is standard practice. Variations in HER2-positivity rates suggest factors affecting test quality; however, the influence of patient-, tumor-, and laboratory-related factors on HER2-positivity rates remains unknown. This observational, prospective study collected routine HER2 testing data from 50 pathology centers in Germany (January 2013–December 2015). For each sample, HER2 status, primary tumor location, method of sample retrieval, and other patient- and tumor-related parameters were recorded. A model for predicting the probability of HER2-positivity was developed using stepwise multiple logistic regression to identify influencing factors. Documented positivity rates and corresponding predicted HER2-positivity probabilities were compared to identify institutes with deviations in HER2-positivity. Data from 2761 mGC/mGEJC routine diagnostic specimens included 2033 with HER2 test results (1554 mGC, 479 mGEJC); overall HER2-positivity rates across centers were 19.8% and 30.5%, respectively. HER2-positivity correlated most with Lauren classification, then HER2 testing rate, primary tumor location, sample type, and testing method (all $p < 0.05$). Three institutes had model-predicted HER2-positivity rates outside the 95% confidence interval of their documented rate, which could not be explained by sample and center characteristics. Results demonstrated the high quality of routine HER2 testing in the mGC/mGEJC cohort analyzed. This is the first study investigating parameters impacting on HER2-positivity rates in mGC/mGEJC in routine practice and suggests that assessment of HER2 testing quality should consider primary tumor location, testing method and rate, and tumor characteristics. Accurate identification of patients with HER2-positive mGC/mGEJC is essential for appropriate use of HER2-targeted therapies.

The investigators in the NIU HER2 study group are listed in Online Resource 1.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00428-019-02541-9>) contains supplementary material, which is available to authorized users.

✉ Gustavo Baretton
Gustavo.Baretton@Uniklinikum-Dresden.de

¹ Institut für Pathologie, Universitätsklinikum Carl Gustav Carus an der Technischen Universität, Fetscherstraße 74, 01307 Dresden, Germany

² Institut für Pathologie, Medizinische Hochschule Hannover, Carl-Neuberg-Strasse 1, 30625 Hannover, Germany

³ Pathologische Institut, Universitätsklinik Heidelberg, Im Neuenheimer Feld 224, 69120 Heidelberg, Germany

⁴ Pathologisches Institut Mannheim, Medizinische Fakultät Mannheim der Universität Heidelberg, Theodor-Kutzer-Ufer 1–3, 68167 Mannheim, Germany

⁵ Interdisziplinäres Tumorzentrum, Universitätsmedizin Mannheim, Theodor-Kutzer Ufer 1–3, 68167 Mannheim, Germany

⁶ Thüringen-Kliniken “Georgius Agricola” GmbH, Rainweg 68, 07318 Saalfeld/Saale, Germany

⁷ Present address: Zentrum für ambulante Medizin, Uniklinikum Jena gGmbH, Carl-Zeiss-Platz 8, 07743 Jena, Germany

⁸ BDS Koch, Bibienstraße 5, 68723 Schwetzingen, Germany

⁹ Roche Pharma AG, Emil-Barell-Straße 1, 79639 Grenzach-Wyhlen, Germany

¹⁰ Institut für Pathologie Nordhessen und Targos Molecular Pathology GmbH, Germaniastr 7, 34119 Kassel, Germany

Keywords HER2-positivity · HER2 testing · Gastric cancer · Gastroesophageal junction cancer · Statistical model

Introduction

Approximately 6–32% of all metastatic gastric or gastroesophageal junction adenocarcinomas (mGC/mGEJC) overexpress human epidermal growth factor receptor 2 (HER2) protein or have *HER2* gene amplification [1–4]. HER2 is a prognostic biomarker predictive for response to targeted therapies, including the recombinant monoclonal antibody trastuzumab, in various cancer types [1, 2, 5–10]. The results of the “Trastuzumab for Gastric Cancer (ToGA)” trial showed that addition of trastuzumab to chemotherapy improved survival in patients with advanced GC/GEJC vs chemotherapy alone, especially for patients with high HER2 expression (median overall survival 16.0 vs 11.8 months, respectively) [2]. Based on these findings, trastuzumab plus chemotherapy is considered the standard of care for first-line treatment of HER2-positive mGC/mGEJC [11].

Selection of patients with HER2-overexpression is an essential prerequisite for HER2-targeted therapy. Wide variation in HER2-positivity rates presents a challenge for selecting effective treatment for patients with mGC/mGEJC in the first-line setting [3, 4, 12–18]. Despite clear guidelines and > 10 years of routine HER2 testing in different cancer types [2, 4, 12–14], there is still variability in HER2-positivity assessments between laboratories or testing centers [1, 12, 17, 19].

Variation in HER2-positivity between testing sites could be due to testing quality [12, 19–21] and thus evaluating HER2-positivity rates may be useful in quality assessment [13, 19–21]. Potential variables that might further influence HER2-positivity rates have not been investigated systematically, due to lack of appropriate data.

The main objectives of the study were: (1) to identify patient- or tumor-related characteristics that are associated with the probability of HER2-positivity; (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.

In the present study, a representative investigation of routine HER2 testing of mGC/mGEJC in Germany, we identified patient-, tumor-, and laboratory-related characteristics that influenced HER2-positivity rates. Overall, a high testing performance was noted, with deviations from the expected positivity rates explainable by sample and center characteristics.

Materials and methods

Study design

This was a prospective, observational study of HER2 testing performed in Germany between 2013 and 2015. Data were collected by pathologists at 50 institutes of pathology (centers), and documented data from up to 250 consecutive mGC/mGEJC cases were provided by each center via an electronic case report form. HER2 testing of tumor samples from patients with histologically confirmed mGC/mGEJC with known TNM classification was part of routine diagnostic practice. The following information was provided for each sample: HER2 test result (positive or negative), patient age, tumor-related characteristics (localization of carcinoma, Lauren classification, TNM, sample origin [primary tumor site or metastatic lesion]), method of sample retrieval (biopsy, resection, or unknown), overall number of biopsies, and first method of determining HER2 status (immunohistochemistry [IHC] or in situ hybridization [ISH]). The following laboratory-related characteristics were recorded: type of institution (university, private practice, or other), commercial kit, manufacturer of the antibody, and regular participation in round-robin tests. Information on fixative, test platforms, and scoring algorithms was also recorded, although it was not mandatory to do so. Neither test results obtained before the start of the study nor retrospective documentation were included. All patient information and test results were anonymized.

HER2-positivity was defined as IHC 3+, or IHC 2+ and ISH-positive. HER2-positivity rate was estimated across all centers, evaluated at individual centers, and assessed for variability between centers. Patient-, tumor-, and laboratory-related characteristics and their association with HER2-positivity were assessed, and a statistical model of their influence on HER2-positivity was developed.

This study was conducted in accordance with the guidelines for Good Clinical Practice and the Declaration of Helsinki. 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.

Statistical methods

Data were analyzed descriptively with standard summary statistics, including absolute and relative frequencies of HER2-positivity, 95% Wilson score-based confidence intervals (CIs), and graphical methods. Chi-square tests were used to examine

the associations of individual variables with HER2-positivity. Multiple logistic regression (MLR) was used to determine and model the combined influence of key variables on the probability of HER2-positivity. A variable was considered a candidate for inclusion in the modeling if $p < 0.2$ in its bivariate association with HER2-positivity or if it was a covariate in the corresponding breast cancer (BC) study model [22]. All candidates were included into a stepwise MLR procedure using a mixed forward inclusion and backward elimination strategy. The model with the lowest small-sample-size corrected Akaike information criterion (AICc) value was considered as the primary model from the statistical perspective.

Influential covariates in the model were ranked by size of p values and by their relative contribution to the variability of HER2-positivity predicted by the model. The receiver operating characteristic (ROC) area under the curve (AUC) was used to assess the prognostic strength of the model. Sensitivity and specificity were determined at the optimal cut-point of the experimental ROC curve.

Prediction profiles were used to visualize the influence of each individual covariate on the probability of HER2-positivity, standardized across the levels of all other covariates in the statistical model.

The generated statistical model was used to predict the probability of HER2-positivity for each individual sample and for the centers. Deviations from these model-predicted probabilities of positivity for centers were considered unexplained center effects and were assessed descriptively by comparing the model-predicted probabilities with documented HER2-positivity rates of centers, including their 95% CIs, and with the overall HER2-positivity rate. For the analysis of center effects, centers with fewer than 10 test results were pooled into one center, termed “Center 99.” A sensitivity analysis was conducted excluding 21 samples with minor issues in the documentation of the HER2 status.

Statistical analyses were performed using SAS® JMP® version 12.2.0 or higher (SAS Institute Inc., NC, USA).

Results

Patient- and tumor-related characteristics

Data were collected from 2761 mGC/mGEJC samples (January 2013–December 2015); HER2 status was available for 2077 samples. Since the study focused on pathology samples and not on patients, patient information is limited (e.g., samples were obtained from an unknown number of patients). The primary analysis set included 2033 mGC/mGEJC samples, with exclusions explained in Fig. 1. Overall, 22.3% with CI (20.5%, 24.1%) of cases were HER2-positive, including 19.8% of mGC and 30.5% of mGEJC cases. HER2-positivity

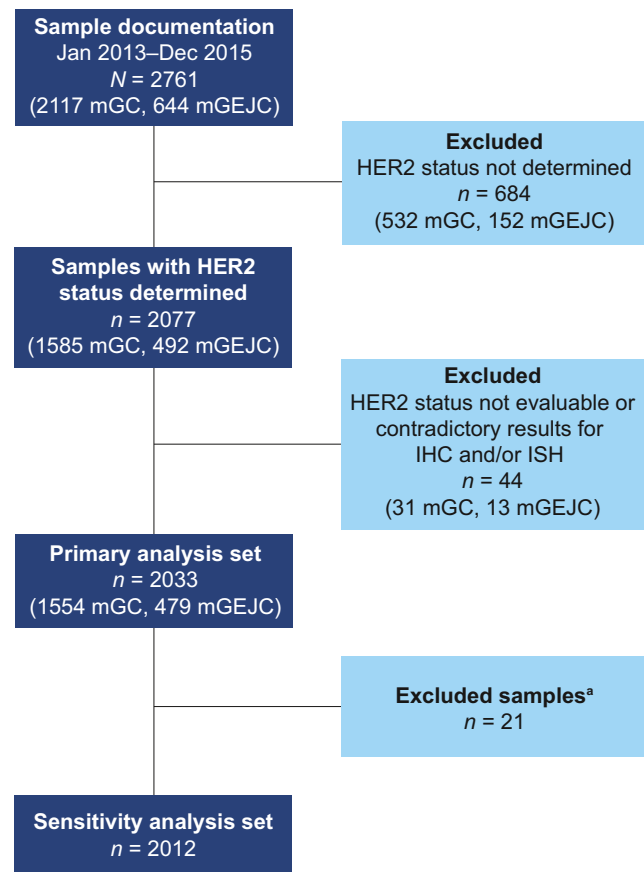


Fig. 1 Flow diagram of main analyses. ^aDocumented HER2 status not fully supported by IHC and/or ISH results. Additional sensitivity analyses were conducted and are reported in Online Resource 2. *HER2* human epidermal growth factor receptor 2, *IHC* immunohistochemistry, *ISH* in situ hybridization, *mGC* metastatic gastric adenocarcinoma, *mGEJC* metastatic gastroesophageal junction adenocarcinoma

rates were highest in patients aged ≤ 39 years, or where ISH was the first method of determining HER2 status (Fig. 2). Rates were lower for patients with diffuse mGC/mGEJC compared with other Lauren classifications, and relative HER2-positivity rates were consistent for mGC and mGEJC between biopsy and resection samples (Fig. 2).

The statistical influence of independent variables on HER2-positivity

In the MLR model, the influence of patient-, tumor-, and laboratory-related variables on HER2-positivity was analyzed. Five of these covariates (Lauren classification, HER2 testing rate per year, localization of carcinoma, method of sample retrieval [biopsy or resection], and first method of determining HER2 status) had a statistically significant influence on HER2-positivity (Table 1). The ranking of covariate importance was the same for all three measures used (p value, main effect, and total effect).

The ROC AUC of this five-parameter model was 0.66 (Supplementary Fig. S1; Online Resource 2). At the optimal

cut-point of 0.26, the sensitivity of model-based prediction was 0.60 and the specificity was 0.66; the Youden Index was 0.26.

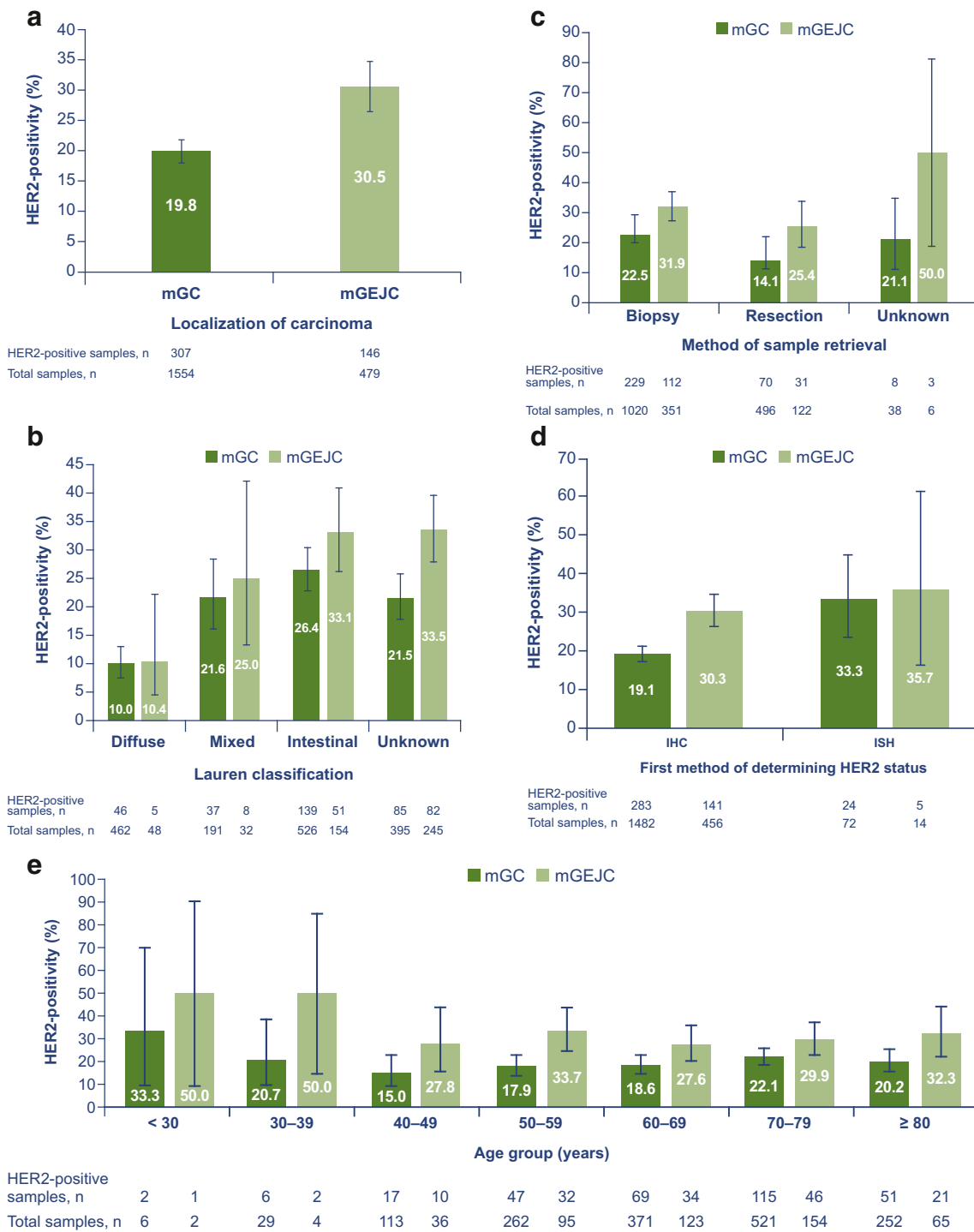


Fig. 2 Rates of HER2-positivity with 95% confidence intervals by individual covariates. The overall rates of HER2-positivity are shown according to **a** localization of carcinoma^a, **b** Lauren classification^b, **c** method of sample retrieval (biopsy or resection)^b, **d** first method of determining HER2 status^c, and **e** age group^d. ^aLikelihood ratio Chi-square test for comparison of rates of HER2-positivity; $p < 0.0001$. ^bLikelihood ratio Chi-square test for HER2-positivity across both localizations of

carcinoma; $p < 0.0001$. ^cLikelihood ratio Chi-square test for HER2-positivity across both localizations of carcinoma; $p = 0.0130$. ^dLikelihood ratio Chi-square test for HER2-positivity across both localizations of carcinoma; $p = 0.4329$. HER2 human epidermal growth factor receptor 2, mGC metastatic gastric adenocarcinoma, mGEJC metastatic gastroesophageal junction adenocarcinoma

Table 1 Likelihood ratio Chi-square test results of covariates and their relative importance in predicting HER2-positivity

Variable	Likelihood ratio Chi-square test of effect of independent variables on HER2-positivity (p value ^b)	Covariate importance in predicting HER2-positivity	
		Main effect ^a	Total effect ^a
Lauren classification	< 0.0001	0.48	0.548
HER2 testing rate for mGC/mGEJC	0.0001	0.138	0.197
Localization of carcinoma	0.0002	0.107	0.163
Method of sample retrieval	0.0284	0.049	0.087
First method of determining HER2 status	0.0593	0.021	0.041

HER2 human epidermal growth factor receptor 2, mGC metastatic gastric adenocarcinoma, mGEJC metastatic gastroesophageal junction adenocarcinoma

^a The main effect indicates the relative contribution by a factor of itself. The total effect indicates the relative contribution of a factor both by itself and in combination with other factors

^b Adjusted for the influence of all other independent variables

Initially, gender appeared to have a statistically significant influence on HER2-positivity ($p = 0.0069$), with a higher rate of HER2-positivity in male vs female patients (24.3% vs 18.4%). However, after adjustment for other covariates, particularly carcinoma type, the influence of gender was no longer statistically significant ($p = 0.43$). Further analysis demonstrated that significantly fewer female patients had mGEJC, which explained the apparent initial association of gender with HER2-positivity. Therefore, gender was not included in the final model.

“Age” was omitted from the final model because, in contrast to the preceding BC study, it did not show a statistical trend in the bivariate analysis ($p = 0.43$). Origin of sample (metastatic or primary tumor), TNM stage, and signet ring cell carcinoma did not show a significant impact in the stepwise MLR procedure, potentially partly due to high percentages of missing values.

In an additional statistical analysis on center level, cut-off value and manufacturer had statistically significant influence ($p < 0.05$), but their influence was strongly determined by a few outlying centers with unknown values for these variables; therefore, they were not further considered. The following parameters did not show an influence on center level: type of institution (university, private practice, or other), fixation method, commercial kit, antibody manufacturer, number of tumor-carrying documented samples, or regular participation in round-robin tests. The sensitivity analysis confirmed the results of the main statistical MLR model analysis showing that five tumor-, patient-, and laboratory-related covariates had a statistically significant influence on HER2-positivity (Supplementary Table S1; Online Resource 2).

Prediction profile for HER2-positivity

The magnitude of change in the prediction profile trace allowed visualization of the relative importance of covariates

in predicting the probability of HER2-positivity; the variability of each parameter was adjusted for the influence of all others at the level of the vertical dotted lines (Fig. 3). For a sample patient (Lauren classification: intestinal; HER2 testing rate: 60%; localization of carcinoma: gastric; method of sample retrieval: biopsy; first method of determining HER2 status: IHC), the model predicted the probability of HER2-positivity as 28.9%. Since the formula of the MLR model looks rather complex, a comprehensive set of 240 combinations of covariate factor levels and their predicted positivity probabilities is provided for illustration purposes in Supplementary Table S2 (Online Resource 2). For practical use, the application of such a table should be facilitated by an appropriate software tool.

Assessment of HER2-positivity by number of biopsies

In an additional analysis of biopsies, the probability of HER2-positivity did not significantly increase with the number of biopsy samples taken from each patient, $p = 0.27$ (Table S3). For a sample patient (localization of carcinoma: stomach; number of biopsies 8), the model predicted the probability of HER2-positivity as 27.2% (Fig. 3b); the result for the number of biopsies was adjusted for the influence of the other three variables as indicated by the vertical dotted lines.

Assessment of center effects

HER2-positivity probability was predicted for every center in the study based on the relevant patient-, tumor-, and laboratory-related characteristics of its samples (Fig. 4). These covariate-predicted HER2-positivity probabilities were sorted in ascending order and plotted together with documented HER2-positivity rates of centers and their 95% CIs. The predicted HER2-positivity probabilities for centers reflected the distribution of Lauren classification, HER2 testing rate, localization of carcinoma, method of sample retrieval, and

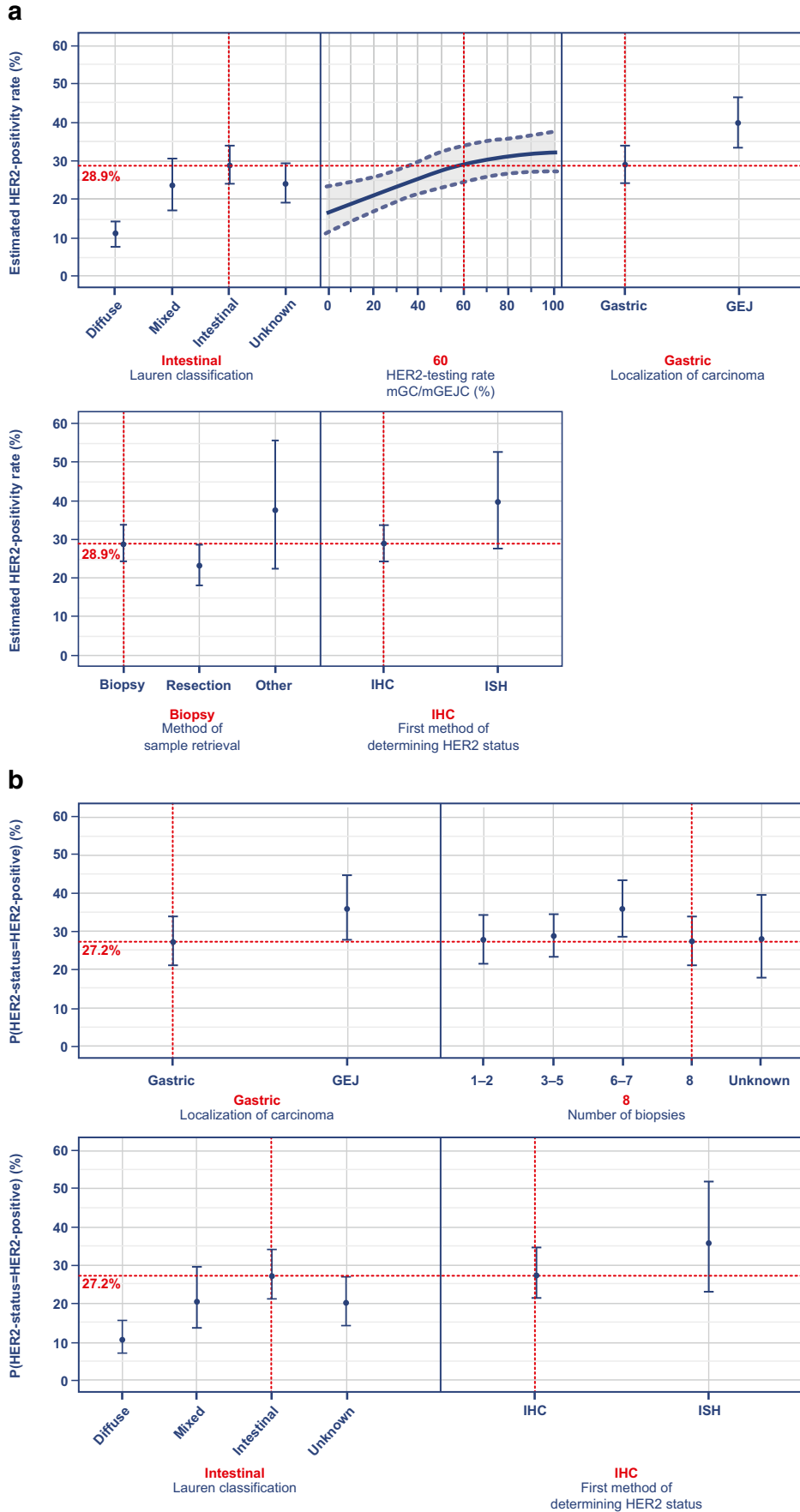


Fig. 3 Prediction profiles for the probability of HER2-positivity. **a** Prediction profile for each covariate and predicted probability of HER2-positivity for a combination of selected levels. The relative importance of individual covariates after adjustment in predicting HER2-positivity is demonstrated by the magnitude of change shown in the compartments by the prediction profiler traces. The dependence of HER2-positivity from the levels of one covariate were standardized at a certain level of all other covariates as indicated by the vertical dotted lines (Lauren classification: intestinal; HER2 testing rate: 60%; localization of carcinoma: gastric; method of sample retrieval: biopsy; first method of determining HER2 status: IHC). The predicted mean positivity probability for this combination of levels was 28.9% (horizontal dotted line; 95% CI). **b** Prediction profile in a submodel for biopsy samples including the number of biopsies used. Patient example: localization of carcinoma: gastric; number of biopsies: 8; Lauren classification: intestinal; first method of determining HER2 status: IHC. The predicted mean positivity probability for this combination of levels was 27.9% (horizontal dashed line; 95% CI). *CI* confidence interval, *GEJ* gastroesophageal junction adenocarcinoma, *HER2* human epidermal growth factor receptor 2, *IHC* immunohistochemistry, *ISH* in situ hybridization, *mGC* metastatic gastric adenocarcinoma, *mGEJC* metastatic gastroesophageal junction adenocarcinoma

first method of determining HER2 status across centers. The overall predicted HER2-positivity probability across all centers was 22.28% and thus the same as the overall documented positivity rate.

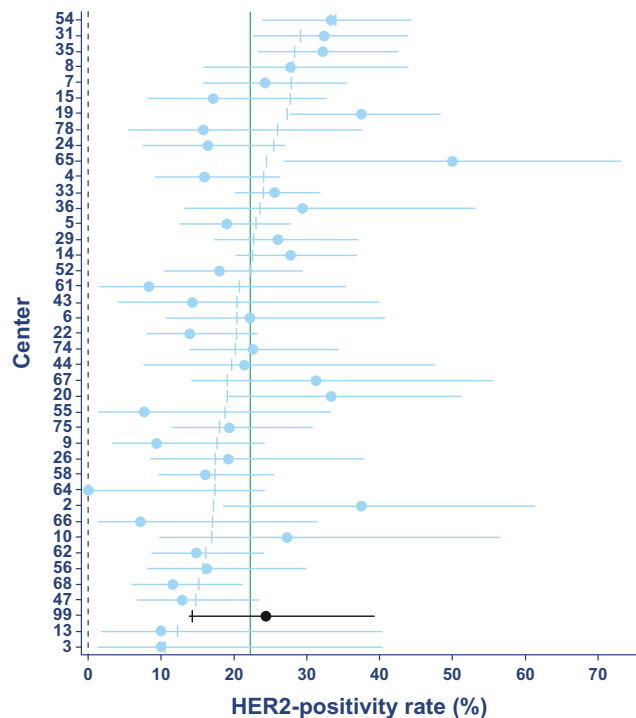


Fig. 4 Documented and model-predicted HER2-positivity. The forest plot for the extended model shows the predicted positivity probabilities (●) for all centers that contributed at least 10 samples; Center 99 pools the samples of all centers with fewer than 10 samples. Centers were sorted by increasing model-predicted positivity probabilities. Dots indicate the positivity rates as documented by the centers, together with their 95% CIs. The solid vertical line indicates the overall mean positivity rate of 22.3%. *CI* confidence interval, *HER2* human epidermal growth factor receptor 2

For most centers, the 95% CI of the documented positivity rate included the model-predicted positivity probabilities. For centers 2, 19, and 65, the lower limits of the 95% CI of the documented positivity rates were higher than their predicted positivity probabilities, indicating a statistically significant center effect, which cannot be explained by their sample and patient characteristics, unless these are adjusted for multiplicity; no statistically significant center effect was found when adjustment for multiplicity by Bonferroni-Holm correction was applied. The relatively wide CIs in Fig. 4, especially for centers with low numbers of patients, present the limitations of discovering center effects based on this approach.

Exclusion of the samples from centers 2, 9, and 65 resulted in only a slight decrease in the overall HER2-positivity rate from 22.3% to 21.3% for the remaining 1923 samples, including a HER2-positivity rate of 18.5% for mGC and 30.2% for mGEJC cases.

There were four more centers (31, 35, 54, and 68) for which the 95% CI of their documented rate included the model-predicted positivity probability but did not include the overall mean positivity rate of 22.3%; these are not considered as significant center effects because their deviation from the overall mean positivity rate can be explained by their sample and patient characteristics.

Discussion

Accurate HER2 testing is critical to limit false-positive/negative results and to select patients who might benefit from HER2-targeted treatments. 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 for BC [12, 23]. However, monitoring HER2 testing accuracy by HER2-positivity rate alone does not account for patient- or tumor-related factors that may influence overall HER2-positivity. Thus, this study aimed to identify patient- or tumor-related characteristics that influence HER2-positivity rate in mGC/mGEJC samples as determined by routine practice and develop a statistical model to predict the probability of HER2-positivity. As a real-world study, rather than ensuring that all centers used the same optimized techniques, our investigation was also intended to examine how differences in techniques between the centers affected HER2-positivity rates.

The moderately sized dataset allowed us to identify factors that affect HER2 testing results, and this is the first study to test for—and report—the combined influence of multiple parameters on HER2-positivity probabilities in the routine diagnosis of mGC/mGEJC.

The influence of tumor location and Lauren classification on HER2-positivity is well established, with GEJC tumors and diffuse tumors having higher and lower HER2-positivity,

respectively [3, 4]. The other covariates identified as having a statistically significant influence on HER2-positivity in this study were somewhat unexpected. The higher HER2-positivity rate in biopsies might be explained by different scoring criteria used for biopsies and resections. Differences in fixation time—known from BC biopsies—might contribute to this difference. In addition, Tominaga et al. [24] could show that tumor surface and borders, which are usually biopsied, more often exhibit an intestinal differentiation (commonly HER2-positive) compared to the center/invasive front of mGC (more often HER2-negative).

In contrast to BC, a reasonable number of IHC 0/1+, ISH-positive cases can be found in mGC/mGEJC [3]. Results from ToGA showed higher HER2-positivity when primary HER2 testing was determined by ISH (23.1%) compared with IHC then ISH (17.8%), due to the proportion of patients with low level amplification but equivocal protein overexpression [3]. Thus, higher overall HER2-positivity rates are observed when the rate of HER2 testing by ISH is higher. This explains why the highest HER2-positivity rate in our trial was reported by the center where ISH was performed first; overall HER2-positivity was 33.7% when HER2 testing was by ISH first and 21.8% when using IHC first, although the number of samples tested using ISH first in this trial was limited. These results indicate a higher risk of obtaining false-negative results if HER2 testing is performed by IHC first. However, ToGA data show that the best responses were observed in patients with significant HER2 protein expression in addition to being ISH-positive, which is the main reason that the CAP-ASCP-ASCO guidelines clearly recommend IHC to be performed first [4].

The higher HER2-positivity rate in high-volume centers might be due to a higher number of biopsies tested (see above) and putatively more experience in HER2 testing; however, no clear explanation for this correlation can be found in the literature.

Of the laboratory parameters tested, HER2-testing rate and first method of determining HER2-positivity were found to have the strongest influences on HER2-positivity rate and were included in the model. Additional laboratory parameters showed a statistically significant influence (e.g., the cutoff value for HER2-positivity), but these were not thought to be relevant. Although it would be interesting to understand whether variation between laboratories would decrease if they all used the same optimized techniques, this is outside the scope of our study as a retrospective analysis of real-world data.

In our study, the initial observed gender effect on HER2-positivity was attributed to the higher proportion of male patients with mGEJC. The increased incidence of GEJC in male patients is in line with previous observations of the overall population [25] and may be due to the increased occurrence of pathologic diseases associated with gastroesophageal reflux in men vs women [26].

The use of these covariates in a statistical model allowed prediction of HER2-positivity probabilities for each center and their comparison with documented rates. One possible explanation for the statistically significant deviation of three centers from the predicted positivity rates could be selection bias as no precise instructions for consecutive sampling were provided in this study. The performance characteristics of the model might be improved in a future study by more complete collection of covariate data, especially regarding Lauren classification.

Our findings are consistent with the equivalent BC study, which found that covariates must be taken into account in order to properly assess variations in HER2-positivity and evaluate HER2 testing quality [22], especially when the covariate levels of a center deviated from those expected. Assessments of positivity rates of centers or studies without adjustment for covariates cannot be considered as appropriate.

The overall mean HER2-positivity rate in this study is similar to that observed in ToGA (22.1%) [3] and falls within the HER2-positivity range of 6–32% reported in other studies of mGC/mGEJC [1, 2, 4]. Therefore, our study is in agreement with published data and is likely to be representative of HER2-positivity rates in German pathology centers.

For most covariates, the proportion of data classified as “missing” or “not evaluable” was relatively low (<5%), except for Lauren classification (31.5%). The ROC AUC of the primary statistical model and the sensitivity and specificity at the optimal cut-point were considered as acceptable performance measures, since the current focus was more on explanation of variation than on prediction. The study is also limited as consecutive sampling was not mandatory. This study was conducted using similar standards to the larger BC study, in which the centers were retrospectively questioned regarding their documentation of samples; from 48 centers, 32 indicated consecutive sampling, 2 indicated selected sampling, and 14 did not provide an answer. There was no statistically significant tendency toward higher deviations of reported positivity from model-predicted positivity in centers which did not provide feedback. Yet consecutive sampling should be implemented in future studies. Sensitivity analyses were conducted that excluded 21 samples with slightly contradictory data in measurements of HER2 status, and the statistical models were confirmed by these sensitivity analyses. Indeed, the borderline p value of the covariate “first method of determining HER2 status” was calculated as $p = 0.0394$ in the sensitivity analysis and thus further supported the inclusion of the parameter in the model.

Despite these limitations, the data collected for this study provide a valuable, real-world view of HER2 testing. Based on data collected from 50 centers during routine practice, and not as part of a patient-selective clinical study, we developed a statistical model that predicts HER2-positivity of a sample based on the evaluation of five covariates. Given the heterogeneity of HER2 staining in mGC/mGEJC and the range of published HER2-positivity rates, this model provides participating

centers with the opportunity to assess their overall HER2 testing performance by determining the deviation of their HER2-positivity rate from the probability of positivity predicted for their samples by the model. Yet further steps will need to be taken before applying this model into clinical practice. Firstly, data from another study are required to validate the statistical model independently. Subsequently, implementation of such a model as a user-friendly software application may be helpful for centers to assess their HER2 testing performance in the future (Online Resource 2). To allow centers to incorporate this model into quality control in the future, a set of samples (preferably consecutive) should be collected and the HER2-testing result, including 95% CIs, and other patient characteristics outlined in the model, should be determined. The model formula can then be used to predict the probability of HER2-positivity for both the individual samples and for the center for comparison with the HER2-testing result. HER2-testing results can also be compared with the centers included in this study and the overall HER2-positivity rate of 22.3%.

Centers that deviated from the expected HER2-positivity testing rate in this study should attempt to identify and, if necessary, control the cause of variability to improve HER2 testing. Centers should follow published HER2 testing guidelines for mGC/mGEJC and establish appropriate quality control procedures to reduce variation at all stages of the testing and reporting process [4]. Participation in a formal proficiency testing program is also advised [4].

In conclusion, this is the first time that a statistical modeling approach has been used to assess and explain variability in HER2-positivity rates in mGC/mGEJC. Our model is based on data collected during routine practice and provides a predicted HER2-positivity probability for any combination of patient- or tumor-related characteristics included as covariates in the model. The model could also be used to identify participating centers that deviated from the expected HER2-positivity rate. As therapy options for HER2-positive mGC/mGEJC continue to evolve, the ability to reliably identify patients likely to respond to a given therapy is essential.

Acknowledgments We would like to thank the patients, their families, the nurses, and the investigators who participated in this study. We would like to thank the 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 Online Resource 1). In addition, we would like to thank Johannes Ammann for operative and scientific support. Support for third-party writing assistance for this manuscript, furnished by Helen Keyworth, PhD, of Health Interactions, was provided by Roche Pharma AG, Grenzach-Wyhlen, Germany.

Author contributions RH, SM, and JR contributed to the conception and study design. HHK, TG, RH, KB, and JR contributed to the collection and assembly of data. GB, PS, TG, RH, WK, CK, and JR contributed to the analysis and interpretation of the data. WK performed the statistical analysis. All the authors contributed to the drafting of the article, provided critical revision for important intellectual content, and approved the final version. CK provided administrative, technical, or logistic support for this article.

Funding This study was funded by Roche Pharma AG, Grenzach-Wyhlen, Germany.

Compliance with ethical standards

Ethical approval Experimental protocols were reviewed and approved by the local ethical committee of the principal investigator and were available for submission to the local ethical committees of the participating centers.

Conflict of interest GB and PS have received reimbursement and honoraria for attending symposia and advisory boards, and supplies for research from Roche Pharma AG. HK has received honoraria and reimbursement from, and acted in an advisory and/or consulting role for, Roche Pharma, Genomic Health, and AstraZeneca. TG has received honoraria for lectures and/or consultant activities from Roche. RH has received honoraria and research funding from, and acted in an advisory role for, Roche. WK is employed by BDS Koch, which receives payment from Roche for statistical services. CK is an employee of Roche Pharma AG, Germany. SM is an employee of Roche. JR has acted in an advisory and/or consulting role for Roche, MSD, and BMS, and received research funding from Roche. KB has declared no conflicts of interest.

Data sharing statement Qualified researchers may request access to individual patient level data through the clinical study data request platform (www.clinicalstudydatarequest.com). Further details on Roche's criteria for eligible studies are available here: <https://clinicalstudydatarequest.com/Study-Sponsors/Study-Sponsors-Roche.aspx>. For further detail on Roche's Global Policy on the Sharing of Clinical Information and how to request access to related clinical study documents, see here: https://www.roche.com/research_and_development/who_we_are_how_we_work/clinical_trials/our_commitment_to_data_sharing.htm.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Boku N (2014) HER2-positive gastric cancer. *Gastric Cancer* 17:1–12. <https://doi.org/10.1007/s10120-013-0252-z>
2. 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. [https://doi.org/10.1016/S0140-6736\(10\)61121-X](https://doi.org/10.1016/S0140-6736(10)61121-X)
3. Van Cutsem E, Bang Y, Feng-yi F et al (2015) HER2 screening data from ToGA: targeting HER2 in gastric and gastroesophageal junction cancer. *Gastric Cancer* 18:476–484. <https://doi.org/10.1007/s10120-014-0402-y>
4. Bartley AN, Washington MK, Colasacco C, Ventura CB, Ismaila N, Benson AB III, Carrato A, Gulley ML, Jain D, Kakar S, Mackay HJ, Streutker C, Tang L, Troxell M, Ajani JA (2017) HER2 testing and clinical decision making in gastroesophageal adenocarcinoma: guideline from the College of American Pathologists, American Society for Clinical Pathology, and the American Society of Clinical Oncology. *J Clin Oncol* 35:446–464. <https://doi.org/10.1200/JCO.2016.69.4836>
5. Vogel CL, Cobleigh MA, Tripathy D, Gutheil JC, Harris LN, Fehrenbacher L, Slamon DJ, Murphy M, Novotny WF, Burchmore M, Shak S, Stewart SJ, Press M (2002) Efficacy and

- safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *J Clin Oncol* 20: 719–726. <https://doi.org/10.1200/JCO.2002.20.3.719>
6. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, Fleming T, Eiermann W, Wolter J, Pegram M, Baselga J, Norton L (2001) Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 344:783–792. <https://doi.org/10.1056/NEJM200103153441101>
 7. Baselga J, Cortés J, Kim SB, Im SA, Hegg R, Im YH, Roman L, Pedrini JL, Pienkowski T, Knott A, Clark E, Benyunes MC, Ross G, Swain SM (2012) Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. *N Engl J Med* 366:109–119. <https://doi.org/10.1056/NEJMoa1113216>
 8. Gianni L, Pienkowski T, Im YH, Roman L, Tseng LM, Liu MC, Lluch A, Staroslawska E, de la Haba-Rodriguez J, Im SA, Pedrini JL, Poirier B, Morandi P, Semiglazov V, Srimuninnimit V, Bianchi G, Szado T, Ratnayake J, Ross G, Valagussa P (2012) 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* 13:25–32. [https://doi.org/10.1016/S1470-2045\(11\)70336-9](https://doi.org/10.1016/S1470-2045(11)70336-9)
 9. Swain SM, Baselga J, Kim SB, Ro J, Semiglazov V, Campone M, Ciruelos E, Ferrero JM, Schneeweiss A, Heeson S, Clark E, Ross G, Benyunes MC, Cortés J (2015) Pertuzumab, trastuzumab, and docetaxel in HER2-positive metastatic breast cancer. *N Engl J Med* 372:724–734. <https://doi.org/10.1056/NEJMoa1413513>
 10. Marty M, Cognetti F, Maraninchi D, Snyder R, Mauriac L, Tubiana-Hulin M, Chan S, Grimes D, Antón A, Lluch A, Kennedy J, O’Byrne K, Conte PF, Green M, Ward C, Mayne K, Extra JM (2005) Randomized phase II trial of the efficacy and safety of trastuzumab combined with docetaxel in patients with human epidermal growth factor receptor 2-positive metastatic breast cancer administered as first-line treatment: the M77001 study group. *J Clin Oncol* 23:4265–4274. <https://doi.org/10.1200/JCO.2005.04.173>
 11. Smyth EC, Verheij M, Allum W, Cunningham D, Cervantes A, Arnold D, ESMO Guidelines Committee (2016) Gastric cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 27:v38–v49. <https://doi.org/10.1093/annonc/mdw350>
 12. Choritz H, Büsche G, Kreipe H, Study Group HER2 Monitor (2011) Quality assessment of HER2 testing by monitoring of positivity rates. *Virchows Arch* 459:283–289. <https://doi.org/10.1007/s00428-011-1132-8>
 13. Rüschoff J, Hanna W, Bilous M, Hofmann M, Osamura RY, Penault-Llorca F, van de Vijver M, Viale G (2012) HER2 testing in gastric cancer: a practical approach. *Mod Pathol* 25:637–650. <https://doi.org/10.1038/modpathol.2011.198>
 14. Cunningham D, Shah MA, Smith D et al (2015) False-negative rate for HER2 testing in 738 gastric and gastroesophageal junction cancers (GEC) from two global randomized clinical trials. *J Clin Oncol* 33(Suppl):Abstract 16. https://doi.org/10.1200/jco.2015.33.3_suppl.16
 15. Baretton G, Dietel M, Gaiser T, Kirchner T, Kreipe HH, Quaas A, Röcken C, Rüschoff J, Tannapfel A, Lordick F, al-Batran S, Hofheinz R, Lorenzen S, Moehler M, Thuss-Patience P (2016) HER2 testing in gastric cancer: results of a meeting of German experts [article in German]. *Pathologe* 37:361–366. <https://doi.org/10.1007/s00292-016-0179-3>
 16. Rüschoff J, Nagelmeier I, Baretton G, Dietel M, Höfler H, Schildhaus HU, Büttner R, Schlake W, Stoss O, Kreipe HH (2010) Her2 testing in gastric cancer. What is different in comparison to breast cancer? [article in German]. *Pathologe* 31:208–217. <https://doi.org/10.1007/s00292-010-1278-1>
 17. Rüschoff J, Dietel M, Baretton G, Arbogast S, Walch A, Monges G, Chenard MP, Penault-Llorca F, Nagelmeier I, Schlake W, Höfler H, Kreipe HH (2010) HER2 diagnostics in gastric cancer—guideline validation and development of standardized immunohistochemical testing. *Virchows Arch* 457:299–307. <https://doi.org/10.1007/s00428-010-0952-2>
 18. Hofmann M, Stoss O, Shi D, Büttner R, van de Vijver M, Kim W, Ochiai A, Rüschoff J, Henkel T (2008) Assessment of a HER2 scoring system for gastric cancer: results from a validation study. *Histopathology* 52:797–805. <https://doi.org/10.1111/j.1365-2559.2008.03028.x>
 19. Bianchi S, Caini S, Paglierani M et al (2015) 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* 21: 477–485. <https://doi.org/10.1007/s12253-014-9852-0>
 20. Perez EA, Suman VJ, Davidson NE, Martino S, Kaufman PA, Lingle WL, Flynn PJ, Ingle JN, Visscher D, Jenkins RB (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. <https://doi.org/10.1200/JCO.2005.03.4744>
 21. Roche PC, Suman VJ, Jenkins RB, Davidson NE, Martino S, Kaufman PA, Addo FK, Murphy B, Ingle JN, Perez EA (2002) Concordance between local and central laboratory HER2 testing in the breast intergroup trial N9831. *J Natl Cancer Inst* 94:855–857. <https://doi.org/10.1093/jnci/94.11.855>
 22. Rüschoff J, Lebeau A, Kreipe H et al (2017) Assessing HER2 testing quality in breast cancer: variables that influence HER2 positivity rate from a large, multicenter, observational study in Germany. *Mod Pathol* 30:217–226. <https://doi.org/10.1038/modpathol.2016.164>
 23. Wolff AC, Hammond ME, Hicks DG et al (2013) 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* 31:3997–4013. <https://doi.org/10.1200/JCO.2013.50.9984>
 24. Tominaga N, Gotoda T, Hara M, Hale MD, Tsuchiya T, Matsubayashi J, Kono S, Kusano C, Itoi T, Fujimoto K, Moriyasu F, Grabsch HI (2016) Five biopsy specimens from the proximal part of the tumor reliably determine HER2 protein expression status in gastric cancer. *Gastric Cancer* 19:553–560. <https://doi.org/10.1007/s10120-015-0502-3>
 25. Buas MF, Vaughan TL (2013) Epidemiology and risk factors for gastroesophageal junction tumors: understanding the rising incidence of this disease. *Semin Radiat Oncol* 23:3–9. <https://doi.org/10.1016/j.semradonc.2012.09.008>
 26. Kim YS, Kim N, Kim GH (2016) Sex and gender differences in gastroesophageal reflux disease. *J Neurogastroenterol Motil* 22: 575–588. <https://doi.org/10.5056/jnm16138>