The identification of predictive factors for perioperative chemotherapy in esophago-gastric cancer

Resection without macroscopic and microscopic residual tumor is the prerequisite for achieving cure in localized esophago-gastric cancer (EGC). But even after R0 resection with adequate lymphadenectomy, the overall prognosis of resected EGC is critical, because minimal residual disease may cause relapse which ultimately leads to death. To prevent recurrence and help patients to survive, perioperative treatment has been studied in randomized, controlled trials and has become standard of care. While Asians prefer postoperative (adjuvant) chemotherapy and some US centers are in favor of adjuvant chemoradiotherapy according to the Intergroup 0116 study, the majority of European, Australian and Canadian centers recommend perioperative chemotherapy or preoperative chemoradiation, depending on the exact localization and the histological subtype of the tumor [1, 2].

Despite radical resection and perioperative treatment, 50%–65% of patients with stage II and III EGC recur and die within 5 years. Obviously, in these patients chemotherapy was unable to eradicate minimal residual disease. On the other hand, between 25% and 50% of patients do recur after surgery alone, indicating that chemotherapy is not needed in a considerable number of patients [3–5]. In brief, unnecessary as well as ineffective treatments are two unresolved issues in EGC.

Biomarker testing has become an integral part of research in oncology. Prognostic biomarkers are investigated to clarify the need for more or less treatment. Predictive biomarkers are needed to select the right drugs for a specific patient. This sounds easier than it is. Postoperative nomograms and modified pathological staging systems may provide more robust prognostic information than classical TNM staging in localized EGC [6, 7]. But the major problem of systems based on the analysis of resection specimens is that prognostic information is unavailable at the crucial time point of treatment planning. Second, unfavorable prognostic biomarkers do not necessarily indicate whether patients will benefit from more intensive treatment and which particular treatment could improve the prognosis. In brief, we do not yet have sufficiently validated prognostic biomarkers in EGC and we completely lack predictive markers in the setting of localized disease eligible for preoperative chemotherapy.

Several hypothesis-generating studies have been published on biomarkers potentially predicting the sensitivity to chemotherapy in EGC. Expression and polymorphisms of 5-fluorouracil (5-FU) targets and platinum detoxification enzymes are among the most frequently investigated molecules. But study results from different groups were often conflicting and inconsistent [8, 9]. Microsatellite instability may describe a distinct biological subgroup of EGC associated with a different prognosis and response to chemotherapy, but findings are not yet robust [10].

One of the most promising hypotheses was the potential association of HER2 amplification with a benefit from epirubicin-containing chemotherapy. As elaborated by Okines et al. in their article, in breast cancer HER2 amplification is a potential predictive biomarker of epirubicin sensitivity, probably due to co-amplification with the topoisomerase IIα gene, located close to HER2 on chromosome 17q21 [11]. This prompted the authors to investigate this question in EGC.

The authors sampled tumor material from the prospective, randomized, controlled, MAGIC trial. This trial randomly assigned patients with localized EGC, clinical stage II and III, to receive surgery alone or surgery with perioperative epirubicin, cisplatin and 5-FU [3]. Perioperative chemotherapy led to a meaningful and statistically significant survival benefit. Paraﬃn-embedded tumor blocks from the diagnostic biopsy, resection specimen or both were received from 415 of 503 patients. Tissue microarrays (TMAs) were constructed. HER2 protein expression was assessed by immunohistochemistry. In addition, gene amplification was investigated using bright-ﬁeld dual in situ-hybridization. The HER2 positivity rate was 10.9% in the whole cohort with a concordance between HER2 protein expression and gene amplification of 96%. A major result of this study was that HER2 was not found to be prognostic. Even more important, pretreatment biopsies did not indicate an enhanced beneﬁt for HER2 positive tumors from epirubicin-containing chemotherapy [11].

Therefore, the principal hypothesis of this study, which was that HER2 expression may predict the beneﬁt of chemotherapy, could not be veriﬁed. One major problem of the analysis may be the limited power of testing for heterogeneity, given that in only 15 patients a positive HER2 status was assessed from pre-treatment biopsies. Another potential problem is the doubtful contribution of epirubicin to the beneﬁt of perioperative chemotherapy of EGC in general [11]. Of note, colleagues from France who conducted a trial in parallel to the UK MAGIC study used cisplatin and 5-FU (without epirubicin!) and came up with the same hazard ratio in favor of perioperative chemotherapy versus surgery alone [5]. A recently presented adjuvant study from the United States found no difference in survival of patients treated with adjuvant radiation combined with 5-FU versus adjuvant radiation combined with epirubicin, cisplatin and 5-FU [12]. Moreover, a systematic review indicating that epirubicin may be effective in advanced gastric cancer has been criticized for...
methodological drawbacks [13]. If gastric cancer is
only moderately sensitive to anthracyclines in general,
any predictive marker for anthracycline sensitivity will
eventually fail.

Almost 20 years of research on HER2-expression in EGC
has generated a plethora of results which can be summarized
as follows [14]: HER2 overexpression correlates significantly
with HER2-gene amplification; HER2 is heterogeneously
distributed in the primary tumor as well as in metastases; and
finally, HER2 is more prevalent in proximal and intestinal type
gastric cancers, respectively. It is currently impossible to
provide definitive conclusions from the literature on the
association of HER2-expression with patient survival, local
tumor growth (T-category), nodal spread (N-category) or
tumor stage according to Unio Internationale Contra Cancrum
(UICC) [14].

The assessment of HER2 overexpression is more
complicated in EGC compared with breast cancer. It
necessitated the development of a novel scoring system, which
is different from the breast cancer scoring [15]. Gastric cancer
cells more commonly harbor basolateral expression and rarely
circumferential HER2-staining [15, 16]. The gastric cancer
scoring-system takes into account the unique expression
patterns (including heterogeneity) of HER2 in EGC and
increased the comparability of study results obtained from
different regions (Asian vs. European/North-American).
However, a major problem remained unanswered, i.e. which
effect perioperative treatment has on HER2 expression and
hence its correlation with prognosis. Okines et al. are among
the first to address systematically this important issue, and
provide evidence that HER2 has no prognostic or predictive
value in patients receiving perioperative chemotherapy in the
MAGIC study [11]. In this respect, European GC study
cohorts are different from many Asian study cohorts, where
adjuvant chemotherapy with fluoropyrimidines for stage II and
III EGCS has been standard of care for a longer period of time
than in Europe. This may have had ill-defined impacts on
studies, evaluating the correlation between HER2-expression
and prognosis. Why is the correlation between tumor HER2-
expression and patient survival inconsistent among different
studies? Apart from sometimes ill-defined confounding
variables such as different treatment modalities (e.g. with and
without adjuvant/perioperative chemotherapy or radiation), it
has to be kept in mind that intestinal and non-intestinal type
EGCs show significant differences with regard to HER2
expression as well as prognosis [14]. Statistical analyses which
do not take into account these facts may generate misleading
results. The seemingly prognostic influence of HER2 may also
be related to the phenotype (i.e. intestinal versus. diffuse).
However, given the overall low prevalence of HER2, subgroup
analyses often suffer from the small patient numbers, and
meta-analyses are hampered by the different staining and
scoring systems used in the diverse studies [14, 17]. This
problem is illustrated by the study of Okines et al. [11]. They
confirm the relatively low prevalence of HER2 overexpression
in a large group of 415 EGC patients. Two recently published
reviews on the overall prevalence of HER2-expression reported
on prevalence ranging from 4.4% to 53.4% [14, 17]. These
differences are certainly due to different scoring systems and
different staining protocols. More recent studies using the
gastric cancer scoring-system (immunohistochemistry coupled
with in situ hybridization) harvest smaller differences,
commonly ranging between 5% and 29% HER2 positivity [18].
Given the low prevalence of HER2 overexpression in EGC,
correct classification of EGC patients as HER2-positive or –
negative is becoming a major issue for the resulting medical
treatment, and hence particular attention should be paid to
study conduction. Okines et al. used TMAs throughout their
study, which were obtained from biopsies and resection
specimens [11]. Our own previous studies have shown that the
generation of TMAs in itself carries the risk of a sampling
error [18]. The comparison of whole tissue sections with the
corresponding TMAs, the latter generated from the same
paraffin blocks used for the assessment of whole tissue
sections, showed a false-negative rate of 24% and a false-
positive rate of 3% for TMAs [18]. Even sampling for the
generation of TMAs carries the risk of a study bias and should
be kept in mind. Okines et al. generated TMAs from biopsy
and resection specimens [11]. The core cylinders collected
from biopsies had a smaller diameter (0.5 mm) than those
obtained from resection specimens (1.0 mm). This adds further
risks of sampling errors and has to be considered, when
interpreting their results.

HER2-expression in EGC teaches many lessons. Research
carried out over a period of 20 years with different scoring
systems limits the comparability of study results [14, 17]. This
is further fraught with the risk of ill-defined therapy regimens
and standards of care used in different parts of the World and
different countries. Not all are clearly outlined in the
publications. Complexity is increased by the heterogeneity of
tissue samples used in the diverse studies (biopsy before or
after therapy; resection specimens before or after chemotherapy
or radiation) and different staining and scoring systems
applied. Studies on HER2 in EGC underscore the importance
of standardizing staining protocols, evaluation systems, study
designs, study conduct and the plenty-full pitfalls, when
standards are lacking. The only chance to advance the
development of companion diagnostics and their translation
into routine clinical practice is standardization.

What is the future of biomarker testing in EGC? HER2
highlights several major advancements and problems of
targeted therapy and the development and validation of
companion diagnostics. The ToGA study [19] has spurred an
abundant and fruitful international research effort on HER2 in
EGC, which is unlikely to have taken place without the results
gov/pubmed for the terms ‘Her2’, ‘ErbB2’, ‘stomach’, ‘gastric’
and ‘cancer’ generated 748 hits for the time period spanning
1986 through 2012. Between 1986 and 2011 the average
number of publications was 17.6/year and increased during the
last 3 years to 108.3/year (Figure 1). The expectation of
improved patient outcomes significantly urges research on
predictive and prognostic biomarkers. These have shown that
(i) study populations are becoming increasingly heterogeneous
with regard to oncological treatment protocols applied to
individual patients and patient subgroups before study.
inclusion. (ii) The number of patients available for studies on targeted therapeutics is small given the low prevalence of the target molecule in EGC subtypes (usually <20% of a tumor type). (iii) Cohorts of treatment naïve patients are becoming increasingly difficult to collect, particularly when perioperative and adjuvant (radio-) chemotherapy is standard of care. (4) National and international guidelines on cancer diagnosis and treatment do not specify the amount of tumor tissue necessary for reaching a diagnosis.

The diagnosis of cancer can be achieved on a fairly small amount of tissue. However, this amount may be unsuitable for the examination of prognostic or predictive biomarkers, as it carries the risk of sampling errors. The only way out of this quandary is an international effort aiming for a rigorous standardization of sampling procedures, thereby reducing the influence of confounding pre-analytical factors (including tissue sampling), standardization of analytical tests controlled by external quality assurance systems, standardization of scoring systems early in the development of companion diagnostics, and robust validation of the diagnostic algorithms by independent study cohorts [20]. These efforts reach beyond REMARK and STARD as we now need standards not only for reporting, but also for the conduct of these studies [20, 21].

F. Lordick1* & C. Röcken2

1University Cancer Center Leipzig, University of Leipzig, 2Institute of Pathology, University of Kiel, Germany (*E-mail: florian.lordick@medizin.uni-leipzig.de)

disclosure

FL worked as an advisor for Roche, Amgen, Lilly, Fresenius Biotech and Ganymed. He received lecture honoraria from Roche, Merck, Amgen and Fresenius Biotech. He got research support from Merck, Sanoﬁ, GSK and Fresenius Biotech.

references


Figure 1. Number of publications found in http://www.ncbi.nlm.nih.gov/pubmed using the terms 'HER2', 'ErbB2', 'gastric', 'stomach' and 'cancer' during the period 1986–2012. The arrow indicates the publication year of the ToGA-study [22].
How to assess assessments?

The main principle of regulatory drug evaluation is relatively simple: the marketing authorisation of a medicinal product relies on the demonstrated efficacy, safety and quality of the product. To reach a positive opinion for marketing authorisation, the applicant has to provide evidence that the benefits of the product in question outweigh the risks in the proposed indication. Although the procedures and legal basis may differ across the regulatory agencies such as the US Food and Drug Administration (FDA), Health Canada (HC) and European Union’s European Medicines Agency (EMA), all the Western countries follow the principle of ‘positive benefit-risk ratio’ for drug approval and utilise harmonised Internal Conference of Harmonisation guidelines.

From 2001 through 2010, 186 applications for novel therapeutic agents have been reported to have been approved by the EMA, 99 by HC and 225 by the FDA, respectively [1]. Specifically for anti-cancer products, 42 approvals, corresponding to 100 indications, were granted by the EMA between 1995 and 2008 [2]. Along with increasing scientific and financial interests in the field of anti-cancer products, these numbers are expected to rise in the near future—highlighted by the fact that oncology is the therapeutic area with the highest number of scientific advice requests received by the EMA [3]. Considering the impact that marketing authorisations and other regulatory decisions have, not only on the clinical practice but also on the conduct of clinical trials, they are quite infrequently analysed or reviewed in the scientific literature. To be provocative, they may often be criticised, but not that often systematically studied. Several papers have been recently published, but they are mostly descriptive in nature or focus on procedural issues or legal framework. For instance, the review timelines have been quite thoroughly analysed [1, 2, 4], but few of the papers attempt to analyse the reasons behind the review times—that are often seen as too long and to delay the clinical use of novel agents.

Many of the studies appear to be comparative, i.e. they seek to find differences between the decisions of the regulatory agencies, such as EMA and FDA. Notably though, Trotta et al. [2] have recently analysed differences in the indications of novel anti-cancer products as approved by the EMA and FDA. More importantly, they went further in their analysis by considering the clinical implications of these differences—which, interestingly, were found to be clinically relevant in 10 cases of 100 [2].

In this issue of Annals of Oncology, Ito et al. provide a critical review of regulatory approvals based on the so-called ‘Public domain applications’ in Japan, with a focus on anti-cancer drugs [5]. ‘Public domain application’ is a flexible regulatory process, by which a currently off-label use of a well-known drug can be approved without further clinical trials. In this procedure, the approval may be granted by the Japanese Pharmaceuticals and Medical Devices Agency (PMDA), if either sufficient data of usage experience outside Japan is provided, if scientific literature demonstrates the safety and efficacy or if the existing results of clinical trials otherwise show sufficient data. The process appears to share some analogy with the Article 505b(2) applications of FDA, as well as with the application based on ‘well-established use’ (WEU) in the EU. The controversial aspect of the paper by Ito et al., however, is not the legal background or discussion...