Minimum biopsy set for HER2 evaluation in gastric and gastro-esophageal junction cancer

Background and study aims: The HER2 status of small endoscopic biopsies is important for predicting the eligibility of patients with metastatic HER2-positive gastric cancer or gastro-esophageal junction (GEJ) cancer for anti-HER2 therapy approved by the U.S. Food and Drug Administration. The aim of this study was to identify the minimum biopsy set required to evaluate the HER2 status with confidence.

Patients and methods: A total of 103 consecutive patients with resected gastric cancer or GEJ cancer were retrospectively selected; 2 formalin-fixed, paraffin-embedded samples of each surgical specimen and all paired endoscopic biopsies were analyzed for HER2 status with both immunohistochemistry (IHC) and fluorescent in situ hybridization (FISH) methods. A total of 10 virtual biopsies were constructed by selecting areas 2.6 mm in diameter on the luminal side of digitalized slides obtained from the surgical specimens. The results of evaluating HER2 status in virtual biopsies, slides containing complete surgical specimens, and endoscopic biopsies were compared.

Results: The resulting minimum biopsy set was applied to the endoscopic biopsy series for validation. Results: A biopsy set containing a minimum of 5 samples was identified as the most accurate in predicting HER2 status (sensitivity, 92%; specificity, 97%). In only 3 of the 103 cases (2.9%) did a comparison of the HER2 evaluation of virtual biopsies and that of entire slides show inconsistent results. Overall agreement between the endoscopic biopsies and surgical samples for HER2 IHC status increased from 78.4% to 92.3% when biopsy sets containing 4 or fewer samples were compared with biopsy sets containing 5 or more samples.

Conclusions: Although the recommendations suggest that 8 to 10 biopsies are necessary, the results show that a minimum set of 5 biopsies may be sufficient for reliable HER2 assessment in gastric cancer and GEJ cancer. However, endoscopists should be aware that a smaller sample size may be less accurate in selecting patients eligible for anti-HER2 therapy.

Introduction

Gastric cancer and gastro-esophageal junction (GEJ) cancer are global health concerns. Gastric cancer is one of the leading causes of cancer mortality; it is the second most common cause of cancer-related deaths worldwide and results in approximately 750,000 deaths per year [1]. The incidence of GEJ cancer has risen in the last three decades, especially in developed countries; unfortunately, outcomes remain very poor, with 5-year survival rates of less than 10% [2, 3]. Owing to this dismal prognosis, novel therapeutic strategies and molecular targeted therapies are under intensive investigation.

The most promising agent in recent years has been trastuzumab (Herceptin; Hoffmann-La Roche, Basel, Switzerland). This is a monoclonal humanized antibody specific for the human epidermal receptor 2 (HER2), a transmembrane tyrosine kinase member of the epidermal growth factor receptor (EGFR) superfamily [4]. Amplification of the HER2 gene has been observed in 20% to 30% of gastric cancers and GEJ cancers [5–9] and has negative prognostic significance, as recently highlighted by a systematic meta-analysis [10]. In addition to its prognostic implications, an important clinical feature of HER2 overexpression/amplification is its predictive role in patients with advanced disease. In 2010, the phase III Trastuzumab for Gastric Cancer (ToGA) study showed a significant survival advantage (overall survival, 16.0 vs 11.8 months) in patients who had gastric...
cancer or GEJ cancer with HER2 immunohistochemistry (IHC) 3+ positivity or HER2 IHC 2+ positivity coupled with fluorescence in situ hybridization (FISH) HER2 amplification (HER2: CEP17 ratio ≥ 2) receiving trastuzumab plus chemotherapy (capecitabine/cisplatin or fluorouracil/cisplatin) compared with those who received chemotherapy alone, with no significant increase in toxic side effects [11]. Because of these results, the U.S. Food and Drug Administration (FDA) and the European Medicine Agency (EMA) approved anti-HER2 therapy for patients with metastatic HER2-positive gastric cancer or GEJ cancer [12, 13].

In such a context, correct evaluation of the HER2 status in gastric cancer and GEJ cancer is an essential component of the diagnostic process in order to predict therapeutic response to anti-HER2 agents. However, two critical points have come to light. First, unlike HER2 expression in breast carcinoma, HER2 expression in gastric cancer and GEJ cancer is highly heterogeneous (with values ranging from 5% to 78% in various studies) [6, 14–19], and membrane immunoreactivity is often incomplete [14]. Such peculiarities led to the proposal of a new scoring system for HER2 expression specifically for the stomach [14, 20, 21]. Second, in many patients with unresectable or metastatic disease who are possible candidates for trastuzumab therapy, only small biopsy samples are available. This tissue may not be representative of HER2 expression in a whole sample [16]. These two critical points underscore the importance of defining the predictive accuracy of endoscopic biopsies in the evaluation of HER2 status. Indeed, biopsy samples probably provide suitable and reliable tissue for the accurate prediction of HER2 status only if the number of biopsy samples evaluated is adequate [16]. In current clinical practice, gastroenterologists perform a variable number of biopsies, and the aim of this study was to identify the minimum endoscopic biopsy set required to evaluate HER2 status in gastric cancer and GEJ cancer with confidence.

Materials and methods

Study assessment

The study included 103 consecutive cases: 50 resected gastric cancers retrospectively selected and retrieved from the Pathology Unit, Department of Surgical and Diagnostic Sciences, University of Genoa, between 2004 and 2009, and 53 consecutive GEJ cancers collected from the Surgical Pathology and Cytopathology Unit, Department of Medicine, University of Padua, between 2006 and 2010.

Selection criteria and methods have previously been detailed [16]. Briefly, cases were selected in which formalin-fixed, paraffin-embedded material from biopsy and subsequent surgical resection was available (50 cases of gastric cancer and 53 cases of GEJ cancer). The median age of the patients was 69 years (range, 37–90), and 73% (75/103) were male. The type of surgical procedure was determined from each patient’s medical records. All cases were reviewed and reclassified based on histopathology and the TNM staging system. Of the 103 tumors, 87 were classified as intestinal (84.5%) and 16 as diffuse according to the criteria of Lauren. For each patient, all diagnostic biopsy samples and 2 representative neoplastic samples from the surgical resections were retrieved. The 2 samples were chosen so that both represented full-thickness slices with evident inked serosal surface/radial margins. The median number of available biopsy samples for each case was 5 (range, 2–13), with a total number of 504 samples. Of these, 302 contained invasive adenocarcinoma (60%), and they were recruited for HER2 status evaluation. From each paraffin block, 5 serial sections, each with a thickness of 4 μm, were cut; 1 section was stained with hematoxylin and eosin, and the other 4 sections were mounted on SuperFrost Plus slides (Thermo Scientific, Braunschweig, Germany) for IHC and FISH analysis.

Immunohistochemistry and immunohistochemical evaluation

Sections were stained with PATHWAY anti-HER2 (clone 4B5) rabbit monoclonal primary antibody (Ventana Medical Systems, Oro Valley, Arizona, USA); IHC was performed with Ventana BenchMark XT automated immunostainer according to the manufacturer’s guidelines. Tissue sections were de-paraffinized and rehydrated. After antigen retrieval, sections were incubated with primary antibodies against HER2, and 3,3′-diaminobenzidine (DAB) was used as a chromogen. Finally, the slides were counterstained with hematoxylin, and coverslips were placed.

The IHC evaluation of HER2 expression on immunostained glass slides was jointly performed by four expert pathologists, who reached a consensus for each case. The consensus HER2 immunoreactivity was scored as 0, 1+, 2+, or 3+ by light microscopy, according to the validated scoring system for HER2 assessment in gastric cancer [20–22]. Score 2+ HER2 immunostaining is considered equivocal, and this finding, as suggested by published diagnostic flowcharts, requires further demonstration of amplification by in situ hybridization techniques [21].

Fluorescence in situ hybridization

HER2 amplification status was investigated by FISH with the PathVysion HER2 DNA Probe kit (Vysis; Downers Grove, Illinois, USA). Methods have previously been detailed [16]. An HER2/CEP17 probe mix was used, and evaluation was performed with a fluorescence microscope (BX61; Olympus, Hicksville, New York, USA); image capture was performed with CytoVision 3.93 software (Applied Imaging, Pittsburgh, Pennsylvania, USA). FISH analysis was performed on 1 sample of all surgical resections; in case of heterogeneous HER2 IHC expression, the field with the highest IHC score was chosen. The average HER2 : CEP17 ratio was calculated in each sample. A cutoff value of 2 distinguishes between amplification (≥ 2) and non-amplification (< 2) [20].

Construction of virtual biopsies

The HER2-stained slides of surgical specimens were digitally scanned with a 40 × objective (Nikon 40 × /0.75 NA Plan Apo; Nikon Instruments Europe, Amsterdam, Netherlands) and the Aperio ScanScope XT system (Aperio Technologies, Vista, California, USA) and digitally saved with an .svs extension (ScanScope Virtual Slides). File sizes ranged between 130645 and 1 225227 kB. Images were then visualized with Aperio ImageScope Viewing Software, downloadable as freeware.

A dedicated Ellipse tool was used to select circular areas, corresponding to ‘virtual biopsies’ (Fig. 1a), on both of the 2 surgical specimen slides selected for every case. These areas were 2.6 mm in diameter, which is estimated to be the average diameter of endoscopic gastric biopsies, as previously published [23]. The selected virtual biopsy area was drawn on the luminal part of the sample, thus simulating superficial biopsy samples obtained at endoscopy. In order to prevent selection bias (secondary to
the selection of IHC-positive areas during the virtual sampling process), the virtual biopsies were outlined by a person with no diagnostic experience and not aware of the study purpose. Colored circles were used to outline 5 randomly spaced virtual biopsy areas on the luminal surface of the digital slide, located on the side opposite the serosal ink markings (● Fig. 1b). A total of 10 virtual biopsies (5 on each of the 2 digital slides available for each case) were selected for each tumor.

For each virtual biopsy, the HER2 IHC evaluation was performed by someone blinded to the HER2 status in the rest of the slide, and the validated scoring system for HER2 assessment in gastric cancer biopsy samples was used [20]. Each virtual biopsy was therefore evaluated as negative (IHC score of 0 or 1+), equivocal (IHC score of 2+), positive (IHC score of 3+), or not assessable (if the sample was composed solely of non-neoplastic tissue, such as necrotic material, granulation tissue, or normal gastric mucosa). Finally, in order to ensure randomness of the selected areas, as happens during endoscopic sampling, each virtual biopsy was numbered with a random number table. In detail, each progressively numbered virtual biopsy was reassigned a number according to the random number table to avoid selection bias. In this way, any random biopsy could be part of the progressive biopsy set (1 biopsy set, 2 biopsy set, 3 biopsy set, and so forth).

Statistical analysis
The HER2 IHC status evaluated in the virtual biopsies was compared with (1) HER2 IHC overexpression in the surgical samples, (2) HER2 FISH amplification in the surgical samples, and (3) HER2 IHC overexpression in the endoscopic biopsies from the same patient. The Poisson regression model was used to establish the minimum biopsy set that could be used to predict overall HER2 status. Briefly, we calculated for each number of virtual biopsies the probability that at least one would be positive if the surgical specimens were positive. Therefore, the HER2 status, defined by biopsy sets composed of a progressively increasing number of virtual biopsies, was compared with the overall value of HER2 expression initially assessed for each case in the surgical specimens. This statistical analysis defines, for each biopsy set, both the specificity and sensitivity of the HER2 status assessment. Finally, in order to compare HER2 status evaluation in virtual biopsies and endoscopic biopsies, Student’s t test for paired data was used. Data analysis was performed with Stata Statistical Software: Release 13 (StataCorp, College Station, Texas, USA), and a P value below 0.05 was considered significant. In order to calculate sensitivity and specificity, both a score of 3+ and a score of 2+ were used to define IHC positivity for HER2 overexpression. This enabled the selection of both HER2 IHC 3+ (positive) and HER2 IHC 2+ (equivocal, with a requirement for a demonstration of actual gene amplification by FISH analysis) cases.

In addition, the resulting minimum biopsy set was applied to the endoscopic biopsy series to determine the effectiveness of this protocol by calculating the overall agreement between the endoscopic biopsies and the surgical samples for HER2 status.

Results
\* For all 103 gastric cancer/GEJ cancer surgical samples, we assessed HER2 status in 10 virtual biopsies, for a total number of 1030 virtual biopsies.

Comparison of HER2 IHC expression in virtual biopsies and expression in corresponding surgical samples
The evaluation of HER2 status in virtual biopsies was compared with IHC scoring in surgical samples. Statistical analysis, performed by analyzing biopsy sets composed of a progressively increasing number of virtual biopsies, identified a minimum biopsy set of 5 samples as the most accurate in predicting HER2 status (sensitivity of 91.9% and specificity of 97%) ( ● Fig. 2). In detail, sensitivity progressively increased from 61.5% for 1 biopsy to 91.9% for 5 biopsies, and no further increase was seen with more than 5 biopsies. With regard to specificity, it did not differ significantly, varying from 95.5% to 97.6% in all biopsy sets.
Comprehensive comparison of HER2 IHC expression in virtual biopsies and expression in IHC and FISH analysis of corresponding surgical samples

A comparison between the IHC evaluation of HER2 expression in virtual biopsies and expression on FISH analysis of surgical samples was key to verifying the reliability of the results of the present study because FISH analysis is considered the gold standard for HER2 status assessment [22, 24]. A comprehensive comparison, schematically shown in Table 1, was performed of (1) IHC scoring of a biopsy set of 5 samples, (2) IHC scoring of surgical samples, and (3) FISH analysis of surgical samples. This comparison demonstrated that IHC evaluation of HER2 expression in the virtual biopsy set of 5 samples predicted HER2 status in the surgical samples, as evaluated by IHC and FISH analysis, with an overall agreement rate as high as 97.1%. Only 3 cases (2.9%) showed inconsistencies. In detail, 1 case was IHC 0 on both the virtual biopsies and the whole surgical sample but amplified on FISH; the other 2 cases were IHC 0 on the virtual biopsies but IHC 2+ (equivocal) on the surgical samples and amplified on FISH.

Comparison of HER2 IHC expression in virtual biopsies and expression in corresponding endoscopic biopsies

No significant differences between IHC results in virtual biopsies and results in corresponding endoscopic biopsies (P=0.46) were found (Table 2). These results allow us to consider virtual biopsies analogous to real endoscopic biopsies.

Evaluation of proposed biopsy set of 5 in real endoscopic biopsy series

Among 103 patients, 51 had 4 or fewer endoscopic biopsies available for analysis; overall agreement between HER2 IHC status in biopsies and that in surgical samples was 78.4%, with 11 discordant cases. There were 52 patients who had 5 or more biopsies available, and overall agreement rose to 92.3%, with only 4 discordant cases. The differences between concordance and discordance in the group with 4 or fewer endoscopic biopsies and the group with 5 or more endoscopic biopsies are very close to significance (P=0.05), and these values are in line with those obtained with virtual biopsies.

Discussion and conclusions

The HER2 testing of advanced gastric cancer and GEJ cancer has a fundamental predictive role in defining which patients are eligible for trastuzumab therapy, as demonstrated in the Trastuzumab for Gastric Cancer (ToGA) study [11]. International and national guidelines [25, 26] recommend that multiple biopsies be performed on tumors, with the suggested number of samples in biopsy sets ranging from 8 to 10, depending on the size and type of neoplasm. However, an evidence-based definition of the minimum endoscopic biopsy set required to ensure appropriate tumor sampling and guarantee a confident evaluation of HER2 status is currently lacking, although it is of clinical relevance. Endoscopic biopsies may vary greatly in number in routine practice, limiting the use of real sampling for the construction and evaluation of a minimum biopsy set. We therefore used virtual biopsies to collect data from a uniform number of biopsy samples (10 virtual biopsies each).

The present study demonstrates that a minimum biopsy set of 5 samples has the highest sensitivity (91.9%) and specificity (97%) for reliable HER2 testing in gastric cancer and GEJ cancer. No increase in accuracy was seen with more than 5 biopsies, even with biopsy sets containing numerous (up to 10) biopsies. This finding is of importance if cost-effectiveness is to be considered. Moreover, numerous biopsies lengthen endoscopy times, reduce patient tolerance, and increase the risk for complications. Our results support the experience-based recommendation of Warneke et al [27] that 5 biopsies are probably sufficient for HER2 testing; however, our result was reached with significant methodologic differences. In particular, virtual biopsies based on...
tissue microarray were performed in the study of Warneke et al, in which 1.5-mm cores were selected in representative regions of the paraffin donor blocks (presumably to avoid necrotic and superficial areas, which are instead sampled at endoscopy). On the other hand, we decided to construct virtual biopsies exclusively at the luminal edges of tumors, simulating the topography of endoscopic samples. Virtual biopsies were not discarded even if they did not contain viable cancer, to simulate routine sampling more closely. Furthermore, our virtual biopsy dimensions were larger (2.6 mm vs 1.5 mm) and more closely simulated real-life dimensions [23].

Our study indicates that it is not necessary for a biopsy set with 5 samples to consist exclusively of neoplastic tissue. Indeed, 27% of the virtual biopsies and 40% of the endoscopic biopsies were classified as not assessable for HER2 status because they were taken from non-neoplastic areas, such as normal gastric mucosa, inflammatory/necrotic material, or non-invasive neoplasia. Possible explanations for the higher percentage of biopsies in our endoscopy series that were not assessable are the following: (1) biopsies were routinely performed at the gastric ulcer borders, where normal gastric mucosa can cover a neoplastic lesion; (2) the samples were composed of muco-necrotic material, which is not present in surgical samples because the surface is cleaned before being cut. These factors may possibly limit the application of virtual biopsy results in daily practice. However, when a biopsy set containing 5 samples was applied to our endoscopic biopsy series, it proved to be the best minimum protocol in real-life practice.

Unlike published guidelines [25,26], which propose 8 to 10 biopsies, our findings suggest that a more conservative biopsy protocol, consisting of 5 biopsies, should be sufficient for HER2 testing. Apart from heterogeneity, one of the justifications for such a high number of biopsies is that the diffuse type of gastric cancer may easily be missed in biopsy material. However, this is probably not crucial if one considers that the diffuse type of gastric cancer is less common than the intestinal type and is more often HER2 negative.

Possible limitations to this study are that the evaluation of the minimum biopsy set was performed in virtual biopsies derived from retrospective surgical material from two centers. However, this was necessary in order to obtain a large and constant number of biopsies, which were not available in our real-life biopsy cases.

In conclusion, the present study demonstrates that virtual biopsies performed on surgical samples of gastric cancer and GEJ cancer can be compared with corresponding endoscopic biopsies. On this basis, evaluating a progressively increasing number of virtual biopsies for HER2 status, we defined a minimum set of 5 biopsies required for a reliable HER2 assessment in gastric cancer and GEJ cancer. However, endoscopists should be aware that a smaller sample size may not be as accurate in selecting patients eligible for anti-HER2 therapy.

Competing interests: None

Acknowledgments

We thank Simona Pigozzi and Silvia Bonadio for their technical support.

References

24 Park YS, Hwang HS, Park HJ et al. Comprehensive analysis of HER2 expression and gene amplification in gastric cancers using immunohistochemistry and in situ hybridization: which scoring system should we use? Hum Pathol 2012; 43: 413 – 422