Elsevier Editorial System(tm) for European Journal Of Cancer

Manuscript Draft

Manuscript Number: EJC-D-19-01143R1

Title: Disease-free survival as a surrogate for overall survival in neoadjuvant trials of gastroesophageal adenocarcinoma: pooled analysis of individual patient data from randomized controlled trials

Article Type: Original Research Article

Keywords: Gastroesophageal adenocarcinoma; neoadjuvant chemotherapy, neoadjuvant chemoradiotherapy, overall survival, disease-free-survival, individual patient data meta-analysis

Corresponding Author: Dr. Ulrich Ronellenfitsch, M.D.

Corresponding Author's Institution: University Hospital Halle

First Author: Ulrich Ronellenfitsch, M.D.

Order of Authors: Ulrich Ronellenfitsch, M.D.; Katrin Jensen, PhD; Svenja Seide, MSc; Meinhard Kieser, PhD; Matthias Schwarzbach, MD; Tracy E Slanger, PhD; Bryan Burmeister, MD; David Kelsen, MD, PhD; Donna Niedzwiecki, PhD; Guillaume Piessen, MD; Christoph Schuhmacher, MD; Susan Urba, MD, PhD; Cornelis van de Velde, MD, PhD; Marc Ychou, MD; Ralf Hofheinz, MD; Sylvie Lorenzen, MD

Abstract: Introduction
Disease-free survival (DFS) is increasingly being used as surrogate endpoint for overall sur-vival (OS) in cancer trials. So far, there has been no validation of the surrogacy of DFS for OS for neoadjuvant treatment of gastroesophageal adenocarcinoma.
Methods
The study uses individual patient data (IPD) from eight randomized controlled trials (n=1,126 patients) comparing neoadjuvant therapy followed by surgery with surgery alone for gas-troesophageal adenocarcinoma. Correlation between OS-time and DFS-time was calculated to evaluate individual-level surrogacy. For each trial, survival curves using the Kaplan-Meier method were plotted and hazard ratios (HRs) on the treatment effects were calculated for OS and DFS separately. Those HRs were pooled in a random-effects meta-analysis. Observed were compared with predicted HRs for OS using results from an error-in-variables linear re-gression model accounting for the uncertainty about the estimated effect. The strength of the association was quantified by the coefficient of determination to assess trial-level surrogacy. The surrogate threshold effect was calculated, to determine the minimum treatment effect on DFS necessary to predict a non-zero treatment effect on OS.
Results
A strong correlation between OS-time and DFS-time was observed, indicating a high individu-al-level surrogacy. For all RCTs, estimated HRs for OS and DFS were highly similar. In the meta-analysis, the overall HR for OS was virtually identical to that for DFS. The estimated coefficient of determination r^2 for the association between HRs for OS and DFS was 0.912 (95% confidence interval 0.75-1.0), indicating a very good fit of the regression model and thus a strong trial-level surrogacy

between OS and DFS. The surrogate threshold effect based on the regression analysis was 0.79.

Discussion

Based on strong correlations between DFS and OS, as well as a strong correlation of the treatment effects of the two endpoints in the error-in-variable regression, DFS seems an ap-propriate surrogate marker for OS in randomized trials of neoadjuvant chemotherapy or chemoradiotherapy for gastroesophageal adenocarcinoma.

Cover Letter

Dear editor,

on behalf of all co-authors, I would like to thank you for giving us the opportunity to submit a revised version of our manuscript

Disease-free survival as a surrogate for overall survival in neoadjuvant trials of gas-troesophageal adenocarcinoma: pooled analysis of individual patient data from ran-domized controlled trials

We have thoroughly addressed all reviewers' comments and provided a detailed point-by-point answer in the attached document. We trust that the manuscript now fulfils all requirements for publication in the European Journal of Cancer.

We would also like to add a previously not involved author to the manuscript. The reason for this amendment is that Dr Katrin Jensen, who is the lead biometrician of the study, is currently sick and unable to work for a longer time. Therefore, Ms Svenja Seide acted as a substitute and performed the extensive re-analysis, which were requested by the reviewers. Consequently, we think she fulfills all ICMJE requirements for being an author on this manuscript.

Kind regards

Ulrich Ronellenfitsch, MD

*Author Form

European Journal of Cancer
Author Form

All manuscripts submitted to the *EJC* must be accompanied by this form. Please scan the form and transmit it to the Editorial Office via EES with the manuscript. If you are unable to do this, please contact the Editorial Office at ejcancer@elsevier.com to organise an alternative way of sending the form to the *EJC*.

Title of Manuscript:
Disease-free survival as a surrogate for overall survival in neoadjuvant trials of gastroesophageal adenocarcinoma: pooled analysis of individual...

Contribution	**Author(s)**
Study concepts:	UR, KJ, MK, RH, SL
Study design:	UR, KJ, MK
Data acquisition:	all authors
Quality control of data and algorithms:	UR, KJ
Data analysis and interpretation:	all authors
Statistical analysis:	KJ, MK
Manuscript preparation:	UR, KJ
Manuscript editing:	UR, KJ
Manuscript review:	all authors

Ethical Approval for Research: No (Yes) / N.A.

External Funding: No / (Yes)
Source of Funding: German Federal Ministry of Education and Research
Name of Principal Investigator: Ulrich Ronellenfitsch
(If funded, please include a statement as to the role of the study sponsor at end of manuscript under a heading 'Role of the Funding Source')

Possible Conflict of Interest: (No) / Yes
(Please ensure that a 'Conflict of Interest' statement is included in your manuscript)

Number of Tables: 2 **Number of Figures:** 3

Name and Title of Corresponding Author: Ulrich Ronellenfitsch, MD
Address: Universitätsklinikum Halle (Saale), Klinik u. Poliklinik für VGEC
Address: Ernst-Grube-Str. 40
Postcode and country: 06120 Halle (Saale), Germany
Tel No: +49-345-557-2327
Fax No: +49-345-557-2823
Email: ulrich.ronellenfitsch@uk-halle.de

"I confirm that all the authors have made a significant contribution to this manuscript, have seen and approved the final manuscript, and have agreed to its submission to the *European Journal of Cancer*".

Signed (corresponding author): U. C.
Date: June 21, 2019

***Revision Notes**

We thank the reviewers for their constructive comments. Below, we provide a point-by-point answer detailing our response to the comments.

Reviewer 1:

- *Statistical analysis:*

  - *The results do not show strong heterogeneity in the treatment effect (I2=16%), therefore the use of random effect model is probably not completely justified.*

The question whether or not between-trial heterogeneity should always be assumed is a highly debated topic in meta-analysis. In our opinion, in the present case particular heterogeneity between trials must be expected, as the treatment schemes used in the trials (perioperative chemoradiotherapy or perioperative chemotherapy, different combinations of chemotherapeutical agents etc.) and probably also the study populations differ from each other. We therefore agree with Julian Higgins "Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified, International Journal of Epidemiology, Volume 37, Issue 5, October 2008, Pages 1158–1160, https://doi.org/10.1093/ije/dyn204" in the aspect that the decision to perform common-effect or random-effects meta-analysis should not be based solely on the I^2-statistic. Furthermore, as I^2 is defined as ratio of variance proportions $I^2 = \frac{\text{Variance between trials}}{\text{Variance between trials+Mean variance within trials}}$, I^2 may be low not only because between-trial heterogeneity is low, but also because within-trial variation is high relative to between-trial variation. Instead of using I^2 as a means of deciding which statistical model to use, we therefore estimate a random-effects model in all cases, but report on the I^2 as additional hint on heterogeneity. In case of homogeneity, the estimate of the between-trial heterogeneity would then be zero. Considering that the difference in the combined effect between both statistical models lies in the fact that this heterogeneity estimate is added in the calculation of the uncertainty of the combined effect, both models would then result in the same estimates.

The rationale why we assume clinical heterogeneity and therefore use random-effects models is explained in the "material and methods" section and has now been amended with concrete examples. After careful consideration, we felt that an in-depth discussion of the statistical background of these two models and the usage of I^2 as a criterion to decide between statistical models would not improve the readability of our manuscript.

  - *The authors may clarify that they used independent model for estimating the treatment effect on DFS and on OS, and then do not account for the fact that both measures were carried out on the same patients. This may then lead to some bias on the variances estimates. Alternative approaches, such as copula introduces extra assumptions, which might raise additional concerns; nevertheless, either both analyses should be carried out or this should be clarified and discussed*

Copula models are indeed a valuable option in surrogate endpoint validation. Following the reviewers suggestions, we added the copula approach following the tutorial by Rotolo and colleagues on using copula in the estimation of surrogacy in time-to-event data “surrosurv: An R package for the evaluation of failure time surrogate endpoints in individual patient data meta-analyses of randomized clinical trials, Computer Methods and Programs in Biomedicine, Volume 155, 2018, https://doi.org/10.1016/j.cmpb.2017.12.005” and now in addition estimate the level of surrogacy by means of a (unadjusted) Clayton copula. We have added the (very similar) results of the copula model to the results section and present them now at the end of the methods and results paragraphs:

“To account for the repeated use of data stemming from the same patients for the two endpoints, we used the copula approach as described in Rotolo et al. (2018) as sensitivity analysis.”

“In addition to the regression analyses, copula estimation was performed to account for correlation between the treatment effects of the two outcomes within patients. Results using an unadjusted Clayton copula are very similar to those of the regression analysis. The coefficient of determination for the copula analysis is 0.95 while the surrogate threshold effect is 0.74. As convergence could not be achieved, adjustment for the second-step linear regression for measurement-error in the copula model was not performed.”

- *What measurement error model was used in the regression analysis? Please provide some reference to be able to reproduce the results.*

We used the Stata command “eivreg” for estimation. The implemented model in Stata is specified as follows:

$$Y = X\beta + e$$
$$X = X^{*} + U$$

In this model, the outcome Y (in our case the treatment effects on overall survival from the individual trials) is explained through the variable X (in our case the treatment effects on disease free survival from the individual trials), to model trial-level surrogacy. Considering that observed values of X will include a measurement error, this error is explicitly modelled in a second level of the regression by adding the error term U as an additive measurement error in one observed variable.

Draper NR, Smith H. Applied regression analysis, third edition. New York, NY: John Wiley & Sons; 1998.

Kmenta J. Elements of econometrics: Second edition. Ann Arbor, MI: University of Michigan Press; 1997.

Treiman DJ. Quantitative Data Analysis: Doing Social Research to Test Ideas. San Francisco, CA: Jossey-Bass; 2008.

Buyse, M, Burzykowski T, Michiels S & Carroll K. (2008). Individual- and trial-level surrogacy in colorectal cancer. Statistical Methods in Medical Research, 17(5), 467–475. https://doi.org/10.1177/0962280207081864

We have added the pertinent explanation including the references to the “material and methods” section of the manuscript.

- *Likewise, please provide clarification on how the confidence interval for R2 was calculated*

The standard error of R^2 was bootstrapped (1000 repetitions) and the 95% confidence interval was then calculated using the bootstrapped standard error and quantiles based on the student’s‘ t-distribution. We have added this description to the “material and methods” section.

- *Results:*

  - *Please summarize the type of treatments included in the analysis from table 1 in the text as this gives the area of applicability of the conclusions.*

We have added this information to the first paragraph of the “results” section.

  - *Mean follow-up is provided. Due to the high censoring rate, median follow-up would be more appropriate.*

We now report the median follow-up time instead of the mean follow-up time.

  - *More than 80% of the patients are described to have died without documented recurrence. This seems to be quite high. Is it related to insufficient monitoring of recurrence? Please clarify.*

The description how and when recurrences were detected was somewhat ambiguous. In fact, in several patients recurrence was diagnosed at the time of death, most often as the immediate cause of death. This has been amended in the manuscript. Consequently, the notion that 80% of the patients are described to have died without documented recurrence is an overestimation. In fact, this holds true for only 57% of patients, as stated in the discussion. Although the consistency of the correlation between DFS and OS across trials does not suggest large differences in follow-up schemes and DFS determination across the different trials, the ascertainment of DFS in some trials might have been suboptimal. This fact had already been addressed in the “discussion” section.

  - *The word cumulative hazard ratio is ambiguous. Do author refer to "overall HR"?*

Thank you for pointing out this semantical mistake. We have now replaced the term “cumulative hazard ratio” by “overall hazard ratio” throughout the manuscript.

- *The sentence "the deviation of observed DFS from what would be expected based on OS is small and within the confidence limits." is awkward as we would rather expect to have prediction of HR(OS) based on HR(DFS)*

Thank you for spotting this mistake. We have now interchanged OS and DFS in this sentence as well as in other sentences where these terms were used.

- *At several places in the text, DFS and OS should be replaced by treatment effect on DFS and on OS (or HRs).*

As suggested, we have replaced DFS/OS by “the treatment effect on DFS/OS”.

- *Please clarify whether the individual or the trial-level surrogacy is considered.*

Indeed, we may need to state more clearly to which level of surrogacy we are referring in the respective paragraphs. We performed both analyses (as suggested by Buyse M, Moeberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomised experiments. Biostatistics 2000;1: 49–68). We used the OS-time and DFS-time of all patients in all trials to assess individual level-surrogacy. We additionally calculated the treatment effect on OS and DFS separately by estimating the two HR in each trial and provide two combined effects (or overall effects) by means of random-effects meta-analyses. Using the estimated HR for both OS and DFS from the individual trials, we estimated trial-level surrogacy using an error-in-variable regression with an additive measurement error. Both analyses showed a high level of surrogacy which we now clarified in methods and results section by adding the terms ‘individual-level surrogacy’ and ‘trial-level surrogacy’ at the respective passages in the text.

- *Due to the very large fraction of death counted in the DFS, it is not clear what would be the benefit of using DFS instead of OS in RCTs. An additional analysis that would report the correlation between HR(DFS) at 2, 3, 4 years and HR(OS) at 5 years (for instance), would help the reader to evaluate the median follow-up requested for future trials would DFS be used as a primary endpoint.*

Following your suggestion, we have now also calculated the correlation coefficient between DFS-times and OS-times at varying time-points. This is described in the “material and methods” section, and the results are presented in the “results” section.

- *Theoretically, STE is computed for trials of infinite size. Considering that typical RCTs in this indication are of medium size, STE might be over-estimated.*

We have added a pertinent statement in the “discussion” section.

- *Figure 1: Please add number of patients at risk below the survival curves*

We changed the figure accordingly. The number at risk have now been added to the survival curves.

- *Figure 3: Please use different codes for identifying CT and RT-CT trials.*

Figure 3 has been amended accordingly with codes identifying the trials.

- *Conclusion:*
- *"results show a considerable correlation between the two outcomes in patients from the included trials both on the individual trial level and in the pooled population. " this sentence is a bit unclear. One cannot compute correlation between single trials. Please clarify*

We have modified this misleading sentence.

- *Under the assumption of proportional hazards, HRs are not influenced by intensity of follow-up in randomized trials if follow-up is similar. Please clarify this part of discussion.*

We have amended the discussion to that regard.

- *The authors are correct when they state that there is no clear-cut threshold for R2 above which an intermediate endpoint would qualify for surrogacy. However, following Shi et al. (JCO 2017), the value of 0.8 is often accepted.*

Thank you for providing this suggested threshold value along with its reference. We have added it to the discussion.

Reviewer 2:

Minor

1. *Introduction.*

a. *The authors state that multimodal treatment has shown improved outcome. There are some relevant trials lacking, e.g. Al-Batran Lancet 2019; 393: 1948-57. In addition, new neoadjuvant treatment strategies in recruiting trials are numerous, e.g. INNOVATION-trial, CRITICS-trial, and were not mentioned.*

Thank you for this valuable remark. In addition to a rather general review on novel treatments for gastroesophageal adenocarcinoma, we have now also specifically mentioned and cited the INNOVATION and CRITICS trials.

b. *Please reconsider the term 'advanced disease' at the end of the Introduction, since you intend to state that neoadjuvant treatment is generally considered in non-metastatic disease.*

Following your suggestion, we have replaced the term ‘advanced’ with ‘non-metastatic’.

2. *Material and methods.*

a. *Is there any information on the distribution of tumor subtypes, i.e. Lauren classification or based on Cancer Genome Atlas' molecular classification? Of note, the predicted survival and response on therapy is highly depending on these subtypes.*

Unfortunately, histological and molecular subtypes are not consistently reported in the included trials. Therefore, these data were not available for our analyses. Although you are absolutely right that response to therapy depends on these characteristics, there is no evidence suggesting that the association between DFS and OS is systematically influenced by the histological or molecular subtype. Nevertheless, we do now address this possible shortcoming of our study in the discussion.

b. *The neoadjuvant treatment strategies should be explained in a more comprehensive way, i.e. design of treatment, specifying (combination of) chemotherapeutic agent(s), number of cycles, radiotherapy dose and number of fractions.*

We have added detailed information on the neoadjuvant treatment schemes to table 1.

c. *In addition, you interchange neoadjuvant (Material and methods section) with perioperative (Results section). Reconsider terminology, and specify in the Material and methods section.*

To avoid misunderstandings and increase clarity for the reader, we have replaced the term “perioperative” and “preoperative” with “neoadjuvant”, so that only “neoadjuvant” is used throughout the manuscript.

d. *The authors define DFS as time from a landmark of six months after randomization to recurrence or death. I agree that you have to set a surrogate starting point. However, six months is a rather long interval from randomization. The rate of chemo(radio)therapy-related death and postoperative mortality are of interest in implementing a neoadjuvant treatment strategy, and by using six months landmark you introduce bias. There are trials included which investigated a perioperative strategy, suggesting these patients would have longer duration of treatment. Could these patients have affected outcome?*

The definition of disease-free survival in the given context is somewhat difficult, as it cannot be specified from the same starting time as overall survival. While overall survival covers the survival during the whole observation period, possible disease progression during the first months may not be meaningfully evaluated when considering differences between treatment and control group. Inherent to the design of all included RCTs, progression might be detected earlier, i.e. at the time of the operation, in patients undergoing upfront surgery.

Due to this reason we decided to use the time-point t0 + 0.5 years as starting point for DFS in our analysis (“landmark analysis”). Obviously, this choice of the landmark is somewhat arbitrary, and early therapy-related and postoperative deaths might not be mirrored exactly by the landmark analysis. On the other hand side, these deaths are all accounted for in the OS analysis. There is no commonly agreed landmark time for such analyses, but six months have been used in an RCT on neoadjuvant chemotherapy for esophageal squamous cell cancer (Boonstra JJ et al. BMC Cancer. 2011;11:181) and in our previous meta-analyses (Ronellenfitsch U et al. Eur J Cancer 2013;49:3149-58; Ronellenfitsch U et al. EJSO 2017;43:1550-1558), so for the sake of consistency we chose the same value.

We have now addressed this issue more in-depth in the discussion.

4. *Figure 1.*

a. *Define the treatment arm in the legend: 'treat_arm=0 DFS' represents the DFS of the group of patients treated with upfront surgery etc.*

We have labelled the curves in the figure legend.

b. *Figure 1 is presented as a Kaplan Meier curve. However, due to the landmark analysis t=0 for the DFS curves are not starting at 1.00, which is incorrect. All patients at risk at t=0 (landmark six months) should be included (=1.00).*

To account for the fact that the patients are under risk for both events simultaneously, we have now changed the graph such that event times for both outcomes can be directly compared. Event times for DFS are now set to 0.5 if events happened during the first 6 months of follow-up. Thus, respective patients are included in the graph but only the relevant times are displayed.

5. *Figure 2.*

a. *In this Forest plot the term 'Favours peri-op chemo' is being used. Please substitute for 'Favours pre-op chemotherapy' or 'Favours neoadjuvant therapy', since there were also radiotherapy studies included.*

We have changed the annotation accordingly.

b. *There is no clear order in the presented trials; please align with Table 1.*

We have changed the order of trials in figure 2 so that it matches the order in table 1. For the sake of consistency, we have also readjusted the order in which the trials are presented in table 2.

6. *Figure 3.*

a. *This 'calibration plot' should be clarified, whereas a label on the y-axis is missing (observed/expected DFS/OS), and the different trials should be denominated.*

We have added a label on the y-axis as requested and labelled the single trials in the figure.

Highlights

- Overall survival correlates strongly with disease-free survival in randomized trials of neoadjuvant chemotherapy or chemoradiotherapy for gastroesophageal adenocarcinoma.
- The vast majority of variation in overall survival can be explained by variation in disease-free survival.
- Disease-free survival seems an appropriate surrogate marker for overall survival, and future trials could consider using the former as primary endpoint.

Disease-free survival as a surrogate for overall survival in neoadjuvant trials of gastroesophageal adenocarcinoma: pooled analysis of individual patient data from randomized controlled trials

Ulrich Ronellenfitsch[a], Katrin Jensen[b], Svenja Seide[b], Meinhard Kieser[b], Matthias Schwarzbach[c], Tracy E Slanger[d], Bryan Burmeister[e], David Kelsen[f], Donna Niedzwiecki[g], Guillaume Piessen[h], Christoph Schuhmacher[i], Susan Urba[j], Cornelis van de Velde[k], Marc Ychou[l], Ralf Hofheinz[m], Sylvie Lorenzen[n]

[a]Medical Faculty of the Martin Luther University Halle-Wittenberg and University Hospital Halle (Saale), Department of Visceral, Vascular and Endocrine Surgery, Ernst-Grube-Str. 40, 06120 Halle (Saale), Germany, ulrich.ronellenfitsch@uk-halle.de

[b]Institute of Medical Biometry and Informatics, University of Heidelberg, Im Neuenheimer Feld 130.3, 69120 Heidelberg, Germany, jensen@imbi.uni-heidelberg.de, ~~kieser@imbi.uni-heidelberg.de~~meinhard.kieser@imbi.uni-heidelberg.de, seide@imbi.uni-heidelberg.de

[c]Department of General, Visceral, Vascular, and Thoracic Surgery, Klinikum Frankfurt Höchst, Gotenstraße 6-8, 65929 Frankfurt am Main, Germany matthias.schwarzbach@klinikumfrankfurt.de

[d]Institute for Quality and Efficiency in Health Care (IQWiG), Im Mediapark 8, 50670 Cologne, Germany, t.slanger@hotmail.com

[e]University of Queensland, Princess Alexandra Hospital, Brisbane, Australia, Princess Alexandra Hospital, Brisbane, QLD 4102, Australia, bryan.burmeister@health.qld.gov.au

[f]Memorial Sloan-Kettering Cancer Center and Weill Cornell Medical College, 1275 York Avenue, New York, NY 10065, USA, kelsend@mskcc.org

[g]The Alliance for Clinical Trials in Oncology (Alliance) Statistics and Data Center, Duke University Medical Center, Hock Plaza, 2424 Erwin Rd, Room 8040, Durham, NC 27705, USA, donna.niedzwiecki@duke.edu

[h]Department of Digestive and Oncological Surgery, University Hospital C. Huriez, Inserm UMR-S 1172, Jean-Pierre Aubert Research Center (JPARC) Team, 1, rue Polonovski, 59037 Lille cedex, France, guillaume.piessen@chru-lille.fr

[i]Department of Surgery, Klinikum rechts der Isar, Technical University Munich, Ismaninger Str. 22, 81675 Munich, Germany, christoph.schuhmacher@tum.de

[j]Division of Hematology/Oncology, University of Michigan Medical Center, 1500 E Medical Center Drive, C347, SPC 5848 Ann Arbor, MI 48109, USA, surba@med.umich.edu

[k]Department of Surgery, Leiden University Medical Center, K6-R, P.O. Box 9600, 2300 RC Leiden, The Netherlands, c.j.h.van_de_velde@lumc.nl

[l]Centre Régional de Lutte Contre le Cancer, Val d'Aurelle, Montpellier Cedex 05, France, marc.ychou@icm.unicancer.fr

[m]Day Treatment Center (TTZ), Interdisciplinary Tumor Center Mannheim (ITM) & 3rd Department of Medicine, University Medical Centre Mannheim, Medical Faculty Mannheim of the University of Heidelberg, Theodor-Kutzer-Ufer 1-3, 68167 Mannheim, Germany, ralf.hofheinz@umm.de

[n]3rd Department of Internal Medicine (Hematology/Medical Oncology), Klinikum rechts der Isar, Technische Universität München, Ismaninger Str. 22, 81675 Munich, Germany, sylvielorenzen@gmx.de

Corresponding author:

Ulrich Ronellenfitsch, MD

Medical Faculty of the Martin Luther University Halle-Wittenberg and

University Hospital Halle (Saale)

Department of Visceral, Vascular and Endocrine Surgery

Ernst-Grube-Str. 40

06120 Halle (Saale), Germany

phone: +49-345-557-2327

fax: +49-345-557-2823

e-mail: ulrich.ronellenfitsch@uk-halle.de

Parts of this work were supported by the German Federal Ministry of Education and Research (grant number 01 KG 0807) and intramural research funds of the University of Heidelberg (Foundation for Cancer Research).

Abstract

Introduction

Disease-free survival (DFS) is increasingly being used as surrogate endpoint for overall survival (OS) in cancer trials. So far, there has been no validation of the surrogacy of DFS for OS for neoadjuvant treatment of gastroesophageal adenocarcinoma.

Methods

The study uses individual patient data (IPD) from eight randomized controlled trials (n=1,126 patients) comparing neoadjuvant therapy followed by surgery with surgery alone for gastroesophageal adenocarcinoma. Correlation between OS-time and DFS-time was calculated to evaluate individual-level surrogacy. For each trial, survival curves using the Kaplan-Meier method were plotted and hazard ratios (HRs) on the treatment effects were calculated for OS and DFS. separately. Those HRs were pooled in a random-effects meta-analysis. Observed were compared with predicted HRs for DFS inOS using results from an error-in-variables linear regression model accounting for the uncertainty about the estimated effect. The strength of the association was quantified by the coefficient of determination to assess trial-level surrogacy. The surrogate threshold effect was calculated, to determine the minimum treatment effect on DFS necessary to predict a non-zero treatment effect on OS.

Results

A strong correlation between OS-time and DFS-time was observed, indicating a high individual-level surrogacy. For all RCTs, estimated HRs for OS and DFS were highly similar. The cumulativeIn the meta-analysis, the overall HR for OS was virtually identical to that for DFS. The estimated coefficient of determination r^2 for the association between HRs for OS and DFS was 0.912 (95% confidence interval 0.75-1.0), indicating a very good fit of the regression model and thus a strong correlationtrial-level surrogacy between OS and DFS. The surrogate threshold effect based on the regression analysis was 0.79.

Discussion

Based on strong correlations between DFS and OS, as well as a strong correlation ~~between DFS and OS on both~~ of the ~~individual patient and trial level and on~~treatment effects of the ~~finding that~~ two endpoints in the ~~vast majority of variation in OS can be explained by variation error~~ in ~~DFS~~-variable regression, DFS seems an appropriate surrogate marker for OS in randomized trials of neoadjuvant chemotherapy or chemoradiotherapy for gastroesophageal adenocarcinoma.

Keywords

Gastroesophageal adenocarcinoma; neoadjuvant chemotherapy, neoadjuvant chemoradiotherapy, overall survival, disease-free-survival, individual patient data meta-analysis

Introduction

The majority of patients with gastroesophageal adenocarcinoma, i.e. adenocarcinoma of the esophagus, gastroesophageal junction or stomach, present with advanced disease [(1, 2]). In the absence of distant metastasis, oncological resection is the only potentially curative modality. However, the prognosis after surgery alone in locally advanced disease is poor with 5-year survival rates of only about 20-30%. A considerable proportion of patients who undergo upfront resection will eventually relapse and die as a result of their disease. Multimodal treatment concepts comprising chemotherapy or chemoradiotherapy have shown survival benefits [(3-6]7). Because of difficulties in administering chemotherapy or radiotherapy soon after surgical procedures, major efforts have been undertaken to explore different neoadjuvant treatment strategies to improve outcomes. Several randomized trials and meta-analyses have demonstrated that preoperativeneoadjuvant chemotherapy and preoperativeneoadjuvant chemoradiotherapy followed by surgery are associated with longer survival compared to surgery alone [(3, 4]). Therefore, both strategies are now recommended by guidelines for the treatment of locally advancednon-metastatic gastroesophageal adenocarcinoma [7, (8], 9). Despite this approach the rate of recurrence remains high, and new neoadjuvant treatment-streatment concepts including postoperative chemoradiation after neoadjuvant chemotherapy like in the CRITICS trial (10), biologicals such as dual HER2 targeting in the INNOVATION trial (11), or immunotherapy are being explored [9](12).

The conduction of trials assessing such new treatments is often difficult, as they require a large number of patients in order to detect the relatively small incremental benefits of a new treatment. Overall survival (OS) has been considered the gold standard endpoint for cancer clinical trials. However, OS requires an extended follow-up period and thus trial duration, which leads to higher costs and a long delay until results become available. This has resulted in an increasing number of cancer trials using actuarial endpoints, which can be ascertained sooner, such as disease-free survival (DFS) [10](13). In locally advancednon-metastatic gastroesophageal adenocarcinoma, a large proportion of relapses occur before 3 years [(3]).

However, post-recurrence treatment could dilute or even eliminate an apparent improvement in tumor control achieved by neoadjuvant treatment. Therefore, the validity of DFS as surrogate endpoint for OS in gastroesophageal adenocarcinoma remains controversial.

A meta-analysis concluded that DFS is an appropriate endpoint for OS in studies of gastric cancer in the adjuvant setting, showing that the effect of treatment on OS is largely predictable from its effect on DFS ~~[11]~~(14), whereas another analysis found this not to be true in the palliative setting ~~[12]~~(15). However, most evidence from large-scale randomized trials in resectable gastroesophageal adenocarcinoma exists for neoadjuvant treatment; which has now, as opposed to adjuvant treatment, become standard of care for ~~advanced~~non-metastatic disease. Therefore, the present analysis assesses if DFS is a valid surrogate endpoint for OS in trials using neoadjuvant chemotherapy or chemoradiotherapy for gastroesophageal adenocarcinoma.

Material and methods

Trial and patient selection

We used individual patient data (IPD) from randomized controlled trials (RCTs) comparing neoadjuvant chemotherapy or chemoradiotherapy (i.e. therapy administered at least partially prior to surgery) with surgery alone for gastroesophageal adenocarcinoma. All RCTs including patients with resectable, non-metastatic tumors without prior treatment and providing information on both OS and DFS were potentially eligible. There were no exclusion criteria regarding specific treatment regimens. Trials were identified by a systematic literature review covering publications until 2011, the details of which have been previously published [(3, 13]16). All trialists from eligible trials were solicited to provide IPD, and trials were only included in the analyses in case of an affirmative response. Upon data collection, trialists had been asked to provide most recent follow-up data, even if follow-up was longer than that reported in respective publications. Some of the IPD had already been used in a previous meta-analysis comparing treatment effects of neoadjuvant chemotherapy with surgery alone and in a secondary analysis exploring predictors of postoperative survival [(3, 14]17). From two eligible RCTs [(4, 15]18), the final results were only published after completion of that meta-analysis. One of these trials [15](18) provided IPD for the present analysis. This resulted in IPD from eight trials [15-22](18-25), which comprise 1,126 patients, entering the analysis (table 1). All included RCTs had been approved by the respective competent ethical committee.

Definition of outcomes

Overall survival (OS) was defined as time from randomization to death or to the last documented follow-up. Disease-free survival (DFS) was defined as time from a landmark six months after randomization to recurrence or death, whichever occurred first, or to the last documented follow-up. The purpose of this landmark was to account for differences in timing between randomization and surgery between trial arms. Recurrence and death within the first six months were considered events at the landmark.

Statistical analyses

Characteristics of patients were compared between groups using the chi-square test for discrete variables, and the Wilcoxon-Mann-Whitney test for continuous variables. Correlation between OS-time and DFS-time was assessed by means of the Spearman rank correlation coefficient to assess individual-level surrogacy. OS and DFS were calculated according to the Kaplan-Meier method separately for patients who received neoadjuvant therapy and patients who underwent surgery alone. This was done in the entire study population and for patients from each single RCT. Survival in the single strata was compared using the log-rank test. Hazard ratios with 95% confidence intervals were calculated for the comparison of neoadjuvant therapy and surgery alone for each RCT for the treatment effect on both OS and DFS. These hazard ratios were pooled in ~~a~~two separate meta-~~analysis~~analyses to provide a combined effect of the estimated hazard ratios. Random-effects models were used for calculation of point estimates and confidence intervals because heterogeneity between the 'true' effects of the different regimens (neoadjuvant chemoradiotherapy or neoadjuvant chemotherapy, different combinations of chemotherapeutical agents etc.) used in the trials was assumed. ~~All~~Additionally, all results were investigated for statistical heterogeneity by I^2 statistics, without using this measure to choose between meta-analytic models.

To compare the observed with the predicted hazard ratio for ~~DFS~~the treatment effect on OS, a linear regression model accounting for the uncertainty about the estimated effects ~~by using~~was used. There, the treatment effects on DFS were included as predictors in an error-in-variables linear regression model with 95% prediction limits ~~was used.~~ to predict the treatment effects on OS. The strength of the association was quantified by the coefficient of determination r^2 to assess trial-level surrogacy. Considering that the estimated treatment effects from the individual trials on DFS will include a measurement error, we added an additive measurement error in the observed variable (26-29). The standard error of R^2 was boot-

strapped (1000 repetitions) and the 95% confidence interval was then calculated using the bootstrapped standard error and quantiles based on the student's‘ t-distribution.

In order to better evaluate the median follow-up requested for future trials using DFS as primary endpoint, the correlation coefficient between DFS-time and OS-time at varying time points (one, two, three, and fourin years of follow-up) was additionally calculated. To account for the repeated use of data stemming from the same patients for the two endpoints, we used the copula approach as described in Rotolo et al. as sensitivity analysis (30).

All significance tests were two-sided with p=0.05 as cutoff. IPD were analysed using SAS 9.4 (SAS Institute Inc.). Meta-analyses were conducted with Stata 14.2 (Stata Corp~~.~~).) and the copula estimation was performed in R version 3.5.1 using the extension surrosurv (30).

Results

Table 1 provides an overview of the characteristics of the eight included RCTs. All but one trial were multi-center trials. Three trials were carried out in the USA, two in France, one in the Netherlands, and three in several countries in Europe, North Africa or Australasia. Four RCTs comprised a neoadjuvant chemoradiotherapy scheme and four a neoadjuvant chemotherapy scheme in their experimental arm. All trials used 5-fluorouracil, and seven out of eight trials cisplatin as chemotherapeutical backbone. Table 2 shows demographic and clinical characteristics of the 1,126 patients included in the analysis, both for the entire study population and separately for patients from the neoadjuvant therapy and surgery alone arms. Histological and molecular tumor subtypes were not consistently reported in the included trials. Therefore, these data were not available. There were no relevant differences in demographic and preoperative clinical characteristics between the pooled populations from the two study arms. Most patients were male, had a good performance status and a tumor location at the esophagus or gastroesophageal junction. Postoperatively, patients who had undergone neoadjuvant chemotherapy had significantly less often advanced T and N stages and a significantly higher rate of complete resection.

~~The mean follow-up for all patients included in the analysis was 4.6 years.~~OS-time and DFS-time were highly correlated with a Spearman rank correlation coefficient of 0.8943, indicating a good individual-level surrogacy. If follow-up in the investigated trials had been only one, two, three, or four years for DFS, the corresponding correlation coefficients would have been 0.68, 0.77, 0.82 and 0.85.The median follow-up for all patients included in the analysis was 2.10 years (95% confidence interval 1.92-2.29 years). During follow-up, in the neoadjuvant treatment arms 389 patients had a recurrence or died without documented prior recurrence, counting as event in the DFS analysis. 393 patients died, counting as event in the OS analysis. In the surgery alone arms, 419 patients had an event counting in the DFS analysis and 423 patients in the OS analysis. For 361 patients in the neoadjuvant treatment arms and 398 patients in the surgery alone arms, recurrence was documented as death of the patient, i.e.

either no recurrence was diagnosed prior to the death of the patient or recurrence was diagnosed at the time of death. In these cases, DFS and OS were the same.

In figure 1, curves for OS and DFS, calculated according to the landmark method, are presented stratified by treatment arm. Both OS and DFS are longer in patients who had received neoadjuvant treatment compared to those who had undergone surgery alone. OS and DFS curves run largely parallel in patients who had received neoadjuvant treatment as well as in patients who had undergone surgery alone.

Figure 2 shows the Forest plot of hazard ratios for the comparison of treatment effects on OS and DFS between neoadjuvant therapy and surgery alone. There are differences between the absolute values of the hazard ratios from the single RCTs, with the ~~cumulative~~ hazard ratio indicating a survival benefit for ~~perioperative~~neoadjuvant chemotherapy. For ~~all RCTs~~each separate RCT, the point ~~estimates~~estimate of the hazard ~~ratios~~ratio for the treatment effect on OS and DFS ~~are~~is highly similar. The point estimate and confidence interval of the ~~cumulative~~ hazard ratio for the treatment effect on OS ~~are~~is virtually identical to those of the hazard ratio for the treatment effect on DFS.

Results of the error-in-variable regression are shown in figure 3. For one relatively small RCT comparing neoadjuvant chemoradiotherapy (cisplatin/5-fluorouracil with 50.4 Gray) with surgery alone the observed hazard ratio for ~~DFS~~the treatment effect on OS was lower than expected based on the hazard ratio for ~~OS~~ the treatment effect on DFS ~~[20]~~(23). For all other RCTs, the deviation of observed ~~DFS~~OS from what would be expected based on ~~OS~~DFS is small and within the confidence limits. The coefficient of determination r^2 for the association between the HRs for the treatment effects on OS and DFS is 0.912 (95% confidence interval 0.75-1.0), indicating a very good fit of the regression model and thus a strong ~~correlation~~trial-level surrogacy between OS and DFS. The surrogate threshold effect based on the regression analysis was 0.79. This means that a future trial yielding a hazard ratio for the treatment effect on DFS below 0.79 could be expected with a 95% probability to yield a hazard ratio for the treatment effect on OS below one.

In addition to the regression analyses, copula estimation was performed to account for correlation between the treatment effects of the two outcomes. Results using an unadjusted Clayton copula are very similar to those of the regression analysis. The coefficient of determination for the copula analysis is 0.95 while the surrogate threshold effect is 0.74. As convergence could not be achieved, adjustment for the second-step linear regression for measurement-error in the copula model was not performed.

Discussion

The aim of this individual patient data analysis was to assess ~~how strong OS was correlated to,~~individual- and trial-level surrogacy between OS and DFS or, in other words, how predictable OS was by DFS in randomized trials comparing neoadjuvant treatment to surgery alone for gastroesophageal adenocarcinoma. In case of a strong and reliable prediction, DFS could be used as a valid surrogate endpoint, shortening overall trial duration and providing trial results faster.

The results show a considerable correlation between the two outcomes in patients from the included trials, both ~~on~~between the two outcomes themselves (individual ~~trial -~~level surrogacy) and between the treatment effects on the outcomes estimated in the ~~pooled population.~~individual trials (trial-level surrogacy). This observation can partially be explained by the fact that 57% of patients died without prior diagnosis of recurrence, which led to their DFS being identical to their OS. The standard definition of DFS used in oncological trials comprises that deaths without documented prior recurrence are counted as events in DFS analyses ~~[23].~~(31). Thus, this finding is externally valid with regard to other trials. Most patients with disease recurrence, however, died several months or few years after diagnosis of the recurrence. The length of the time interval between recurrence and death is potentially influenced by chemotherapeutic, radiotherapeutic or even surgical treatment. None of the trials provided information on an individual patient level if and what treatment patients received. Therefore, one can only speculate about its possible effects. In general, it might be assumed that patients who underwent surgery alone and are thus chemotherapy-naïve receive more dose-intense chemotherapy than patients who had already undergone ~~preoperative~~neoadjuvant chemotherapy. Likewise, radiotherapy for locoregional recurrence can usually only be administered in patients who had not been treated with neoadjuvant irradiation. However, given the biological complexity of the disease, it cannot be readily concluded that the time between recurrence and death is indeed longer in patients without prior neoadjuvant therapy.

The correct determination of DFS strongly depends on follow-up intervals and the specific kind of clinical, radiological and histopathological examinations carried out in order to detect disease recurrence. These inevitably vary across the RCTs included in our analyses as they were conducted in different settings and during different time periods. However, as visual inspection resulted in no violation of the proportional hazards assumption, and hazard ratios can therefore be assumed to be time-independent, length, intensity and frequency of follow-up will not influence the estimation. OS, on the other hand side, is a very stable indicator because ascertainment of death during follow-up is supposedly very accurate. The consistency of the correlation between DFS and OS across trials does not suggest large differences in DFS determination across the different trials. The choice of the landmark time at six months after randomisation for DFS analyses is somewhat arbitrary. Early therapy-related and post-operative deaths as well as deaths during the early phase of postoperative continuation of chemotherapy might not be mirrored exactly by this approach. On the other hand side, these deaths are all accounted for in the OS analysis. There is no commonly agreed landmark time for such analyses, but six months have been used in an RCT on neoadjuvant chemotherapy for esophageal squamous cell cancer (32) and in our previous meta-analyses (3, 17), which led us to choose the same value for the present analysis.

In the meta-analysis, the ~~cumulative~~overall hazard ratios for the treatment effects on OS and DFS are virtually identical. Likewise, the coefficient of determination in the error-in-variable regression is close to one. This indicates a very good model fit which reflects a strong correlation between treatment effects on OS and DFS~~. The~~ (trial-level surrogacy). Furthermore, the correlation ~~is~~between the two outcomes themselves are was also high, indicating a good individual-level surrogacy. These results are consistent across all different included trials, regardless of the specificities of the applied ~~preoperative~~neoadjuvant chemotherapy and regardless if patients had received combined chemoradiotherapy or merely chemotherapy. Only one small trial using radiotherapy along with cisplatin/5-fluorouracil doublet chemotherapy constituted an outlier, as the observed HR for the treatment effect on DFS was lower than what was expected based on OS in the trial ~~[20]~~(23).

The results indicate that the majority of variations in OS can be explained by the effect of the respective ~~preoperative~~neoadjuvant treatment on DFS. This speaks in favor of DFS serving as appropriate surrogate marker for OS in trials evaluating neoadjuvant chemotherapy or chemoradiotherapy in patients with gastroesophageal adenocarcinoma. As a limitation, none of the included trials used targeted therapy with biologicals or ~~monocolonal~~monoclonal antibodies, or immunotherapy, and therefore deductions regarding these more recent treatment concepts ~~are difficult to make~~(11, 12) are difficult to make. Moreover, histological and molecular tumor subtypes were not consistently reported in the included trials and therefore not available for analysis. Response to therapy depends on these characteristics, and although there is no direct evidence to that regard, the association between DFS and OS might be influenced by the histological or molecular subtype of the tumor.

The calculated surrogate threshold effect of 0.79 means that in future trials, a treatment yielding a reduction in the hazards of disease recurrence of at least 21% can be assumed to have a beneficial effect also on overall survival. Comparable analyses have found a surrogate threshold effect of 0.92 for adjuvant chemotherapy in gastric cancer ~~[11]~~(14) and of 0.56 for chemotherapy in metastatic or irresectable gastric cancer ~~[12]~~(15). There is no clearly established limit above which a surrogate threshold effect would qualify an outcome as appropriate surrogate outcome for another one. However, most randomized controlled trials in oncology are powered to detect effects in the magnitude of a 20%-30% reduction in the hazards of relapse or death. Moreover, based on an individual patient-level analysis of multiple randomized trials by Shi et al, 0.8 is often regarded as a meaningful boundary (33). Therefore, with a surrogate threshold effect of 0.79 in the present analysis, DFS can be regarded as a reasonably appropriate surrogate endpoint for OS. It must be noted, however, that methodologically, the surrogate threshold effect is computed for trials of infinite size. Given that the included RCTs were all of medium size, the surrogate threshold effect might be overestimated.

In summary, based on a strong correlation between DFS and OS on both the individual patient and trial level, as well as on the finding that the vast majority of variation in OS can be explained by variation in DFS, DFS seems to be an appropriate surrogate marker for OS in randomized trials of neoadjuvant chemotherapy or chemoradiotherapy for gastroesophageal adenocarcinoma. However, as novel treatment concepts with substances other than cytotoxic compounds keep evolving, this finding requires continued validation.

Acknowledgements

Parts of this work were supported by the German Federal Ministry of Education and Research (grant number 01 KG 0807~~). The funding source~~) and intramural research funds of the University of Heidelberg (Foundation for Cancer Research). The funding sources had no role in the study design; collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication. In memory of Prof. Christophe Mariette who initiated the study FFCD9901. We thank Emilie Barbier, FFCD Statistician, for her support.

Conflicts of interest

None declared.

Table 1: Randomized controlled trials meeting the inclusion criteria, from which IPD were used in the analysis.

Trial acronym/ first author	Accrual period	Countries	Main inclusion criteria	~~Regimen~~Chemotherapy / chemoradiotherapy regimen
ACCORD 07 (~~22~~25)	1995-2003	France (multi-centre)	adenocarcinoma of lower third of esophagus or GE junction or stomach; UICC stage II or greater; suitable for curative resection; PS 0/1; 18-75 years	~~cisplatin/5-fluorouracil pre- and postoperatively~~2 to 3 cycles (cisplatin 100mg/m² on day 1 or 2; 5-fluorouracil 4000mg/m² cumulative dose over 5 days, then 22 days break) preoperatively; surgery 4 to 6 weeks after last chemotherapy dose, 3 to 4 cycles (see above) postop. 4 to 6 weeks after surgery for patients who had R0 resection, no progression or major toxicity during preop. therapy and at least T3 or N+ tumor in histopathology
CALGB 9781 (~~20~~23)	1997-2000	USA (multi-center)	squamous cell or adenocarcinoma of thoracic esophagus or GE junction, resectable (T1-3, Nx), including regional thoracic lymph node (N1) metastases, supraclavicular lymph node metastasis <1.5cm, lymph node metastases to levels 15-20 <1.5cm; no age limit	~~cisplatin, 5-fluorouracil with concurrent radiotherapy preoperatively~~1 cycle (cisplatin 200 mg/m² cumulative dose on days 1 and 29, 5-fluorouracil 8000 mg/m² cumulative dose on days 1 to 4 and 29 to 32, radiotherapy (1.8 Gy/5 d/wk) begun within 24 hours of the chemotherapy administration, continued for 5.5 weeks, final 5.4 Gy given as a boost (total dose 50.4 Gy)
EORTC 40954 (~~19~~22)	1999-2004	several European countries, Egypt (multi-centre)	adenocarcinoma of stomach or GE junction, cT3/4 Nx M0/M1(lymph); PS 0-1; 18-70 years	~~cisplatin, 5-Fluorouracil, folinic acid preoperatively~~1 cycle (cisplatin 150 mg/m² cumulative dose on days 1, 15 and 29; 5-fluorouracil 12,000 mg/m² cumulative dose on days 1, 8, 15, 22, 29 and 36; folinic acid 3000 mg/m² cumulative dose on days 1, 8, 15, 22, 29 and 36); re-

				staging, if no progression or toxicity 1 more cycle as described above restarting on day 50; surgery on days 57 to 63 of the second cycle
FAMTX (~~17~~20)	1993-1996	Netherlands (multi-centre)	adenocarcinoma of the stomach (not cardia); >cT1; resectable with no evidence of distant metastases; PS 0-2; <75 years	~~5-fluorouracil, leucovorin, doxorubicin, methotrexate preoperatively~~2 cycles (methotrexate 1500 mg/m² on day 2; 5-fluorouracil 1500 mg/m² on day 2; leucovorin 240 or 480 mg (depending on MTX level) cumulative dose on days 3 to 4; doxorubicin 30 mg/m² on day 15; 13 days break); re-staging; in case of response or stable disease another 2 cycles (see above);
FFCD 9901 (~~15~~18)	2000-2009	France (multi-centre)	thoracic esophageal adenocarcinoma or squamous cell carcinoma; suitable for curative resection; cT1/2N0/1 or cT3N0; PS 0-1; <75 years	~~cisplatin, 5-Fluorouracil, with concurrent radiotherapy preoperatively~~2 cycles (fluorouracil and cisplatin. FU 800 mg/m2 per 24 hours was administered as a continuous infusion from days 1 to 4 and 29 to 32. Cisplatin 75 mg/m^2 on day 1 or 2 and day 29 or 30 or, alternatively, 15 mg/m^2 from days 1 to 5 and 29 to 33), concomitant radiotherapy (45 Gy five fractions per week over 5 weeks).
RTOG 8911 (~~18~~21)	1990-1995	USA (multi-center)	squamous cell or adenocarcinoma of thoracic esophagus or GE junction; stage I-III excluding T4 tumors; absence of supraclavicular or distant metastases; fit for surgery; at least 18 years;	~~cisplatin, 5-fluorouracil pre- and postoperatively~~ 3 cycles (cisplatin 100 mg/m² on day 1; 5-fluorouracil 1000 mg/m² cumulative dose on days 1 to 5, 23 days break); operation 2 to 4 weeks after the end of the last cycle; in case of stable or responsive disease upon surgery 2 postoperative cycles (see above, except cisplatin dose re-

				duced to 75 mg/m²) starting 2 to 6 weeks after surgery
TROG-AGITG (~~16~~19)	1994-2000	Australia, New Zealand, Singapore (multi-center)	invasive cancer of thoracic oesophagus; cT1-3 cN0-1; no involvement of cervical esophagus or celiac nodes; PS 0 or 1; no age limit	~~cisplatin, 5-fluorouracil with concurrent radiotherapy preoperatively~~1 cycle (cisplatin 80 mg/m² on day 1; 5-fluorouracil 3200 mg/m² cumulative dose on days 1 to 4) with 35 Gy radiotherapy in 15 fractions over 3 weeks, starting concurrently with chemotherapy; surgery 3 to 6 weeks after completion of radiotherapy; postoperative radiotherapy permitted for patients with residual disease after surgery if indicated clinically for patients assigned to surgery alone
Urba (~~21~~24)	1989-1994	USA (single-center)	squamous cell, adenocarcinoma or mixed adenosquamous carcinoma of esophagus or GE junction, limited to esophagus and regional lymph nodes (including celiac nodes); Karnofsky index >=60%; ≤75 years	~~cisplatin, 5-fluorouracil, vinblastine with concurrent radiotherapy preoperatively~~ 1 cycle (cisplatin 200 mg/m² cumulative dose on days 1 through 5 and 17 through 21, 5-fluorouracil 6300 mg/m² cumulative on days 1 through 21, vinblastin 8 mg/m² on days 1 through 4 and 17 through 20, radiotherapy in fractions of 1.5 Gy twice a day, on days 1 through 5, 8 through 12, and 15 through 19, to a total dose of 45 Gy)

PS: performance status (ECOG/WHO)

Table 2: Demographic and clinical characteristics of patients included in the analysis, both for the entire study population and separately for patients from the neoadjuvant therapy and surgery alone arms

	~~Perioperative~~Neoadjuvant Chemotherapy N=562	Surgery alone N=564	Total N=1126	p-value
Trial				
- ACCORD	113 (20.1%)	111 (19.7%)	224 (19.9%)	0.998
- CALGB	23 (4.1%)	19 (3.4%)	42 (3.7%)	~~0.998~~
- EORTC	69 (12.3%)	71 (12.6%)	140 (12.4%)	
- FAMTX	27 (4.8%)	29 (5.1%)	56 (5.0%)	
- FFCD	98 (17.4%)	97 (17.2%)	195 (17.3%)	
- RTOG	115 (20.5%)	121 (21.5%)	236 (21.0%)	
- TROG-AGITG	80 (14.2%)	78 (13.8%)	158 (14.0%)	
~~- FAMTX~~	~~27 (4.8%)~~	~~29 (5.1%)~~	~~56 (5.0%)~~	
~~- RTOG~~	~~115 (20.5%)~~	~~121 (21.5%)~~	~~236 (21.0%)~~	
~~- EORTC~~	~~69 (12.3%)~~	~~71 (12.6%)~~	~~140 (12.4%)~~	
- URBA	37 (6.6%)	38 (6.7%)	75 (6.7%)	
~~- ACCORD~~	~~113 (20.1%)~~	~~111 (19.7%)~~	~~224 (19.9%)~~	
~~- FFCD~~	~~98 (17.4%)~~	~~97 (17.2%)~~	~~195 (17.3%)~~	
Gender				
- male	483 (85.9%)	467 (82.8%)	950 (84.4%)	0.147
- female	79 (14.1%)	97 (17.2%)	176 (15.6%)	
Age [years]				
- N	562	564	1126	0.908
- Mean +/- SD	59.8 +/-9.3	59.6 +/-9.4	59.7 +/-9.3	
- p5, p25, p75, p95	44.0, 53.2, 67.0, 73.2	43.0, 53.3, 67.0, 73.0	44.0, 53.2, 67.0, 73.1	
- Median	60.8	61.0	61.0	
- Min, Max	23.0, 78.0	26.1, 80.5	23.0, 80.5	
Age				
- < 65 years	366 (65.1%)	377 (66.8%)	743 (66.0%)	0.827
- 65 - 75 years	184 (32.7%)	176 (31.2%)	360 (32.0%)	
- > 75 years	12 (2.1%)	11 (2.0%)	23 (2.0%)	
Tumor location				
- Stomach	88 (15.7%)	89 (15.8%)	177 (15.7%)	0.984
- GE junction	153 (27.2%)	158 (28.0%)	311 (27.6%)	
- Esophagus	261 (46.4%)	260 (46.1%)	521 (46.3%)	
- Esophagus / GE junction (no further specification)	60 (10.7%)	57 (10.1%)	117 (10.4%)	
Performance status				
- 0	373 (71.5%)	365 (71.3%)	738 (71.4%)	0.263
- 1	144 (27.6%)	146 (28.5%)	290 (28.0%)	
- 2	5 (1.0%)	1 (0.2%)	6 (0.6%)	

	~~Perioperative~~Neoadjuvant Chemotherapy N=562	Surgery alone N=564	Total N=1126	p-value
- missing	40	52	92	
T stage [preoperative, clinical]				
- T0	1 (0.5%)	0 (0.0%)	1 (0.2%)	0.867
- T1	22 (10.7%)	16 (7.7%)	38 (9.2%)	
- T2	57 (27.7%)	56 (27.0%)	113 (27.4%)	
- T3	121 (58.7%)	130 (62.8%)	251 (60.8%)	
- T4	5 (2.4%)	5 (2.4%)	10 (2.4%)	
- missing	357	361	718	
N stage [preoperative, clinical]				
- N0	61 (80,8%)	47 (63.5%)	108 (70.1%)	0.178
- N1	18 (22.5%)	27 (36.5%)	45 (29.2%)	
- N2	1 (1.3%)	0 (0.0%)	1 (0.6%)	
- missing	482	490	972	
T stage [postoperative, histopathological]				
- T0	53 (13.2%)	2 (0.5%)	55 (6.7%)	<.001
- T1	63 (15.7%)	64 (15.2%)	127 (15.5%)	
- T2	112 (27.9%)	106 (25.2%)	218 (30.2%)	
- T3	156 (38.9%)	207 (49.2%)	363 (50.3%)	
- T4	17 (4.2%)	42 (10.0%)	59 (8.2%)	
- missing	161	143	304	
N stage [postoperative, histopathological]				
- N0	181 (45.3%)	110 (26.4%)	291 (35.6%)	<.001
- N1	171 (42.8%)	210 (50.4%)	381 (46.6%)	
- N2	35 (8.8%)	59 (14.1%)	94 (11.5%)	
- N3	13 (3.3%)	38 (9.1%)	51 (6.2%)	
- missing	162	147	309	
Margin status				
- R0	395 (91.2%)	374 (82.3%)	769 (86.7%)	0.001
- R1	18 (4.2%)	35 (7.4%)	53 (6.0%)	
- R2	20 (4.6%)	45 (9.5%)	65 (7.3%)	
- missing	129	110	239	

Figure 1: Time-to-event curves for OS and DFS, calculated according to the landmark method, stratified by treatment arm. Treat_arm=0 OS: overall survival in the upfront surgery arms, treat_arm=1 OS: overall survival in the neoadjuvant therapy arms, treat_arm=0 DFS: disease-free survival in the up-front surgery arms, treat_arm=1 DFS: disease-free survival in the neoadjuvant therapy arms.

Figure 2: Forest plot of hazard ratios for the comparison of OS and DFS between neoadjuvant therapy and surgery alone

Figure 3: Error-in-variable regression, observed and predicted HRs for OS and DFS with 95% prediction limits. RT-CT: neoadjuvant chemoradiotherapy, CT: neoadjuvant chemotherapy.

References

1. Rubenstein JH, Shaheen NJ. Epidemiology, Diagnosis, and Management of Esophageal Adenocarcinoma. Gastroenterology. 2015;149(2):302-17.e1.

2. Jim MA, Pinheiro PS, Carreira H, Espey DK, Wiggins CL, Weir HK. Stomach cancer survival in the United States by race and stage (2001-2009): Findings from the CONCORD-2 study. Cancer. 2017;123 Suppl 24:4994-5013.

3. Ronellenfitsch U, Schwarzbach M, Hofheinz R, Kienle P, Kieser M, Slanger TE, et al. Preoperative chemo(radio)therapy versus primary surgery for gastroesophageal adenocarcinoma: systematic review with meta-analysis combining individual patient and aggregate data. European journal of cancer~~.~~ (Oxford, England : 1990). 2013;49(15):3149-58.

4. van Hagen P, Hulshof MC, van Lanschot JJ, Steyerberg EW, van Berge Henegouwen MI, Wijnhoven BP, et al. Preoperative chemoradiotherapy for esophageal or junctional cancer. The New England journal of medicine. 2012;366(22):2074-84.

5. Schirren R, Reim D, Novotny AR. Adjuvant and/or neoadjuvant therapy for gastric cancer? A perspective review. Therapeutic advances in medical oncology. 2015;7(1):39-48.

6. Egyud MR, Tseng JF, Suzuki K. Multidisciplinary Therapy of Esophageal Cancer. The Surgical clinics of North America. 2019;99(3):419-37.

7. Al-Batran SE, Homann N, Pauligk C, Goetze TO, Meiler J, Kasper S, et al. Perioperative chemotherapy with fluorouracil plus leucovorin, oxaliplatin, and docetaxel versus fluorouracil or capecitabine plus cisplatin and epirubicin for locally advanced, resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4): a randomised, phase 2/3 trial. Lancet (London, England). 2019;393(10184):1948-57.

8. Lordick F, Mariette C, Haustermans K, Obermannova R, Arnold D. Oesophageal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of oncology : official journal of the European Society for Medical Oncology. 2016;27(suppl 5):v50-v7.

~~8~~9. Smyth EC, Verheij M, Allum W, Cunningham D, Cervantes A, Arnold D. Gastric cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.

Annals of oncology : official journal of the European Society for Medical Oncology. 2016;27(suppl 5):v38-v49.

~~9~~10. Cats A, Jansen EPM, van Grieken NCT, Sikorska K, Lind P, Nordsmark M, et al. Chemotherapy versus chemoradiotherapy after surgery and preoperative chemotherapy for resectable gastric cancer (CRITICS): an international, open-label, randomised phase 3 trial. The Lancet Oncology. 2018;19(5):616-28.

11. Wagner AD, Grabsch HI, Mauer M, Marreaud S, Caballero C, Thuss-Patience P, et al. EORTC-1203-GITCG - the "INNOVATION"-trial: Effect of chemotherapy alone versus chemotherapy plus trastuzumab, versus chemotherapy plus trastuzumab plus pertuzumab, in the perioperative treatment of HER2 positive, gastric and gastroesophageal junction adenocarcinoma on pathologic response rate: a randomized phase II-intergroup trial of the EORTC-Gastrointestinal Tract Cancer Group, Korean Cancer Study Group and Dutch Upper GI-Cancer group. BMC cancer. 2019;19(1):494.

12. Maron SB, Catenacci DVT. Update on Gastroesophageal Adenocarcinoma Targeted Therapies. Hematology/oncology clinics of North America. 2017;31(3):511-27.

~~10~~13. Mathoulin-Pelissier S, Gourgou-Bourgade S, Bonnetain F, Kramar A. Survival end point reporting in randomized cancer clinical trials: a review of major journals. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2008;26(22):3721-6.

~~11~~14. Oba K, Paoletti X, Alberts S, Bang YJ, Benedetti J, Bleiberg H, et al. Disease-free survival as a surrogate for overall survival in adjuvant trials of gastric cancer: a meta-analysis. Journal of the National Cancer Institute. 2013;105(21):1600-7.

~~12~~15. Paoletti X, Oba K, Bang YJ, Bleiberg H, Boku N, Bouche O, et al. Progression-free survival as a surrogate for overall survival in advanced/recurrent gastric cancer trials: a meta-analysis. Journal of the National Cancer Institute. 2013;105(21):1667-70.

~~13~~16. Ronellenfitsch U, Schwarzbach M, Hofheinz R, Kienle P, Kieser M, Slanger TE, et al. Perioperative chemo(radio)therapy versus primary surgery for resectable adenocarcinoma of the stomach, gastroesophageal junction, and lower esophagus. The Cochrane database of systematic reviews. 2013(5):Cd008107.

~~14~~17. Ronellenfitsch U, Schwarzbach M, Hofheinz R, Kienle P, Nowak K, Kieser M, et al. Predictors of overall and recurrence-free survival after neoadjuvant chemotherapy for

gastroesophageal adenocarcinoma: Pooled analysis of individual patient data (IPD) from randomized controlled trials (RCTs). European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 2017;43(8):1550-8.

~~15~~18. Mariette C, Dahan L, Mornex F, Maillard E, Thomas PA, Meunier B, et al. Surgery alone versus chemoradiotherapy followed by surgery for stage I and II esophageal cancer: final analysis of randomized controlled phase III trial FFCD 9901. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2014;32(23):2416-22.

~~16~~19. Burmeister BH, Smithers BM, Gebski V, Fitzgerald L, Simes RJ, Devitt P, et al. Surgery alone versus chemoradiotherapy followed by surgery for resectable cancer of the oesophagus: a randomised controlled phase III trial. The Lancet Oncology. 2005;6(9):659-68.

~~17~~20. Hartgrink HH, van de Velde CJ, Putter H, Songun I, Tesselaar ME, Kranenbarg EK, et al. Neo-adjuvant chemotherapy for operable gastric cancer: long term results of the Dutch randomised FAMTX trial. European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 2004;30(6):643-9.

~~18~~21. Kelsen DP, Winter KA, Gunderson LL, Mortimer J, Estes NC, Haller DG, et al. Long-term results of RTOG trial 8911 (USA Intergroup 113): a random assignment trial comparison of chemotherapy followed by surgery compared with surgery alone for esophageal cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2007;25(24):3719-25.

~~19~~22. Schuhmacher C, Gretschel S, Lordick F, Reichardt P, Hohenberger W, Eisenberger CF, et al. Neoadjuvant chemotherapy compared with surgery alone for locally advanced cancer of the stomach and cardia: European Organisation for Research and Treatment of Cancer randomized trial 40954. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2010;28(35):5210-8.

~~20~~23. Tepper J, Krasna MJ, Niedzwiecki D, Hollis D, Reed CE, Goldberg R, et al. Phase III trial of trimodality therapy with cisplatin, fluorouracil, radiotherapy, and surgery compared with surgery alone for esophageal cancer: CALGB 9781. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2008;26(7):1086-92.

~~21~~24. Urba SG, Orringer MB, Turrisi A, Iannettoni M, Forastiere A, Strawderman M. Randomized trial of preoperative chemoradiation versus surgery alone in patients with locoregional esophageal carcinoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2001;19(2):305-13.

~~22~~25. Ychou M, Boige V, Pignon JP, Conroy T, Bouche O, Lebreton G, et al. Perioperative chemotherapy compared with surgery alone for resectable gastroesophageal adenocarcinoma: an FNCLCC and FFCD multicenter phase III trial. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2011;29(13):1715-21.

~~23~~26. Draper N, Smith, H. Applied regression analysis, third edition. New York, NY: John Wiley & Sons; 1998.

27. Kmenta J. Elements of econometrics: Second edition. Ann Arbor, MI: University of Michigan Press; 1997.

28. Treiman D. Quantitative Data Analysis: Doing Social Research to Test Ideas. San Francisco, CA: Jossey-Bass; 2008.

29. Buyse M, Burzykowski T, Michiels S, Carroll K. Individual- and trial-level surrogacy in colorectal cancer. Statistical methods in medical research. 2008;17(5):467-75.

30. Rotolo F, Paoletti X, Michiels S. surrosurv: An R package for the evaluation of failure time surrogate endpoints in individual patient data meta-analyses of randomized clinical trials. Computer methods and programs in biomedicine. 2018;155:189-98.

31. Panageas KS, Knezevic A. Selection of endpoints. In: Halabi S, Michiels S, editors. Textbook of Clinical Trials in Oncology: A Statistical Perspective. Boca Raton, FL: CRC Press; 2019.

32. Boonstra JJ, Kok TC, Wijnhoven BP, van Heijl M, van Berge Henegouwen MI, Ten Kate FJ, et al. Chemotherapy followed by surgery versus surgery alone in patients with resectable oesophageal squamous cell carcinoma: long-term results of a randomized controlled trial. BMC cancer. 2011;11:181.

33. Shi Q, Schmitz N, Ou FS, Dixon JG, Cunningham D, Pfreundschuh M, et al. Progression-Free Survival as a Surrogate End Point for Overall Survival in First-Line Diffuse Large B-Cell Lymphoma: An Individual Patient-Level Analysis of Multiple Randomized Trials (SEAL). Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2018;36(25):2593-602.

Disease-free survival as a surrogate for overall survival in neoadjuvant trials of gastroesophageal adenocarcinoma: pooled analysis of individual patient data from randomized controlled trials

Ulrich Ronellenfitsch[a], Katrin Jensen[b], Svenja Seide[b], Meinhard Kieser[b], Matthias Schwarzbach[c], Tracy E Slanger[d], Bryan Burmeister[e], David Kelsen[f], Donna Niedzwiecki[g], Guillaume Piessen[h], Christoph Schuhmacher[i], Susan Urba[j], Cornelis van de Velde[k], Marc Ychou[l], Ralf Hofheinz[m], Sylvie Lorenzen[n]

[a]Medical Faculty of the Martin Luther University Halle-Wittenberg and University Hospital Halle (Saale), Department of Visceral, Vascular and Endocrine Surgery, Ernst-Grube-Str. 40, 06120 Halle (Saale), Germany, ulrich.ronellenfitsch@uk-halle.de

[b]Institute of Medical Biometry and Informatics, University of Heidelberg, Im Neuenheimer Feld 130.3, 69120 Heidelberg, Germany, jensen@imbi.uni-heidelberg.de, meinhard.kieser@imbi.uni-heidelberg.de, seide@imbi.uni-heidelberg.de

[c]Department of General, Visceral, Vascular, and Thoracic Surgery, Klinikum Frankfurt Höchst, Gotenstraße 6-8, 65929 Frankfurt am Main, Germany matthias.schwarzbach@klinikumfrankfurt.de

[d]Institute for Quality and Efficiency in Health Care (IQWiG), Im Mediapark 8, 50670 Cologne, Germany, t.slanger@hotmail.com

[e]University of Queensland, Princess Alexandra Hospital, Brisbane, Australia, Princess Alexandra Hospital, Brisbane, QLD 4102, Australia, bryan.burmeister@health.qld.gov.au

[f]Memorial Sloan-Kettering Cancer Center and Weill Cornell Medical College, 1275 York Avenue, New York, NY 10065, USA, kelsend@mskcc.org

[g]The Alliance for Clinical Trials in Oncology (Alliance) Statistics and Data Center, Duke University Medical Center, Hock Plaza, 2424 Erwin Rd, Room 8040, Durham, NC 27705, USA, donna.niedzwiecki@duke.edu

[h]Department of Digestive and Oncological Surgery, University Hospital C. Huriez, Inserm UMR-S 1172, Jean-Pierre Aubert Research Center (JPARC) Team, 1, rue Polonovski, 59037 Lille cedex, France, guillaume.piessen@chru-lille.fr

[i]Department of Surgery, Klinikum rechts der Isar, Technical University Munich, Ismaninger Str. 22, 81675 Munich, Germany, christoph.schuhmacher@tum.de

[j]Division of Hematology/Oncology, University of Michigan Medical Center, 1500 E Medical Center Drive, C347, SPC 5848 Ann Arbor, MI 48109, USA, surba@med.umich.edu

[k]Department of Surgery, Leiden University Medical Center, K6-R, P.O. Box 9600, 2300 RC Leiden, The Netherlands, c.j.h.van_de_velde@lumc.nl

[l]Centre Régional de Lutte Contre le Cancer, Val d'Aurelle, Montpellier Cedex 05, France, marc.ychou@icm.unicancer.fr

[m]Day Treatment Center (TTZ), Interdisciplinary Tumor Center Mannheim (ITM) & 3rd Department of Medicine, University Medical Centre Mannheim, Medical Faculty Mannheim of the University of Heidelberg, Theodor-Kutzer-Ufer 1-3, 68167 Mannheim, Germany, ralf.hofheinz@umm.de

[n]3rd Department of Internal Medicine (Hematology/Medical Oncology), Klinikum rechts der Isar, Technische Universität München, Ismaninger Str. 22, 81675 Munich, Germany, sylvielorenzen@gmx.de

Corresponding author:

Ulrich Ronellenfitsch, MD

Medical Faculty of the Martin Luther University Halle-Wittenberg and

University Hospital Halle (Saale)

Department of Visceral, Vascular and Endocrine Surgery

Ernst-Grube-Str. 40

06120 Halle (Saale), Germany

phone: +49-345-557-2327

fax: +49-345-557-2823

e-mail: ulrich.ronellenfitsch@uk-halle.de

Parts of this work were supported by the German Federal Ministry of Education and Research (grant number 01 KG 0807) and intramural research funds of the University of Heidelberg (Foundation for Cancer Research).

Abstract

Introduction

Disease-free survival (DFS) is increasingly being used as surrogate endpoint for overall survival (OS) in cancer trials. So far, there has been no validation of the surrogacy of DFS for OS for neoadjuvant treatment of gastroesophageal adenocarcinoma.

Methods

The study uses individual patient data (IPD) from eight randomized controlled trials (n=1,126 patients) comparing neoadjuvant therapy followed by surgery with surgery alone for gastroesophageal adenocarcinoma. Correlation between OS-time and DFS-time was calculated to evaluate individual-level surrogacy. For each trial, survival curves using the Kaplan-Meier method were plotted and hazard ratios (HRs) on the treatment effects were calculated for OS and DFS separately. Those HRs were pooled in a random-effects meta-analysis. Observed were compared with predicted HRs for OS using results from an error-in-variables linear regression model accounting for the uncertainty about the estimated effect. The strength of the association was quantified by the coefficient of determination to assess trial-level surrogacy. The surrogate threshold effect was calculated, to determine the minimum treatment effect on DFS necessary to predict a non-zero treatment effect on OS.

Results

A strong correlation between OS-time and DFS-time was observed, indicating a high individual-level surrogacy. For all RCTs, estimated HRs for OS and DFS were highly similar. In the meta-analysis, the overall HR for OS was virtually identical to that for DFS. The estimated coefficient of determination r^2 for the association between HRs for OS and DFS was 0.912 (95% confidence interval 0.75-1.0), indicating a very good fit of the regression model and thus a strong trial-level surrogacy between OS and DFS. The surrogate threshold effect based on the regression analysis was 0.79.

Discussion

Based on strong correlations between DFS and OS, as well as a strong correlation of the treatment effects of the two endpoints in the error-in-variable regression, DFS seems an appropriate surrogate marker for OS in randomized trials of neoadjuvant chemotherapy or chemoradiotherapy for gastroesophageal adenocarcinoma.

Keywords

Gastroesophageal adenocarcinoma; neoadjuvant chemotherapy, neoadjuvant chemoradiotherapy, overall survival, disease-free-survival, individual patient data meta-analysis

Introduction

The majority of patients with gastroesophageal adenocarcinoma, i.e. adenocarcinoma of the esophagus, gastroesophageal junction or stomach, present with advanced disease [(1, 2). In the absence of distant metastasis, oncological resection is the only potentially curative modality. However, the prognosis after surgery alone in locally advanced disease is poor with 5-year survival rates of only about 20-30%. A considerable proportion of patients who undergo upfront resection will eventually relapse and die as a result of their disease. Multimodal treatment concepts comprising chemotherapy or chemoradiotherapy have shown survival benefits (3-7). Because of difficulties in administering chemotherapy or radiotherapy soon after surgical procedures, major efforts have been undertaken to explore different neoadjuvant treatment strategies to improve outcomes. Several randomized trials and meta-analyses have demonstrated that neoadjuvant chemotherapy and neoadjuvant chemoradiotherapy followed by surgery are associated with longer survival compared to surgery alone (3, 4). Therefore, both strategies are now recommended by guidelines for the treatment of non-metastatic gastroesophageal adenocarcinoma (8, 9). Despite this approach the rate of recurrence remains high, and new neoadjuvant treatment concepts including postoperative chemoradiation after neoadjuvant chemotherapy like in the CRITICS trial (10), biologicals such as dual HER2 targeting in the INNOVATION trial (11), or immunotherapy are being explored (12).

The conduction of trials assessing such new treatments is often difficult, as they require a large number of patients in order to detect the relatively small incremental benefits of a new treatment. Overall survival (OS) has been considered the gold standard endpoint for cancer clinical trials. However, OS requires an extended follow-up period and thus trial duration, which leads to higher costs and a long delay until results become available. This has resulted in an increasing number of cancer trials using actuarial endpoints, which can be ascertained sooner, such as disease-free survival (DFS) (13). In non-metastatic gastroesophageal adenocarcinoma, a large proportion of relapses occur before 3 years (3). However, post-

recurrence treatment could dilute or even eliminate an apparent improvement in tumor control achieved by neoadjuvant treatment. Therefore, the validity of DFS as surrogate endpoint for OS in gastroesophageal adenocarcinoma remains controversial.

A meta-analysis concluded that DFS is an appropriate endpoint for OS in studies of gastric cancer in the adjuvant setting, showing that the effect of treatment on OS is largely predictable from its effect on DFS (14), whereas another analysis found this not to be true in the palliative setting (15). However, most evidence from large-scale randomized trials in resectable gastroesophageal adenocarcinoma exists for neoadjuvant treatment; which has now, as opposed to adjuvant treatment, become standard of care for non-metastatic disease. Therefore, the present analysis assesses if DFS is a valid surrogate endpoint for OS in trials using neoadjuvant chemotherapy or chemoradiotherapy for gastroesophageal adenocarcinoma.

Material and methods

Trial and patient selection

We used individual patient data (IPD) from randomized controlled trials (RCTs) comparing neoadjuvant chemotherapy or chemoradiotherapy (i.e. therapy administered at least partially prior to surgery) with surgery alone for gastroesophageal adenocarcinoma. All RCTs including patients with resectable, non-metastatic tumors without prior treatment and providing information on both OS and DFS were potentially eligible. There were no exclusion criteria regarding specific treatment regimens. Trials were identified by a systematic literature review covering publications until 2011, the details of which have been previously published (3, 16). All trialists from eligible trials were solicited to provide IPD, and trials were only included in the analyses in case of an affirmative response. Upon data collection, trialists had been asked to provide most recent follow-up data, even if follow-up was longer than that reported in respective publications. Some of the IPD had already been used in a previous meta-analysis comparing treatment effects of neoadjuvant chemotherapy with surgery alone and in a secondary analysis exploring predictors of postoperative survival (3, 17). From two eligible RCTs (4, 18), the final results were only published after completion of that meta-analysis. One of these trials (18) provided IPD for the present analysis. This resulted in IPD from eight trials (18-25), which comprise 1,126 patients, entering the analysis (table 1). All included RCTs had been approved by the respective competent ethical committee.

Definition of outcomes

Overall survival (OS) was defined as time from randomization to death or to the last documented follow-up. Disease-free survival (DFS) was defined as time from a landmark six months after randomization to recurrence or death, whichever occurred first, or to the last documented follow-up. The purpose of this landmark was to account for differences in timing between randomization and surgery between trial arms. Recurrence and death within the first six months were considered events at the landmark.

Statistical analyses

Characteristics of patients were compared between groups using the chi-square test for discrete variables, and the Wilcoxon-Mann-Whitney test for continuous variables. Correlation between OS-time and DFS-time was assessed by means of the Spearman rank correlation coefficient to assess individual-level surrogacy. OS and DFS were calculated according to the Kaplan-Meier method separately for patients who received neoadjuvant therapy and patients who underwent surgery alone. This was done in the entire study population and for patients from each single RCT. Survival in the single strata was compared using the log-rank test. Hazard ratios with 95% confidence intervals were calculated for the comparison of neoadjuvant therapy and surgery alone for each RCT for the treatment effect on both OS and DFS. These hazard ratios were pooled in two separate meta-analyses to provide a combined effect of the estimated hazard ratios. Random-effects models were used for calculation of point estimates and confidence intervals because heterogeneity between the 'true' effects of the different regimens (neoadjuvant chemoradiotherapy or neoadjuvant chemotherapy, different combinations of chemotherapeutical agents etc.) used in the trials was assumed. Additionally, all results were investigated for statistical heterogeneity by I^2 statistics, without using this measure to choose between meta-analytic models.

To compare the observed with the predicted hazard ratio for the treatment effect on OS, a linear regression model accounting for the uncertainty about the estimated effects was used. There, the treatment effects on DFS were included as predictors in an error-in-variables linear regression model with 95% prediction limits to predict the treatment effects on OS. The strength of the association was quantified by the coefficient of determination r^2 to assess trial-level surrogacy. Considering that the estimated treatment effects from the individual trials on DFS will include a measurement error, we added an additive measurement error in the observed variable (26-29). The standard error of R^2 was bootstrapped (1000 repetitions) and

the 95% confidence interval was then calculated using the bootstrapped standard error and quantiles based on the student's' t-distribution.

In order to better evaluate the median follow-up requested for future trials using DFS as primary endpoint, the correlation coefficient between DFS-time and OS-time at varying time points (one, two, three, and fourin years of follow-up) was additionally calculated. To account for the repeated use of data stemming from the same patients for the two endpoints, we used the copula approach as described in Rotolo et al. as sensitivity analysis (30).

All significance tests were two-sided with p=0.05 as cutoff. IPD were analysed using SAS 9.4 (SAS Institute Inc.). Meta-analyses were conducted with Stata 14.2 (Stata Corp.) and the copula estimation was performed in R version 3.5.1 using the extension surrosurv (30).

Results

Table 1 provides an overview of the characteristics of the eight included RCTs. All but one trial were multi-center trials. Three trials were carried out in the USA, two in France, one in the Netherlands, and three in several countries in Europe, North Africa or Australasia. Four RCTs comprised a neoadjuvant chemoradiotherapy scheme and four a neoadjuvant chemotherapy scheme in their experimental arm. All trials used 5-fluorouracil, and seven out of eight trials cisplatin as chemotherapeutical backbone. Table 2 shows demographic and clinical characteristics of the 1,126 patients included in the analysis, both for the entire study population and separately for patients from the neoadjuvant therapy and surgery alone arms. Histological and molecular tumor subtypes were not consistently reported in the included trials. Therefore, these data were not available. There were no relevant differences in demographic and preoperative clinical characteristics between the pooled populations from the two study arms. Most patients were male, had a good performance status and a tumor location at the esophagus or gastroesophageal junction. Postoperatively, patients who had undergone neoadjuvant chemotherapy had significantly less often advanced T and N stages and a significantly higher rate of complete resection.

OS-time and DFS-time were highly correlated with a Spearman rank correlation coefficient of 0.8943, indicating a good individual-level surrogacy. If follow-up in the investigated trials had been only one, two, three, or four years for DFS, the corresponding correlation coefficients would have been 0.68, 0.77, 0.82 and 0.85.The median follow-up for all patients included in the analysis was 2.10 years (95% confidence interval 1.92-2.29 years). During follow-up, in the neoadjuvant treatment arms 389 patients had a recurrence or died without documented prior recurrence, counting as event in the DFS analysis. 393 patients died, counting as event in the OS analysis. In the surgery alone arms, 419 patients had an event counting in the DFS analysis and 423 patients in the OS analysis. For 361 patients in the neoadjuvant treatment arms and 398 patients in the surgery alone arms, recurrence was documented as death of the patient, i.e. either no recurrence was diagnosed prior to the death of the patient or recurrence was diagnosed at the time of death. In these cases, DFS and OS were the same.

In figure 1, curves for OS and DFS, calculated according to the landmark method, are presented stratified by treatment arm. Both OS and DFS are longer in patients who had received neoadjuvant treatment compared to those who had undergone surgery alone. OS and DFS curves run largely parallel in patients who had received neoadjuvant treatment as well as in patients who had undergone surgery alone.

Figure 2 shows the Forest plot of hazard ratios for the comparison of treatment effects on OS and DFS between neoadjuvant therapy and surgery alone. There are differences between the absolute values of the hazard ratios from the single RCTs, with the hazard ratio indicating a survival benefit for neoadjuvant chemotherapy. For each separate RCT, the point estimate of the hazard ratio for the treatment effect on OS and DFS is highly similar. The point estimate and confidence interval of the hazard ratio for the treatment effect on OS is virtually identical to those of the hazard ratio for the treatment effect on DFS.

Results of the error-in-variable regression are shown in figure 3. For one relatively small RCT comparing neoadjuvant chemoradiotherapy (cisplatin/5-fluorouracil with 50.4 Gray) with surgery alone the observed hazard ratio for the treatment effect on OS was lower than expected based on the hazard ratio for the treatment effect on DFS (23). For all other RCTs, the deviation of observed OS from what would be expected based on DFS is small and within the confidence limits. The coefficient of determination r^2 for the association between the HRs for the treatment effects on OS and DFS is 0.912 (95% confidence interval 0.75-1.0), indicating a very good fit of the regression model and thus a strong trial-level surrogacy between OS and DFS. The surrogate threshold effect based on the regression analysis was 0.79. This means that a future trial yielding a hazard ratio for the treatment effect on DFS below 0.79 could be expected with a 95% probability to yield a hazard ratio for the treatment effect on OS below one.

In addition to the regression analyses, copula estimation was performed to account for correlation between the treatment effects of the two outcomes. Results using an unadjusted Clayton copula are very similar to those of the regression analysis. The coefficient of determina-

tion for the copula analysis is 0.95 while the surrogate threshold effect is 0.74. As convergence could not be achieved, adjustment for the second-step linear regression for measurement-error in the copula model was not performed.

Discussion

The aim of this individual patient data analysis was to assess individual- and trial-level surrogacy between OS and DFS or, in other words, how predictable OS was by DFS in randomized trials comparing neoadjuvant treatment to surgery alone for gastroesophageal adenocarcinoma. In case of a strong and reliable prediction, DFS could be used as a valid surrogate endpoint, shortening overall trial duration and providing trial results faster.

The results show a considerable correlation between the two outcomes in patients from the included trials, both between the two outcomes themselves (individual-level surrogacy) and between the treatment effects on the outcomes estimated in the individual trials (trial-level surrogacy). This observation can partially be explained by the fact that 57% of patients died without prior diagnosis of recurrence, which led to their DFS being identical to their OS. The standard definition of DFS used in oncological trials comprises that deaths without documented prior recurrence are counted as events in DFS analyses (31). Thus, this finding is externally valid with regard to other trials. Most patients with disease recurrence, however, died several months or few years after diagnosis of the recurrence. The length of the time interval between recurrence and death is potentially influenced by chemotherapeutic, radiotherapeutic or even surgical treatment. None of the trials provided information on an individual patient level if and what treatment patients received. Therefore, one can only speculate about its possible effects. In general, it might be assumed that patients who underwent surgery alone and are thus chemotherapy-naïve receive more dose-intense chemotherapy than patients who had already undergone neoadjuvant chemotherapy. Likewise, radiotherapy for locoregional recurrence can usually only be administered in patients who had not been treated with neoadjuvant irradiation. However, given the biological complexity of the disease, it cannot be readily concluded that the time between recurrence and death is indeed longer in patients without prior neoadjuvant therapy.

The correct determination of DFS strongly depends on follow-up intervals and the specific kind of clinical, radiological and histopathological examinations carried out in order to detect

disease recurrence. These inevitably vary across the RCTs included in our analyses as they were conducted in different settings and during different time periods. However, as visual inspection resulted in no violation of the proportional hazards assumption, and hazard ratios can therefore be assumed to be time-independent, length, intensity and frequency of follow-up will not influence the estimation. OS, on the other hand side, is a very stable indicator because ascertainment of death during follow-up is supposedly very accurate. The consistency of the correlation between DFS and OS across trials does not suggest large differences in DFS determination across the different trials. The choice of the landmark time at six months after randomisation for DFS analyses is somewhat arbitrary. Early therapy-related and post-operative deaths as well as deaths during the early phase of postoperative continuation of chemotherapy might not be mirrored exactly by this approach. On the other hand side, these deaths are all accounted for in the OS analysis. There is no commonly agreed landmark time for such analyses, but six months have been used in an RCT on neoadjuvant chemotherapy for esophageal squamous cell cancer (32) and in our previous meta-analyses (3, 17), which led us to choose the same value for the present analysis.

In the meta-analysis, the overall hazard ratios for the treatment effects on OS and DFS are virtually identical. Likewise, the coefficient of determination in the error-in-variable regression is close to one. This indicates a very good model fit which reflects a strong correlation between treatment effects on OS and DFS (trial-level surrogacy). Furthermore, the correlation between the two outcomes themselves are was also high, indicating a good individual-level surrogacy. These results are consistent across all different included trials, regardless of the specificities of the applied neoadjuvant chemotherapy and regardless if patients had received combined chemoradiotherapy or merely chemotherapy. Only one small trial using radiotherapy along with cisplatin/5-fluorouracil doublet chemotherapy constituted an outlier, as the observed HR for the treatment effect on DFS was lower than what was expected based on OS in the trial (23).

The results indicate that the majority of variations in OS can be explained by the effect of the respective neoadjuvant treatment on DFS. This speaks in favor of DFS serving as appropriate surrogate marker for OS in trials evaluating neoadjuvant chemotherapy or chemoradiotherapy in patients with gastroesophageal adenocarcinoma. As a limitation, none of the included trials used targeted therapy with biologicals or monoclonal antibodies, or immunotherapy, and therefore deductions regarding these more recent treatment concepts (11, 12) are difficult to make. Moreover, histological and molecular tumor subtypes were not consistently reported in the included trials and therefore not available for analysis. Response to therapy depends on these characteristics, and although there is no direct evidence to that regard, the association between DFS and OS might be influenced by the histological or molecular subtype of the tumor.

The calculated surrogate threshold effect of 0.79 means that in future trials, a treatment yielding a reduction in the hazards of disease recurrence of at least 21% can be assumed to have a beneficial effect also on overall survival. Comparable analyses have found a surrogate threshold effect of 0.92 for adjuvant chemotherapy in gastric cancer (14) and of 0.56 for chemotherapy in metastatic or irresectable gastric cancer (15). There is no clearly established limit above which a surrogate threshold effect would qualify an outcome as appropriate surrogate outcome for another one. However, most randomized controlled trials in oncology are powered to detect effects in the magnitude of a 20%-30% reduction in the hazards of relapse or death. Moreover, based on an individual patient-level analysis of multiple randomized trials by Shi et al, 0.8 is often regarded as a meaningful boundary (33). Therefore, with a surrogate threshold effect of 0.79 in the present analysis, DFS can be regarded as a reasonably appropriate surrogate endpoint for OS. It must be noted, however, that methodologically, the surrogate threshold effect is computed for trials of infinite size. Given that the included RCTs were all of medium size, the surrogate threshold effect might be overestimated.

In summary, based on a strong correlation between DFS and OS on both the individual patient and trial level, as well as on the finding that the vast majority of variation in OS can be

explained by variation in DFS, DFS seems to be an appropriate surrogate marker for OS in randomized trials of neoadjuvant chemotherapy or chemoradiotherapy for gastroesophageal adenocarcinoma. However, as novel treatment concepts with substances other than cytotoxic compounds keep evolving, this finding requires continued validation.

Acknowledgements

Parts of this work were supported by the German Federal Ministry of Education and Research (grant number 01 KG 0807) and intramural research funds of the University of Heidelberg (Foundation for Cancer Research). The funding sources had no role in the study design; collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication. In memory of Prof. Christophe Mariette who initiated the study FFCD9901. We thank Emilie Barbier, FFCD Statistician, for her support.

Conflicts of interest

None declared.

Table 1: Randomized controlled trials meeting the inclusion criteria, from which IPD were used in the analysis.

Trial acronym/ first author	Accrual period	Countries	Main inclusion criteria	Chemotherapy / chemoradiotherapy regimen
ACCORD 07 (25)	1995-2003	France (multi-centre)	adenocarcinoma of lower third of esophagus or GE junction or stomach; UICC stage II or greater; suitable for curative resection; PS 0/1; 18-75 years	2 to 3 cycles (cisplatin 100mg/m² on day 1 or 2; 5-fluorouracil 4000mg/m² cumulative dose over 5 days, then 22 days break) preoperatively; surgery 4 to 6 weeks after last chemotherapy dose, 3 to 4 cycles (see above) postop. 4 to 6 weeks after surgery for patients who had R0 resection, no progression or major toxicity during preop. therapy and at least T3 or N+ tumor in histopathology
CALGB 9781 (23)	1997-2000	USA (multi-center)	squamous cell or adenocarcinoma of thoracic esophagus or GE junction, resectable (T1-3, Nx), including regional thoracic lymph node (N1) metastases, supraclavicular lymph node metastasis <1.5cm, lymph node metastases to levels 15-20 <1.5cm; no age limit	1 cycle (cisplatin 200 mg/m² cumulative dose on days 1 and 29, 5-fluorouracil 8000 mg/m² cumulative dose on days 1 to 4 and 29 to 32, radiotherapy (1.8 Gy/5 d/wk) begun within 24 hours of the chemotherapy administration, continued for 5.5 weeks, final 5.4 Gy given as a boost (total dose 50.4 Gy)
EORTC 40954 (22)	1999-2004	several European countries, Egypt (multi-centre)	adenocarcinoma of stomach or GE junction, cT3/4 Nx M0/M1(lymph); PS 0-1; 18-70 years	1 cycle (cisplatin 150 mg/m² cumulative dose on days 1, 15 and 29; 5-fluorouracil 12,000 mg/m² cumulative dose on days 1, 8, 15, 22, 29 and 36; folinic acid 3000 mg/m² cumulative dose on days 1, 8, 15, 22, 29 and 36); restaging, if no progression or toxicity 1 more cycle as described above restarting on day 50; surgery on days 57 to 63 of the second cycle

FAMTX (20)	1993-1996	Netherlands (multi-centre)	adenocarcinoma of the stomach (not cardia); >cT1; resectable with no evidence of distant metastases; PS 0-2; <75 years	2 cycles (methotrexate 1500 mg/m² on day 2; 5-fluorouracil 1500 mg/m² on day 2; leucovorin 240 or 480 mg (depending on MTX level) cumulative dose on days 3 to 4; doxorubicin 30 mg/m² on day 15; 13 days break); re-staging; in case of response or stable disease another 2 cycles (see above);
FFCD 9901 (18)	2000-2009	France (multi-centre)	thoracic esophageal adenocarcinoma or squamous cell carcinoma; suitable for curative resection; cT1/2N0/1 or cT3N0; PS 0-1; <75 years	2 cycles (fluorouracil and cisplatin. FU 800 mg/m2 per 24 hours was administered as a continuous infusion from days 1 to 4 and 29 to 32. Cisplatin 75 mg/m^2 on day 1 or 2 and day 29 or 30 or, alternatively, 15 mg/m^2 from days 1 to 5 and 29 to 33), concomitant radiotherapy (45 Gy five fractions per week over 5 weeks).
RTOG 8911 (21)	1990-1995	USA (multi-center)	squamous cell or adenocarcinoma of thoracic esophagus or GE junction; stage I-III excluding T4 tumors; absence of supraclavicular or distant metastases; fit for surgery; at least 18 years;	3 cycles (cisplatin 100 mg/m² on day 1; 5-fluorouracil 1000 mg/m² cumulative dose on days 1 to 5, 23 days break); operation 2 to 4 weeks after the end of the last cycle; in case of stable or responsive disease upon surgery 2 postoperative cycles (see above, except cisplatin dose reduced to 75 mg/m²) starting 2 to 6 weeks after surgery
TROG-AGITG (19)	1994-2000	Australia, New Zealand, Singapore (multi-center)	invasive cancer of thoracic oesophagus; cT1-3 cN0-1; no involvement of cervical esophagus or celiac nodes; PS 0 or 1; no age limit	1 cycle (cisplatin 80 mg/m² on day 1; 5-fluorouracil 3200 mg/m² cumulative dose on days 1 to 4) with 35 Gy radiotherapy in 15 fractions over 3 weeks, starting concurrently with chemotherapy; surgery 3 to 6 weeks after completion of radiotherapy; postoperative radiotherapy permitted for patients with residual disease

				after surgery if indicated clinically for patients assigned to surgery alone
Urba (24)	1989-1994	USA (single-center)	squamous cell, adenocarcinoma or mixed adenosquamous carcinoma of esophagus or GE junction, limited to esophagus and regional lymph nodes (including celiac nodes); Karnofsky index >=60%; ≤75 years	1 cycle (cisplatin 200 mg/m² cumulative dose on days 1 through 5 and 17 through 21, 5-fluorouracil 6300 mg/m² cumulative on days 1 through 21, vinblastin 8 mg/m² on days 1 through 4 and 17 through 20, radiotherapy in fractions of 1.5 Gy twice a day, on days 1 through 5, 8 through 12, and 15 through 19, to a total dose of 45 Gy)

PS: performance status (ECOG/WHO)

Table 2: Demographic and clinical characteristics of patients included in the analysis, both for the entire study population and separately for patients from the neoadjuvant therapy and surgery alone arms

	Neoadjuvant Chemotherapy N=562	Surgery alone N=564	Total N=1126	p-value
Trial				
- ACCORD	113 (20.1%)	111 (19.7%)	224 (19.9%)	0.998
- CALGB	23 (4.1%)	19 (3.4%)	42 (3.7%)	
- EORTC	69 (12.3%)	71 (12.6%)	140 (12.4%)	
- FAMTX	27 (4.8%)	29 (5.1%)	56 (5.0%)	
- FFCD	98 (17.4%)	97 (17.2%)	195 (17.3%)	
- RTOG	115 (20.5%)	121 (21.5%)	236 (21.0%)	
- TROG-AGITG	80 (14.2%)	78 (13.8%)	158 (14.0%)	
- URBA	37 (6.6%)	38 (6.7%)	75 (6.7%)	
Gender				
- male	483 (85.9%)	467 (82.8%)	950 (84.4%)	0.147
- female	79 (14.1%)	97 (17.2%)	176 (15.6%)	
Age [years]				
- N	562	564	1126	0.908
- Mean +/- SD	59.8 +/-9.3	59.6 +/-9.4	59.7 +/-9.3	
- p5, p25, p75, p95	44.0, 53.2, 67.0, 73.2	43.0, 53.3, 67.0, 73.0	44.0, 53.2, 67.0, 73.1	
- Median	60.8	61.0	61.0	
- Min, Max	23.0, 78.0	26.1, 80.5	23.0, 80.5	
Age				
- < 65 years	366 (65.1%)	377 (66.8%)	743 (66.0%)	0.827
- 65 - 75 years	184 (32.7%)	176 (31.2%)	360 (32.0%)	
- > 75 years	12 (2.1%)	11 (2.0%)	23 (2.0%)	
Tumor location				
- Stomach	88 (15.7%)	89 (15.8%)	177 (15.7%)	0.984
- GE junction	153 (27.2%)	158 (28.0%)	311 (27.6%)	
- Esophagus	261 (46.4%)	260 (46.1%)	521 (46.3%)	
- Esophagus / GE junction (no further specification)	60 (10.7%)	57 (10.1%)	117 (10.4%)	
Performance status				
- 0	373 (71.5%)	365 (71.3%)	738 (71.4%)	0.263
- 1	144 (27.6%)	146 (28.5%)	290 (28.0%)	
- 2	5 (1.0%)	1 (0.2%)	6 (0.6%)	
- missing	40	52	92	
T stage [preoperative, clinical]				
- T0	1 (0.5%)	0 (0.0%)	1 (0.2%)	0.867

	Neoadjuvant Chemotherapy N=562	Surgery alone N=564	Total N=1126	p-value
- T1	22 (10.7%)	16 (7.7%)	38 (9.2%)	
- T2	57 (27.7%)	56 (27.0%)	113 (27.4%)	
- T3	121 (58.7%)	130 (62.8%)	251 (60.8%)	
- T4	5 (2.4%)	5 (2.4%)	10 (2.4%)	
- missing	357	361	718	
N stage [preoperative, clinical]				
- N0	61 (80,8%)	47 (63.5%)	108 (70.1%)	0.178
- N1	18 (22.5%)	27 (36.5%)	45 (29.2%)	
- N2	1 (1.3%)	0 (0.0%)	1 (0.6%)	
- missing	482	490	972	
T stage [postoperative, histopathological]				
- T0	53 (13.2%)	2 (0.5%)	55 (6.7%)	<.001
- T1	63 (15.7%)	64 (15.2%)	127 (15.5%)	
- T2	112 (27.9%)	106 (25.2%)	218 (30.2%)	
- T3	156 (38.9%)	207 (49.2%)	363 (50.3%)	
- T4	17 (4.2%)	42 (10.0%)	59 (8.2%)	
- missing	161	143	304	
N stage [postoperative, histopathological]				
- N0	181 (45.3%)	110 (26.4%)	291 (35.6%)	<.001
- N1	171 (42.8%)	210 (50.4%)	381 (46.6%)	
- N2	35 (8.8%)	59 (14.1%)	94 (11.5%)	
- N3	13 (3.3%)	38 (9.1%)	51 (6.2%)	
- missing	162	147	309	
Margin status				
- R0	395 (91.2%)	374 (82.3%)	769 (86.7%)	0.001
- R1	18 (4.2%)	35 (7.4%)	53 (6.0%)	
- R2	20 (4.6%)	45 (9.5%)	65 (7.3%)	
- missing	129	110	239	

Figure 1: Time-to-event curves for OS and DFS, calculated according to the landmark method, stratified by treatment arm. Treat_arm=0 OS: overall survival in the upfront surgery arms, treat_arm=1 OS: overall survival in the neoadjuvant therapy arms, treat_arm=0 DFS: disease-free survival in the up-front surgery arms, treat_arm=1 DFS: disease-free survival in the neoadjuvant therapy arms.

Figure 2: Forest plot of hazard ratios for the comparison of OS and DFS between neoadjuvant therapy and surgery alone

Figure 3: Error-in-variable regression, observed and predicted HRs for OS and DFS with 95% prediction limits. RT-CT: neoadjuvant chemoradiotherapy, CT: neoadjuvant chemotherapy.

References

1. Rubenstein JH, Shaheen NJ. Epidemiology, Diagnosis, and Management of Esophageal Adenocarcinoma. Gastroenterology. 2015;149(2):302-17.e1.

2. Jim MA, Pinheiro PS, Carreira H, Espey DK, Wiggins CL, Weir HK. Stomach cancer survival in the United States by race and stage (2001-2009): Findings from the CONCORD-2 study. Cancer. 2017;123 Suppl 24:4994-5013.

3. Ronellenfitsch U, Schwarzbach M, Hofheinz R, Kienle P, Kieser M, Slanger TE, et al. Preoperative chemo(radio)therapy versus primary surgery for gastroesophageal adenocarcinoma: systematic review with meta-analysis combining individual patient and aggregate data. European journal of cancer (Oxford, England : 1990). 2013;49(15):3149-58.

4. van Hagen P, Hulshof MC, van Lanschot JJ, Steyerberg EW, van Berge Henegouwen MI, Wijnhoven BP, et al. Preoperative chemoradiotherapy for esophageal or junctional cancer. The New England journal of medicine. 2012;366(22):2074-84.

5. Schirren R, Reim D, Novotny AR. Adjuvant and/or neoadjuvant therapy for gastric cancer? A perspective review. Therapeutic advances in medical oncology. 2015;7(1):39-48.

6. Egyud MR, Tseng JF, Suzuki K. Multidisciplinary Therapy of Esophageal Cancer. The Surgical clinics of North America. 2019;99(3):419-37.

7. Al-Batran SE, Homann N, Pauligk C, Goetze TO, Meiler J, Kasper S, et al. Perioperative chemotherapy with fluorouracil plus leucovorin, oxaliplatin, and docetaxel versus fluorouracil or capecitabine plus cisplatin and epirubicin for locally advanced, resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4): a randomised, phase 2/3 trial. Lancet (London, England). 2019;393(10184):1948-57.

8. Lordick F, Mariette C, Haustermans K, Obermannova R, Arnold D. Oesophageal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of oncology : official journal of the European Society for Medical Oncology. 2016;27(suppl 5):v50-v7.

9. Smyth EC, Verheij M, Allum W, Cunningham D, Cervantes A, Arnold D. Gastric cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.

Annals of oncology : official journal of the European Society for Medical Oncology. 2016;27(suppl 5):v38-v49.

10. Cats A, Jansen EPM, van Grieken NCT, Sikorska K, Lind P, Nordsmark M, et al. Chemotherapy versus chemoradiotherapy after surgery and preoperative chemotherapy for resectable gastric cancer (CRITICS): an international, open-label, randomised phase 3 trial. The Lancet Oncology. 2018;19(5):616-28.

11. Wagner AD, Grabsch HI, Mauer M, Marreaud S, Caballero C, Thuss-Patience P, et al. EORTC-1203-GITCG - the "INNOVATION"-trial: Effect of chemotherapy alone versus chemotherapy plus trastuzumab, versus chemotherapy plus trastuzumab plus pertuzumab, in the perioperative treatment of HER2 positive, gastric and gastroesophageal junction adenocarcinoma on pathologic response rate: a randomized phase II-intergroup trial of the EORTC-Gastrointestinal Tract Cancer Group, Korean Cancer Study Group and Dutch Upper GI-Cancer group. BMC cancer. 2019;19(1):494.

12. Maron SB, Catenacci DVT. Update on Gastroesophageal Adenocarcinoma Targeted Therapies. Hematology/oncology clinics of North America. 2017;31(3):511-27.

13. Mathoulin-Pelissier S, Gourgou-Bourgade S, Bonnetain F, Kramar A. Survival end point reporting in randomized cancer clinical trials: a review of major journals. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2008;26(22):3721-6.

14. Oba K, Paoletti X, Alberts S, Bang YJ, Benedetti J, Bleiberg H, et al. Disease-free survival as a surrogate for overall survival in adjuvant trials of gastric cancer: a meta-analysis. Journal of the National Cancer Institute. 2013;105(21):1600-7.

15. Paoletti X, Oba K, Bang YJ, Bleiberg H, Boku N, Bouche O, et al. Progression-free survival as a surrogate for overall survival in advanced/recurrent gastric cancer trials: a meta-analysis. Journal of the National Cancer Institute. 2013;105(21):1667-70.

16. Ronellenfitsch U, Schwarzbach M, Hofheinz R, Kienle P, Kieser M, Slanger TE, et al. Perioperative chemo(radio)therapy versus primary surgery for resectable adenocarcinoma of the stomach, gastroesophageal junction, and lower esophagus. The Cochrane database of systematic reviews. 2013(5):Cd008107.

17. Ronellenfitsch U, Schwarzbach M, Hofheinz R, Kienle P, Nowak K, Kieser M, et al. Predictors of overall and recurrence-free survival after neoadjuvant chemotherapy for

gastroesophageal adenocarcinoma: Pooled analysis of individual patient data (IPD) from randomized controlled trials (RCTs). European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 2017;43(8):1550-8.

18. Mariette C, Dahan L, Mornex F, Maillard E, Thomas PA, Meunier B, et al. Surgery alone versus chemoradiotherapy followed by surgery for stage I and II esophageal cancer: final analysis of randomized controlled phase III trial FFCD 9901. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2014;32(23):2416-22.

19. Burmeister BH, Smithers BM, Gebski V, Fitzgerald L, Simes RJ, Devitt P, et al. Surgery alone versus chemoradiotherapy followed by surgery for resectable cancer of the oesophagus: a randomised controlled phase III trial. The Lancet Oncology. 2005;6(9):659-68.

20. Hartgrink HH, van de Velde CJ, Putter H, Songun I, Tesselaar ME, Kranenbarg EK, et al. Neo-adjuvant chemotherapy for operable gastric cancer: long term results of the Dutch randomised FAMTX trial. European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 2004;30(6):643-9.

21. Kelsen DP, Winter KA, Gunderson LL, Mortimer J, Estes NC, Haller DG, et al. Long-term results of RTOG trial 8911 (USA Intergroup 113): a random assignment trial comparison of chemotherapy followed by surgery compared with surgery alone for esophageal cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2007;25(24):3719-25.

22. Schuhmacher C, Gretschel S, Lordick F, Reichardt P, Hohenberger W, Eisenberger CF, et al. Neoadjuvant chemotherapy compared with surgery alone for locally advanced cancer of the stomach and cardia: European Organisation for Research and Treatment of Cancer randomized trial 40954. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2010;28(35):5210-8.

23. Tepper J, Krasna MJ, Niedzwiecki D, Hollis D, Reed CE, Goldberg R, et al. Phase III trial of trimodality therapy with cisplatin, fluorouracil, radiotherapy, and surgery compared with surgery alone for esophageal cancer: CALGB 9781. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2008;26(7):1086-92.

24. Urba SG, Orringer MB, Turrisi A, Iannettoni M, Forastiere A, Strawderman M. Randomized trial of preoperative chemoradiation versus surgery alone in patients with locoregional esophageal carcinoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2001;19(2):305-13.

25. Ychou M, Boige V, Pignon JP, Conroy T, Bouche O, Lebreton G, et al. Perioperative chemotherapy compared with surgery alone for resectable gastroesophageal adenocarcinoma: an FNCLCC and FFCD multicenter phase III trial. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2011;29(13):1715-21.

26. Draper N, Smith, H. Applied regression analysis, third edition. New York, NY: John Wiley & Sons; 1998.

27. Kmenta J. Elements of econometrics: Second edition. Ann Arbor, MI: University of Michigan Press; 1997.

28. Treiman D. Quantitative Data Analysis: Doing Social Research to Test Ideas. San Francisco, CA: Jossey-Bass; 2008.

29. Buyse M, Burzykowski T, Michiels S, Carroll K. Individual- and trial-level surrogacy in colorectal cancer. Statistical methods in medical research. 2008;17(5):467-75.

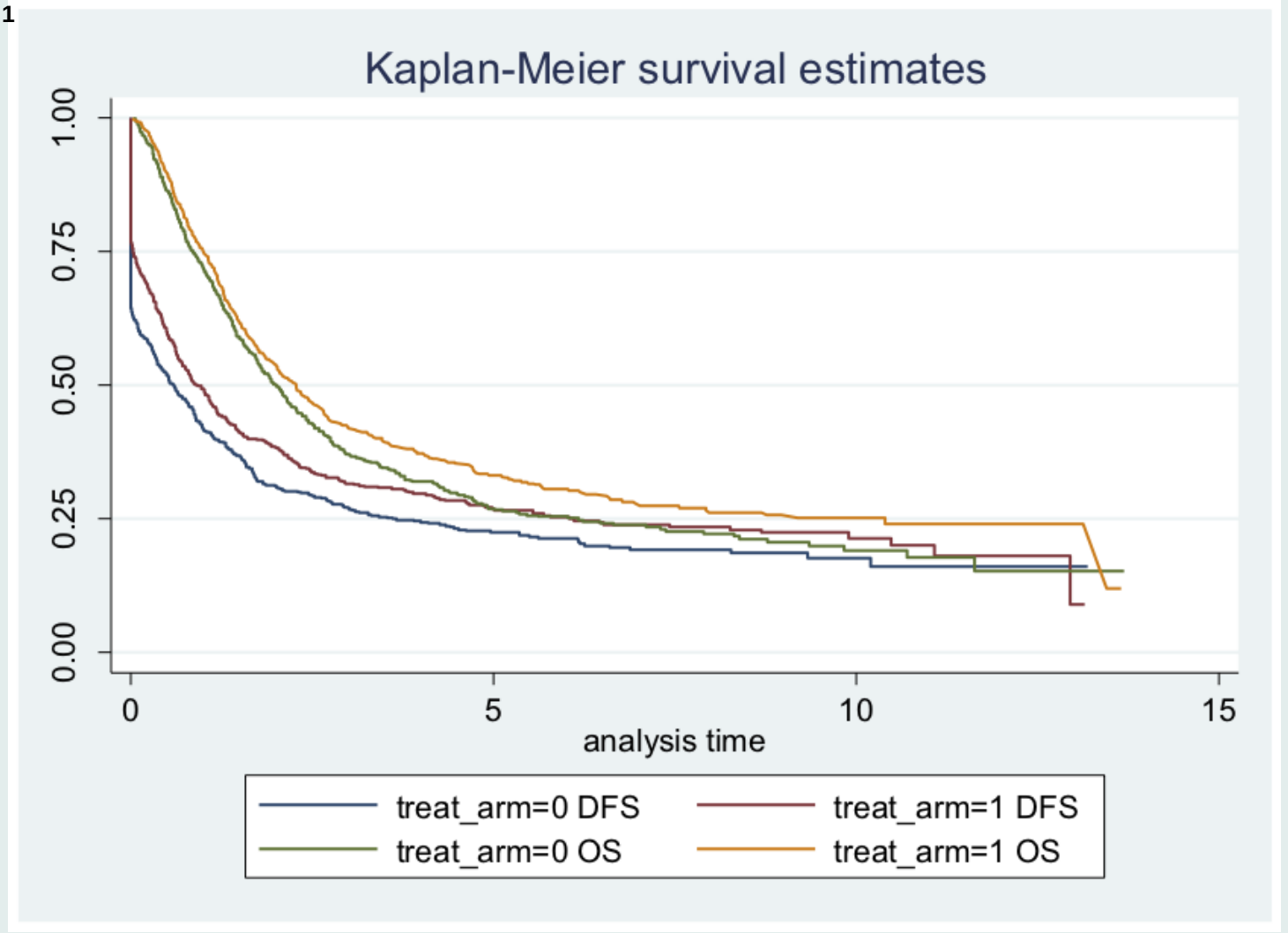
30. Rotolo F, Paoletti X, Michiels S. surrosurv: An R package for the evaluation of failure time surrogate endpoints in individual patient data meta-analyses of randomized clinical trials. Computer methods and programs in biomedicine. 2018;155:189-98.

31. Panageas KS, Knezevic A. Selection of endpoints. In: Halabi S, Michiels S, editors. Textbook of Clinical Trials in Oncology: A Statistical Perspective. Boca Raton, FL: CRC Press; 2019.

32. Boonstra JJ, Kok TC, Wijnhoven BP, van Heijl M, van Berge Henegouwen MI, Ten Kate FJ, et al. Chemotherapy followed by surgery versus surgery alone in patients with resectable oesophageal squamous cell carcinoma: long-term results of a randomized controlled trial. BMC cancer. 2011;11:181.

33. Shi Q, Schmitz N, Ou FS, Dixon JG, Cunningham D, Pfreundschuh M, et al. Progression-Free Survival as a Surrogate End Point for Overall Survival in First-Line Diffuse Large B-Cell Lymphoma: An Individual Patient-Level Analysis of Multiple Randomized Trials (SEAL). Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2018;36(25):2593-602.

Figure 1



Patients at risk

0 DFS	564	99	16	0
1 DFS	562	121	33	0
0 OS	620	119	21	0
1 OS	621	162	49	4

Figure 2

Click here to download high resolution image

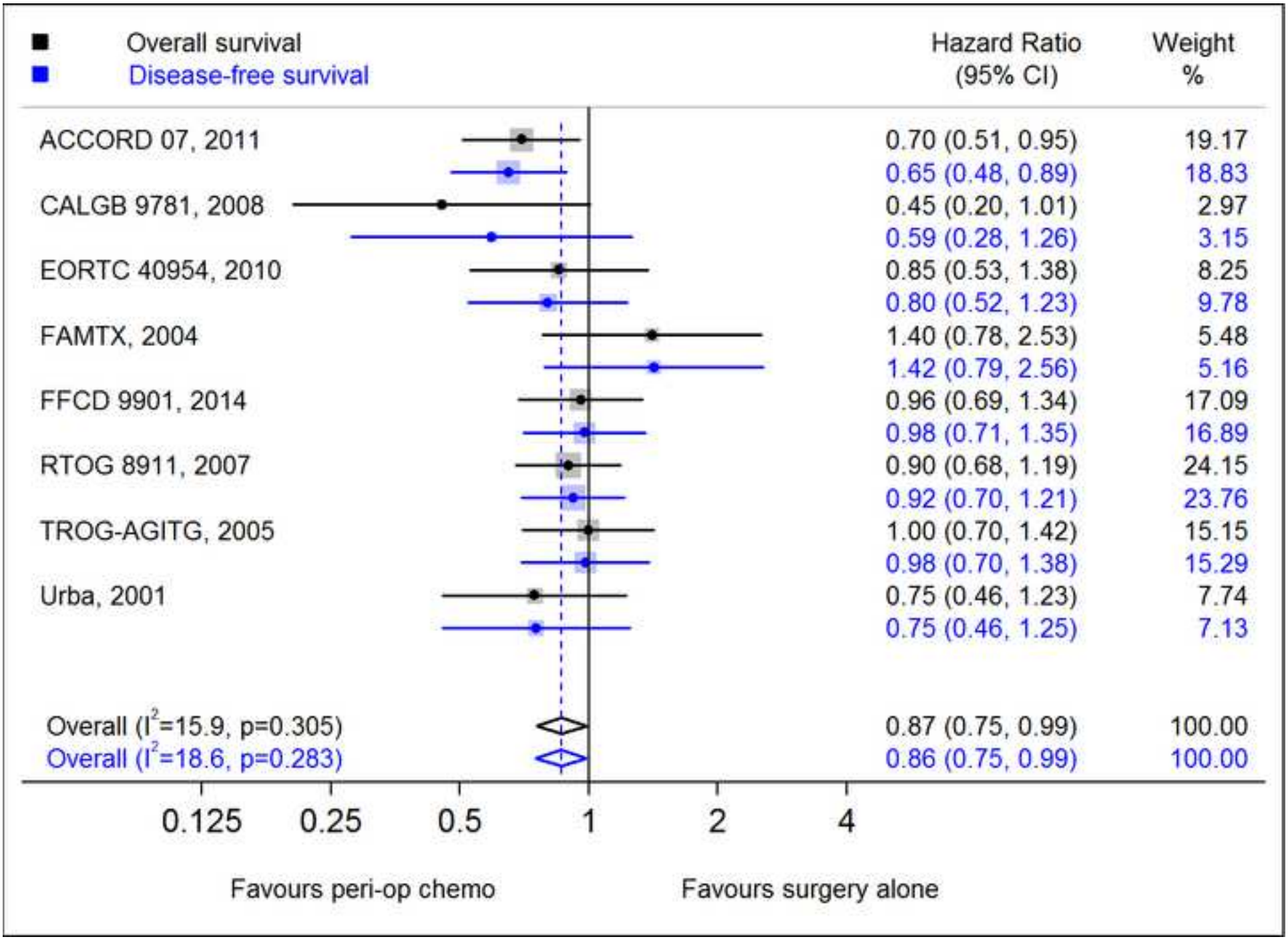
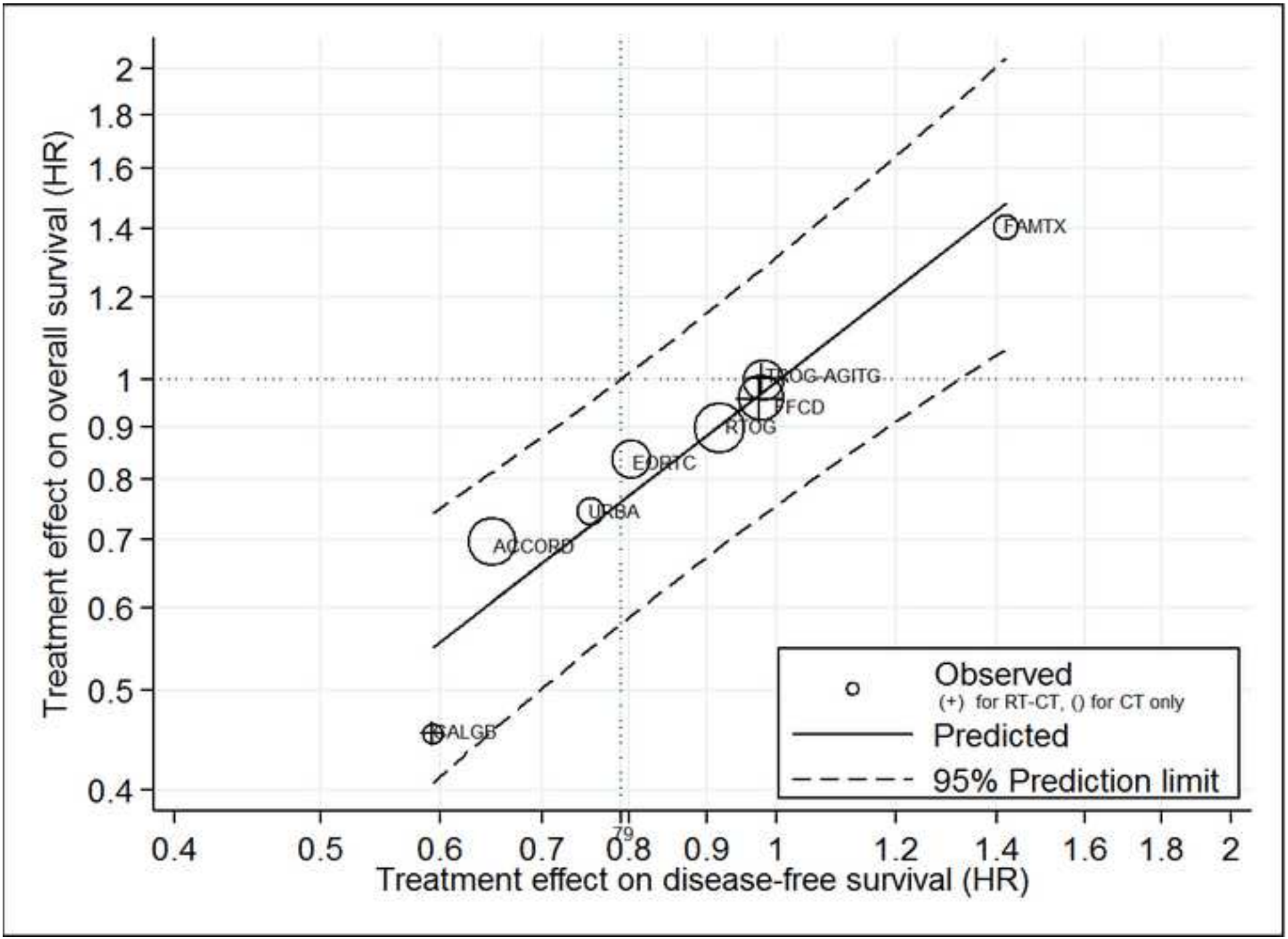


Figure 3

Click here to download high resolution image



Disease-free survival as a surrogate for overall survival in neoadjuvant trials of gastroesophageal adenocarcinoma: pooled analysis of individual patient data from randomized controlled trials

Ulrich Ronellenfitsch, Katrin Jensen, Svenja Seide, Meinhard Kieser, Matthias Schwarzbach, Tracy E Slanger, Bryan Burmeister, David Kelsen, Donna Niedzwiecki, Guillaume Piessen, Christoph Schuhmacher, Susan Urba, Cornelis van de Velde, Marc Ychou, Ralf Hofheinz, Sylvie Lorenzen

Conflicts of interest statement

The authors declare that they have no conflict of interest.