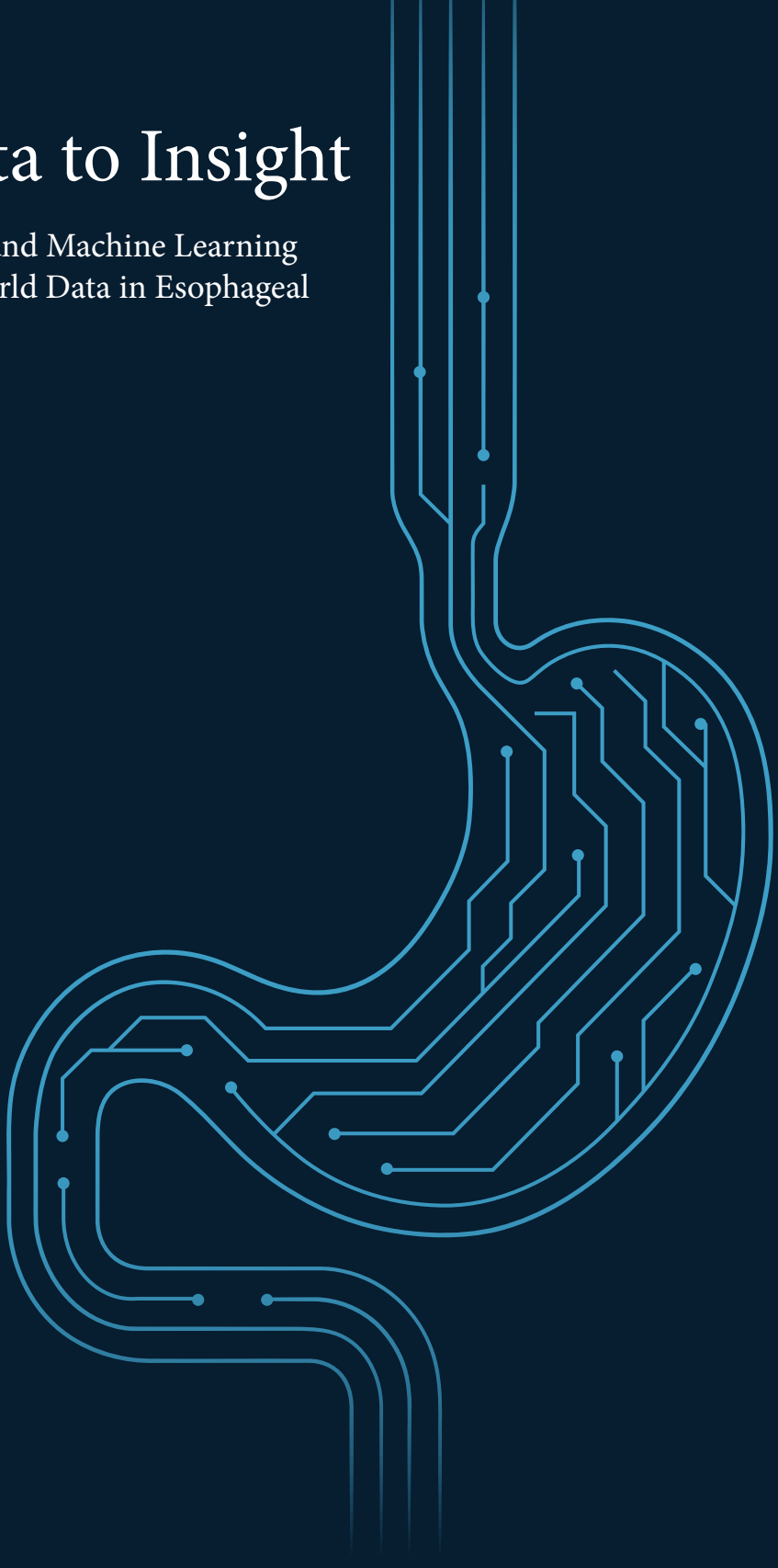
From Data to Insight

Applying Statistical and Machine Learning Methods to Real-World Data in Esophageal and Gastric Cancer



Steven C. Kuijper

From Data to Insight: Applying Statistical and Machine Learning Methods to Real-world Data in Esophageal and Gastric Cancer

Steven C. Kuijper

Colophon

Cover art	S.C. Kuijper
Lay-out	S.C. Kuijper
Printed by	Ridderprint, www.ridderprint.nl
ISBN	978-94-6522-661-3

Financial support for this thesis was kindly provided by the Cancer Center Amsterdam.

From Data to Insight: Applying Statistical and Machine Learning Methods to Real-world Data in Esophageal and Gastric Cancer

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. P.P.C.C. Verbeek
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op dinsdag 14 oktober 2025, te 13.00 uur

door Steven Coenrad Kuijper
geboren te Haarlem

Promotiecommissie

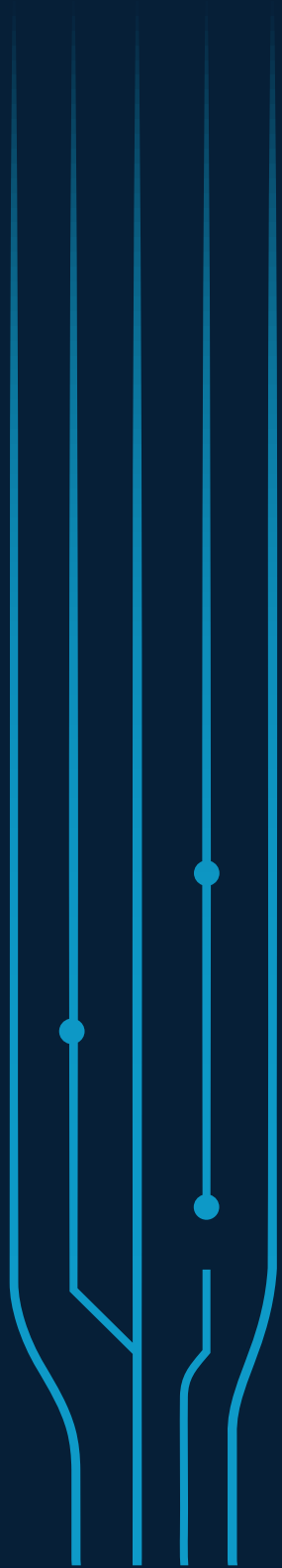
Promotor:	prof. dr. H.W.M. van Laarhoven	AMC-UvA
Copromotor:	dr. R.H.A. Verhoeven	IKNL
Overige leden:	prof. dr. A. Abu-Hanna	AMC-UvA
	prof. dr. J.J.G.H.M. Bergman	AMC-UvA
	prof. dr. V.M.H. Coupé	Vrije Universiteit Amsterdam
	prof. dr. E.M.A. Smets	AMC-UvA
	dr. P.S.N. van Rossum	Vrije Universiteit Amsterdam
	dr. M.J. Bijlsma	Organon

Faculteit der Geneeskunde

Table of contents

Part IV: Prediction of survival outcomes

General Introduction



General introduction

Esophageal and gastric cancer are highly deadly diseases. In 2022, these cancer types ranked 11th and 5th worldwide, with 510,716 and 968,350 diagnoses, respectively.[1] In the Netherlands, there were 2,717 new diagnoses of esophageal cancer and 1,620 new diagnoses of gastric cancer, including junction/cardia cancer, in 2023.[2] Despite advances in the detection and treatment of these cancer types, survival for patients with these cancer types remains to be generally poor. To illustrate this, the relative 5-year overall survival of patients diagnosed with esophageal and gastric cancer that has not metastasized is roughly 40%.[2] For patients with metastasized esophageal and gastric cancer the median overall survival in the general population is around 7 months.[2]

There is a variety of different treatment options available for patients with these cancer types. Generally, these can be classified into treatment with curative intent and palliative treatment. Surgical removal of the tumor remains the key component in curative treatment strategies.[3,4] Depending on the location of the tumor, surgical resection can be preceded and/or followed by chemotherapy or a combination of chemotherapy and radiotherapy. For patients who are less fit for surgery, definitive chemoradiotherapy can be an alternative potentially curable treatment. While potentially curable treatment regimens usually consist of a combination of a surgical resection and chemo(radio)therapy, palliative treatment consists primarily of systemic therapy.[3,4] This could include chemotherapy containing different compounds, targeted therapy, and/or immunotherapy. Alternative palliative treatment options include palliative surgery, radiotherapy, an esophageal stent and best supportive care.

Despite the variety of treatment options, it is not always obvious which treatment is most appropriate for patients. Patients with esophageal and gastric cancer are generally relatively old, often have comorbidities and are often diagnosed with distant metastases.[5] The aim of treating patients' cancer is therefore not always at prolonging life but could also be to sustain or improve health-related quality of life. While treatment may affect health-related quality of life negatively, studies have shown that there is a large heterogeneity in the course of health-related quality of life after treatment.[6,7] In fact, treatment could also sustain or even improve health-related quality of life. The choice of treatment is therefore not trivial and requires carefull considerations from physicians and patients as it may have consequences for life expectancy and health-related quality of life.

The context in which such treatment-related decisions are discussed and between patients and physicians is commonly referred to as shared decision making.[8] In this process it is important that patients and physicians explore the wishes of patients in terms of treatment outcomes such as life expectancy and health-related quality of life. This requires good communication between patients and physicians but also good, reliable and ideally personalized information for the physician to communicate about the efficacy of different treatment options and how these treatment options could affect patient's life expectancy and health-related quality of life. In this thesis, we will explore statistical and machine learning methods applied to real-world data from patients with esophageal and gastric cancer to achieve this goal.

Statistical models and machine learning are powerful methods that can obtain valuable information from data. They can, for example, be used for the modeling or prediction of treatment outcomes in clinical settings. While statistical models and machine learning models share many similarities, they are—in fact—different. Statistical models typically have a predetermined structure and are relatively easy to interpret, whereas machine learning models have no predetermined structure, but are more difficult to interpret. There are numerous different applications of statistics and machine learning in the field of (esophagogastric) oncology that provide treatment-related information, both on data from randomized clinical trials as well as real-world data.

Data from randomized clinical trials, are the gold-standard data in medical sciences as they enable causal treatment effect estimation. However, clinical trials often maintain strict inclusion criteria and consequentially only include a fraction of the total patient population.[9] Clinical trials generally have a high internal validity, which is to say that clinical trials estimate treatment effect with little bias due to the randomization, but can lack external validity, which refers to the generalizability or transportability of the results to the total population of patients with a particular disease. For example, the randomized controlled FLOT4-AIO trial showed improved survival for patients with gastric cancer treated with perioperative fluorouracil, leucovorin, oxaliplatin, and docetaxel (FLOT) compared to anthracyclin triplets. This difference in survival was not observed in real-world data, showing that results from clinical trials are not always transportable to real-world populations.[10]

To this end, real-world data can be used, which is observational in nature and typically includes all patients in daily clinical practice.[11] As such, it usually includes all patients and not merely a fraction of the patients, thereby providing a more realistic view on the treatment and outcomes in daily clinical practice.

In recent years, the use of real-world data has increased and is a growing area of research within medicine. Statistical and machine learning methods can be used to extract valuable information from real-world data that would otherwise have not been accessible through clinical trials. Large-scale nation-wide databases such as the Netherlands Cancer Registry (NCR) provide a rich infrastructure to perform real-world data studies, and has been used for national and international comparisons in the field of esophagogastric oncology. For example, data from the NCR has been used to identify treatment patterns and outcomes for patients with esophageal and gastric cancer in daily clinical practice,[5,12,13] to investigate the relationship between hospital volume and treatment outcomes,[14] to estimate the relationship between effectiveness of first-line treatment and second-line treatment,[15] and international comparisons between the Netherlands and other countries.[16,17]

While these examples primarily involve modelling of treatment outcomes in order to understand the relationship between variables of interest and the outcome, in recent years the NCR has also been used to develop prediction models.[18] The SOURCE prediction models were published in 2021 and consists of four different prediction models: for patients with potentially curable esophageal and gastric cancer and palliative esophageal and gastric cancer, which was integrated into an online support tool. This tool allows physicians and patients to view personalized prediction models (including survival rates) and overall evidence-based outcomes (such as HRQoL and side effects/complications).

The subsequent SOURCE trial investigated the impact of incorporating the SOURCE tool into physician-patient consultations, along with comprehensive training, on factors such as the accuracy of outcome information, patient satisfaction, understanding of information, and treatment decision-making. First results from the SOURCE trial showed an overall improvement in the precision with which physicians conveyed information about treatment-related outcomes, underscoring the importance of good data-driven models.[19]

Aim and outline of the thesis

All the afore mentioned examples have advanced understanding of treatment outcomes within the field of esophagogastric oncology from a clinical and a public health perspective. However, when it concerns the application of statistical and machine learning methods to real-world data, there are areas of improvement that could lead to better or more nuanced understanding and treatment outcomes which can be clinically useful. The general aim of this thesis is to apply statistical and machine learning methods to real-world data from patients with esophageal and gastric cancer, to advance understanding on the application of these methods to real-world data and to advance understanding of treatment outcomes.

Part I. Epidemiology of esophageal and gastric cancer

In order to understand treatment outcomes on a global scale, direct international comparisons of treatment outcomes can be very insightful. In Chapter 1, we demonstrate a methodologically rigorous method with which the comparison between countries can be made while controlling for baseline population mortality by capitalizing on the power of parametric survival modeling. To this end, we estimate and compare survival of patients with esophageal and gastric cancer of the Netherlands and Belgium and investigate to what extend potential survival differences can be explained by differences in treatment.

In survival analyses, estimates such as overall median survival or x-year survival have been traditionally used to quantify survival. While informative, they fail to demonstrate the heterogeneity of patients' survival since they capture survival in a single number. In Chapter 2 we apply a method called survival scenarios, that enables investigation of the survival curve beyond median survival only. Thereby proving a more nuanced understanding which patients have reaped the most benefits from treatment advances over the past years.

Furthermore, traditional survival metrics such as overall median survival fail to capture the evolving risk profiles of survivors over time. For example, suppose a patient has a 60% probability of surviving 3 years past diagnosis. Given that this patient has already survived three years, what is the probability that this patient will survive another x-years. This is called conditional survival, and conditional survival estimates can be very insightful for patients and physicians. Therefore, in Chapter 3, we estimate conditional relative survival for patients with esophageal and gastric cancer in the metastatic setting. We demonstrate that This metric provides a more nuanced understanding of survival probabilities as patients survive beyond initial diagnosis periods, offering valuable insights into the changing risks and prognosis over time.

Part II. Health-related quality of life

Part II of this thesis deals with the application of statistical models in health-related quality of life research. In recent years, large scale cohort studies such as the Prospective Observational Cohort Study of Oesophageal-gastric cancer Patients (POCOP) have been initiated to gather patients reported outcome measures on health-related quality of life.[20] Such large scale cohort studies have in recent years been used to model health-related quality of life as treatment outcome. An important potential limitation of such large-scale cohort studies is that there is an inherent risk of selection bias, which implies that some patients may be more likely to participate than other patients. This may affect the degree to which such cohorts are reflective of the total real-world population. Currently, there is no streamlined method within the field of oncology to investigate the real-world representativeness of patients in health-related quality of life cohort studies. In Chapter 4 we address this challenge by demonstrating a novel method with which the real-world representativeness of health-related quality of life longitudinal cohort studies can be assessed.

Furthermore, there are currently no published prediction models on health-related quality of life for patients with esophageal and gastric cancer. Such prediction models on personalized health-related quality of life outcomes would be extremely useful in informing patients how their health-related quality of life could be impacted by treatment. In Chapter 5 we will develop prediction models using statistical and machine learning models based on data from the POCOP cohort study. In this study, we will develop risk prediction model to estimate the risk of a significant deterioration in health-related quality of life using elastic-net regression modelling. We also develop a sequential score prediction model with XGboost, where the actual values of the questionnaires can be predicted across a continuous time scale.

Part III. Clinical trials and real-world data

In Part III two of this thesis, we demonstrate how data from clinical trials and real-world data can be used to complement each other and to investigate treatment effects for which no clinical trial is available. There have been many established methods do deal with this problem. Examples are propensity score methods such as matching and weighting that can deal with the observational nature of data while still being able to estimate causal effects.

In Chapter 6, we will combine trial data with real-world data by demonstrating how real-world data can be used to create a control arm for comparison with one-armed trial. We perform this in the context of the one-armed DESTINY-Gastric02 study, where the survival of patients with non-metastatic HER2-positive esophageal adenocarcinoma treated with trastuzumab-deruxtecan was evaluated.[21] As this one-armed study had no control arm, we will use data from the Netherlands Cancer Registry to form a control arm that satisfies the necessary requirement for causal inference to be able to make valid comparisons.

In Chapter 7, we apply causal inference techniques in a different clinical setting and exclusively with real-world data. In this chapter, we show how real-world data can be used to mimic clinical trials and apply a rigorous methodological approach to estimate the treatment effect for a new immunotherapy for patients with esophageal cancer. Using data from the Netherlands Cancer Registry, we create two comparable sets of patients who have been treated according to the neoadjuvant chemoradiotherapy regimen. and one arm was additionally treated with adjuvant nivolumab. In this chapter we apply causal inference methods to estimate survival of patients treated with adjuvant nivolumab.

Part VI. Prediction of survival outcomes

Though the enterprise of prediction modeling is very promising for daily clinical practice, there are opportunities to further the development of existing prediction models in the field of esophagogastric cancer. In this final part of the thesis, Part IV, we deal with the development and application of prediction models within esophageal and gastric cancer.

The utility of existing prediction models in clinical practice could be explored further. For example, the utility of prediction models in aiding clinicians and patients in shared decision making, but also the utility of prediction models to be used for different ends, such as a tool for including patients in clinical trials. In Chapter 8 we demonstrate the SOURCE survival prediction model can add to the survival expectation of the clinician in the context of the POLDER study. The POLDER trial investigated the effect of external beam radiotherapy to relieve trouble with swallowing for patients with esophageal cancer. Patients were required to be expected to live for at least three months to be included in the study, but a large proportion did not reach this point. We demonstrate a method to investigate if the SOURCE prediction model could have added to the decision of the clinician to include patients into the study.

With the existing SOURCE models, clinicians can make predictions for patients who will start potentially curable treatment or first-line systemic treatment. However, when patients with metastatic esophageal and gastric cancer are faced with the decision to continue with a second-line systemic therapy or opt for best-supportive care after failure of the first-line systemic therapy, no prediction models are available. Therefore, in Chapter 9, we present a prediction model for patients with esophagogastric cancer that have completed first-line systemic therapy and have the option to continue beyond-first line palliative treatment or best supportive care called SOURCE Beyond First-Line.

Within the field of prediction modeling, new methods have been developed that have made it possible to add biomarkers to existing clinical prediction models without the need of a total refitting of the model and re-estimation of the model coefficients. This has opened the door for investigation of the added value of using biomarkers in existing prediction models such as SOURCE. Information from PET-CT scans and circulating-free tumor DNA could potentially improve personalized prediction models. In Chapter 10 we investigate if, how and to what extent radiomics and circulating free tumor DNA can improve the existing SOURCE prediction models, by applying state of the art methodologies.

Finally, causality is currently a problem within the realm of clinical prediction modelling. In daily clinical practice, clinicians and patients might be interested in what the effect of a particular treatment is on survival and use prediction models to help determine a treatment strategy with optimal survival outcomes. However, due to the observational nature of the data on which the models were trained, this is unfortunately not possible. In the data on which the model was trained it is only known which treatment the patient actually received. Hence it is unknown what the effect of a different treatment would have been if the patient in the training data had been given a different treatment, which is commonly referred to as a counterfactual.

In the final chapter of this thesis, Chapter 11, we introduce a new method of prediction that combines the field of causal inference with personalized prediction called counterfactual prediction. We demonstrate this in a proof of concept study where we create a fully parametric counterfactual survival prediction model based on observational data for patients with esophageal cancer who have the option to either follow potentially curative treatment according to the CROSS treatment regimen or definitive chemoradiotherapy.

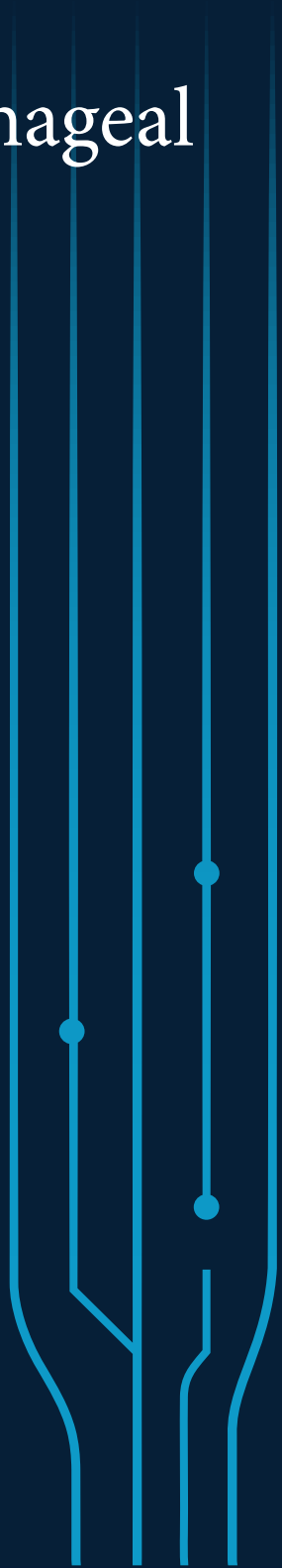
References

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229-263. doi:10.3322/caac.21834
2. IKNL. Nederlandse Kankerregistratie (NKR). iknl.nl/nkr-cijfers. Published online 2021.
3. Obermannová R, Alsina M, Cervantes A, et al. Oesophageal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Annals of Oncology. 2022;33(10):992-1004. doi:10.1016/j.annonc.2022.07.003
4. Lordick F, Carneiro F, Cascinu S, et al. Gastric cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Annals of Oncology. 2022;33(10):1005-1020. doi:10.1016/j.annonc.2022.07.004
5. Pape M, Vissers PAJ, Dijksterhuis WPM, et al. Comparing treatment and outcomes in advanced esophageal, gastroesophageal junction, and gastric adenocarcinomas: a population-based study. Ther Adv Med Oncol. 2023;15. doi:10.1177/17588359231162576
6. van den Boorn HG, Stroes CI, Zwinderman AH, et al. Health-related quality of life in curatively-treated patients with esophageal or gastric cancer: A systematic review and meta-analysis. Crit Rev Oncol Hematol. 2020;154(August):103069. doi:10.1016/j.critrevonc.2020.103069
7. van Kleef JJ, ter Veer E, van den Boorn HG, et al. Quality of Life During Palliative Systemic Therapy for Esophagogastric Cancer: Systematic Review and Meta-Analysis. JNCI: Journal of the National Cancer Institute. 2020;112(1):12-29. doi:10.1093/jnci/djz133
8. Stiggelbout AM, Van Der Weijden T, De Wit MPT, et al. Shared decision making: Really putting patients at the centre of healthcare. BMJ (Online). 2012;344(7842). doi:10.1136/bmj.e256
9. Donnelly CB, Wotherspoon AC, Morris M, et al. A population-level investigation of cancer clinical trials participation in a UK region. European Journal of Cancer Prevention. Published online 2017:229-235. doi:10.1097/CEJ.0000000000000373
10. Geerts JFM, van der Zijden CJ, van der Sluis PC, et al. Perioperative Chemotherapy for Gastro-Esophageal or Gastric Cancer: Anthracyclin Triplets versus FLOT. Cancers (Basel). 2024;16(7). doi:10.3390/cancers16071291
11. Mahon P, Hall G, Dekker A, Vehreschild J, Tonon G. Harnessing oncology real-world data with AI. Nat Cancer. 2023;4(12):1627-1629. doi:10.1038/s43018-023-00689-7
12. Pape M, Vissers PAJ, de Vos-Geelen J, et al. Treatment patterns and survival in advanced unresectable esophageal squamous cell cancer: A population-based study. Cancer Sci. 2022;113(3):1038-1046. doi:10.1111/cas.15262
13. Dijksterhuis WPM, Verhoeven RHA, Slingerland M, et al. Heterogeneity of first-line palliative systemic treatment in synchronous metastatic esophagogastric cancer patients: A real-world evidence study. Int J Cancer. 2020;146(7):1889-1901. doi:10.1002/ijc.32580
14. Dijksterhuis WPM, Verhoeven RHA, Pape M, et al. Hospital volume and beyond first-line palliative systemic treatment in metastatic oesophagogastric adenocarcinoma: A population-based study. Eur J Cancer. 2020;139:107-118. doi:10.1016/j.ejca.2020.08.010
15. van Velzen MJM, Pape M, Dijksterhuis WPM, et al. The association between effectiveness of first-line treatment and second-line treatment in gastro-oesophageal cancer. Eur J Cancer. 2021;156:60-69. doi:10.1016/j.ejca.2021.07.026
16. Pape M, Vissers PAJ, Kato K, et al. A population-based comparison of patients with metastatic esophagogastric carcinoma between Japan and the Netherlands. J Cancer Res Clin Oncol. 2023;149(14):13323-13330. doi:10.1007/s00432-023-05111-4
17. Kalff MC, Gottlieb-Vedi E, Verhoeven RHA, et al. Presentation, Treatment, and Prognosis of Esophageal Carcinoma in a Nationwide Comparison of Sweden and the Netherlands. Ann Surg. 2021;274(5). https://journals.lww.com/annalsofsurgery/fulltext/2021/11000/presentation,_treatment,_and_prognosis_of.10.aspx

18. van den Boorn HG, Abu-Hanna A, Haj Mohammad N, et al. SOURCE: Prediction Models for Overall Survival in Patients With Metastatic and Potentially Curable Esophageal and Gastric Cancer. J Natl Compr Canc Netw. 2021;19(4):403-410. doi:10.6004/jnccn.2020.7631
19. van de Water LF, Kuijper SC, Henselmans I, et al. Effect of a prediction tool and communication skills training on communication of treatment outcomes: a multicenter stepped wedge clinical trial (the SOURCE trial). EClinicalMedicine. 2023;64. doi:10.1016/j.eclinm.2023.102244
20. Coebergh van den Braak RRJ, van Rijssen LB, van Kleef JJ, et al. Nationwide comprehensive gastro-intestinal cancer cohorts: the 3P initiative. Acta Oncol (Madr). 2018;57(2):195-202. doi:10.1080/0284186X.2017.1346381
21. G.Y. Ku, M. Di Bartolomeo, E. Smyth, I. Chau, H. Park, S. Siena, S. Lonardi, Z.A. Wainberg, J.A. Ajani, J. Chao10, F. Barlaskar, Y. Kawaguchi, A. Qin, J. Singh, G. Meinhardt EVC. Updated analysis of DESTINY-Gastric02: A phase II single-arm trial of trastuzumab deruxtecan (T-DXd) in western patients (Pts) with HER2-positive (HER2+) unresectable/metastatic gastric/gastroesophageal junction (GEJ) cancer who progressed on or after tra. Annals of Oncology (2022) 33 (suppl_7): S555-S580 101016/annonc/annonc1065. Published online 2022.

Part I

Epidemiology of esophageal and gastric cancer



Chapter 1

Treatment and survival of patients with gastric and esophageal cancer in the Netherlands and Belgium: a population-based comparison

Steven C. Kuijper*, Benthe H. Doeve*, Geert Silversmit, Lien van Walle, Philippe Nafteux, Camiel Rosman, Pauline A. J. Vissers, Paul Jeene, Laurens Beerepoot, Sarah Derks, Amine Karimi, Maarten F. Bijlsma, Hanneke W.M. van Laarhoven, Rob H.A. Verhoeven.

*Shared first authorship

Under review at the International Journal of Cancer

Abstract

Background. Despite demographic and socio-economic similarities between the Netherlands and Belgium, differences in survival for patients with esophageal and gastric cancer between the two countries exist. This population-based study investigated these survival differences with the aim to identify explaining factors.
Methods. We analyzed data from the Netherlands Cancer Registry (N=26,980) and the Belgian Cancer Registry (N=15,097) for patients diagnosed with esophageal or gastric cancer between 2010-2016. Survival differences were examined using relative survival, stratified for tumor location and stage. Factors that potentially mediated the survival differences were tested via multivariable excess hazard models controlling for baseline population mortalities.
Results. Five-year relative survival probability was significantly lower for patients from the Netherlands with esophageal cancer (EC) or gastric cancer (GC) (EC: 21% (20%-22%), GC: 20% (19%-21%)) compared to Belgian patients (EC: 24% (23%-25%), GC: 27% (25%-29%)). Across tumor stages, relative survival of patients with esophageal and gastric cancer was lower among patients from the Netherlands. Multivariable excess hazard modeling showed that survival differences of patients with stage IV disease disappeared when adjusting for treatment. This was not observed among patients with stage I – III gastric cancer, where survival differences remained after adjustment for treatment.
Conclusion. Survival of patients with esophageal cancer and gastric cancer is generally higher in Belgium. For patients with stage IV disease, this difference can most likely be attributed to differences in treatment. Although treatment explains part of the survival differences for patients with curable disease, the cause remains largely unknown.

Introduction

Survival of patients diagnosed with esophageal and gastric cancer is highly heterogenous. In patients with non-metastatic gastroesophageal cancer, median survival time is around two years in the Netherlands and varies depending on tumor-intrinsic factors, clinical characteristics such as performance status, and treatment. [1] For patients with metastatic disease, median survival time is around 4 months but can improve to 1,5 years depending on tumor-intrinsic factors, performance status and treatment. [2] In addition to the large heterogeneity of survival among patients, the EUROCARE-5 study demonstrated that there is also a large international variability in survival. Five-year age-standardized relative survival ranged from 5.7% in Lithuania to 21.8% in Belgium for patients with esophageal cancer and from 11.9% in Bulgaria to 34.5% in Iceland for patients with gastric cancer. In the Netherlands, relative survival of patients with esophageal cancer was 8.8 percentage point lower than in Belgium. Similarly, the 5-year age-standardized relative survival for gastric cancer patients was 30.5% in Belgium compared to 20.4% in the Netherlands. [3]
A study by Claassen et al [4] also found relatively superior survival of patients with metastasized gastric cancer in Belgium compared to the Netherlands. In non-metastasized gastric cancer, median overall survival was 18 months in the Netherlands compared to 28 months in Belgium. [5] Interestingly, the Netherlands and Belgium are two western European countries that are similar in a number of demographic characteristics such as socio-economic status [6], tobacco consumption [7], obesity and diet [8]. The observed differences in survival can potentially be explained by several other factors, such as differences in the treatment received, histology, genetics, or tumor stages between the countries. [3] Currently, it is unknown which factors contribute to the observed differences in survival between the Netherlands and Belgium.

This study aimed to compare survival using real-world, population-based, individual-level data of patients with esophageal and gastric cancer diagnosed and treated in the Netherlands and Belgium, and to identify key factors that could explain differences in survival.

Methods

Patients

Patients diagnosed with gastric, gastro-esophageal junction, and esophageal carcinoma between 2010-2016 in the Netherlands and Belgium were identified from the Netherlands Cancer Registry (NCR) and the Belgium Cancer Registry (BCR) respectively. In total, 26,980 individual patients were identified from the NCR and 15,097 individual patients were identified in the BCR.

The NCR is a nationwide database that covers the total population of 17 million in 2016 and includes all patients diagnosed with cancer. Patient identification relies on notifications from the Netherlands' national network and registry of histopathology and cytopathology (PALGA). [9] Additionally, any non-pathological verified tumors were included via the Dutch Hospital Database. Trained data managers of the Netherlands Comprehensive Cancer Organization (IKNL) regularly extract information on diagnosis, tumor stage, and treatment directly from each patients' electronic medical records and add this to the NCR.

The BCR is also a nationwide cancer registry, covering the total population of 11.4 million. Data managers hired by hospitals extract the required data from the medical files and routinely report every new diagnosis to the BCR. Independently, the pathology laboratories supply the BCR with their data via the pathology network. Patients with cancer are registered with their unique national Social Security Identification Number, which enables linkage with vital status and date of death, as well as linkage with administrative databases such as the Intermutualistic Agency (IMA) to obtain details on reimbursed diagnostic/therapeutic procedures and pharmaceuticals. IMA data are not directly linked to a specific diagnosis, therefore specific timeframes around the diagnosis date are applied to consider the reimbursed treatment as treatment for the cancer.

Coding and classifications

Primary tumor location was coded as either stomach (C16.1-C16.9) or esophagus (including the gastro-esophageal-junction; international classification of diseases for oncology: C15.0-C16.0). Both adenocarcinomas and squamous cell carcinomas were included as well as carcinoma not otherwise specified. Carcinoma in situ was excluded for this study. Tumors were staged using the International Union Against Cancer (UICC) TNM classification 7th edition. Simplified clinical tumor stages (I, II, III, IV, X) were used in all analyses and were used in addition to tumor location as a stratifying factor. For patients with gastric cancer, a cT4 tumor without specification of a and b, was coded as a cT4a tumor.

For patients with esophageal cancer, treatment was mutually exclusively classified as endoscopic resection, surgical resection only, surgical resection with (neo)adjuvant chemotherapy, neoadjuvant chemoradiotherapy followed by resection with or without adjuvant therapy, chemoradiotherapy not followed by resection, systemic therapy only, radiotherapy only and other or unknown. For patients with gastric cancer, treatment was classified as endoscopic resection, surgery only, neoadjuvant chemotherapy and surgery (with or without adjuvant chemotherapy), systemic therapy only and other or unknown. Classification was decided on the most invasive treatment. For example, if patients both underwent endoscopic resection and surgical resection, treatment was classified as surgical resection. The category

'Other or unknown' included patients that could not be categorized in one of the beforementioned treatments, including best supportive care.

Statistical analysis

Differences in treatment between the Netherlands and Belgium were visualized in bar plots per tumor stage and were interpreted based on visual inspection. We included a category of unknown treatment since any treatment, including best supportive care, will impact survival despite it not being available to correct for in our analysis. We estimated 5-year survival using relative survival framework. [10] This method adjusts for the baseline population mortality conditional on sex, year of birth, and country. Life tables for the relevant population demographics were used to control for these factors. Differences between relative survival curves of countries were statistically tested with a log-rank-type test. [11] We opted to include all patients despite missing variables to limit selection bias and to increase representativeness of these real-world data, particularly given the difference in proportion of missing data between both countries and a previous study that showed a difference in survival in a cohort of only known stages compared to the cohort that included all patients. [5]

We further investigated differences in relative survival between the Netherlands and Belgium through excess hazard modeling. Excess hazard modeling can be used to compare the impact of a specific disease on survival between two countries by quantifying the additional risk of death attributable to the disease in each country, controlling for the baseline risk of mortality in the general populations. [12] This allowed us to estimate the extent to which the survival differences could be explained by demographic, clinical, and treatment variables while controlling for baseline population mortality. An excess hazard ratio of 0.89 in favor of Belgium indicates that the excess mortality rate was 11% lower in Belgium compared to the Netherlands.

Flexible parametric models were applied to model the excess hazard up to 8 year since diagnosis, using the R package mexhaz. [11] B-spline were used to specify the baseline excess hazard curve. In a first step, an appropriate baseline excess hazard function was constructed by optimizing the parameters for the B-spline (spline degree, number of spline knots). The flexible curve was visually compared to a step function, in order to exclude overfitting or a too flexible function. This was performed for the total population, stratified to tumor sublocation and stratified to tumor stages.

Three different excess hazard models were fitted to the data to determine factors which mediated survival differences. The first model only contained country as a predictor variable in the model. The second model contained country and clinical variables (sex, age, and for esophageal cancer also histological subtype), and the third model included country, clinical variables, and treatment (categories as specified in section 'Coding and classifications'). Age was modelled as a non-proportional effect. The excess hazard model estimated excess hazard ratios for all the parameters. By sequentially fitting these three models and testing the respective excess hazard ratios for significance, we could observe which factor had the largest impact on the difference in relative survival between the Netherlands and Belgium. Forest plots were created to visualize the excess hazard ratios.

All reported confidence intervals were two-sided 95% confidence intervals and all p-values <0.05 were considered statistically significant. All analyses were performed in R version 4.2.2.[13]

Results

Baseline characteristics

Baseline characteristics per disease and country are listed in Table 1. There were fewer patients from the Netherlands with esophageal cancer under 59 years of age and there were relatively fewer patients from the Netherlands with gastric cancer that were over 80 years of age. The proportion of squamous cell carcinomas of the esophagus was considerably lower in the Netherlands (NL: 25.0% vs BE: 37.5%). The number of patients with a T1 tumor was significantly lower in the Netherlands compared to Belgium (esophageal cancer: NL: 4.4% vs BE: 7.6%, gastric cancer: NL 3.7% vs BE: 5.8%) (Table 1) and the number of patients with T1N0 esophageal cancer was also lower (NL: 27%; BE: 50%) whilst the proportion of patients with T2N0 was higher (NL: 73%; BE: 50%) (Supplementary Table 1). In addition, the number of patients with an unknown cT, cN and cM was significantly lower in the Netherlands. The difference was particularly large for cM; 0.3% of patients with esophageal cancer and 0.8% of patients with gastric cancer in the Netherlands had an unknown cM-stage compared to 32.5% and 41.9% in Belgium.

Treatment

There were notable differences in the distribution of type of treatment. For patients with esophageal cancer in the Netherlands, fewer patients received surgery only compared to patients in Belgium with stage I (NL: 13%; BE: 39%), II (NL: 4%; BE: 16%), III (NL: 2%; BE: 6%), IV (NL: 0.1%; BE: 1%) and X (NL: 4%; BE:11%). In stage I, a lower percentage of patients with esophageal cancer was treated with (neo)adjuvant chemotherapy with surgery in the Netherlands (NL: 3%; BE: 15%). Concurrently, there was a higher proportion of patients who received neoadjuvant chemoradiotherapy followed by surgery with(out) adjuvant therapy in clinical stage I (NL: 33%; BE: 3%), II (NL: 53%; BE: 24%) and III (NL: 39%; BE: 25%) compared with Belgium (Figure 1A, Supplementary Table 2). The same trend could be observed in patients with stage I gastric cancer; there was a lower frequency of surgical resection only in the Netherlands compared to Belgium (NL: 37; BE: 50%) and a higher frequency of (neo)adjuvant chemotherapy with surgery (NL: 36%; BE: 28%) (Figure 1B, Supplementary Table 3). On the other hand, there was a lower frequency of (neo)adjuvant chemotherapy with surgery in the Netherlands compared to Belgium in gastric cancer stage II (NL: 47%; BE 51%) and stage III (NL: 28%; BE: 50%). There was a concurrent increase in the patients treated with Other or unknown treatment in the Netherlands (Stage II: NL: 20%; BE: 11%, Stage III: NL: 40%; BE: 12%).

Patients with stage IV disease in the Netherlands received less systemic therapy only compared to patients in Belgium (esophageal cancer: NL: 33%; BE: 50%, gastric cancer: NL 35%; BE: 53%). Patients with stage IV esophageal and gastric cancer received less (neo) adjuvant chemotherapy with surgery (esophageal cancer: NL: 1%; BE: 3%, gastric cancer: NL 2%; BE 5%). Patients with stage IV esophageal cancer in the Netherlands received less chemoradiotherapy not followed by resection (NL: 6%; BE: 13%).

Concurrently, patients from the Netherlands with stage IV gastric cancer received more Other or unknown treatment (gastric cancer: NL: 60% vs BE: 37%), whilst patients in the Netherlands with stage IV esophageal cancer more often received only radiotherapy (NL: 25%; BE: 2%) or Other or unknown treatment (NL: 34%; BE: 26%). In Supplementary Figure 1 we show that in the Netherlands the latter treatment category largely consists of Best Supportive Care.

Table 1. Characteristics of patients with Esophageal and Gastric cancer from the Netherlands and Belgium identified in the Netherlands Cancer Registry and the Belgium Cancer Registry, respectively.

	Esophagus			Gastric		
	Netherlands	**Belgium**	**p**	**Netherlands**	**Belgium**	**p**
N	18294	9617		8686	5480	
Sex (%)						
Male	13652 (74.6)	7206 (74.9)	0.588	5258 (60.5)	3183 (58.1)	0.004
Female	4642 (25.4)	2411 (25.1)		3428 (39.5)	2297 (41.9)	
Age (%)						
<=59	3742 (20.5)	2315 (24.1)	<0.001	1497 (17.2)	910 (16.6)	<0.001
60-79	11445 (62.6)	5481 (57.0)		4874 (56.1)	2593 (47.3)	
>=80	3107 (17.0)	1821 (18.9)		2315 (26.7)	1977 (36.1)	
Histology (%)						
Adenocarcinoma	13062 (71.4)	5818 (60.5)	<0.001	8363 (96.3)	5344 (97.5)	<0.001
Squamous cell	4567 (25.0)	3603 (37.5)		323 (3.7)	136 (2.5)	
Other	665 (3.6)	196 (2.0)		-	-	
cT (%)						
1	799 (4.4)	727 (7.6)	<0.001	321 (3.7)	318 (5.8)	<0.001
2	4425 (24.2)	1192 (12.4)		2144 (24.7)	527 (9.6)	
3	6759 (36.9)	3568 (37.1)		1293 (14.9)	1223 (22.3)	
4	1216 (6.6)	522 (5.4)		-	-	
4a	-	-		303 (3.5)	344 (6.3)	
4b	-	-		794 (9.1)	55 (1.0)	
x	5095 (27.9)	3608 (37.5)		3831 (44.1)	3013 (55.0)	
cN (%)						
0	5246 (28.7)	1776 (18.5)	<0.001	3758 (43.3)	1165 (21.3)	<0.001
1	6031 (33.0)	2906 (30.2)		1772 (20.4)	992 (18.1)	
2	4122 (22.5)	1061 (11.0)		1309 (15.1)	371 (6.8)	
3	935 (5.1)	363 (3.8)		172 (2.0)	145 (2.6)	
x	1960 (10.7)	3511 (36.5)		1675 (19.3)	2807 (51.2)	
cM (%)						
0	11740 (64.2)	4414 (45.9)	<0.001	5128 (59.0)	1948 (35.5)	<0.001
1	6495 (35.5)	2081 (21.6)		3491 (40.2)	1237 (22.6)	
x	59 (0.3)	3122 (32.5)		67 (0.8)	2295 (41.9)	

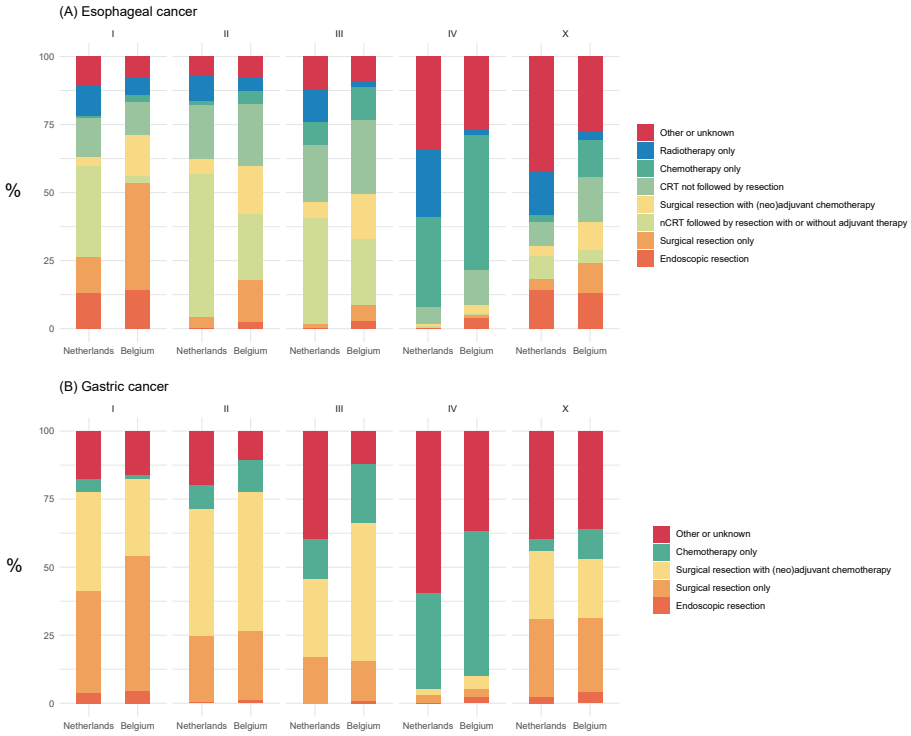


Figure 1. Distribution of treatment categories per clinical stage for patients (A) esophageal cancer and (B) gastric cancer in the Netherlands and Belgium.

Relative survival

Five-year relative survival of all patients with esophageal cancer was significantly lower in the Netherlands (21%, 95% CI: (20%-22%)) compared to Belgium (24%, 95% CI: (23%-25%); p<0.001) (Figure 2A). Among patients in stage I (5-year RS: NL: 47% (44%-50%); BE: 53% (48%-58%); p=0.03), stage IV (1-year RS: NL: 21% (20%-23%); BE: 33% (31%-35%); p<0.001) and stage X (5-year RS: NL: 24% (22%-26%); BE: 25% (23%-26%); p < 0.001) relative survival of patients from the Netherlands was also significantly lower compared to patients from Belgium (Figure 2B, 2E, 2F). Among patients in stage II and stage III, no significant differences in 5-year relative survival were observed, although the relative survival curve of patients with stage II disease in the Netherlands was consistently above the curve of patients from Belgium (Figure 2C, 2D).

Patients from the Netherlands with gastric cancer had a significantly lower 5-year relative survival compared to patients from Belgium (5-year RS: NL: 20% (19%-21%); BE: 27% (25%-29%); p<0.001)(Figure 3A). Across all known tumor stages, survival for patients from the Netherlands was lower than for patients from Belgium: stage I (5-year RS: NL: 50% (46%-54%); BE: 69% (62%-77%); p<0.001), stage II (5-year RS: NL: 31% (27%-35%); BE: 38% (33%-43%); p=0.001), stage III (5-year RS: NL: 12% (0.09%-15%); BE: 26% (19%-35%); p<0.001), stage IV (1-year RS: NL: 15% (14%-17%); BE: 27% (24%-29%); p<0.001) (Figure 3B-E). In contrast, patients with gastric cancer in an unknown stage from the Netherlands had a higher survival (5-year RS: NL: 30% (28%- 33%); BE: 28% (26%- 31%); p=0.036).

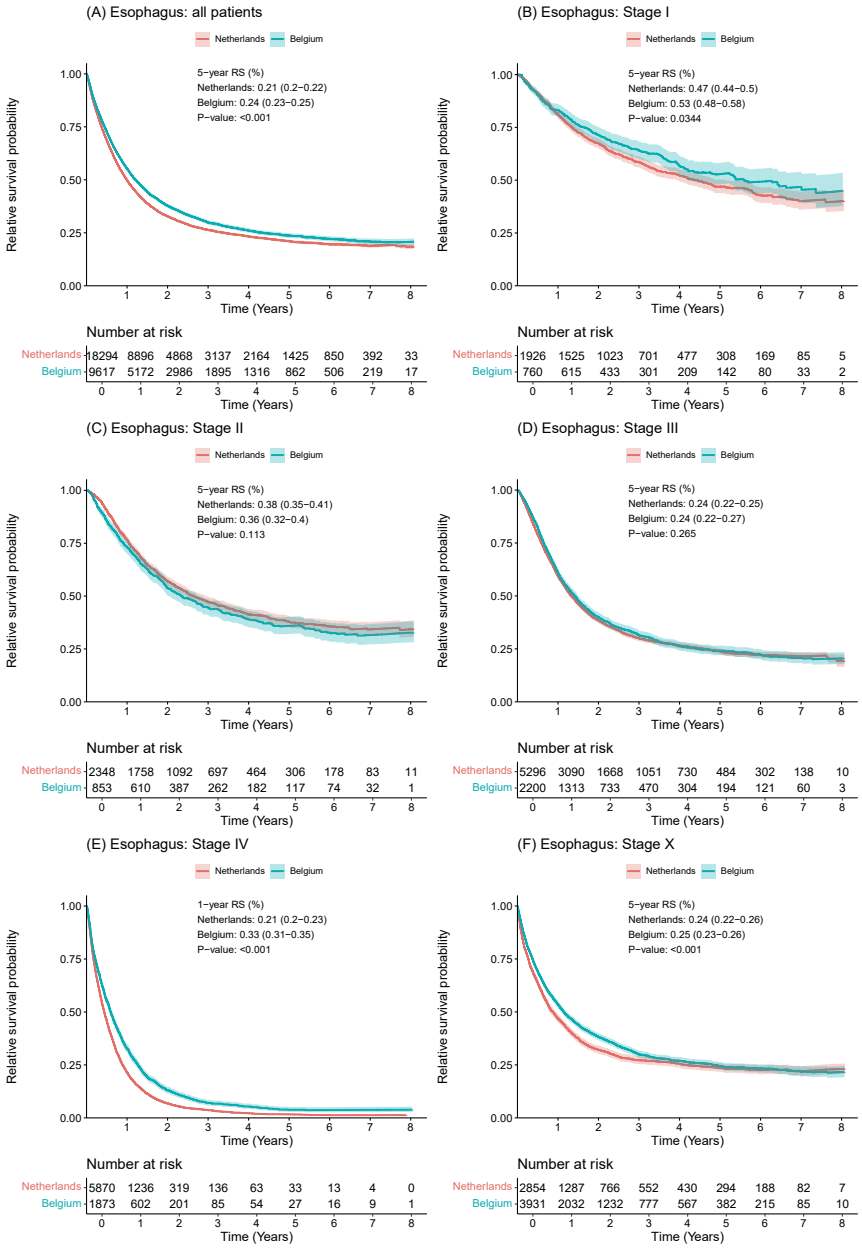


Figure 2. Relative survival probability curves for all patients with esophageal cancer (A), and for patients with stage I (B), stage II (C), stage III (D), stage IV (E), and stage X (F). Relative survival was adjusted for baseline population mortality conditional on sex, year of birth, and country. P-value denotes a log-rank-type test.

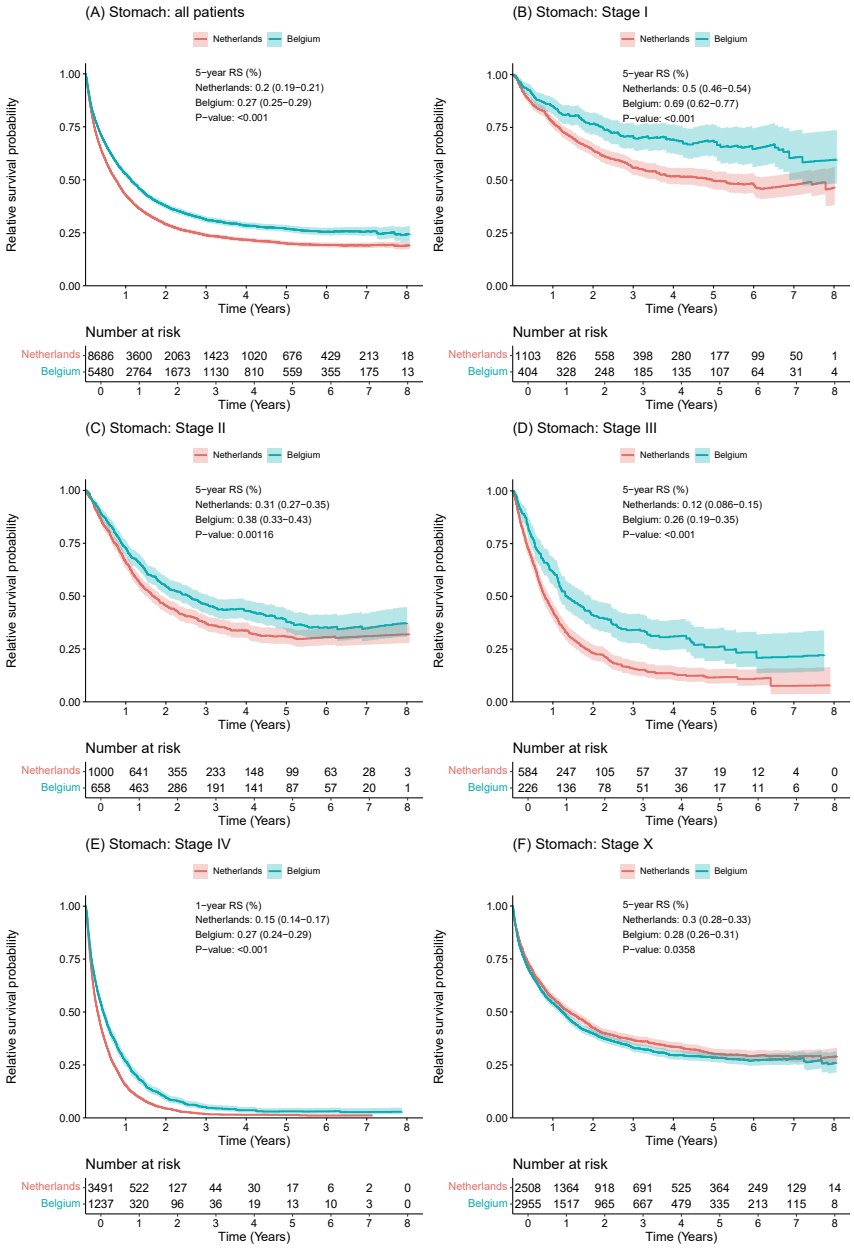


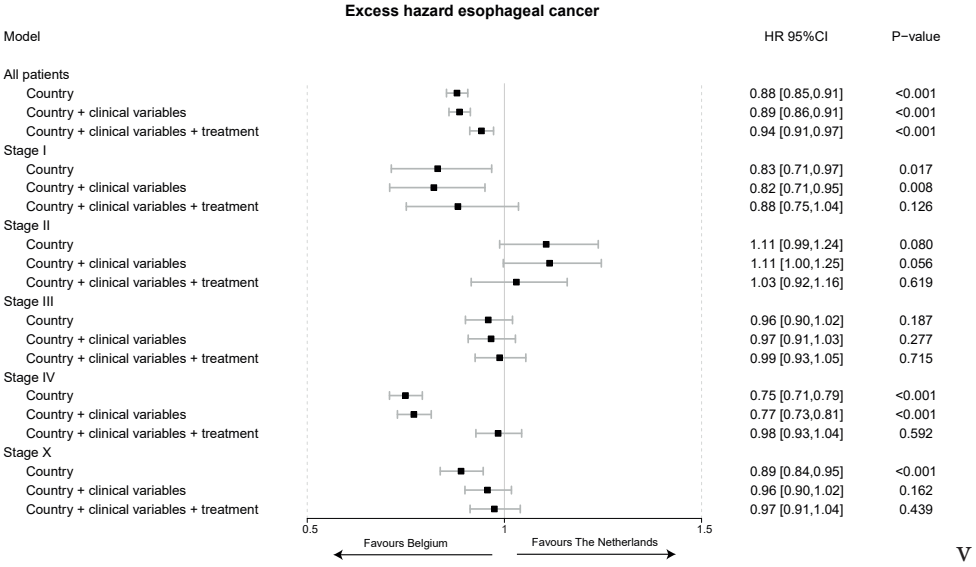
Figure 3. Relative survival probability curves for all patients with gastric cancer (A), and for patients with stage I (B), stage II (C), stage III (D), stage IV (E), and stage X (F). Relative survival was adjusted for baseline population mortality conditional on sex, year of birth , and country. P-value denotes a log-rank-type test.

Excess hazard

The excess hazard baseline function up to 8 years since diagnosis was modelled with B-splines of degree 3. For esophageal cancer all models used 2 knots, at 0.5 and 4 year except for stage IV for which one knot at 2 year sufficed. For gastric cancer, most models used 2 knots at 0.4 and 1 year, except for stage I which needed an additional knot at 3 years to prevent an upward bent of the spline beyond 5 years.

For all patients with esophageal cancer, the excess hazard ratio favored Belgium when controlled for sex, age, histology and treatment. However, while excess mortality rate was 12% lower in Belgium without correction (HR 0.88, 95% CI (0.85-0.91)), this was only 6% when also correcting for treatment. (0.94 95% CI: (0.91-0.97))(Figure 4A). When analyzing stages separately, the excess hazard ratio favored Belgium for patients with stage I disease when controlled for clinical variables (0.82 (0.71-0.95)) and only shifted slightly toward not favoring any country when also correcting for treatment (0.88 (0.75-1.04)). For patients with stage II disease, the excess mortality rate was 11% lower in the Netherlands without correction of any variables, which decreased to 3% when correcting for treatment. The confidence interval of both stage I and stage II were unfortunately quite large. In stage III, there was no difference in excess mortality rates. For patients with stage IV disease, the excess hazard favored Belgium when controlled for clinical variables (0.77 (0.73-0.82)) but not when also correcting for treatment (0.98 (0.93-1.04)). For stage X, the excess hazard ratio favored Belgium (0.89 (0.84-0.95)) but not when correcting for clinical variables (0.96 (0.90-1.02)). For all patients with gastric cancer, the excess hazard ratio favored Belgium when correcting for both clinical variables and treatment. However, while excess mortality rate was 21% lower in Belgium without any correction (0.79(0.76-0.82)), this was 15% lower when correcting for treatment (0.85 (0.81-0.88)) (Figure 4B). When analyzing stages separately, excess hazard ratio also favored Belgium in stage I, II and III with correction for both clinical variables and treatment (Stage I: 0.56 (0.44-0.71); Stage II: 0.83 (0.72-0.95); Stage III: 0.79 (0.64 – 0.97)). For stage III, however, the excess mortality rate without correction in Belgium was 41% lower (0.59 (0.49-0.72)) compared to 21% when also correcting for treatment (0.79 (0.64-0.97)). In stage IV, the excess hazard ratio favored Belgium when correcting for clinical variables (Stage IV: 0.72 (0.68-0.77)) but not when also correcting for treatment (0.95 (0.89-1.02)). In stage X, the excess hazard ratio favored the Netherlands when correcting for clinical variables (1.08 (1.01-1.16)) but not when also correcting for treatment (1.01 (0.94-1.08)).

(A)



V

(B)

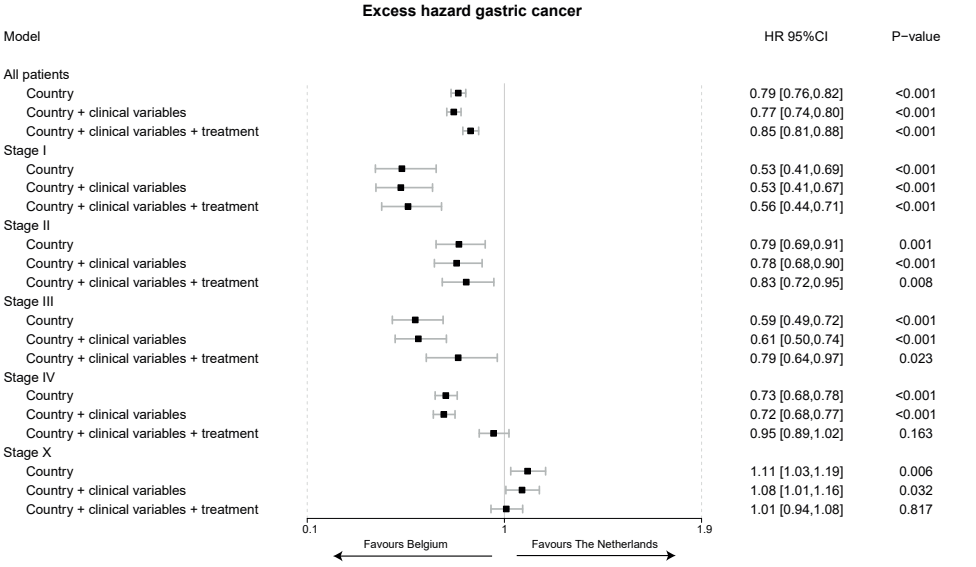


Figure 4. Forest plot depicting excess hazard ratios for (A) esophageal cancer and (B) gastric cancer. The sequential models first included country with the Netherlands as a reference, then country and clinical variables (EC: sex, age, histology, GC: sex, age) sex and lastly country, clinical variables, and the treatment categories as defined in figure 1.

Discussion

This study described notable differences in treatment and relative survival of patients with esophageal and gastric cancer between the Netherlands and Belgium. Survival was lower in the Netherlands for patients with gastric cancer in any stage and for patients with stage I and IV esophageal cancer. For patients with stage IV disease, the differences in relative survival appeared to be largely explained by differences in treatment. Treatment only partly mediated a difference in survival between the Netherlands and Belgium for patients with stage I and stage II esophageal cancer, and stage III gastric cancer. Despite correction for clinical variables and treatment, survival of patients with potentially curable gastric cancer remained lower in the Netherlands.

Relative survival was lower for patients from the Netherlands with both esophageal and gastric cancer in stage IV, but these differences disappeared when correcting for treatment. Patients with stage IV disease in Belgium received more local radical treatment such as chemoradiotherapy not followed by resection and surgical resection with (neo)adjuvant chemotherapy. In addition, patients were less likely to be treated with systemic therapy in the Netherlands. A similar low frequency of treatment with systemic therapy in the Netherlands compared to Belgium was previously observed in patients with malignant mesothelioma.[14] In the Netherlands, the lower frequency of systemic therapy was accompanied with a higher frequency of best supportive care for both gastric and esophageal metastatic disease. Best supportive care includes 'radiotherapy only' as this does not improve survival and is generally used for relieving dysphagia in esophageal cancer. [15-17] The high frequency in treatment with systemic therapy could account for the observed better survival of patients with stage IV disease in Belgium, as systemic therapy has been shown to improve survival compared to best supportive care for patients with gastric and esophageal cancer. [18-21]

In the Netherlands, significant emphasis is placed on "appropriate care" which aims to deliver effective patient care by avoiding unnecessary interventions, particularly in case cure cannot be achieved. This approach seeks to minimize treatments that offer limited benefits and pose potential side effects which affect quality of life of the patients, while also considering the financial implications for society. [22] As part of this approach, end of life discussions occur relatively more frequently in the Netherlands compared to for example Belgium.[23] These end of life discussions are associated with a choice for less invasive treatment, including less palliative systemic therapy. [24] However, given the observed differences in relative survival between the Netherlands and Belgium, the question arises whether the Dutch caution to prescribe palliative systemic therapy is unduly restrictive. Unfortunately, in our study we do not have data available on differences in quality of life between patients from the Netherlands and Belgium. However, two systematic reviews demonstrated that quality of life remains stable during an extended period of time during chemotherapy for metastatic esophagogastric cancer compared to best supportive care. [25, 26] Real-world data from the Netherlands has actually shown an improvement of quality of life during chemotherapy, while quality of life deteriorated at disease progression irrespective of treatment with chemotherapy. [27] Recent treatment advances with addition of nivolumab to chemotherapy in the first line has shown stable or even improved health-related quality of life, with a lower chance of definitive deterioration. [28] Also, importantly, real-world data indicate that patients treated at Dutch centers with a high volume of systemic therapy prescriptions experience better survival outcomes, while these centers administer less systemic therapy in

the last three months of life. [29-31] This observation suggests that these centers are proficient in selecting patients who are most likely to derive significant survival benefits from chemotherapy. Taken together, the reported survival benefits of systemic therapy and its minimal negative impact on quality of life challenge the Dutch reserve regarding the prescription of palliative systemic therapy in metastatic gastric and esophageal cancer.

Relative survival was higher for patients with stage I esophageal cancer in Belgium, a difference that decreased when correcting for treatment. We observed that patients with stage I disease in Belgium were more frequently treated with surgery only, whilst in the Netherlands these patients were also treated with neoadjuvant chemoradiotherapy followed by resection. The different treatment regimens can possibly be explained by a difference in incidence of T1N0 and T2N0 tumors, which together make up stage I tumors. In Belgium, proportionally more T1N0 tumors were diagnosed, for which the standard of care is endoscopic resection or surgical resection only, whereas in the Netherlands more T2N0 tumors were diagnosed. [32] The standard of care for this subgroup is controversial, with evidence for and against neoadjuvant treatment in addition to surgery. It seems in the Netherlands neoadjuvant treatment for this subgroup is frequently prescribed. Literature suggests that T2 tumors have a worse prognosis than T1 tumors. [33, 34] This may be an explanation for the poorer survival of patients with stage I esophageal cancer in the Netherlands. Alternatively, patients with T2N0 could experience an increase in postoperative mortality after treatment with neoadjuvant chemoradiotherapy, as suggested by Marriette et al. [35] It is important to note that differences in treatment do not fully mediate the difference in survival we found. There was still a decrease of excess mortality of 12% in Belgium compared to the Netherlands after correcting for treatment. The large confidence intervals, likely a result of fewer survival events in a smaller subgroup of stage I patients, make it harder to draw definitive conclusions.

The Dutch and Belgian healthcare systems, while both highly regarded, exhibit distinct differences in structure, operation, accessibility, and specialist referral processes. Accessibility in the Netherlands requires patients to first see a general practitioner (GP) who acts as a gatekeeper and provides referrals to specialists if needed. [36] In contrast, accessibility in Belgium is more direct, as patients have the freedom to consult specialists without a referral from a GP. Possibly, this quick access may result in less diagnostic delay in Belgium, which could serve as a potential explanation for the higher frequency of T1 tumors observed. It is noteworthy that centralized surgical care in high-volume specialized centers was implemented for esophagus surgery in July 2019 in Belgium compared to 2013 in the Netherlands. [37, 38] This makes the higher survival in Belgium of curable disease even more striking.

Survival of patients with gastric cancer was higher in Belgium across curable stages I, and II irrespective of clinical characteristics and treatment. For stage III, treatment only partly mediated the difference in survival between the two countries. A similar result was found in an article investigating overall survival in multiple countries over time. [5] Here correction for surgery did not improve overall survival in the Netherlands and Belgium after 2010. Therefore, the factor that mediated the higher relative survival in Belgium compared to the Netherlands remains unknown. One possible explanation is a tumor-intrinsic factor causing patients in the Netherlands to be more resistant to treatment. Recent research has shown that the proportion of diffuse subtype of gastric cancer is increasing in the Netherlands. [39, 40] Survival for this subtype is poor, and has not improved much in recent years

despite treatment advances. Currently, it is unknown what the incidence is of this subgroup in Belgium.

In the context of this study, it is important to consider that a large number of patients from Belgium had an unknown tumor stage (X) compared to patients from the Netherlands. Almost 33% of the Belgian patients with esophageal cancer and 42% for gastric cancer had a unknown cM category. This is most likely the result of differences in method of cancer registration and data retrieval. In the stage X groups, survival was lower for patients with gastric cancer in Belgium, but higher among patients with esophageal cancer. Had the tumor stage been known, it would likely not have changed outcomes for the other staging subgroups of patients from the Netherlands, as only a very small fraction was unknown. For Belgian patients, however, outcomes might have changed due to the relatively large number of unknowns. For Belgian patients with esophageal and gastric cancer with stage X, five-year relative survival corresponded roughly with Stage III. This might indicate that survival of potentially curable clinical stage was overestimated whilst survival of palliative clinical stage was underestimated by the exclusion of the stage X patients. It also calls into question whether diagnostic approach to staging in the Netherlands and Belgium can be compared accurately, which is important to consider for our results per stage subgroup.

This study is the first population-based direct comparison between the Netherlands and Belgium regarding survival and treatment using real-world data of patients with esophageal and gastric cancer. The relative survival and excess hazard modelling allowed us to investigate survival differences and factors that mediated them while correcting for baseline population mortality, conditional on sex, age, and country. A limitation of this study lies in the limited number of clinical variables available for correction in the excess hazard models such as performance status or her2neu positivity between countries. These factors are known to impact survival of patients with esophageal and gastric cancer. [41] The lack of smoking status in the baseline characteristics is another limitation. The proportion of patients with esophageal squamous cell carcinoma was higher in Belgium. This type of cancer is notoriously related to smoking. Smoking status influences the expected mortality rate, and therefore the survival of a population with a high proportion of smokers is lower than the survival of the general population which is used in relative survival analysis. [42] Interestingly enough, for our results that would mean that the higher survival for patients in Belgium is underestimated compared to patients in the Netherlands. However, it has been reported that the difference between overall survival and cancer-specific survival in esophageal cancer is very small [43], which disputes the impact of smoking status on survival in patients with esophageal cancer. In any case, future studies should aim to incorporate more clinically relevant factors to investigate survival discrepancies.

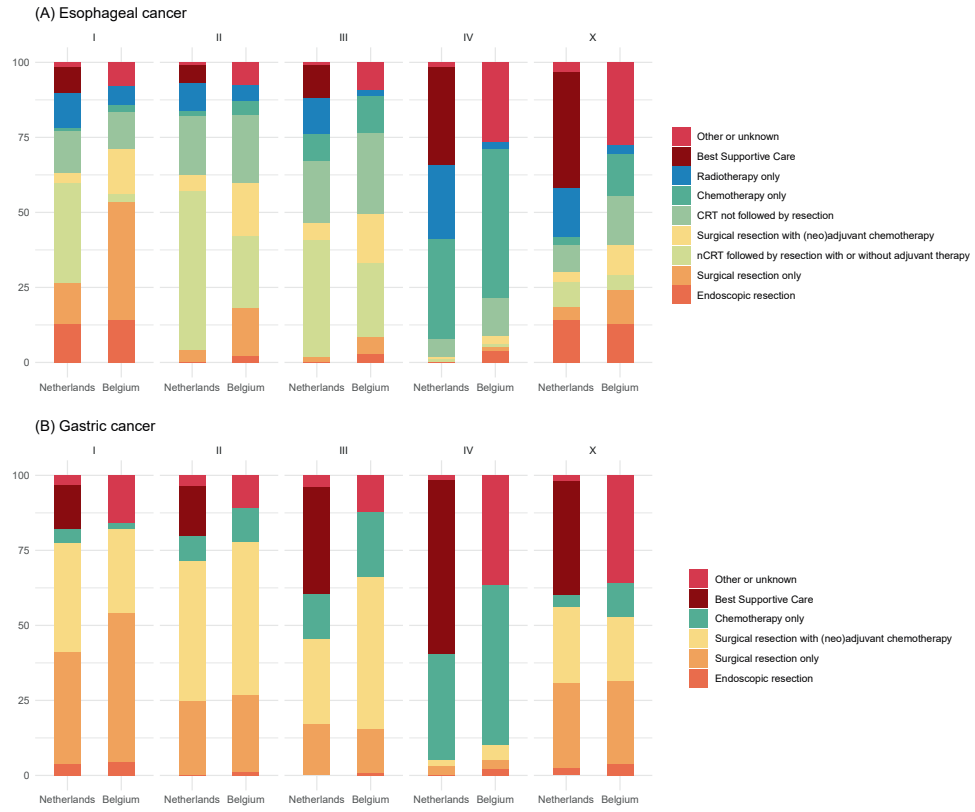
In conclusion, survival of patients with esophageal cancer and gastric cancer differs between the Netherlands and Belgium. For patients with stage IV disease, the difference in survival can most likely be attributed to differences in treatment. Although treatment explains part of the survival differences for patients with curable disease, the cause remains largely unknown.

References

1. Kuijper, S.C., et al., Trends in best-case, typical and worst-case survival scenarios of patients with non-metastatic esophagogastric cancer between 2006 and 2020: A population-based study. International journal of cancer, 2023. 153(1): p. 33-43.
2. Pape, M., et al., Beyond median overall survival: estimating trends for multiple survival scenarios in patients with metastatic esophagogastric cancer. Journal of the National Comprehensive Cancer Network, 2022. 20(12): p. 1321-1329. e4.
3. Anderson, L.A., et al., Survival for oesophageal, stomach and small intestine cancers in Europe 1999–2007: results from EUROCARE-5. European Journal of Cancer, 2015. 51(15): p. 2144-2157.
4. Claassen, Y., et al., International comparison of treatment strategy and survival in metastatic gastric cancer. BJS open, 2019. 3(1): p. 56-61.
5. Huang, L., et al., Survival trends of patients with non-metastatic gastric adenocarcinoma in the US and European countries: the impact of decreasing resection rates. Cancer Communications, 2022. 42(7): p. 648-662.
6. Eurostat. Real GDP per capita. 2024 09/07/2024 [cited 2024 12-07-2024]; Available from: https://ec.europa.eu/eurostat/databrowser/view/sdg_08_10/default/table.
7. Eurostat. Tobacco consumption statistics. 2019 [cited 2024 12-07-2024]; Available from: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Tobacco_consumption_statistics.
8. Eurostat. Overweight and obesity - BMI statistics. 2019 [cited 2024 12-07-2024]; Available from: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Overweight_and_obesity_-_BMI_statistics.
9. Casparie, M., et al., Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. Analytical Cellular Pathology, 2007. 29(1): p. 19-24.
10. Perme, M.P., J. Stare, and J. Estève, On estimation in relative survival. Biometrics, 2012. 68(1): p. 113-120.
11. Grafféo, N., et al., A log-rank-type test to compare net survival distributions. Biometrics, 2016. 72(3): p. 760-769.
12. Charvat, H. and A. Belot, Mexhaz: An R package for fitting flexible hazard-based regression models for overall and excess mortality with a random effect. Journal of Statistical Software, 2021. 98: p. 1-36.
13. Team, R.C., R: A language and environment for statistical computing. 2013.
14. Damhuis, R., et al., Treatment patterns and survival analysis in 9014 patients with malignant pleural mesothelioma from Belgium, the Netherlands and England. Lung Cancer, 2015. 89(2): p. 212-217.
15. Yang, C., et al., Interventions for dysphagia in oesophageal cancer. Cochrane Database of Systematic Reviews, 2014(10).
16. Lordick, F., et al., Gastric cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up . Annals of Oncology, 2022. 33(10): p. 1005-1020.
17. Guttmann, D.M., et al., Improved overall survival with aggressive primary tumor radiotherapy for patients with metastatic esophageal cancer. Journal of Thoracic Oncology, 2017. 12(7): p. 1131-1142.
18. Glimelius, B., et al., Randomized comparison between chemotherapy plus best supportive care with best supportive care in advanced gastric cancer. Annals of Oncology, 1997. 8(2): p. 163-168.
19. Murad, A.M., et al., Modified therapy with 5-fluorouracil, doxorubicin, and methotrexate in advanced gastric cancer. Cancer, 1993. 72(1): p. 37-41.
20. Pyrhönen, S., et al., Randomised comparison of fluorouracil, epidoxorubicin and methotrexate (FEMTX) plus supportive care with supportive care alone in patients with non-resectable gastric cancer. British journal of cancer, 1995. 71(3): p. 587-591.
21. Cunningham, D., et al., Capecitabine and oxaliplatin for advanced esophagogastric cancer.

New England Journal of Medicine, 2008. 358(1): p. 36-46.
22. Bolt, E.E., et al., Appropriate and inappropriate care in the last phase of life: an explorative study among patients and relatives. BMC health services research, 2016. 16: p. 1-11.
23. Evans, N., et al., End-of-life decisions: a cross-national study of treatment preference discussions and surrogate decision-maker appointments. PloS one, 2013. 8(3): p. e57965.
24. Wright, A.A., et al., Associations between end-of-life discussions, patient mental health, medical care near death, and caregiver bereavement adjustment. Jama, 2008. 300(14): p. 1665-1673.
25. Al-Batran, S.E. and J.A. Ajani, Impact of chemotherapy on quality of life in patients with metastatic esophagogastric cancer. Cancer: Interdisciplinary International Journal of the American Cancer Society, 2010. 116(11): p. 2511-2518.
26. Van Kleef, J.J., et al., Quality of life during palliative systemic therapy for esophagogastric cancer: systematic review and meta-analysis. JNCI: Journal of the National Cancer Institute, 2020. 112(1): p. 12-29.
27. Pape, M., et al., Long-term health-related quality of life in patients with advanced esophagogastric cancer receiving first-line systemic therapy. Supportive Care in Cancer, 2023. 31(9): p. 520.
28. Moehler, M., et al., Health-related quality of life with nivolumab plus chemotherapy versus chemotherapy in patients with advanced gastric/gastroesophageal junction cancer or esophageal adenocarcinoma from CheckMate 649. Journal of Clinical Oncology, 2023. 41(35): p. 5388-5399.
29. Dijksterhuis, W.P., et al., Hospital volume and beyond first-line palliative systemic treatment in metastatic oesophagogastric adenocarcinoma: a population-based study. European Journal of Cancer, 2020. 139: p. 107-118.
30. Mohammad, N.H., et al., Volume-outcome relation in palliative systemic treatment of metastatic oesophagogastric cancer. European Journal of Cancer, 2017. 78: p. 28-36.
31. Besseling, J., et al., Use of palliative chemotherapy and ICU admissions in gastric and esophageal cancer patients in the last phase of life: a nationwide observational study. Cancers, 2021. 13(1): p. 145.
32. Obermannová, R., et al., Oesophageal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up . Annals of Oncology, 2022. 33(10): p. 992-1004.
33. Igaki, H., et al., Prognostic evaluation of patients with clinical T1 and T2 squamous cell carcinomas of the thoracic esophagus after 3-field lymph node dissection. Surgery, 2003. 133(4): p. 368-374.
34. Ikoma, N., et al., Survival rates in T1 and T2 gastric cancer: A Western report. Journal of surgical oncology, 2016. 114(5): p. 602-606.
35. Mariette, C., 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, 2014. 32(23): p. 2416-2422.
36. Kroneman, M., et al., Netherlands: health system review. 2016.
37. van Walle, L., et al., A Population-Based Study Using Belgian Cancer Registry Data Supports Centralization of Esophageal Cancer Surgery in Belgium. Annals of surgical oncology, 2023. 30(3): p. 1545-1553.
38. SONCOS, Multidisciplinaire normering oncologische zorg in Nederland. 2012.
39. Koemans, W., et al., The metastatic pattern of intestinal and diffuse type gastric carcinoma–A Dutch national cohort study. Cancer epidemiology, 2020. 69: p. 101846.
40. van Der Kaaij, R.T., et al., A population-based study on intestinal and diffuse type adenocarcinoma of the oesophagus and stomach in the Netherlands between 1989 and 2015. European Journal of Cancer, 2020. 130: p. 23-31.
41. van den Boorn, H.G., et al., SOURCE: prediction models for overall survival in patients with metastatic and potentially curable esophageal and gastric cancer. Journal of the national comprehensive cancer network, 2021. 19(4): p. 403-410.

42. Blakely, T., et al., Bias in relative survival methods when using incorrect life-tables: Lung and bladder cancer by smoking status and ethnicity in New Zealand. International journal of cancer, 2012. 131(6): p. E974-E982.
43. Vissers, P.A., et al., The association between hospital variation in curative treatment for esophagogastric cancer and health-related quality of life and survival. European Journal of Surgical Oncology, 2023. 49(10): p. 107019.



Supplemental Figure 1. Distribution of treatment categories per clinical stage for patients with esophageal cancer (A) and gastric cancer (B) in the Netherlands and Belgium. For the Netherlands, Best Supportive Care was differentiated from other or unknown.

Supplemental table 1. Distribution of T- and N-category per stage for esophageal cancer

	Stage I		Stage II		Stage III		Stage IV		Stage X	
	Netherlands	Belgium	Netherlands	Belgium	Netherlands	Belgium	Netherlands	Belgium	Netherlands	Belgium
T1 N0	526 (27.3)	381 (50.1)	-	-	-	-	22 (0.4)	14 (0.7)	-	44 (1.1)
T1 N1	-	-	64 (2.7)	83 (9.7)	-	-	20 (0.3)	17 (0.9)	-	20 (0.5)
T1 N2	-	-	-	-	13 (0.2)	15 (0.7)	16 (0.3)	13 (0.7)	-	3 (0.1)
T1 N3	-	-	-	-	4 (0.1)	14 (0.6)	-	-	-	-
T1 Nx	-	-	-	-	-	-	7 (0.1)	13 (0.7)	127 (4.4)	110 (2.8)
T2 N0	1400 (72.7)	379 (49.9)	-	-	-	-	227 (3.9)	25 (1.3)	-	51 (1.3)
T2 N1	-	-	861 (36.7)	315 (36.9)	-	-	597 (10.2)	80 (4.3)	-	66 (1.7)
T2 N2	-	-	-	-	336 (6.3)	78 (3.5)	605 (10.3)	27 (1.4)	-	4 (0.1)
T2 N3	-	-	-	-	180 (3.4)	32 (1.5)	-	-	-	-
T2 Nx	-	-	-	-	-	-	112 (1.9)	20 (1.1)	107 (3.7)	115 (2.9)
T3 N0	-	-	1423 (60.6)	455 (53.3)	-	-	130 (2.2)	44 (2.3)	-	50 (1.3)
T3 N1	-	-	-	-	2262 (42.7)	1143 (52.0)	583 (9.9)	385 (20.6)	-	257 (6.5)
T3 N2	-	-	-	-	1139 (21.5)	404 (18.4)	670 (11.4)	222 (11.9)	-	66 (1.7)
T3 N3	-	-	-	-	412 (7.8)	229 (10.4)	-	-	-	-
T3 Nx	-	-	-	-	-	-	36 (0.6)	65 (3.5)	104 (3.6)	248 (6.3)
T4 N0	-	-	-	-	147 (2.8)	45 (2.0)	45 (0.8)	7 (0.4)	-	5 (0.1)
T4 N1	-	-	-	-	261 (4.9)	109 (5.0)	155 (2.6)	65 (3.5)	1 (0.0)	26 (0.7)
T4 N2	-	-	-	-	203 (3.8)	43 (2.0)	199 (3.4)	67 (3.6)	-	15 (0.4)
T4 N3	-	-	-	-	118 (2.2)	39 (1.8)	-	-	-	-
T4 Nx	-	-	-	-	-	-	35 (0.6)	37 (2.0)	52 (1.8)	64 (1.6)
Tx N0	-	-	-	-	-	-	395 (6.7)	49 (2.6)	931 (32.6)	227 (5.8)
Tx N1	-	-	-	-	-	-	836 (14.2)	180 (9.6)	391 (13.7)	160 (4.1)
Tx N2	-	-	-	-	-	-	778 (13.3)	67 (3.6)	163 (5.7)	37 (0.9)
Tx N3	-	-	-	-	221 (4.2)	49 (2.2)	-	-	-	-
Tx Nx	-	-	-	-	-	-	402 (6.8)	476 (25.4)	978 (34.3)	2363 (60.1)

Supplemental table 2. Distribution of T- and N-category per stage for gastric cancer.

	Stage I		Stage II		Stage III	
	Netherlands	Belgium	Netherlands	Belgium	Netherlands	Netherlands
T1 N0	223 (20.2)	160 (39.6)	-	-	-	-
T1 N1	13 (1.2)	20 (5.0)	-	-	-	-
T1 N2	-	-	-	3 (0.5)	-	-
T1 N3	-	-	1 (0.1)	-	-	-
T1 Nx	-	-	-	-	-	-
T2 N0	867 (78.6)	224 (55.4)	-	-	-	-
T2 N1	-	-	244 (24.4)	75 (11.4)	-	-
T2 N2	-	-	109 (10.9)	18 (2.7)	-	-
T2 N3	-	-	-	-	11 (1.9)	5 (2.2)
T2 Nx	-	-	-	-	-	-
T3 N0	-	-	345 (34.5)	224 (34.0)	-	-
T3 N1	-	-	245 (24.5)	311 (47.3)	-	-
T3 N2	-	-	-	-	147 (25.2)	101 (44.7)
T3 N3	-	-	-	-	18 (3.1)	31 (13.7)
T3 Nx	-	-	-	-	-	-
T4A N0	-	-	56 (5.6)	27 (4.1)	-	-
T4A N1	-	-	-	-	41 (7.0)	48 (21.2)
T4A N2	-	-	-	-	30 (5.1)	21 (9.3)
T4A N3	-	-	-	-	6 (1.0)	4 (1.8)
T4A Nx	-	-	-	-	-	-
T4B N0	-	-	-	-	156 (26.7)	2 (0.9)
T4B N1	-	-	-	-	108 (18.5)	7 (3.1)
T4B N2	-	-	-	-	65 (11.1)	6 (2.7)
T4B N3	-	-	-	-	2 (0.3)	1 (0.4)
T4B Nx	-	-	-	-	-	-
Tx N0	-	-	-	-	-	-
Tx N1	-	-	-	-	-	-
Tx N2	-	-	-	-	-	-
Tx N3	-	-	-	-	-	-
Tx Nx	-	-	-	-	-	-

Supplemental table 2 (Continued). Distribution of T- and N-category per stage for gastric cancer.

Stage IV		Stage X	
Belgium	Netherlands	Belgium	Netherlands
10 (0.3)	10 (0.8)	-	21 (0.7)
14 (0.4)	6 (0.5)	-	7 (0.2)
9 (0.3)	6 (0.5)	-	-
1 (0.0)	6 (0.5)	-	-
4 (0.1)	17 (1.4)	46 (1.8)	62 (2.1)
229 (6.6)	13 (1.1)	-	45 (1.5)
219 (6.3)	21 (1.7)	-	22 (0.7)
218 (6.2)	11 (0.9)	-	5 (0.2)
29 (0.8)	6 (0.5)	-	-
123 (3.5)	20 (1.6)	95 (3.8)	62 (2.1)
105 (3.0)	36 (2.9)	-	36 (1.2)
141 (4.0)	128 (10.3)	-	79 (2.7)
171 (4.9)	66 (5.3)	-	30 (1.0)
22 (0.6)	33 (2.7)	-	5 (0.2)
49 (1.4)	44 (3.6)	50 (2.0)	99 (3.4)
37 (1.1)	20 (1.6)	-	9 (0.3)
41 (1.2)	60 (4.9)	-	17 (0.6)
37 (1.1)	34 (2.7)	-	13 (0.4)
14 (0.4)	23 (1.9)	-	4 (0.1)
24 (0.7)	32 (2.6)	17 (0.7)	32 (1.1)
91 (2.6)	6 (0.5)	-	2 (0.1)
110 (3.2)	4 (0.3)	-	2 (0.1)
95 (2.7)	6 (0.5)	-	-
24 (0.7)	8 (0.6)	-	-
94 (2.7)	5 (0.4)	49 (2.0)	6 (0.2)
410 (11.7)	72 (5.8)	1229 (49.0)	258 (8.7)
350 (10.0)	97 (7.8)	246 (9.8)	88 (3.0)
304 (8.7)	32 (2.6)	124 (4.9)	19 (0.6)
40 (1.1)	17 (1.4)	4 (0.2)	2 (0.1)
476 (13.6)	398 (32.2)	648 (25.8)	2030 (68.7)

Supplemental Table 3. Distribution of treatment for patients with esophageal cancer diagnosed in the Netherlands and Belgium.

	Esophageal cancer									
	I		II		III		IV		X	
	Netherlands	Belgium	Netherlands	Belgium	Netherlands	Belgium	Netherlands	Belgium	Netherlands	Belgium
N	1926	760	2348	853	5296	2200	5870	1873	2854	3931
Chemotherapy only	18 (0.9)	18 (2.4)	39 (1.7)	42 (4.9)	466 (8.8)	271 (12.3)	1950 (33.2)	929 (49.6)	79 (2.8)	540 (13.7)
CRT not followed by resection	275 (14.3)	94 (12.4)	459 (19.5)	192 (22.5)	1111 (21.0)	599 (27.2)	365 (6.2)	241 (12.9)	256 (9.0)	651 (16.6)
Endoscopic resection	248 (12.9)	108 (14.2)	6 (0.3)	19 (2.2)	3 (0.1)	64 (2.9)	1 (0.0)	72 (3.8)	404 (14.2)	506 (12.9)
nCRT followed by resection with or without adjuvant therapy	642 (33.3)	20 (2.6)	1240 (52.8)	205 (24.0)	2054 (38.8)	541 (24.6)	47 (0.8)	19 (1.0)	245 (8.6)	186 (4.7)
Other or unknown	201 (10.4)	60 (7.9)	165 (7.0)	65 (7.6)	630 (11.9)	204 (9.3)	2005 (34.2)	498 (26.6)	1201 (42.1)	1089 (27.7)
Radiotherapy only	219 (11.4)	49 (6.4)	219 (9.3)	44 (5.2)	637 (12.0)	41 (1.9)	1455 (24.8)	45 (2.4)	459 (16.1)	118 (3.0)
Surgical resection only	258 (13.4)	298 (39.2)	91 (3.9)	134 (15.7)	94 (1.8)	123 (5.6)	4 (0.1)	19 (1.0)	115 (4.0)	444 (11.3)
Surgical resection with (neo)adjuvant chemotherapy	65 (3.4)	113 (14.9)	129 (5.5)	152 (17.8)	301 (5.7)	357 (16.2)	43 (0.7)	50 (2.7)	95 (3.3)	397 (10.1)

Supplemental Table 4. Distribution of treatment for patients with gastric cancer diagnosed in the Netherlands and Belgium

	Gastric cancer									
	I		II		III		IV		X	
	Netherlands	Belgium	Netherlands	Belgium	Netherlands	Belgium	Netherlands	Belgium	Netherlands	Belgium
N	1103	404	1000	658	584	226	3491	1237	2508	2955
Chemotherapy only	51 (4.6)	7 (1.7)	86 (8.6)	76 (11.6)	87 (14.9)	49 (21.7)	1229 (35.2)	660 (53.4)	107 (4.3)	334 (11.3)
Endoscopic resection	42 (3.8)	18 (4.5)	2 (0.2)	7 (1.1)	0 (0.0)	2 (0.9)	1 (0.0)	26 (2.1)	58 (2.3)	115 (3.9)
Other or unknown	197 (17.9)	65 (16.1)	201 (20.1)	71 (10.8)	232 (39.7)	28 (12.4)	2081 (59.6)	454 (36.7)	998 (39.8)	1061 (35.9)
Surgical resection only	411 (37.3)	200 (49.5)	246 (24.6)	168 (25.5)	99 (17.0)	33 (14.6)	105 (3.0)	38 (3.1)	716 (28.5)	811 (27.4)
Surgical resection with (neo) adjuvant chemotherapy	402 (36.4)	114 (28.2)	465 (46.5)	336 (51.1)	166 (28.4)	114 (50.4)	75 (2.1)	59 (4.8)	629 (25.1)	634 (21.5)

Chapter 2

Trends in best-case, typical and worst-case survival scenarios of patients with non-metastatic esophagogastric cancer between 2006 and 2020: a population-based study

Steven C. Kuijper, Marieke Pape, Pauline A.J. Vissers, Paul M. Jeene, Ewout A. Kouwenhoven, Nadia Haj Mohammad, Jelle P. Ruurda, Meindert N. Sosef, Rob H.A. Verhoeven, Hanneke W.M. van Laarhoven

Based on:
Kuijper SC, Pape M, Vissers PAJ, et al. Trends in best-case, typical and worst-case survival scenarios of patients with non-metastatic esophagogastric cancer between 2006 and 2020: A population-based study. *Int J Cancer*. 2023

Abstract

Background. New treatment options and centralization of surgery have improved survival for patients with non-metastatic esophageal or gastric cancer. It is unknown, however, which patients benefitted the most from treatment advances. The aim of this study was to identify best-case, typical and worst-case scenarios in terms of survival time, and to assess if survival associated with these scenarios changed over time.
Methods. Patients with non-metastatic potentially resectable esophageal or gastric cancer diagnosed between 2006-2020 were selected from the Netherlands Cancer Registry. Best-case (20th percentile), upper-typical (40th percentile), typical (median), lower-typical (60th percentile) and worst-case (80th percentile) survival scenarios were defined, and regression analysis was used to investigate the change in survival time for each scenario across years.
Results. For patients with esophageal cancer (N=24,352) survival time improved on average 12.0 (until 2011), 1.5 (until 2018), 0.7, 0.4 and 0.2 months per year for the best-case, upper-typical, median, lower-typical and worst-case scenario, respectively. For patients with gastric cancer (N=9,993) survival time of the best-case scenario remained constant, whereas the upper-typical, median, lower-typical and worst-case scenario improved on average with 1.0 (until 2018), 0.5, 0.2, and 0.2 months per year, respectively. Subgroup analyses showed that, survival scenarios improved for nearly all patients across treatment groups and for patients with squamous cell carcinomas or adenocarcinomas.
Conclusion. Survival improved for almost all patients suggesting that in clinical practice the vast majority of patients benefitted from treatment advances. The clinically most meaningful survival advantage was observed for the best-case scenario of esophageal cancer.

Introduction

In the past decades novel treatment options and centralization of surgery have improved survival for patients with non-metastatic esophageal, gastroesophageal junction and gastric carcinoma.[1–5] For patients with non-metastatic esophageal or gastroesophageal junctional cancer, neoadjuvant chemoradiotherapy improved 5-year survival from 34% underwent 47% compared to surgery alone.[2] For patients with unresectable disease or who do not wish to undergo surgery, definitive chemoradiation is an alternative treatment.[6–9]but 5-year overall survival rates have been only 5% to 10%. Wepreviously reported results of a study conducted from January 1986 to April 1990 of combined chemotherapy and RT vs RT alone when an interim analysis revealed sig- nificant benefit for combined therapy. Objective To report the long-term outcomes of a previously reported trial designed to determine if adding chemotherapy during RT improves the survival rate of patients with esophageal carcinoma. Design Randomized controlled trial conducted 1985 to 1990 with follow-up of at least 5 years, followed by a prospective cohort study conducted between May 1990 and April 1991. Setting Multi-institution participation, ranging from tertiary academic referral cen- ters to general community practices. Patients Patients had squamous cell or adenocarcinoma of the esophagus, T1-3 N0-1 M0, adequate renal and bone marrow reserve, and a Karnofsky score of at least 50. Interventions Combined modality therapy (n=134 For patients with resectable gastric cancer, perioperative chemotherapy with epirubicin, cisplatin, and fluorouracil (ECF) improved 5-year survival rate from 26% to 36% compared to surgery alone.[5] Furthermore, perioperative chemotherapy with fluorouracil, oxaliplatin and docetaxel (FLOT) showed a 5-year survival of 45% compared to 36% among patients treated with ECF/epirubicin, cisplatin and capecitabine (ECX).[1] Population-based studies also showed an improved survival of patients with non-metastatic esophageal or gastric cancer.[3,4,10]and has been implemented for gastric cancer since 2012 in the Netherlands. This study evaluated the impact of centralizing gastric cancer surgery on outcomes for all patients with gastric cancer. Methods: Patients diagnosed with non-cardia gastric adenocarcinoma in the intervals 2009–2011 and 2013–2015 were selected from the Netherlands Cancer Registry. Clinicopathological data, treatment characteristics and mortality were assessed for the periods before (2009–2011 In contrast to clinical trials, such analyses include elderly, frail patients and patients with comorbidities, and therefore better reflect the daily clinical practice.[8,11]

In the Netherlands, centralization processes have occurred for both esophageal and gastric cancer and have improved survival outcomes.[4,12–15]blood loss, resection margin, lymphadenectomy, chemotherapy, postoperative complications and hospital stay, and overall and disease-free survival For esophageal cancer, since 2006 a minimum of ten esophageal cancer resections annually per hospital was mandated and since 2011 increased to ≥20 esophagectomies .[16] For gastric cancer, a minimum of ten gastrectomies annually was required since 2011 and was increased to ≥20 in 2013.[4]

Recent studies have expressed overall survival in terms of best, typical and worst-case scenarios which can be used to inform patients about survival expectations.[17–19] This can provide patients with information regarding the bandwidth of survival time. In addition, these studies often show that mainly the best-case scenarios have improved over time. This implies that only the survival of patients with already the best prognosis have improved over time.

To investigate if survival has improved over time we defined best-case, upper-typical, typical, lower-typical and worst-case survival scenarios for patients with non-metastatic esophageal or gastric cancer and assessed if survival for these scenarios have changed over time across treatments and histological subtypes.

Methods

Patients

Patients with non-metastatic esophageal (International classification of diseases for oncology: C15.0-C15.9), gastroesophageal junction/cardia (C16.0), and gastric cancer (C16.1-C16.9) diagnosed in 2006-2020 were selected from the Netherlands Cancer Registry (NCR).[20] The NCR contains patient, tumor and treatment characteristics of all patients diagnosed with cancer in the Netherlands. All newly diagnosed malignancies identified by the national automated pathology archive are included in the NCR. Trained data managers of the NCR routinely extract information on diagnosis, stage and treatment from medical records. Vital status was retrieved through linkage of the NCR with the Dutch Personal Records Database and was updated until 1 February, 2022.

Coding and classifications

Topography and morphology were coded according to the International Classification of Disease for Oncology (ICD-O-3).[20] ICD-O-3 morphology codes were used to classify tumors as adenocarcinoma, squamous cell carcinoma and carcinoma not otherwise specified (NOS). Tumors were staged according to the International Union Against Cancer (UICC) TNM classification valid at the time of diagnosis. Patients diagnosed between 2006-2009, 2010-2016, 2017-2020 were staged according to the sixth, seventh and eighth edition, respectively.[21–23] All patients with non-metastatic disease (cM0) at primary diagnosis were included. In addition, patients with esophageal or gastroesophageal junction cancer with cM1a tumor according to TNM-6 were included, as most patients with a M1a tumor had a distal tumor with coeliac lymph nodes which can be considered cN+ according to TNM-7 and TNM-8.[3] Gastroesophageal junction/cardia tumors were classified as esophageal tumors.

All patients diagnosed between 2009-2020 with an unresectable cT4b tumor were excluded as this was not considered a potentially curable disease. Patients with gastric cancer diagnosed between 2006-2009 (TNM-6) with a cT4 tumor were excluded as these patients would have been staged as cT4b (tumor invades adjacent structures) according to TNM-7 and -8. Patients with esophageal cancer with a cT4 tumor according to TNM-6 were also excluded as there was no TNM-6 equivalent of cT4b for patients with esophageal cancer. To evaluate the effects of excluding these patients, we performed sensitivity analyses without excluding these patients (Supplementary Tables 4-6).

Information on histological subtyping according to Lauren classification before 2015 could not be determined based on the standard data collection of the NCR. For these patients the Lauren classifications were obtained through adopting the approach described by van der Kaaij et al. [24] In short, Lauren classifications were obtained from complete pathology reports of the biopsies available from the Dutch Nationwide Pathology Databank (PALGA), from which a syntax was developed to automatically classify all adenocarcinomas in the data. From 2015 onwards Lauren classification were determined based on the tumor morphology code.

Treatment

Treatment was mutually exclusively classified as resection (endoscopic or surgical resection with or without (neo)adjuvant treatment), chemoradiotherapy (without resection; for patients with esophageal cancer only), treatment without curative potential (for patients with esophageal cancer only), best supportive care (for patients with esophageal cancer only) or no resection (which included treatment without curative potential and best supportive care; for patients with gastric cancer only).

Statistical analysis

Kaplan-Meier curves were constructed for each year of diagnosis to determine the survival scenarios. We partitioned the survival curve according to survival probabilities: the 20th (best-case), 40th (upper-typical), 50th (median), 60th (lower-typical), and 80th (worst-case) percentile. For each of the percentiles (scenarios) the corresponding survival time in months was calculated. The 20th percentile, for example, is interpreted as the 20% longest survivors, i.e. when 80% have died, and was therefore classified as the best-case scenario. Contrary to prior studies using this approach,[17,19] we could not include the 10th percentile, as a 10% survival probability in the Kaplan-Meier curve was not observed in the majority of years. Therefore, we adjusted the percentiles to be able to estimate five survival scenarios as prior research demonstrated that patients preferred to be informed using such scenarios.[18] In addition, subgroup analyses were performed based on treatment and histology (esophageal only).

The estimated survival times of each scenario were plotted and a linear regression line with 95% confidence intervals were fitted to the data points using the ggplot2 package for R. Weighted least squares estimation was used to fit the regression line, where the number of patients per year of diagnosis was used as regression weight. We tested the slope for significance (two-sided) and considered $p \leq .05$ as statistically significant. The slope reflected the average change in months survival per year. Linear regression analysis was not performed for scenarios with less than four survival estimates. Trends in categorical variables over time, e.g. frequency of treatments, were tested with the multinomial Cochran-Armitage trend test.[25] All analyses were conducted with R studio version 4.0.3 and R version 3.6.1.

Results

We identified 24,352 patients with non-metastatic esophageal cancer and 9,993 patients with non-metastatic gastric cancer (Table 1). Median overall survival of patients with esophageal cancer was 20.4 (95% CI: 19.9-20.8) months and for patients with gastric cancer was 20.1 (19.3-21.1) months (Supplementary Table 1).

Trends in treatment

For patients with esophageal cancer, the percentage of resections remained constant around 50% (p=.200), while the percentage of patients that received chemoradiotherapy increased significantly from 3% in 2006 to 28% in 2020 (p<.001) (Figure 1). The percentage of patients that received treatment without curative intent (p<.001) and best supportive care (p<.001) decreased significantly 23% in 2006 to 12% and 15% percent for treatment without curative potential and best supportive care, respectively. For patients with gastric cancer, the percentage of patients that underwent a resection (p=.300) or received no resection (p=.300) remained constant around 65% and 35%, respectively, between 2006-2020 (Figure 1).

Among patients with esophageal cancer who underwent resection, the percentage of patients who received neoadjuvant treatment increased from 21% in 2006 to 93% in 2020 (p<.001) (Supplementary Figure 1). Among patients with gastric cancer who underwent a resection, the percentage of patients receiving neoadjuvant treatment (with or without adjuvant treatment) increased from 8% in 2006 to 60% in 2020 (p<.001), and the percentage of patients who underwent neo- and adjuvant treatment increased from 9% in 2006 to 36% in 2020 (p<.001). The proportion male to female, median age, and proportion of different histological subtypes remained constant over time (Supplementary Figure 2).

Trends of survival scenarios

Total population

All slopes of the survival scenarios for patients with esophageal or gastric cancer were significantly positive (Figure 2; Table 2). For patients with esophageal cancer, on average survival increased with 12.0, 1.5, 0.7, 0.4 and 0.2 months per year for the best-case (p20; until 2011 after which the p20 was not observed), upper-typical (p40; until 2018), median (p50), lower-typical (p60) and worst-case scenario (p80), respectively. For patients with gastric cancer, the survival time per year increased on average with 1.0, 0.5, 0.2 and 0.2 months for the upper-typical (p40; until 2018), median (p50), lower-typical (p60) and worst-case scenario (p80), respectively, while the best-case (p20) scenario did not significantly increase.

Per treatment

For patients with esophageal or gastric cancer who underwent a resection, the best-case scenario (p20) could not be calculated due to limited follow-up time. That is to say, the p20 percentile was not yet completed because these patients were still alive at the end of follow-up. All other scenarios of resected patients improved significantly over time (Figure 2, Table 2). For patients with esophageal cancer who underwent resection, on average the survival time increased by 7.5, 3.6, 1.7 and 0.6 months per year for the upper-typical (p40; until 2007), median (p50; until 2013), lower-typical (p60; until 2016) and worst-case scenario (p80; until 2018), respectively. For patients with gastric cancer who underwent a resection on average the survival time increased with 2.6, 1.3, 0.9 and 0.4 months per year for the upper-typical (p40; until 2008), median (p50; until 2015), lower-typical (p60; until 2017) and worst-case scenario (p80; until 2018), respectively. For patients with esophageal cancer who received chemoradiotherapy, the median (p50) significantly increased on average with 0.4 months per year between 2006 and 2020, whereas the other scenarios remained unchanged.

Table 1. Patient and tumor characteristics at primary diagnosis. NOS = none otherwise specified.

		Eesophagus (N=24,352)		Stomach (N=9,993)
	All	Adenocarcinoma	Squamous cell carcinoma	
Sex				
Male	17,704 (72.7%)	13,910 (79.3%)	3,449 (54.8%)	5,972 (59.8%)
Female	6,648 (27.3%)	3,625 (20.7%)	2,849 (45.2%)	4,021 (40.2%)
Age				
Mean (SD)	69.4 (11.0)	69.2 (11.2)	69.6 (10.2)	72.2 (12.6)
Histological subtype				
Adenocarcinoma	17,535 (72.0%)	-	-	9,723 (97.3%)
Squamous cell arcinoma	6,298 (25.9%)	-	-	10 (0.1%)
Carcinoma NOS	519 (2.1%)	-	-	260 (2.6%)
Treatment				
No resection	-	-	-	3,361 (33.6%)
Best supportive care	3,794 (15.6%)	2,458 (14.0%)	1,112 (17.7%)	-
Treatment without curative potential	3,942 (16.2%)	2,470 (14.1%)	1,347 (21.4%)	-
Chemoradiotherapy	3,997 (16.4%)	2,148 (12.2%)	1,790 (28.4%)	-
Resection	12,619 (51.8%)	10,459 (59.6%)	2,049 (32.5%)	6,632 (66.4%)

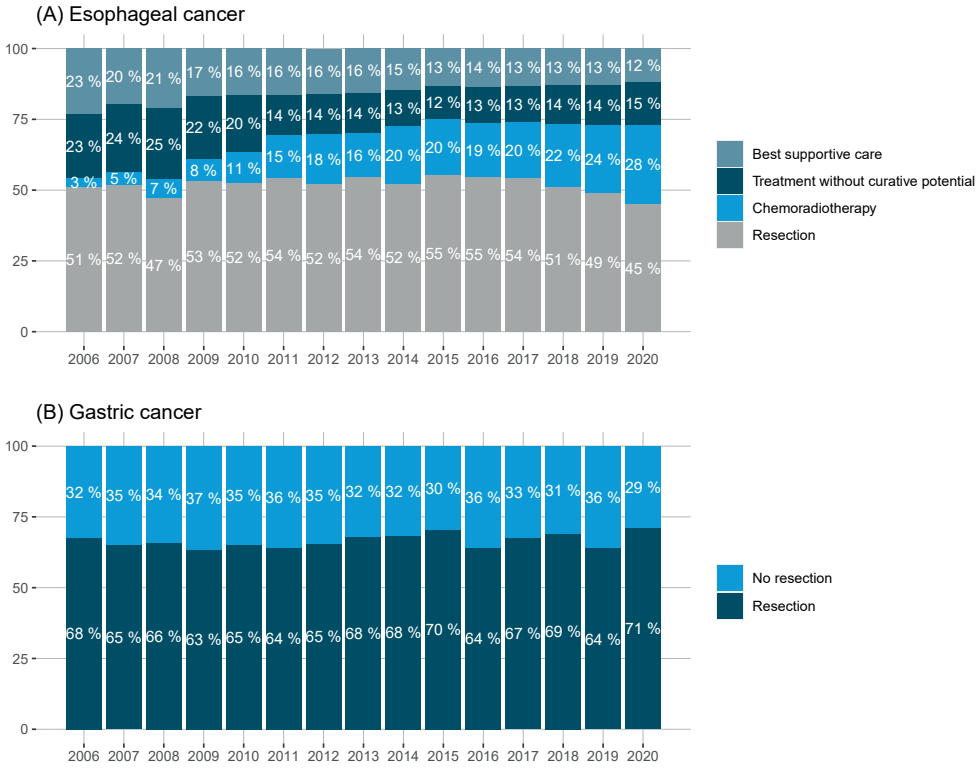


Figure 1. Treatment of patients with non-metastatic esophageal (A) and gastric cancer (B) by incidence year.

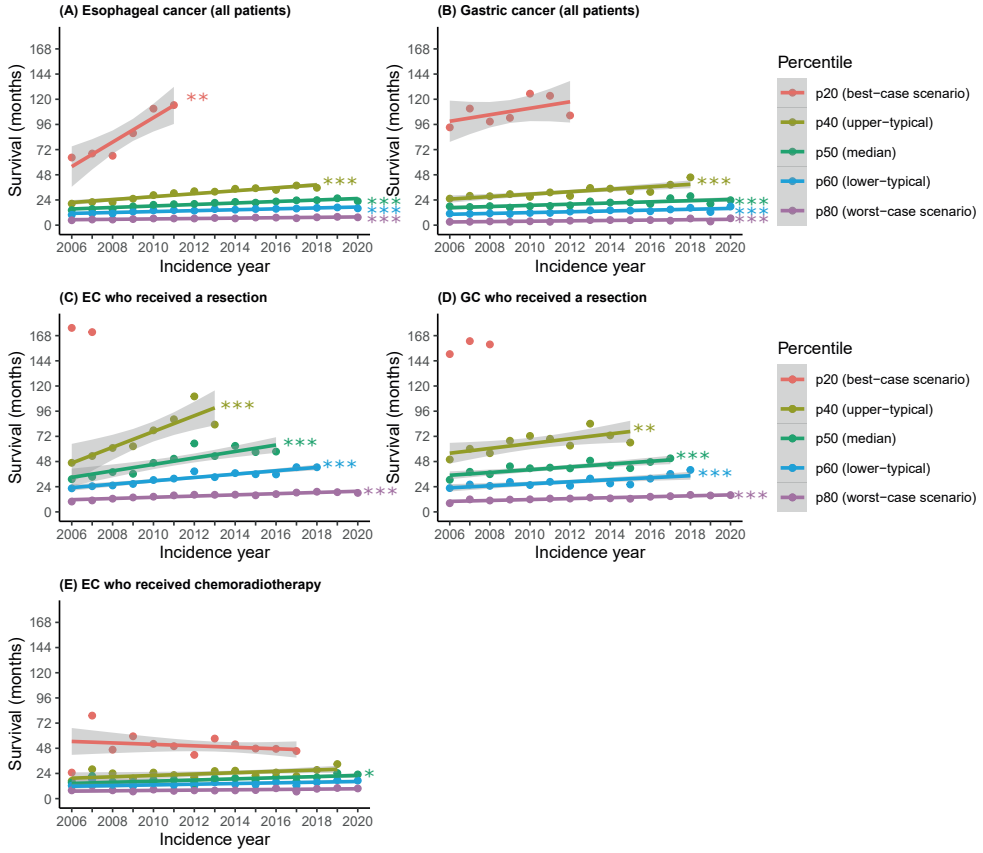


Figure 2. Trends of survival scenarios for patients with esophageal (A,C,E) and gastric cancer (B,D), stratified by primary tumor location and treatment. Regression lines and confidence intervals were only estimated for scenarios with at least four observations. EC = Esophageal cancer, GC = Gastric cancer. * p<.05, ** p<.01, *** <.001.

Table 2. Survival scenarios of patients with esophageal or gastric cancer diagnosed between 2006 and 2020 with slope estimates and 95% CI of the trend lines, stratified to primary tumor location and treatment. The slope reflects the average number of months change in survival per incidence year.

	Esophageal cancer				
	Observed until	OS (months), 2006	OS (months), Last observed year	Slope (95% CI)	p-value
Total population (N)				N=24,352	
p20	2011	64	114	12.04 (7.99-16.09)	.004
p40	2018	20	36	1.51 (1.25-1.77)	<.001
p50	2020	14	23	0.72 (0.59-0.85)	<.001
p60	2020	10	16	0.36 (0.27-0.45)	<.001
p80	2020	5	7	0.20 (0.16-0.24)	<.001
Resection (N)				N=12,619	
p20	A			A	
p40	2007	47	83	7.45 (5.73-9.17)	<.001
p50	2013	31	57	3.56 (2.67-4.45)	<.001
p60	2016	22	43	1.68 (1.37-1.99)	<.001
p80	2018	10	18	0.56 (0.44-0.68)	<.001
Chemoradiotherapy (N)				N=3997	
p20	2017	24	45	-0.93 (-2.01-0.15)	.122
p40	2019	17	33	0.53 (0.05-1.01)	.053
p50	2020	15	23	0.35 (0.06-0.64)	.032
p60	2020	12	17	0.21 (-0.02-0.44)	.088
p80	2020	8	10	0.07 (-0.1-0.24)	.402

[A] The 20th percentile was not observed.[B] Slope was not estimatedsince less han four years were observed. - Group was not analyzed as chemoradiotherapy is not a treatment with curative potential in gastric cancer.

Table 2 (Continued). Survival scenarios of patients with esophageal or gastric cancer diagnosed between 2006 and 2020 with slope estimates and 95% CI of the trend lines, stratified to primary tumor location and treatment. The slope reflects the average number of months change in survival per incidence year.

	Gastric cancer				
	Observed until	OS (months), 2006	OS (months), Last observed year	Slope (95% CI)	p-value
Total population (N)				N=9,993	
p20	2012	93	105	3.64 (-0.27-7.55)	.127
p40	2018	25	46	1.03 (0.70-1.36)	<.001
p50	2020	18	24	0.46 (0.27-0.65)	<.001
p60	2020	11	18	0.23 (0.12-0.34)	<.001
p80	2020	3	7	0.16 (0.10-0.22)	<.001
Resection (N)				N=6,632	
p20	B			B	
p40	2008	50	66	2.62 (1.13-4.11)	.009
p50	2015	31	51	1.27 (0.75-1.79)	.001
p60	2017	23	40	0.88 (0.51-1.25)	<.001
p80	2018	8	16	0.44 (0.35-0.53)	<.001
Chemoradiotherapy (N)					
p20	-			-	-
p40	-			-	-
p50	-			-	-
p60	-			-	-
p80	-			-	-

[A] The 20th percentile was not observed.[B] Slope was not estimatedsince less han four years were observed. - Group was not analyzed as chemoradiotherapy is not a treatment with curative potential in gastric cancer.

Per histological subtype

All survival scenarios increased significantly for all patients with esophageal adenocarcinoma and squamous cell carcinoma. For patients with an adenocarcinoma an average increase of 12.2, 1.6, 0.8, 0.5 and 0.3 months per year for best-case (p20; until 2011), upper-typical (p40; until 2018), median (p50; until 2019), lower-typical (p60) and worst-case scenario (p80) was, respectively (Supplementary Table 2, Figure 3). For patient with a squamous cell carcinoma an average increase of 7.5, 1.4, 0.7, 0.3 and 0.1 months per year was for the best-case (p20; until 2013), upper-typical (p40; until 2019), median (p50), lower-typical (p60) and worst-case scenario (p80), respectively.

Esophageal cancer by histological subtype and treatment

Supplementary Table 2 displays all slope estimates of patients with esophageal cancer by histological subtype and treatment. For patients with esophageal cancer who underwent resection, all survival scenarios for adenocarcinomas and squamous cell carcinomas increased significantly (Figure 3). For patients with esophageal adenocarcinoma receiving chemoradiotherapy, all scenarios but the best-case (p20) scenario and the worst-case scenarios (p80) increased significantly. For patients with esophageal squamous cell carcinoma who received chemoradiotherapy the upper-typical (p40), and lower-typical (p60) scenarios increased significantly.

Esophageal and gastric adenocarcinoma by Lauren classification subtype

For patients with Lauren intestinal and diffuse esophageal adenocarcinoma all scenarios increased significantly (Supplementary Table 3; Figure 4). Among patients with gastric intestinal adenocarcinoma all but the best-case scenario (p20) increased significantly over time, and for patients with diffuse gastric adenocarcinomas only the upper-typical (p40) and lower-typical (p60) scenarios increased significantly annually.

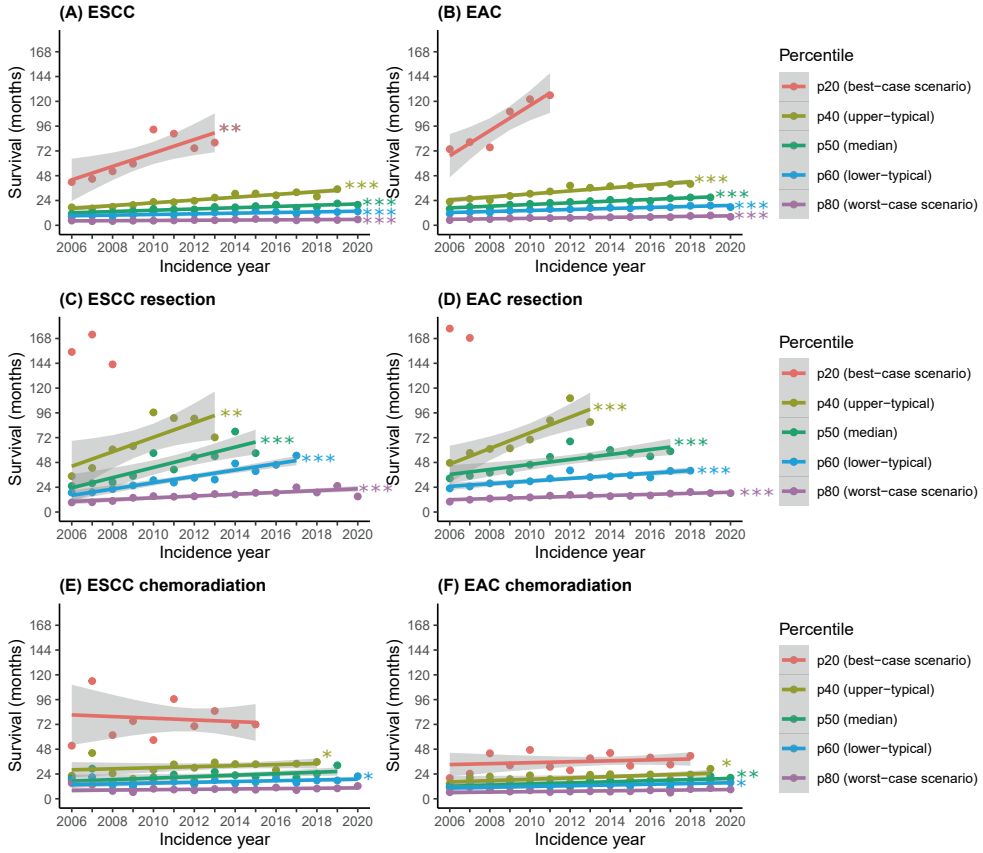


Figure 3. Trends of survival scenarios for patients with esophageal cancer, stratified to patients who received a resection (C,D) and chemoradiation (E,F), and patients with squamous cell carcinoma (A, C, E) and adenocarcinoma (B, D, F). ESCC = Esophageal squamous cell carcinoma, EAC = Esophageal adenocarcinoma. Regression lines and confidence interval were only estimated for scenarios that had minimally four observations. * p<.05, ** p<.01, *** <.001

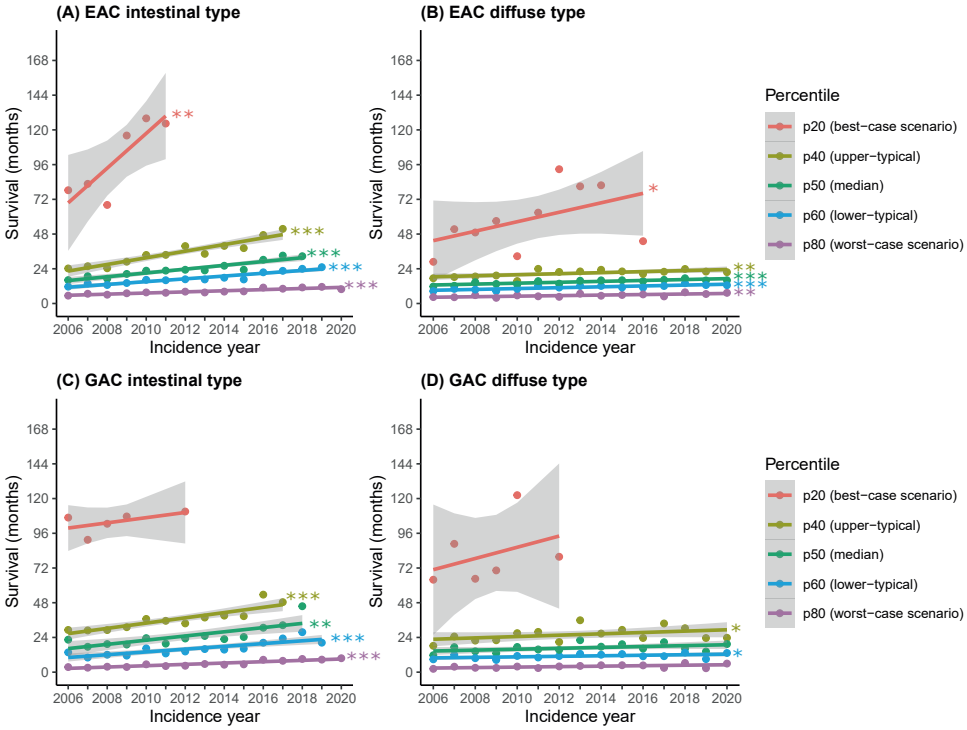


Figure 4. Trends of survival scenarios for patients with esophageal (A, B) and gastric adenocarcinoma (C, D), stratified by Lauren classification diffuse (B, D) and intestinal type (A, C). Regression lines and confidence interval were only estimated for scenarios that had minimally four observations. * p<.05, ** p<.01, *** <.001. EAC=Esophageal adenocarcinoma, GAC=Gastric adenocarcinoma.

Discussion

This study showed that survival for almost all patients with non-metastatic esophageal and gastric cancer improved significantly between 2006 and 2020. Besides the median overall survival improvements reported in earlier population-based studies,[3,4,10,16,24]a minimal volume standard of 10 oesophagectomies per year was introduced in 2006. For gastrectomy, no minimal volume standard was set. Aims of this study were to describe changes in hospital volumes, mortality and survival and to explore if high hospital volume is associated with better outcomes after oesophagectomy and gastrectomy in the Netherlands. Methods: From 1989 to 2009, 24,246 patients underwent oesophagectomy (N = 10,025 this study showed that survival of nearly all patients increased over the past 15 years.

Aforementioned treatment advances for patients with non-metastatic esophageal or gastric cancer could explain the observed survival improvements. Additionally, the obligatory Dutch Upper Gastrointestinal Cancer Audit (DUCA) was introduced in 2011 and led to improvements in post-operative outcomes for patients with esophagogastric cancer.[15,26] Although no clear survival improvements were observed in survival in and around any specific years in addition to the already existing trends, it is nevertheless likely that treatment advances have contributed to the increasing survival trends.

The largest survival benefits were observed in the best-case and upper-typical scenarios, whereas the smallest survival benefits were observed in the worst-case and lower-typical scenarios. This suggests that treatment advances over past years have mainly affected the survival prospects of the best patients. This could potentially be explained by patient and tumor related factors. Patients with a better performance status generally have a better prognosis.[27] For example, patients with gastric cancer with better performance statuses are more likely to continue with adjuvant treatment.[1,28]oxaliplatin and docetaxel In the current study, the proportion of patients who continued with adjuvant treatment after neoadjuvant treatment increased over the years, which may explain the increasing trends in survival. Additionally, tumor biology may also play a role.[29] Tumors that already respond well to a certain type of treatment might respond better to a treatment that has been further improved than a tumor that was resistant already from the very start.[30] Therefore, the best-case scenarios (p20) may be characterized by patients in good physical condition with tumors that respond well to treatment. The generally increasing survival trends could also be explained by stage migration due to improved diagnostic techniques, such as high-quality Positron Emission Tomography-Computed Tomography (PET-CT).[31] Due to these improved techniques, more patients over time are diagnosed with a locally advanced or metastatic disease at diagnosis, through which survival of the best-case scenarios improved over time.

Among different histological subtypes according to Lauren classification for patients with esophageal adenocarcinoma, all scenarios of both intestinal and diffuse adenocarcinomas improved significantly over time. Among patients with diffuse and intestinal gastric adenocarcinoma, although most scenarios increased, the best-case scenario's did not increase over time. Survival of the median scenario (p50) for patients with diffuse type adenocarcinomas also did not increase over time. Despite this, the overall survival of the best-case scenarios were substantially higher compared to other scenarios. These results are in line with earlier findings that show increasing survival trends among intestinal and diffuse type esophageal adenocarcinomas and slower increasing survival trends for intestinal

and diffuse type gastric adenocarcinomas.[24] What is more, although the trends of survival scenarios were relatively similar between diffuse and intestinal type, survival of patients with intestinal type adenocarcinomas was generally higher than patients with intestinal type adenocarcinomas.

Whereas nearly all best-case (p20) scenarios increased, the best-case scenario of patients with esophageal cancer who received chemoradiotherapy (without resection) remained constant. The proportion of patients that received chemoradiotherapy increased from 3% in 2006 to 28% in 2020. A proportion of patients who previously would not have received treatment with curative intent, probably received chemoradiotherapy in later years, thereby decreasing the overall level of fitness of patients receiving chemoradiotherapy over time. The increasing number of patients who receive chemoradiotherapy may be explained by the introduction of definitive chemoradiotherapy for medically inoperable patients or patients who do not wish to undergo a resection, or by patients that received neoadjuvant chemoradiotherapy not followed by a resection.[32,33] Over time more than 75% of patients were treated with definitive chemoradiotherapy.[8] Additionally, median overall survival of patients that had no treatment with curative intent (9.7 months) or received best supportive care (3.4 months) was substantially lower compared to patients that received chemoradiotherapy (19.1 months). Hence, it may be the case that the proportion of patients who received chemoradiotherapy with poor survival prospects due to their physical condition increased over time resulting constant survival trend of the best-case scenario.

In general, the results showed a survival benefit for almost all patients. Results from these analyses can be used by clinicians to communicate life expectancy to patients beyond the median survival estimates reported in clinical trials and population-based studies.[18,19] Such information can be used complementary to the use of individualized clinical prediction models such as the SOURCE model which also takes patient, tumor and treatment information into account.[27,34]however, have mostly been developed for survival prediction after surgery (ie, when treatment has already been completed Whereas prediction models estimate the predicted probability of surviving a given amount of time, the survival scenarios provide a bandwidth of the survival. Together these information sources can provide patients a realistic expectancy of their survival.

This study has a number of strengths. Population-based data was used, which contains a more representative patient sample from the daily clinical practice compared to patients in clinical trials. Compared to previous studies investigating survival scenarios,[17–19] we used linear regression with weighted least squared estimation to investigate general trend over time rather than comparing estimates. Weighted least squared estimation corrects for changing number of observations in later incidence years, compared to ordinary least squares which attributed equal weights to all data points. This study also has a few limitations. Firstly, due to the follow-up time, the upper 20th percentile best-case scenario was frequently not observed beyond 2011. Secondly, point estimates should be interpreted with caution as variance around the regression line showed considerable variability in certain scenarios. Thirdly, data performance status and comorbidities were not available for a large part of the population before 2015, and therefore trends among these characteristics could not be tested. Finally, as it was required to recode TNM 7th and 8th edition into the 6th edition and only potentially curable patients were to be included, we excluded all cT4 tumors as the TNM 6th edition had no cT4b for patients with esophageal cancer. Sensitivity analyses

revealed that the effects of excluding these patients was very small and did not change the conclusions from this study.

Conclusion

Our study showed that for almost all patients with non-metastatic esophageal and gastric cancer survival improved over the past 15 years, with the largest survival improvements for patients with the best prognosis.

References

1. Al-Batran SE, Homann N, Pauligk C, 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 ra. *Lancet*. 2019;393(10184):1948-1957. doi:10.1016/S0140-6736(18)32557-1
2. van Hagen P, Hulshof MCCM, van Lanschot JJB, et al. Preoperative Chemoradiotherapy for Esophageal or Junctional Cancer. *N Engl J Med*. 2012;366(22):2074-2084. doi:10.1056/nejmoa1112088
3. van Putten M, de Vos-Geelen J, Nieuwenhuijzen GAP, et al. Long-term survival improvement in oesophageal cancer in the Netherlands. *Eur J Cancer*. 2018;94:138-147. doi:10.1016/j.ejca.2018.02.025
4. van Putten M, Nelen SD, Lemmens VEPP, et al. Overall survival before and after centralization of gastric cancer surgery in the Netherlands. *Br J Surg*. 2018;105(13):1807-1815. doi:10.1002/bjs.10931
5. Cunningham D, Allum WH, Stenning SP, et al. Perioperative Chemotherapy versus Surgery Alone for Resectable Gastroesophageal Cancer. *N Engl J Med*. 2006;355(1):11-20. doi:10.1056/NEJMoa055531
6. Cooper JS, Guo MD, Herskovic A, et al. Chemoradiotherapy of Locally Advanced. *J Am Med Assoc*. 1999;281(17):1623-1627.
7. Kumar S, Dimri K, Khurana R, Rastogi N, Das KJM, Lal P. A randomised trial of radiotherapy compared with cisplatin chemo-radiotherapy in patients with unresectable squamous cell cancer of the esophagus. *Radiother Oncol*. 2007;83(2):139-147. doi:10.1016/j.radonc.2007.03.013
8. Koëter M, van Putten M, Verhoeven RHA, Lemmens VEPP, Nieuwenhuijzen GAP. Definitive chemoradiation or surgery in elderly patients with potentially curable esophageal cancer in the Netherlands: a nationwide population-based study on patterns of care and survival. *Acta Oncol (Madr)*. 2018;57(9):1192-1200. doi:10.1080/0284186X.2018.1450521
9. Hulshof MCCM, Geijsen ED, Rozema T, et al. Randomized Study on Dose Escalation in Definitive Chemoradiation for Patients With Locally Advanced Esophageal Cancer (ARTDECO Study). *J Clin Oncol*. 2021;39(25):2816-2824. doi:10.1200/JCO.20.03697
10. Arnold M, Rutherford MJ, Bardot A, et al. Progress in cancer survival, mortality, and incidence in seven high-income countries 1995–2014 (ICBP SURVMARK-2): a population-based study. *Lancet Oncol*. 2019;20(11):1493-1505. doi:10.1016/S1470-2045(19)30456-5
11. Donnelly CB, Wotherspoon AC, Morris M, et al. A population-level investigation of cancer clinical trials participation in a UK region. *Eur J Cancer Prev*. Published online 2017:229-235. doi:10.1097/CEJ.0000000000000373
12. Nelen SD, Heuthorst L, Verhoeven RHA, et al. Impact of Centralizing Gastric Cancer Surgery on Treatment, Morbidity, and Mortality. *J Gastrointest Surg*. 2017;21(12):2000-2008. doi:10.1007/s11605-017-3531-x
13. Gooiker GA, Van Der Geest LGM, Wouters MWJM, et al. Quality improvement of pancreatic surgery by centralization in the Western Part of the Netherlands. *Ann Surg Oncol*. 2011;18(7):1821-1829. doi:10.1245/s10434-010-1511-4
14. Güller U, Warschkow R, Ackermann CJ, Schmied BM, Cerny T, Ess S. Lower hospital volume is associated with higher mortality after oesophageal, gastric, pancreatic and rectal cancer resection. *Swiss Med Wkly*. 2017;147(July):1-9. doi:10.4414/smw.2017.14473
15. Voeten DM, Busweiler LAD, van der Werf LR, et al. Outcomes of Esophagogastric Cancer Surgery During Eight Years of Surgical Auditing by the Dutch Upper Gastrointestinal Cancer Audit (DUCA). *Ann Surg*. 2021;274(5):866-873. doi:10.1097/SLA.0000000000005116
16. Dikken JL, Dassen AE, Lemmens VEP, et al. Effect of hospital volume on postoperative mortality and survival after oesophageal and gastric cancer surgery in the Netherlands between 1989 and 2009. *Eur J Cancer*. 2012;48(7):1004-1013. doi:10.1016/j.ejca.2012.02.064

17. Kiely BE, Soon YY, Tattersall MHN, Stockler MR. How Long Have I Got? Estimating Typical, Best-Case, and Worst-Case Scenarios for Patients Starting First-Line Chemotherapy for Metastatic Breast Cancer: A Systematic Review of Recent Randomized Trials. *J Clin Oncol*. 2011;29(4):456-463. doi:10.1200/JCO.2010.30.2174
18. Kiely BE, McCaughan G, Christodoulou S, et al. Using scenarios to explain life expectancy in advanced cancer: Attitudes of people with a cancer experience. *Support Care Cancer*. 2013;21(2):369-376. doi:10.1007/s00520-012-1526-4
19. Hamers PAH, Elferink MAG, Stellato RK, et al. Informing metastatic colorectal cancer patients by quantifying multiple scenarios for survival time based on real-life data. *Int J Cancer*. 2021;148(2):296-306. doi:10.1002/ijc.33200
20. Percy, C.; Jack, A.; Shanmugarathan, S.; Sobin, L.; Parkin DM. *International Classification of Diseases for Oncology: ICD-O. Third Edition*.; 2000.
21. TNM Classification of Malignant Tumours, 8th Edition. *Int Union Against Cancer*. Published online 2014.
22. *TNM Classification of Malignant Tumours, 6th Edition*. Wiley; 2014.
23. *TNM Classification of Malignant Tumours, 7th Edition*. Wiley-Liss; 2009.
24. van der Kaaij RT, Koemans WJ, van Putten M, et al. A population-based study on intestinal and diffuse type adenocarcinoma of the oesophagus and stomach in the Netherlands between 1989 and 2015. *Eur J Cancer*. 2020;130:23-31. doi:10.1016/j.ejca.2020.02.017
25. Szabo A. Test for Trend With a Multinomial Outcome. *Am Stat*. 2019;73(4):313-320. doi:10.1080/00031305.2017.1407823
26. Busweiler LAD, Wijnhoven BPL, van Berge Henegouwen MI, et al. Early outcomes from the Dutch Upper Gastrointestinal Cancer Audit. *Br J Surg*. 2016;103(13):1855-1863. doi:10.1002/bjs.10303
27. van den Boorn HG, Abu-Hanna A, Haj Mohammad N, et al. SOURCE: Prediction Models for Overall Survival in Patients With Metastatic and Potentially Curable Esophageal and Gastric Cancer. *J Natl Compr Canc Netw*. 2021;19(4):403-410. doi:10.6004/jnccn.2020.7631
28. Cats A, Jansen EPM, van Grieken NCT, et al. Chemotherapy versus chemoradiotherapy after surgery and preoperative chemotherapy for resectable gastric cancer (CRITICS): an international, open-label, randomised phase 3 trial. *Lancet Oncol*. 2018;19(5):616-628. doi:10.1016/S1470-2045(18)30132-3
29. Goedegebuure RSA, Harrasser M, de Klerk LK, et al. Pre-treatment tumor-infiltrating T cells influence response to neoadjuvant chemoradiotherapy in esophageal adenocarcinoma. *Oncoimmunology*. 2021;10(1):1-12. doi:10.1080/2162402X.2021.1954807
30. van Velzen MJM, Pape M, Dijksterhuis WPM, et al. The association between effectiveness of first-line treatment and second-line treatment in gastro-oesophageal cancer. *Eur J Cancer*. 2021;156:60-69. doi:10.1016/j.ejca.2021.07.026
31. Kim TJ, Kim HY, Lee KW, Kim MS. Multimodality assessment of esophageal cancer: Preoperative staging and monitoring of response to therapy. *Radiographics*. 2009;29(2):403-421. doi:10.1148/rg.292085106
32. Hulshof MCCM, van Laarhoven HWM. Chemoradiotherapy in tumours of the oesophagus and gastro-oesophageal junction. *Best Pract Res Clin Gastroenterol*. 2016;30(4):551-563. doi:10.1016/j.bpg.2016.06.002
33. Lordick F, Mariette C, Haustermans K, Obermannová R, Arnold D, on behalf of the ESMO Guidelines Committee clinicalguidelines@esmo org. Oesophageal cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2016;27:v50-v57. doi:10.1093/annonc/mdw329
34. Van De Water LF, Van Den Boorn HG, Hoxha F, et al. Informing patients with esophagogastric cancer about treatment outcomes by using a web-based tool and training: Development and evaluation study. *J Med Internet Res*. 2021;23(8). doi:10.2196/27824

Supplementary Table 1. Median survival (months) with 95% confidence intervals of included patients.

		Esophagus		Stomach
	All	Adenocarcinoma	Squamouscell carcinoma	
All patients	20.4 (19.9-20.8)	22.5 (21.9-23.2)	16.3 (15.4-17)	20.1 (19.3-21.1)
Resection	52.5 (50.1-54.7)	51.5 (49.1-54.1)	55.3 (50.7-63.6)	43.8 (41.7-46.8)
Chemoradiotherapy	19.1 (18.2-20.1)	16.9 (16.2-18)	23.1 (21.3-24.9)	-
Treatment without curative potential	9.7 (9.4-10)	9.9 (9.6-10.5)	9.2 (8.7-9.7)	-
Best supportive care	3.4 (3.2-3.5)	3.9 (3.7-4.3)	2.6 (2.3-3)	-
No resection	-	-	-	4.5 (4.2-4.9)

Supplementary Table 2. Survival scenarios of patients with esophageal adenocarcinoma and squamous cell carcinoma between 2006 and 2020.

	Esophageal squamous cell carcinoma				
	Observed until	OS (months), 2006	OS (months), Last observed year	Slope (95% CI)	p-value
Total population (N)				N=6,298	
p20	2013	42	80	7.54 (4.60-10.48)	.002
p40	2019	17	35	1.38 (1.15-1.61)	<.001
p50	2020	11	20	0.69 (0.57-0.81)	<.001
p60	2020	8	14	0.34 (0.21-0.47)	<.001
p80	2020	5	6	0.10 (0.05-0.15)	<.001
Resection (N)				N=2,049	
p20	[A]			[A]	[A]
p40	2013	35	72	9.51 (5.99-13.03)	.002
p50	2015	25	57	4.73 (3.29-6.17)	<.001
p60	2017	19	55	2.71 (2.24-3.18)	<.001
p80	2020	9	15	1.02 (0.83-1.21)	<.001
				N=1,790	
p20	2015	51	72	-0.91 (-3.69-1.87)	.537
p40	2018	22	36	0.98 (0.34-1.62)	.012
p50	2019	20	32	0.57 (-0.01-1.15)	.078
p60	2020	18	22	0.35 (0.05-0.65)	.042
p80	2020	15	12	0.24 (0.00-0.48)	.072

[A] The 20th percentile was not observed.

Supplementary Table 2 (Continued). Survival scenarios of patients with esophageal adenocarcinoma and squamous cell carcinoma between 2006 and 2020.

	Esophageal adenocarcinoma				
	Observed until	OS (months), 2006	OS (months), Last observed year	Slope (95% CI)	p-value
Total population (N)				N=17,535	
p20	2011	74	126	12.19 (7.69-16.69)	.006
p40	2018	23	40	1.59 (1.27-1.91)	<.001
p50	2019	16	27	0.81 (0.68-0.94)	<.001
p60	2020	11	18	0.51 (0.42-0.60)	<.001
p80	2020	5	8	0.27 (0.21-0.33)	<.001
Resection (N)				N=10,495	
p20	A			[A]	[A]
p40	2013	48	87	7.35 (5.72-8.98)	<.001
p50	2017	33	59	2.89 (1.84-3.94)	<.001
p60	2018	23	40	1.32 (0.90-1.74)	<.001
p80	2020	10	18	0.52 (0.40-0.64)	<.001
Chemoradiotherapy (N)				N=2,148	
p20	2018	20	42	0.67 (-0.35-1.69)	.226
p40	2019	15	29	0.48 (0.07-0.89)	.042
p50	2020	12	20	0.40 (0.18-0.62)	.002
p60	2020	11	16	0.22 (0.04-0.40)	.036
p80	2020	6	9	0.08 (-0.10-0.26)	.397

[A] The 20th percentile was not observed.

Supplementary Table 3. Survival scenarios of patients with esophageal adenocarcinoma intestinal and diffuse type 2006 and 2020.

	Esophageal cancer				
	Observed until	OS (months), 2006	OS (months), Last observed year	Slope (95% CI)	p-value
Intestinal type (N)				N=10,715	
p20	2011	78	124	11.88 (4.60-19.16)	.003
p40	2017	24	52	2.16 (1.73-2.59)	<.001
p50	2018	16	33	1.24 (0.99-1.49)	<.001
p60	2019	12	25	0.88 (0.72-1.04)	<.001
p80	2020	5	10	0.39 (0.31-0.47)	<.001
Diffuse type (N)				N=2,613	
p20	2016	29	43	5.5 (2.08-8.92)	.014
p40	2019	17	22	0.36 (0.13-0.59)	.008
p50	2020	12	16	0.3 (0.19-0.41)	<.001
p60	2020	9	13	0.29 (0.18-0.40)	<.001
p80	2020	4	7	0.18 (0.09-0.27)	.002

Supplementary Table 3 (Continued). Survival scenarios of patients with esophageal adenocarcinoma intestinal and diffuse type 2006 and 2020.

	Gastric cancer				
	Observed until	OS (months), 2006	OS (months), Last observed year	Slope (95% CI)	p-value
Intestinal type (N)				N=4,165	
p20	2011	107	111	1.95 (0.54-3.36)	.073
p40	2017	29	48	1.57 (1.10-2.04)	<.001
p50	2018	22	46	1.03 (0.41-1.65)	.007
p60	2019	14	20	0.87 (0.53-1.21)	<.001
p80	2020	4	10	0.47 (0.37-0.57)	<.001
Diffuse type (N)				N=3,517	
p20	2012	64	80	4.88 (-4.03-13.79)	.537
p40	2020	18	24	0.58 (0.06-1.10)	.012
p50	2020	12	19	0.30 (-0.01-0.61)	.078
p60	2020	9	13	0.18 (0.01-0.35)	.042
p80	2020	2	6	0.19 (0.08-0.30)	.072

Supplementary Table 4. Sensitivity analysis of survival scenarios of patients with esophageal or gastric cancer diagnosed between 2006 and 2020 with excluded cT4b patients.

	Esophageal cancer				
	Observed until	OS (months), 2006	OS (months), Last observed year	Slope (95% CI)	p-value
Total population (N)				N=25,400	
p20	2011	55	113	12.45 (8.83-16.07)	.003
p40	2018	18	35	1.5 (1.25-1.75)	<.001
p50	2020	13	23	0.7 (0.58-0.82)	<.001
p60	2020	9	15	0.4 (0.30-0.5)	<.001
p80	2020	4	7	0.17 (0.13-0.21)	<.001
Resection (N)				N=12,760	
p20	A			A	
p40	2007	45	83	8.03 (6.26-9.8)	<.001
p50	2013	29	57	3.72 (2.85-4.59)	<.001
p60	2016	21	43	1.76 (1.43-2.09)	<.001
p80	2018	10	18	0.57 (0.45-0.69)	<.001
Chemoradiotherapy (N)				N=4,237	
p20	2017	69	45	-1.26 (-2.17--0.35)	.022
p40	2019	18	33	0.51 (0.03-0.99)	.057
p50	2020	16	23	0.28 (-0.01-0.57)	.076
p60	2020	13	17	0.16 (-0.06-0.38)	.184
p80	2020	8	10	0.06 (-0.1-0.22)	.507

[A] The 20th percentile was not observed.

Supplementary Table 4 (Continued). Sensitivity analysis of survival scenarios of patients with esophageal or gastric cancer diagnosed between 2006 and 2020 with excluded cT4b patients.

	Gastric cancer				
	Observed until	OS (months), 2006	OS (months), Last observed year	Slope (95% CI)	p-value
Total population (N)				N=10,763	
p20	2012	81	100	5.21 (2.40-8.02)	.015
p40	2018	23	42	1.03 (0.64-1.42)	<.001
p50	2020	15	23	0.63 (0.39-0.87)	<.001
p60	2020	10	17	0.39 (0.26-0.52)	<.001
p80	2020	3	6	0.11 (0.05-0.17)	.002
Resection (N)				N=6,881	
p20	B			B	
p40	2008	44	65	7.49 (-3.37-18.35)	.405
p50	2015	28	49	2.79 (1.29-4.29)	.006
p60	2017	21	39	1.36 (0.8-1.92)	<.001
p80	2018	8	16	0.92 (0.58-1.26)	<.001
Chemoradiotherapy (N)					
p20	-			-	-
p40	-			-	-
p50	-			-	-
p60	-			-	-
p80	-			-	-

[B] Slope was not estimated since less than four years were observed.
- Group was not analyzed as chemoradiotherapy is not a treatment with curative potential in gastric cancer

Supplementary Table 5. Sensitivity analysis of survival scenarios of patients with esophageal adenocarcinoma and squamous cell carcinoma between 2006 and 2020 with excluded cT4b patients.

	Esophageal adenocarcinoma				
	Observed until	OS (months), 2006	OS (months), Last observed year	Slope (95% CI)	p-value
Total population (N)				N=17,891	
p20	2011	69	124	12.85 (9.27-16.43)	.002
p40	2018	21	40	1.59 (1.28-1.9)	<.001
p50	2019	15	24	0.84 (0.69-0.99)	<.001
p60	2020	10	17	0.53 (0.43-0.63)	<.001
p80	2020	5	8	0.27 (0.21-0.33)	<.001
Resection (N)				N=10,536	
p20	A			A	A
p40	2013	47	87	7.52 (5.82-9.22)	<.001
p50	2017	31	59	2.99 (1.92-4.06)	<.001
p60	2018	23	40	1.33 (0.91-1.75)	<.001
p80	2020	10	18	0.54 (0.42-0.66)	<.001
Chemoradiotherapy (N)				N=2,199	
p20	2018	19	42	0.58 (-0.45-1.61)	.295
p40	2019	15	29	0.48 (0.07-0.89)	.044
p50	2020	12	22	0.41 (0.2-0.62)	.002
p60	2020	11	16	0.19 (0.02-0.36)	.040
p80	2020	8	9	0.00 (-0.16-0.16)	.992

[A] The 20th percentile was not observed.

Supplementary Table 5 (Continued). Sensitivity analysis of survival scenarios of patients with esophageal adenocarcinoma and squamous cell carcinoma between 2006 and 2020 with excluded cT4b patients.

	Esophageal squamous cell carcinoma				
	Observed until	OS (months), 2006	OS (months), Last observed year	Slope (95% CI)	p-value
Total population (N)				N=6,952	
p20	2013	38	78	7.19 (4.62-9.76)	.002
p40	2019	15	33	1.31 (1.11-1.51)	<.001
p50	2020	10	17	0.54 (0.39-0.69)	<.001
p60	2020	7	12	0.28 (0.17-0.39)	<.001
p80	2020	4	5	0.11 (0.08-0.14)	<.001
Resection (N)				N=2,111	
p20	A			A	A
p40	2013	34	78	2.48 (-33.01-37.97)	.913
p50	2015	24	57	8.75 (4.91-12.59)	.004
p60	2017	18	55	4.95 (3.61-6.29)	<.001
p80	2020	9	14	2.80 (2.31-3.29)	<.001
Chemoradiotherapy (N)				N=1,975	
p20	2015	172	71	-3.24 (-7.2-0.72)	.148
p40	2018	26	35	0.7 (0.12-1.28)	.039
p50	2019	23	23	0.4 (0.01-0.79)	.063
p60	2020	18	22	0.33 (0.02-0.64)	.053
p80	2020	14	11	0.15 (-0.03-0.33)	.120

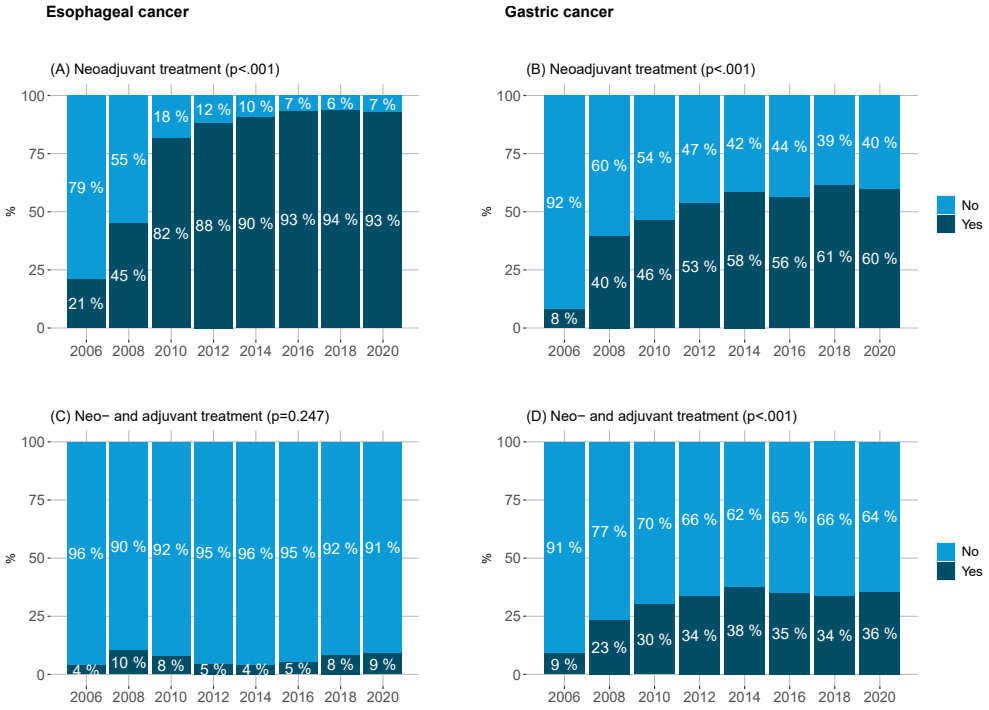
[A] The 20th percentile was not observed.

Supplementary Table 6. Sensitivity analysis of survival scenarios of patients with esophageal adenocarcinoma intestinal and diffuse type 2006 and 2020 with excluded cT4b patients.

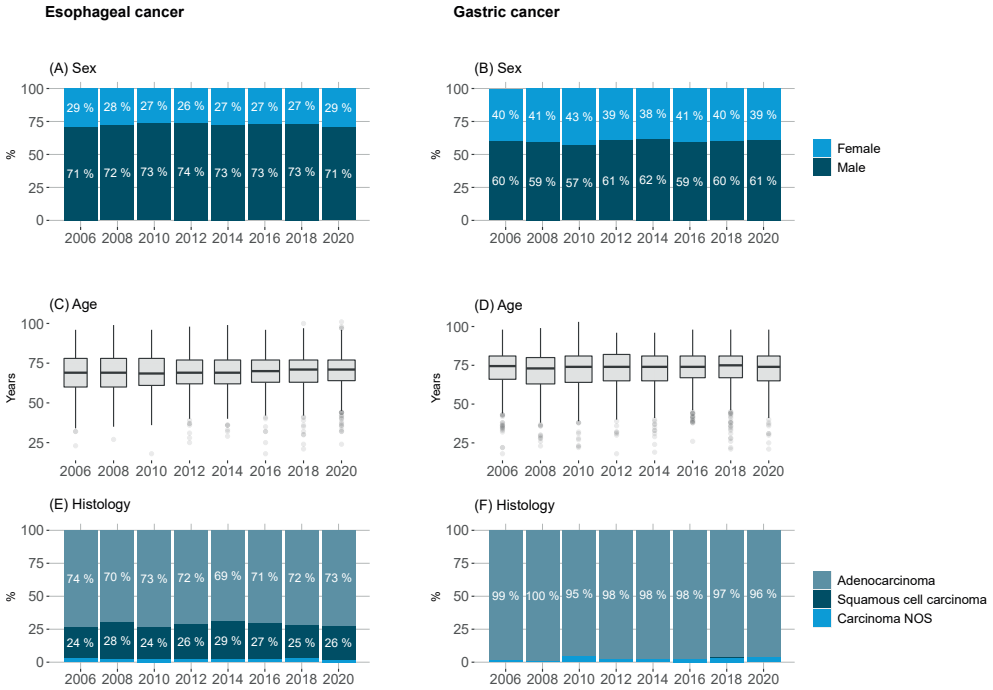
			Esophageal cancer		
	Observed until	OS (months), 2006	OS (months), Last observed year	Slope (95% CI)	p-value
Intestinal type (N)				N=10,900	
p20	2011	73	121	11.82 (4.78-18.86)	.003
p40	2017	23	52	2.14 (1.78-2.5)	<.001
p50	2018	16	33	1.25 (1.01-1.49)	<.001
p60	2019	11	25	0.90 (0.74-1.06)	<.001
p80	2020	5	10	0.40 (0.32-0.48)	<.001
Diffuse type (N)				N=2,690	
p20	2016	29	43	5.39 (2.65-8.13)	.003
p40	2019	14	22	0.38 (0.16-0.6)	.006
p50	2020	11	16	0.29 (0.17-0.41)	<.001
p60	2020	8	12	0.29 (0.18-0.4)	<.001
p80	2020	4	7	0.17 (0.08-0.26)	.002

Supplementary Table 6 (Continued). Sensitivity analysis of survival scenarios of patients with esophageal adenocarcinoma intestinal and diffuse type 2006 and 2020 with excluded cT4b patients.

Gastric cancer				
Observed until	OS (months), 2006	OS (months), Last observed year	Slope (95% CI)	p-value
			N=4,488	
2011	94	109	3.51 (0.91-6.11)	.077
2017	26	42	1.61 (1-2.22)	<.001
2018	20	32	1.33 (0.79-1.87)	<.001
2019	12	20	0.88 (0.57-1.19)	<.001
2020	3	19	0.40 (0.31-0.49)	<.001
			N=3,789	
2012	47	58	2.09 (-3.21-7.39)	.470
2020	14	24	0.69 (0.25-1.13)	.009
2020	11	19	0.42 (0.21-0.63)	.002
2020	8	12	0.23 (0.1-0.36)	.003
2020	2	6	0.21 (0.11-0.31)	.002



Supplementary Figure 1. Proportion of patients with esophageal cancer (left) and gastric cancer (right) who underwent resection and recived neoadjuvant and adjuvant treatment. P-values of the Cochrane-Armitage test are between parentheses.



Supplementary Figure 2. Baseline characteristics of patients over time of patients with esophageal cancer (left) and gastric cancer (right).

Chapter 3

Conditional relative survival in non-metastatic esophagogastric cancer between 2006 and 2020: a population-based study

Marieke Pape, Steven C. Kuijper, Pauline A.J. Vissers, Jelle P. Ruurda, Karen J. Neelis, Hanneke W.M. van Laarhoven, Rob H.A. Verhoeven

Based on:
Pape M, Kuijper SC, Vissers PAJ, et al. Conditional relative survival in nonmetastatic esophagogastric cancer between 2006 and 2020: A population-based study. *Int J Cancer*. 2023

Abstract

Background. Conditional relative survival is useful for communicating prognosis to patients as it provides an estimate of the life expectancy after having already survived a certain time after treatment. This study estimates the three year relative survival conditional on having survived a certain period for patients with esophageal or gastric cancer.
Methods. Patients with non-metastatic esophageal or gastric cancer diagnosed between 2006-2020 treated with curative intent (resection with or without (neo)adjuvant therapy, or chemoradiotherapy) were selected from the Netherlands Cancer Registry. Conditional relative survival was calculated since resection or last day of chemoradiotherapy.
Results. The probability of surviving an additional three years (i.e. three year conditional relative survival), if the patients survived 1,3 and 5 years after diagnosis was 62%, 79%, 87% and 69%, 84%, 90% for esophageal and gastric cancer, respectively. The three year conditional relative survival after having survived 3 years for patients with esophageal cancer who underwent a resection (n=12,204) was 91%, 88%, 77% and 60% for pathological stage 0, I, II and III, and for patients with esophageal cancer who received chemoradiotherapy (n=4,158) was 51% and 66% for clinical stage II and III, respectively. The three year conditional relative survival after having survived after having survived 3 years for patients with gastric cancer who underwent a resection (n=6,531) was 99%, 90%, 73% and 59% for pathological stage 0, I, II and III, respectively.
Conclusion. Despite poor prognosis of patients with esophageal or gastric cancer, life expectancy increases substantially after patients have survived several years after treatment. This study provides valuable information for communication of prognosis to patients during follow-up after treatment.

Introduction

Esophageal and gastric cancer are the sixth and third common cause of death from cancer worldwide, respectively.[1] In the Netherlands, annually approximately 3000 patients are diagnosed with esophageal cancer (including gastroesophageal junction cancer) and approximately 1100 with gastric cancer.[2] In the Netherlands, survival of patients with esophageal and gastric cancer has improved due to novel treatment options, centralization of surgery and installment of an obligatory national surgical audit.[3-9] The introduction of neoadjuvant chemoradiotherapy improved survival in esophageal cancer and neoadjuvant chemotherapy improved survival in gastric cancer.[3, 5, 8, 10] Improved survival was also observed due to a reduction in complications and an improvement in quality of surgery after centralization of esophageal and gastric cancer (annual minimum of 20 esophagectomies and gastrectomies per hospital) and the installment of the Dutch Upper GI Cancer Audit (DUCA).[4, 6, 7, 9, 11] For patients with esophageal cancer in the curative setting, chemoradiotherapy followed by surgery is the standard of care and recently adjuvant nivolumab is recommended in case of an incomplete response.[8, 12] For patients with esophageal cancer, definitive chemoradiotherapy is an alternative treatment option for patients with poor functional status, unwilling to undergo surgery or who have an inoperable tumor.[13, 14] Approximately 50% and 70% of patients with non-metastatic esophageal and gastric cancer receive endoscopic or surgical resection, respectively.[15] Use of definitive chemoradiotherapy has increased over the years and approximately 20% of patients with non-metastatic esophageal cancer receive this type of treatment.[15, 16]

Five-year survival of patients with esophageal or gastric cancer receiving treatment with curative intent is under 50%.[3, 17] These survival estimates are informative at diagnosis but become less relevant over time as prognosis changes for patients still alive after the initial years since diagnosis. In such instances, conditional survival is more meaningful as conditional survival estimates the survival of patients given the fact that patients have already survived a time period after diagnosis. Conditional survival is defined as the probability of surviving an additional number of *y* years on the condition that a patient has already survived *x* years.[18] Conditional relative survival estimates are also corrected for the expected survival of the general population. Previous studies in esophagogastric have focused on the conditional survival instead of the conditional relative survival and therefore these estimates are not corrected for other causes of death other than cancer.[19-22] Physicians can use conditional relative survival rates in communication with patients during follow-up about their current life expectancy.

The aim of this study was to provide insights in the three year conditional relative survival of patients with non-metastatic esophageal or gastric cancer after treatment with curative intent. Additionally, separate analyses were performed for two periods (2006-2012 and 2013-2020) to investigate change in survival over time as survival in esophageal or gastric cancer has improved.

Methods

Study population

Patients diagnosed with non-metastatic esophageal (C15.0-C15.9), gastroesophageal junction/cardia (C16.0) or gastric cancer (C16.1-C16.9) who received endoscopic or surgical resection or chemoradiotherapy (esophageal or junctional cancer only) diagnosed between 2006-2020 were selected from the Netherlands Cancer Registry (NCR).[23] The NCR contains all diagnosed malignancies in the Netherlands and is notified by the national automated pathology archive. Specially trained employees routinely extract information on diagnosis, tumor stage and treatment from medical records. Information on vital status was obtained through annual linkage with the Dutch Personal Records Database and updated until February 1, 2022.

Tumors were staged according to the sixth (2006-2009), seventh (2010-2016) and eight (2017-2020) edition of the Union for International Cancer Control (UICC) TNM classification.[24-26] Patients with gastroesophageal junction/cardia cancer were classified as esophageal cancer. All patients with non-metastatic disease (cM0) were included. Additionally, patients with cM1a stage according to TNM-6 were included, as most patients with a M1a stage had a distal tumor with coeliac lymph nodes which is considered cN+ instead of cM1 according to TNM-7 and TNM-8. Patients with pM1 stage were considered as having metastatic disease and excluded from the analysis (n=203 and n=272 for esophageal and gastric cancer, respectively) (Supplementary figure 1). As tumor stage classification was vastly different in TNM-7 and -8 compared to -6, tumor stage was recoded according to TNM-6. For patients who underwent surgery, pathological stage of the resection specimen was noted if available, otherwise the clinical stage was used. A pT0N0M0 or pT0NXM0 stage was classified as a complete pathological response (stage 0). For patients with esophageal cancer receiving chemoradiotherapy the clinical tumor stage was noted as classification of the pathological tumor stage based on the resection specimen is impossible.

Statistical analysis

Patient and tumor characteristics were displayed with frequencies and percentages for categorical variables, or median and interquartile range for continuous variables. Relative survival (with 95% confidence intervals) was calculated since end of treatment (resection or last day of chemoradiotherapy) using life tables of the general Dutch population matched by age, sex and calendar year with the Ederer II method.[27] The Ederer II method estimates relative survival by comparing survival estimates from patients included in the study to survival estimates of the general population with comparable demographic characteristics (age, sex, and calendar year).[27] Conditional relative survival is defined as the probability of surviving an additional y years on the condition that the patient has survived x years, corrected for the expected survival of the general population.[28] The conditional relative survival is calculated as , where S(x) is the relative survival at time x. Conditional three year relative survival was calculated from 0 to 10 years after treatment (i.e. the probability of surviving an additional 3 years) in half-year intervals as previously described.[28, 29] First, the three year relative survival was calculated and for each half year interval after this date (i.e. 0.5, 1.0, 1.5, up until 10 years), three year relative survival was calculated for patients still alive at the specific time points. Conditional three year relative survival analyses were performed separately for

patients with esophageal and gastric cancer. Analyses were stratified by treatment and tumor stage. For analyses stratified by (clinical or pathological) tumor stage, analyses for patients with an unknown tumor stage (n=441) were not performed due to irrelevance for the clinical practice. To calculate conditional survival a minimum of 25 patients for each (half) year of follow-up was deemed necessary.

If resection date was missing, the average time from diagnosis until resection (separately calculated with or without neoadjuvant therapy when applicable) was subtracted from the follow-up time to estimate survival time (n=47). If last day of chemoradiotherapy was missing and start date was available, average time of duration of chemoradiotherapy used to estimate last day of chemoradiotherapy (n=1027). If both start and end date of chemoradiotherapy were missing, the average time from diagnosis until last day of chemoradiotherapy used to estimate last day of chemoradiotherapy (n=70). Patients who died within the adjusted time interval were excluded (n=23) (Supplementary figure 1).

Separate analyses of the relative and conditional three year relative survival were performed for two periods (2006-2012 and 2013-2020) to investigate change in survival over time. In all analyses non-overlapping 95% confidence intervals were judged as statistically significant. All analyses were performed using STATA/SE version 17.0 (StataCorp, College Station, Texas, USA).

Results

We included 16,362 and 6,531 patients with esophageal and gastric cancer, respectively (Supplementary figure 1, Table 1). In patients with esophageal cancer, 12,204 patients underwent resection and 4,158 patients received chemoradiotherapy not followed by resection. The three year relative survival since end of treatment was 52% and 58% for all patients with esophageal and gastric cancer (Supplementary table 1). The three year relative survival significantly increased from the period 2006-2012 to the period 2013-2020 from 50% to 53% and 56% to 61% for esophageal and gastric cancer, respectively (Supplementary table 1, Figure 1). For patients with esophageal cancer who underwent surgery, the three year relative survival between 2006-2012 was 55% for all stages and 72%, 82%, 54% and 27% for stage 0, stage I, stage II and stage III, respectively (Supplementary table 1, Figure 2A). For patients with esophageal cancer who received chemoradiotherapy, the 3 year relative survival between 2006-2012 was 27% for all stages and 33% and 22% for stage II and stage III, respectively (Figure 2B). For gastric cancer, the three year relative survival between 2006-2012 was 56% for all stages and 93%, 81%, 50% and 25% for stage 0, stage I, stage II, stage III, respectively (Figure 2C). For patients with esophageal cancer who underwent surgery, the three year relative survival between 2013-2020 was 61% for all stages and 72%, 82%, 54% and 27% for stage 0, stage I, stage II and stage III, respectively (Supplementary table 1, Figure 2A). For patients with esophageal cancer who received chemoradiotherapy, the 3 year relative survival between 2013-2020 was 32% for all stages and 37% and 26% for stage II and stage III, respectively (Figure 2B). For gastric cancer, the three year relative survival between 2013-2020 was 61% for all stages and 100%, 83%, 55% and 27% for stage 0, stage I, stage II, stage III, respectively (Figure 2C).

Table 1. Baseline characteristics of patients stratified by primary tumor location.

	Esophageal cancer	Gastric cancer
	N=16,362	N=6,531
Age (years), median (IQR)	67.0 (60.0-73.0)	71.0 (62.0-78.0)
Sex, n (%)		
Male	12,377 (75.6)	4,035 (61.8)
Female	3,985 (24.4)	2,496 (38.2)
Primary tumor location n (%)		
Esophageal	13,681 (83.6)	0 (0.0)
Gastroesophageal junction/Cardia	2,681 (16.4)	0 (0.0)
Gastric	0 (0.0)	6,531 (100.0)
Histology n (%)		
Adenocarcinoma	12,160 (74.3)	6,480 (99.2)
Squamous cell carcinoma	4,031 (24.6)	10 (0.2)
Carcinoma NOS	171 (1.0)	41 (0.6)
Type of treatment n (%)		
Endoscopic resection	1,414 (8.6)	217 (3.3)
Surgical resection	10,790 (65.9)	6,314 (96.7)
No (neo)adjuvant therapy	2,060 (19.5)	3,108 (49.7)
Neoadjuvant therapy only	7,817 (73.8)	1,240 (19.8)
Neoadjuvant and adjuvant therapy	686 (6.5)	1,765 (28.2)
Adjuvant therapy only	26 (0.2)	144 (2.3)
Chemoradiotherapy	4,158 (25.4)	-
Clinical tumor stage for patients receiving chemoradiotherapy, n (%)		
I	39 (0.9)	
II	1,743 (41.9)	
III	2,043 (49.1)	
Unknown	333 (8.0)	
Pathological tumor stage for patients who underwent resection, n (%)1		
0 (complete response)	1,721 (14.1)	229 (3.5)
I	2,768 (22.7)	2,786 (42.7)
II	4,575 (37.5)	1,638 (25.1)
III	3,081 (25.2)	1,353 (20.7)
IV	-	476 (7.3)
Unknown	59 (0.5)	49 (0.8)

[1]Clinical tumor stage instead of pathological tumor stage was used for 203 (1.2%) and 52 (0.8%) patients with esophageal and gastric cancer, respectively. NOS: not otherwise specified

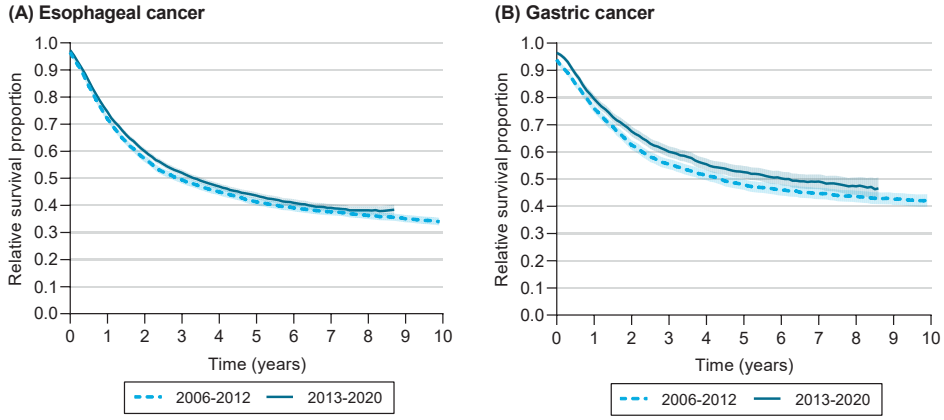


Figure 1. Relative survival with 95% confidence interval of all patients with esophageal (A) or gastric (B) cancer stratified for the period2006-2012 and 2013-2020.

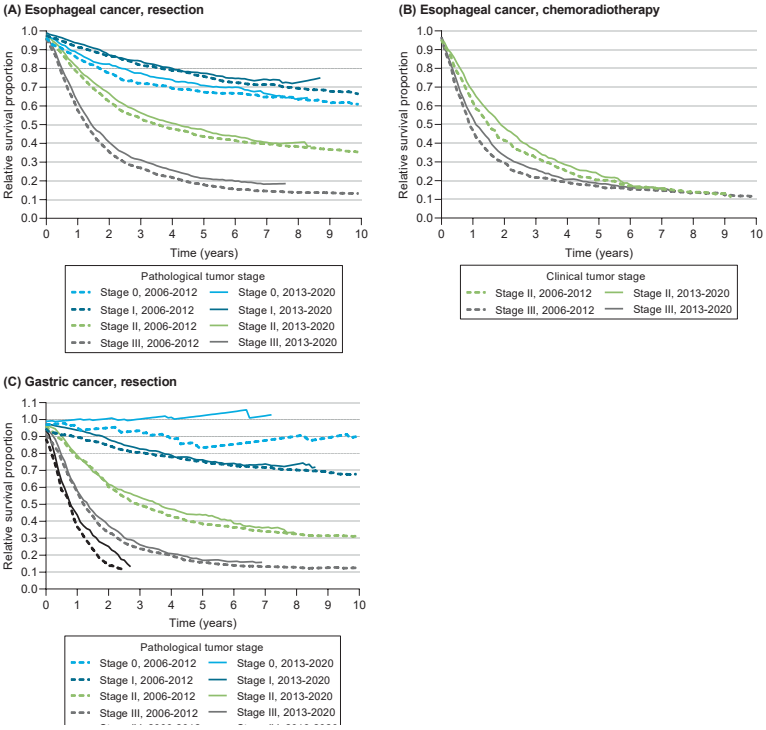


Figure 2. Relative survival of patients with esophageal cancer who underwent resection (A), patients with esophageal cancer receiving chemoradiotherapy (B) and patients with gastric cancer who underwent resection (C) cancer stratified by pathological or clinical tumor stage and the period 2006-2012 and 2013-2020. Patients with an unknown tumor stage were excluded.

Esophageal cancer

The probability of surviving an additional three years (i.e. three year conditional relative survival), if patients with esophageal cancer survived 1 year after treatment was 62% and increased thereafter to 79% 87% and 93% if patients survived 3, 5 and 10 years after treatment (Table 2, Figure 3A). Between the period 2006-2012 and 2013-2020, the three year relative survival conditional on having survived 1 year (61% versus 62%), 3 years (79% versus 79%) and 5 years (87% versus 87%) was similar, respectively (Table 2, Supplementary figure 1A). Conditional relative survival estimates stratified by age and histology are available in Supplementary table 2 and 3.

In patients who underwent surgery, the three year conditional relative survival if patients survived 3 years after treatment was 91%, 88%, 77% and 60% for pathological stage 0 (complete response), stage I, stage II and stage III, respectively (Table 2, Figure 4A).

In patients who received chemoradiotherapy, the three year conditional relative survival if patients survived 3 years after treatment was 51% and 66% for clinical tumor stage II and III, respectively (Table 2, Figure 4B).

Gastric cancer

The probability of surviving an additional three years (i.e. three year conditional relative survival), if patients with gastric cancer survived 1 year after treatment was 68% and increased to 83%, 90% and 95% if patients survived 3, 5 and 10 years after treatment, respectively (Table 2, Figure 3B). Between the period 2006-2012 and 2013-2020, the three year relative survival conditional on having survived 1 year (67% versus 69%), 3 years (83% versus 84%) and 5 years (91% versus 90%) was similar, respectively (Table 2, Supplementary figure 1B). Conditional relative survival estimates stratified by age is available in Supplementary table 2.

The three year conditional relative survival if patients survived 3 years after treatment was 99%, 90%, 73% and 59% in for pathological stage 0, stage I, stage II and stage III, respectively (Table 2, Figure 4C). For pathological stage IV, the three year conditional relative survival if patients survived 1 year after treatment was 19%.

Table 2. Three year conditional relative survival for esophageal and gastric cancer stratified by pathological (resection) or clinical (chemoradiotherapy) tumor stage.

Cancer type	Treatment	Period	Tumor stage	Number of patients	Years already survived, % (95% CI)			
					1	3	5	10
Esophageal	Resection or chemo-radiotherapy	2006-2020	All stages	16,362	62 (61-63)	79 (77-80)	87 (86-89)	93 (90-95)
		2006-2012	All stages	6,385	61 (60-63)	79 (77-80)	87 (86-89)	93 (90-95)
		2013-2020	All stages	9,977	62 (61-63)	79 (77-80)	87 (84-90)	-
	Resection	2006-2020	All stages	12,204	67 (66-68)	81 (80-82)	89 (87-90)	93 (90-96)
			0	1,721	83 (80-85)	91 (88-93)	93 (90-96)	93 (80-101)
			I	2,768	86 (84-88)	88 (86-91)	92 (89-94)	92 (86-97)
			II	4,575	62 (60-64)	77 (75-79)	86 (84-89)	95 (90-98)
			III	3,081	38 (36-40)	60 (56-64)	78 (73-83)	94 (85-100)
	Chemoradiotherapy	2006-2020	All stages	4,158	40 (38-43)	59 (55-63)	70 (64-76)	-
			II	1,743	41 (37-44)	51 (45-58)	66 (55-75)	
			III	2,043	38 (34-41)	66 (60-73)	77 (68-85)	-
Gastric	Resection	2006-2020	All stages	6,531	68 (66-69)	83 (81-85)	90 (88-93)	95 (91-99)
		2006-2012	All stages	3,363	67 (65-69)	83 (80-85)	91 (88-93)	95 (91-99)
		2013-2020	All stages	3,168	69 (66-71)	84 (80-86)	90 (85-94)	-
		2006-2020	0	229	99 (94-102)	99 (93-103)	105 (98-107)	-
			I	2,786	86 (84-88)	90 (88-92)	94 (91-96)	94 (88-99)
			II	1,638	56 (53-59)	73 (68-77)	82 (76-86)	96 (86-104)
			III	1,353	34 (30-38)	59 (52-65)	80 (70-88)	94 (77-104)
			IV	476	19 (13-25)	-	-	-

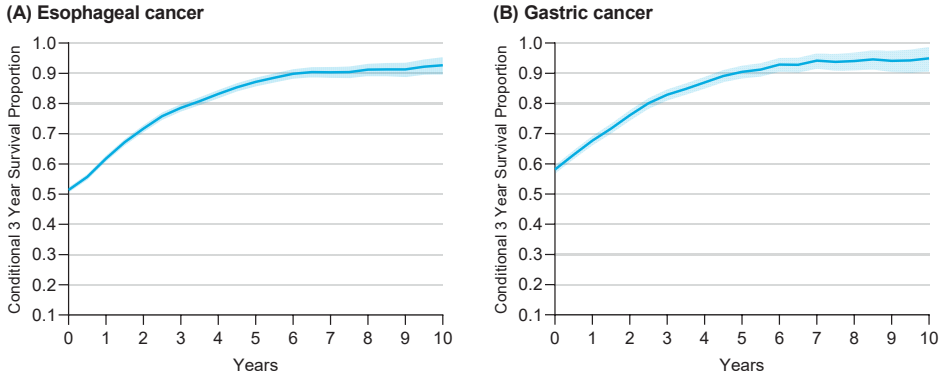


Figure 3. Conditional 3-year relative survival with 95% confidence interval for every year survived for patients with esophageal (A) and gastric cancer (B)

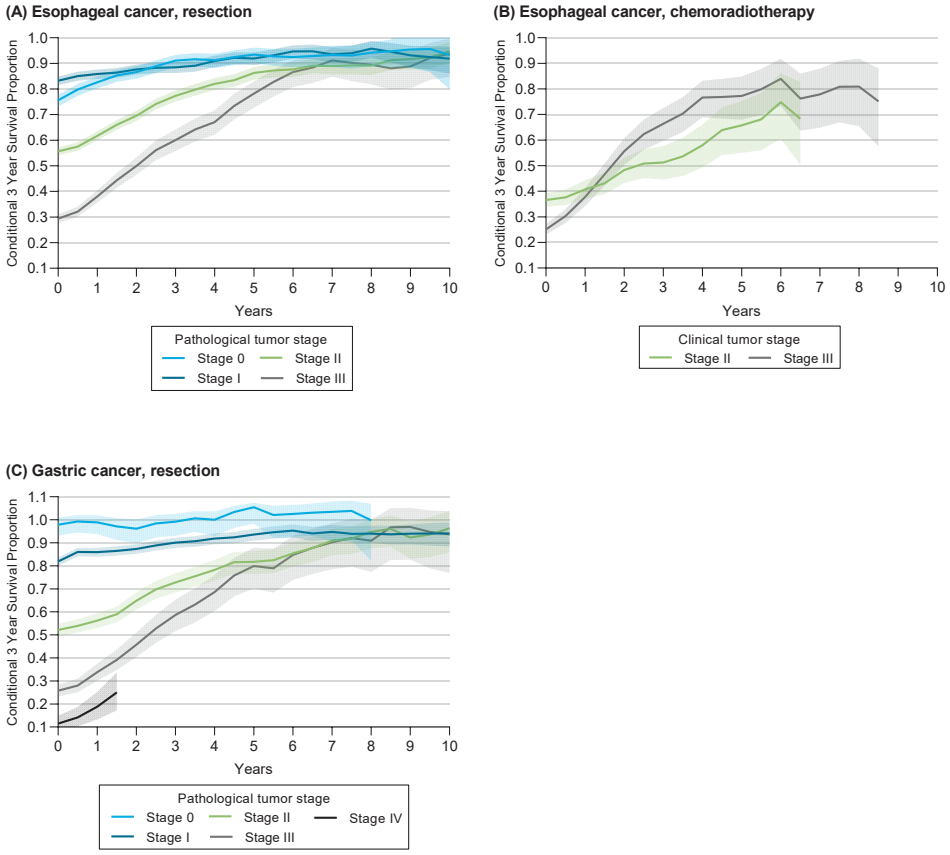


Figure 4. Conditional 3-year relative survival with 95% confidence interval for every year survived for patients with esophageal cancer who underwent resection (A), patients with esophageal cancer receiving chemoradiotherapy (B) and patients with gastric cancer who underwent resection (C) stratified by pathological or clinical tumor stage. Patients with an unknown tumor stage were excluded.

Discussion

This study in esophageal and gastric cancer shows that the three year conditional relative survival on having survived additional years after treatment improves over time. The probability of surviving the next three years for patients with non-metastatic disease who already survived three years after treatment with curative intent was 79% and 83% for esophageal and gastric cancer, respectively.

Three year conditional relative survival on having survived three years of patients with esophageal cancer who underwent resection (81%) was higher compared to a previous study of patients with esophageal cancer who received neoadjuvant chemoradiotherapy followed by surgery reporting three year conditional survival of 70%.[19] In our study conditional relative survival was calculated as compared to conditional survival, which could explain the higher observed survival. Furthermore, a retrospective observational study in the United States reported a five year conditional survival if patients survived 5 years after diagnosis of 67% and 63% for patients with localized and regional esophageal cancer (independent of treatment), respectively.[20] Our study included only patients who received treatment with curative intent (resection or chemoradiotherapy) and reported the probability of surviving an additional three years instead of an additional five years after having already survived 5 years, which probably explains the higher three year conditional relative survival after 5 years in our study (87%).

Surprisingly, for patients with esophageal cancer who received chemoradiotherapy the probability of surviving an additional three years, if patients survived 3 years after treatment was lower in clinical stage II compared to clinical stage III disease (66% versus 51%). There is a level of inaccuracy for clinical staging which could lead to misclassification of stage II and III disease resulting in less accurate survival estimates.30-32 As patients receiving chemoradiotherapy do not undergo resection, classification of the pathological tumor stage based on the resection specimen is impossible.

Three year conditional relative survival on having survived one year after resection for patients with gastric cancer (68%) was similar as compared to a previous study using data from the Surveillance, Epidemiology, and End Results database to calculate conditional survival in the United States (66%).[22] Another study in seven academic institutions in the United States of patients with gastric cancer who underwent resection reported three year conditional survival if patients survived 1, 3 and 5 years after treatment of 56%, 71% and 82%, respectively.[21] Estimates in our study were higher, with 68%, 83% and 90% three year conditional relative survival if patients survived 1, 3 and 5 years after treatment, respectively, which could be due to the fact that our study conditional relative survival was calculated as compared to conditional survival.

Our results are relevant for the clinical practice to more accurately inform patients with esophagogastric cancer of their life expectancy after treatment with curative intent, as their prognosis clearly improves over time. Our results show that conditional relative survival estimates increase up to 10 years after treatment for specific groups. The present study reported an increase in survival after stratification for pathological or clinical tumor for all stages conditional on having survived 1 year as compared to 3 and 5 years, with the exception of stage IV non-metastatic gastric cancer as survival could not be calculated after 1 year due to limited sample size. In addition, for patients with stage II and III esophageal

or gastric cancer, the three year conditional relative survival if patients survived 10 years (94 and 96%) after treatment was higher if survived five years (78 and 86%) after treatment.

The differences in conditional relative survival between the pathological stages decreased over time, with three year conditional relative survival if survived five years exceeding >90% only for stage 0 and I, and if survived 10 years exceeding 90% across all stages for esophageal and gastric cancer who underwent resection. Previous studies considered 'statistical cure', the point after which no excess mortality is reported on-group level, when the five year conditional relative survival exceeded 90% or 95%.[29, 33, 34] In our study we estimated the three year conditional relative survival as prognosis of esophageal or gastric cancer is poor. In patients with colorectal cancer who underwent resection statistical cure of >95% for the 5-year conditional relative survival was previously reported in pathological stages I-III within ten years.[29] In our study, the three year conditional relative survival only reached >95% for gastric cancer stage I (within conditional 1 year survived) and stage III (within 10 years survived) indicating that statistical cure of >95% is difficult to achieve in esophageal and gastric cancer.

For both esophageal and gastric cancer, the three year relative survival improved slightly between 2006-2012 and 2013-2020. However, the three year conditional relative survival if patients survived 1, 3 and 5 years was similar between these two periods. This indicates that survival has mainly improved in the first few years after treatment and more patients will survive the first initial 'crucial' years after treatment. This could amongst other be due to the reported decreased 30-day mortality after an esophagectomy (4.2% to 2.5%) or gastrectomy (7.1% and 4.3%) between 2011-2018 in the Netherlands.[9]

Our study has several limitations. Firstly, information on disease recurrence is not available in the NCR and prognosis after recurrence is poor. Secondly, risk factors for esophagogastric cancer, i.e. smoking, alcohol and obesity, could affect the life expectancy beyond the risk of developing esophagogastric cancer and therefore the relative survival estimates could be overestimated as these were corrected for the general population. Thirdly, due to potential misclassification of clinical stage II and III disease in esophageal cancer, the results of patients receiving chemoradiotherapy stratified by stage should be interpreted with caution. Lastly, all patients who received chemoradiotherapy were included as differentiation between definitive chemoradiotherapy and neoadjuvant chemoradiotherapy not followed by surgery was impossible.

In conclusion, this study reported three year conditional relative survival estimates for patients with esophageal or gastric cancer who received treatment with curative intent. Despite poor prognosis of patients with esophageal or gastric cancer, life expectancy increases substantially after patients having survived several years after treatment. This information is useful to provide relevant prognosis to patients with esophageal or gastric cancer during their follow-up trajectory after initial treatment.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71: 209-49.
2. Netherlands Comprehensive Cancer Organisation (IKNL). NKR cijfers. www.iknl.nl/nkr-cijfers [accessed December 2022].
3 . Al-Batran SE, Homann N, Pauligk C, Goetze TO, Meiler J, Kasper S, Kopp HG, Mayer F, Haag GM, Luley K, Lindig U, Schmiegel W, 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 2019;393: 1948-57.
4. Busweiler LA, Wijnhoven BP, van Berge Henegouwen MI, Henneman D, van Grieken NC, Wouters MW, van Hillegersberg R, van Sandick JW, Dutch Upper Gastrointestinal Cancer Audit G. Early outcomes from the Dutch Upper Gastrointestinal Cancer Audit. Br J Surg 2016;103: 1855-63.
5. Cunningham D, Allum WH, Stenning SP, Thompson JN, Van de Velde CJ, Nicolson M, Scarffe JH, Lofts FJ, Falk SJ, Iveson TJ, Smith DB, Langley RE, et al. Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. N Engl J Med 2006;355: 11-20.
6. Dikken JL, Dassen AE, Lemmens VE, Putter H, Krijnen P, van der Geest L, Bosscha K, Verheij M, van de Velde CJ, Wouters MW. Effect of hospital volume on postoperative mortality and survival after oesophageal and gastric cancer surgery in the Netherlands between 1989 and 2009. Eur J Cancer 2012;48: 1004-13.
7. van de Poll-Franse LV, Lemmens VE, Roukema JA, Coebergh JW, Nieuwenhuijzen GA. Impact of concentration of oesophageal and gastric cardia cancer surgery on long-term population-based survival. Br J Surg 2011;98: 956-63.
8. van Hagen P, Hulshof MC, van Lanschot JJ, Steyerberg EW, van Berge Henegouwen MI, Wijnhoven BP, Richel DJ, Nieuwenhuijzen GA, Hospers GA, Bonenkamp JJ, Cuesta MA, Blaisse RJ, et al. Preoperative chemoradiotherapy for esophageal or junctional cancer. N Engl J Med 2012;366: 2074-84.
9 . Voeten DM, Busweiler LAD, van der Werf LR, Wijnhoven BPL, Verhoeven RHA, van Sandick JW, van Hillegersberg R, van Berge Henegouwen MI, Dutch Upper Gastrointestinal Cancer Audit G. Outcomes of Esophagogastric Cancer Surgery During Eight Years of Surgical Auditing by the Dutch Upper Gastrointestinal Cancer Audit (DUCA). Ann Surg 2021;274: 866-73.
10. van Putten M, de Vos-Geelen J, Nieuwenhuijzen GAP, Siersema PD, Lemmens V, Rosman C, van der Sangen MJC, Verhoeven RHA. Long-term survival improvement in oesophageal cancer in the Netherlands. Eur J Cancer 2018;94: 138-47.
11. van Putten M, Nelen SD, Lemmens V, Stoot J, Hartgrink HH, Gisbertz SS, Spillenaar Bilgen EJ, Heisterkamp J, Verhoeven RHA, Nieuwenhuijzen GAP. Overall survival before and after centralization of gastric cancer surgery in the Netherlands. Br J Surg 2018;105: 1807-15.
12. Kelly RJ, Ajani JA, Kuzdzal J, Zander T, Van Cutsem E, Piessen G, Mendez G, Feliciano J, Motoyama S, Lievre A, Uronis H, Elimova E, et al. Adjuvant Nivolumab in Resected Esophageal or Gastroesophageal Junction Cancer. N Engl J Med 2021;384: 1191-203.
13. Dutch Clinical Practice Guidelines for Esophageal Carcinoma;01-12-2010. Available from: www.richtlijnendatabase.nl [accessed in February 2022].
14. Lordick F, Mariette C, Haustermans K, Obermannova R, Arnold D, Committee EG. Oesophageal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2016;27: v50-v7.
15. Luijten J, Haagsman VC, Luyer MDP, Vissers PAJ, Nederend J, Huysentruyt C, Creemers GJ, Curvers W, van der Sangen M, Heesakkers FBM, Schrauwen RWM, Jurgens MC, et al. Implementation of a regional video multidisciplinary team meeting is associated with an improved prognosis for

patients with oesophageal cancer A mixed methods approach. Eur J Surg Oncol 2021;47: 3088-96.
16. Koeter M, van Putten M, Verhoeven RHA, Lemmens V, Nieuwenhuijzen GAP. Definitive chemoradiation or surgery in elderly patients with potentially curable esophageal cancer in the Netherlands: a nationwide population-based study on patterns of care and survival. Acta Oncol 2018;57: 1192-200.
17. Shapiro J, van Lanschot JJB, Hulshof M, van Hagen P, van Berge Henegouwen MI, Wijnhoven BPL, van Laarhoven HWM, Nieuwenhuijzen GAP, Hospers GAP, Bonenkamp JJ, Cuesta MA, Blaisse RJB, et al. Neoadjuvant chemoradiotherapy plus surgery versus surgery alone for oesophageal or junctional cancer (CROSS): long-term results of a randomised controlled trial. Lancet Oncol 2015;16: 1090-8.
18. Hieke S, Kleber M, Konig C, Engelhardt M, Schumacher M. Conditional Survival: A Useful Concept to Provide Information on How Prognosis Evolves over Time. Clin Cancer Res 2015;21: 1530-6.
19. Hagens ERC, Feenstra ML, Eshuis WJ, Hulshof M, van Laarhoven HWM, van Berge Henegouwen MI, Gisbertz SS. Conditional survival after neoadjuvant chemoradiotherapy and surgery for oesophageal cancer. Br J Surg 2020;107: 1053-61.
20. Kim E, Koroukian S, Thomas CR, Jr. Conditional Survival of Esophageal Cancer: An Analysis from the SEER Registry (1988-2011). J Thorac Oncol 2015;10: 1490-7.
21. Kim Y, Ejaz A, Spolverato G, Squires MH, Poultsides G, Fields RC, Bloomston M, Weber SM, Votanopoulos K, Acher AW, Jin LX, Hawkins WG, et al. Conditional survival after surgical resection of gastric cancer: a multi-institutional analysis of the us gastric cancer collaborative. Ann Surg Oncol 2015;22: 557-64.
22. Zhong Q, Chen QY, Li P, Xie JW, Wang JB, Lin JX, Lu J, Cao LL, Lin M, Tu RH, Zheng CH, Huang CM. Prediction of Conditional Probability of Survival After Surgery for Gastric Cancer: A Study Based on Eastern and Western Large Data Sets. Surgery 2018;163: 1307-16.
23. Fritz A, Percy C, Jack A, Shanmugarathan S, Sobin L, Parkin DM, Whelan S. International classification of diseases for oncology: ICD-O. Third editioned.: World Health Organization, 2000.
24. TNM Classification of Malignant Tumours, 7th Editioned.: Wiley-Liss, 2009., International Union Against Cancer (UICC).
25. TNM Classification of Malignant Tumours, 8th Editioned.: Wiley-Blackwell, 2017., International Union Against Cancer (UICC).
26 . TNM Classification of Malignant Tumours, 6th Editioned.: Wiley, 2002, International Union Against Cancer (UICC).
27. Ederer F, Axtell LM, Cutler SJ. The relative survival rate: a statistical methodology. Natl Cancer Inst Monogr 1961;6: 101-21.
28. Dickman PW. Conditional survival http://pauldickman.com/software/stata/conditional-survival/ [accessed January 2022].
29. Qaderi SM, Dickman PW, de Wilt JHW, Verhoeven RHA. Conditional Survival and Cure of Patients With Colon or Rectal Cancer: A Population-Based Study. J Natl Compr Canc Netw 2020;18: 1230-7.
30. Rasanen JV, Sihvo EI, Knuuti MJ, Minn HR, Luostarinen ME, Laippala P, Viljanen T, Salo JA. Prospective analysis of accuracy of positron emission tomography, computed tomography, and endoscopic ultrasonography in staging of adenocarcinoma of the esophagus and the esophagogastric junction. Ann Surg Oncol 2003;10: 954-60.
31. Wakelin SJ, Deans C, Crofts TJ, Allan PL, Plevris JN, Paterson-Brown S. A comparison of computerised tomography, laparoscopic ultrasound and endoscopic ultrasound in the preoperative staging of oesophago-gastric carcinoma. Eur J Radiol 2002;41: 161-7.
32. Wallace MB, Nietert PJ, Earle C, Krasna MJ, Hawes RH, Hoffman BJ, Reed CE. An analysis of multiple staging management strategies for carcinoma of the esophagus: computed tomography, endoscopic ultrasound, positron emission tomography, and thoracoscopy/laparoscopy. Ann Thorac Surg 2002;74: 1026-32.

33. Dal Maso L, Panato C, Guzzinati S, Serraino D, Francisci S, Botta L, Capocaccia R, Tavilla A, Gigli A, Crocetti E, Rugge M, Tagliabue G, et al. Prognosis and cure of long-term cancer survivors: A population-based estimation. Cancer Med 2019;8: 4497-507.
34. Lambert PC, Thompson JR, Weston CL, Dickman PW. Estimating and modeling the cure fraction in population-based cancer survival analysis. Biostatistics 2007;8: 576-94.

Supplementary table 1. Three year relative survival since end of treatment for esophageal and gastric cancer stratified by pathological (resection) or clinical (chemoradiotherapy) tumor stage.

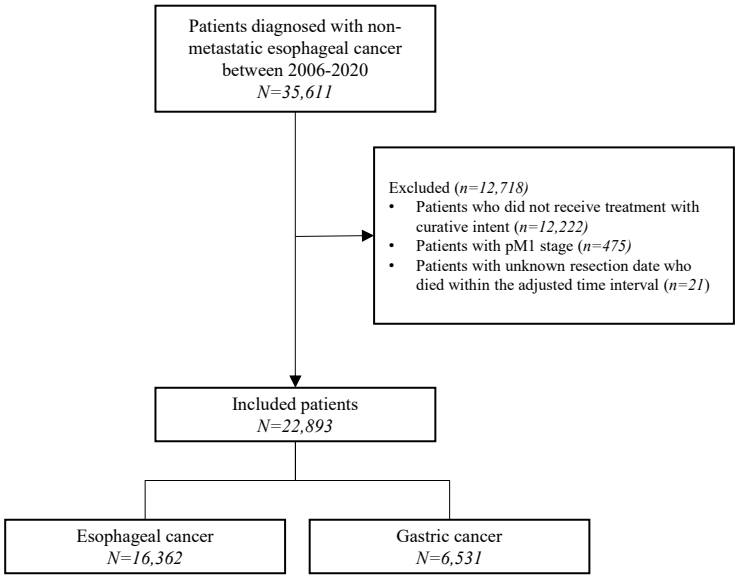
Cancer type	Treatment	Period	Tumor stage	Number of patients	3 year relative survival rate, % (95% CI)
Esophageal	Resection or Chemo-radiotherapy	2006-2002	All stages	16,362	52 (51-52)
		2006-2012	All stages	6,385	50 (49-51)
		2013-2020	All stages	9,977	53 (51-54)
	Resection	2006-2012	All stages	5,295	55 (53-56)
			0	565	72 (68-76)
			I	1,116	82 (79-85)
			II	2,091	54 (52-56)
			III	1,484	27 (25-30)
		2013-2020	All stages	6,909	61 (60-63)
			0	1,156	78 (75-80)
			I	1,652	84 (82-86)
			II	2,484	57 (55-59)
			III	1,597	32 (29-34)
	Chemoradiotherapy	2006-2012	All stages	1,090	27 (24-30)
			II	334	33 (28-39)
			III	574	22 (19-26)
		2013-2020	All stages	3,068	32 (30-34)
			II	1,409	37 (34-40)
			III	1,469	26 (24-29)
Gastric	Resection	2006-2020	All stages	6,531	58 (57-60)
		2006-2012	All stages	3,363	56 (54-58)
			0	64	93 (82-99)
			I	1,433	81 (78-83)
			II	883	50 (47-54)
			III	733	25 (21-28)
			IV	221	-
		2013-2020	All stages	3,168	61 (59-63)
			0	165	100 (94-103)
			I	1,353	83 (80-86)
			II	755	55 (51-59)
			III	620	27 (23-31)
			IV	255	-

Supplementary table 2. Three year conditional relative survival for esophageal and gastric cancer stratified by age (< 65 and ≥65 years) and histology (esophageal cancer only; adenocarcinoma and squamous cell carcinoma).

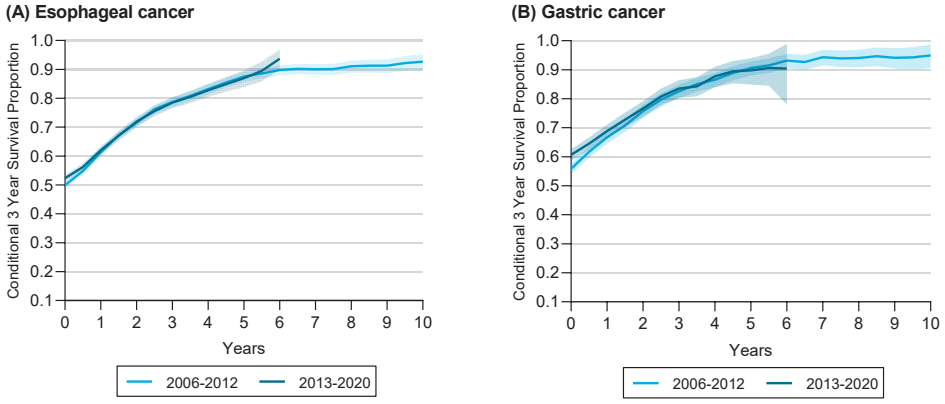
Cancer type	Age group	Number of patients	Years already survived, % (95% CI)			
			1	3	5	10
Esophageal	<65	6,818	62 (61-63)	81 (79-82)	90 (88-91)	95 (92-97)
	≥65	9,544	62 (60-63)	77 (75-78)	84 (81-86)	89 (82-95)
Gastric	<65	1,958	67 (64-69)	80 (77-83)	89 (86-91)	95 (90-98)
	≥65	4,573	68 (66-70)	85 (82-87)	92 (88-94)	95 (88-101)

Supplementary table 3. Three year conditional relative survival for esophageal cancer stratified histology (adenocarcinoma and squamous cell carcinoma).

Histology	Number of patients	Years already survived, % (95% CI)			
		1	3	5	10
Adenocarcinoma	12,160	62 (61-63)	79 (77-80)	88 (87-90)	94 (91-97)
Squamous cell carcinoma	4,031	61 (59-63)	78 (75-80)	83 (79-86)	87 (79-94)



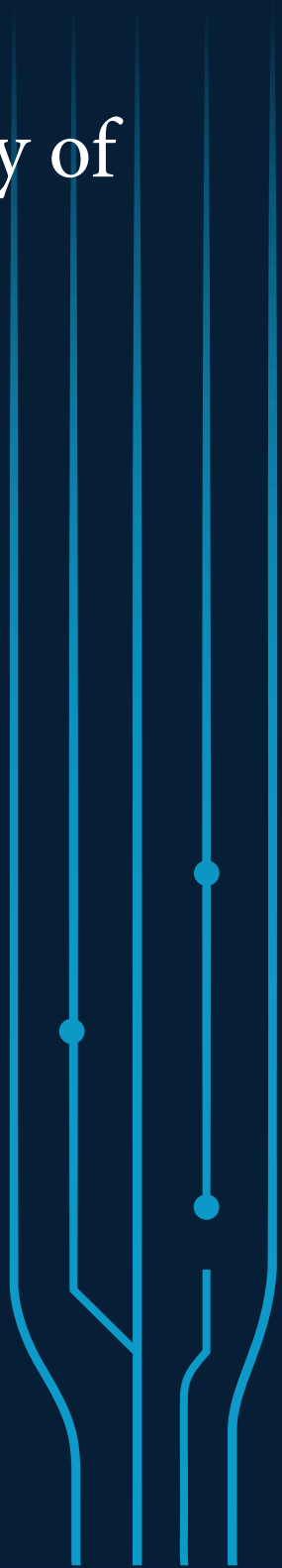
Supplementary figure 1. Flowchart of patient selection.



Supplementary figure 2. Conditional 3 year relative survival with 95% confidence interval for every year survived for patients with esophageal (A) and gastric cancer (B) stratified by period 2006-2012 and 2013-2020.

Part II

Health-Related Quality of Life



Chapter 4

Assessing real-world representativeness of prospective registry cohorts in oncology: insights from patients with esophagogastric cancer

Steven C. Kuijper, Joost Besseling, Thomas Klausch, Marije Slingerland, Charlène J. van der Zijden, Ewout A. Kouwenhoven, Laurens V. Beerepoot, Nadia Haj Mohammad, Bastiaan R. Klarenbeek, Rob H.A.Verhoeven,, Hanneke W.M. van Laarhoven

Based on

Kuijper SC, Besseling J, Klausch T, et al. Assessing real-world representativeness of prospective registry cohorts in oncology: insights from patients with esophagogastric cancer. *J Clin Epidemiol*. 2023

Abstract

Objective. This study aimed to explore the real-world representativeness of a prospective registry cohort with active accrual in oncology, applying a representativeness metric that is novel to health care.
Methods. We used data from the Prospective Observational Cohort Study of Esophageal-Gastric Cancer Patients (POCOP) registry and from the population-based Netherlands Cancer Registry (NCR). We used Representativeness-indicators (R-indicators) and overall survival to investigate the degree to which the POCOP cohort and clinically relevant subgroups were a representative sample compared to the NCR database. Calibration using inverse propensity score weighting was applied to correct differences between POCOP and NCR.
Results. The R-indicator of the entire POCOP registry was 0.72 95%CI[0.71, 0.73]. Representativeness of palliative patients was higher than that of potentially curable patients (R-indicator 0.88 [0.85, 0.90] and 0.70 [0.68, 0.71], respectively). Stratification to clinically relevant subgroups based on treatment resulted in higher R-indicators of the respective subgroups. Both after stratification and calibration weighting survival estimates in the POCOP registry were more similar to that in the NCR population.
Conclusion. This study demonstrated the assessment of real-world representativeness of patients who participated in a prospective registry cohort and showed that real-world representativeness improved when the variability in treatment was accounted for.

Introduction

In oncological research, patient reported outcome measures (PROMs) are a popular method to obtain quality of life and other self-report data from patients.[1–5] When using PROMs, patients are usually requested to complete (digital) questionnaires and data collection thus relies on the active participation of patients. Consequently, some patients may be more inclined to participate than others which could result in a selection bias.[6] For example, patients with high age and higher levels of comorbidities in self-administered health-related quality of life questionnaires may be less likely to participate.[7] Thus, the question arises to what extent patients who are willing to fill out PROMs questionnaires accurately reflect the real-world oncological patient.

Studies on the on the representativeness of prospective cohort studies that collect PROMs compared to the population are still scarce. Recently, a study was published investigating the representativeness of the Prospective Dutch Colorectal Cancer (PLCRC) cohort with respect to the Dutch population of patients with colorectal cancer.[8] In this study, standardized mean differences (SMD) between the prospective cohort and population were used to identify key differences between cohort and population. Although SMDs can provide valuable insights into differences at the variable-level, it does not provide a single intuitive metric capturing the representativeness of a sample. Moreover, because SMDs are calculated for each variable separately, the relative effects of variables on the sample's representativeness cannot be explored.

The aim of this study was to apply and demonstrate the utility of a metric called Representativeness-indicators (R-indicators) for investigating the real-world representativeness of prospective cohort studies to the field of oncology.[9–12] R-indicators have attractive properties which enable researchers to express a sample's representativeness with respect to the population in a single, intuitive metric and allows for the examination of multiple variables simultaneously which can be used to explore variables' relative effect on the representativeness. To this end, we investigated the real-world representativeness of patients included in the Prospective Observational Cohort Study of Esophageal-Gastric Cancer Patients (POCOP) with respect to the Dutch population of patients with esophagogastric cancer.[13] Additionally, we investigated if potential differences between prospective cohort registries and the population could be corrected to produce externally valid results and conclusions.

Methods

Study population

Data of all patients that were diagnosed between 2016 and 2021 with esophageal or gastric cancer (including gastro-esophageal junction carcinoma) and participated in the POCOP registry were used as the sample data. The methods of the POCOP registry have been reported elsewhere.[14] In short, patients are referred to the investigators of the project by someone from the medical team. POCOP investigators then contact potential participants by phone and send the questionnaire by mail or email. When patients provide written informed consent, they are contacted telephonically, via mail or email, every three months during the first year, twice in the second year and annually after that, to collect PROMs.

The reference population consisted of all patients diagnosed between 2016 and 2021 with primary esophageal or gastric cancer in the Netherlands. The data from these patients were obtained from the Netherlands Cancer Registry (NCR). The NCR is an annually updated nation-wide database containing all patients diagnosed with cancer. Trained data managers routinely extract information of the diagnosis, tumor stage and treatment from patients' electronic medical records and add this to the NCR. Identification is mainly based on notification from the nationwide network and registry of histopathology and cytopathology in the Netherlands (PALGA).[15] Patients in the NCR diagnosed in 2016 were staged using TNM 7th edition, patients diagnosed between 2017-2021 were staged using the 8th edition.[16] Patients with TNM-staging based on the 8th edition were converted to the staging of the 7th edition to ensure uniformity of this variable in the analyses. Detailed clinical disease (c-)stages (e.g. IA or IIIB) were simplified to 0, I, II, III, IV, or X, which was then used as categorical variable in all analyses. Finally, all clinical patient data was gathered from the NCR database. By linking the POCOP registry to the NCR we identified which patients participated in POCOP.

Coding and classifications

Patients were defined as either potentially curable (TNM classification $cT_{x\text{-}4a}$ and cM_0) or palliative (cT_{4B} or cM_1). Primary tumor location was defined as stomach or esophagus. Gastro-esophageal junction and cardia carcinomas was coded as esophageal carcinoma if the patient underwent a esophagectomy and coded as gastric carcinoma if the patients underwent a gastrectomy.

Treatments were defined for both the curative and palliative intent, e.g. chemotherapy (either adjuvant, neoadjuvant or palliative), (definitive) chemoradiotherapy and resection (see Supplementary materials Table 1 for all treatments and frequencies). Definitive chemoradiotherapy and neoadjuvant chemoradiotherapy not followed by resection were distinguished from each other based on the dose of radiation. Patients with a dose of 41.4 Gy or less not followed by resection were assumed to have been treated with neoadjuvant chemoradiotherapy. Patients with a doses higher than 41.4 Gy not followed by resection were assumed to have been treated with definitive chemoradiotherapy.[17]

Real-word representativeness

We explored the real-world representativeness of the POCOP registry using the recently developed representativeness indicators (R-indicators).[9–12] This method was developed by the Dutch national bureau of statistics, Statistics Netherlands, and has attracted interest in survey studies.[18,19] The R-indicator quantifies sample representativeness between 0 and 1, with 0 indicating that the sample is not representative at all and 1 that the sample is completely representative for the reference population. The calculation of the R-indicator is based on the variation (i.e. standard deviation) of the propensity that patients participate in the prospective registry cohort, conditional on a set of covariates.

In this analysis using the POCOP registry, response propensities were estimated using a multivariable logistic regression model with all available patient and tumor characteristics as independent variables (treatment, sex, WHO performance status, stage, morphology, age, number of comorbidities, and primary tumor location). In addition to the R-indicator, the so-called partial R-indicators were calculated and reflect which variable in the model contributed the most to the lack of representativeness. The R-indicators and partial R-indicators of the entire POCOP sample with 95% confidence intervals (CI) and the R-indicators for all clinically relevant groups were calculated using the R-indicators code for R.[20] By stratifying the analyses of R-indicators to potentially curable and palliative patients, and treatment groups, we were able to observe if the real-world representativeness was consistent across strata.

Adjusting for differences

In order to correct potential selection bias in the prospective registry cohort and subgroups thereof, we used a calibration weighting (hereafter referred to as calibration) technique based on the Inverse Propensity Weight (IPW). [21] By calibrating the prospective registry cohort to the target population, a pseudo-population is created from the prospective registry cohort with which should more accurately reflect the population data. In this analysis, we calculated the IPW of being included in the POCOP registry with the same multivariable logistic model to estimate the response propensities.

To investigate the degree to which calibration was successful in creating a pseudo population which was better reflective of the NCR population data, we performed a survival analysis in which we constructed Kaplan-Meier (KM) curves of patients in the NCR, POCOP and calibrated POCOP registry. This was performed for the POCOP registry as a whole, and for the subgroups based on treatment intent. Bias was defined as the deviation between the KM-curves of POCOP versus NCR data and calibration POCOP versus NCR data, which was inspected visually. Additionally, median survival and 5-year overall survival was calculated for all analyzed groups. All missing data on variables used in this study were imputed using the random forest imputation implementation of the *missForest* package for R, and the accompanying out-of-bag normalized root mean squared error (NRMSE) was reported to the imputation error.[22] Values very close to zero indicate low imputation error. All analyses were conducted in R version 4.1.0 and R studio version 4.0.3.

Results

Patient population

In total, 2,702 patients were available from POCOP and 16,856 from the NCR (which included all POCOP patients) (Table 1); 65% of patients were treated with curative intent, while 35% of patients were treated with palliative intent (Table 2). The covariate balance after calibrating the POCOP database to the NCR can be found in Table 1. The out-of-bag NRMSE of the imputation was $9.2*10^{-10}$.

Real-world representativeness

Using the R-indicators, we observed that the R-indicator of the total, non-stratified POCOP registry was 0.72 95%CI[0.71,0.73]. Stratified to treatment intent, the R-indicator was 0.88 [0.86, 0.90] for patients treated with palliative intent and 0.70 [0.68, 0.71] for patients treated with curative intent.

Among patients with esophageal cancer treated with curative intent representativeness of the largest group, neoadjuvant chemoradiotherapy followed by a resection, was 0.88 [0.85, 0.92]. For patients who were treated with neoadjuvant chemoradiotherapy not followed by resection representativeness was 0.80 [0.75, 0.86]. In smaller groups, representativeness was 0.90 [0.85, 0.94] for definitive chemoradiation; 1.00 [0.89, 1.00] for neoadjuvant chemotherapy followed by resection (with or without adjuvant chemotherapy); 0.98 [0.98, 1.00] for other treatments; and 0.88 [0.85, 0.92] for endoscopic resection (Figure 1). For patients with gastric cancer, representativeness of patients treated with neoadjuvant chemotherapy followed by resection (with or without adjuvant chemotherapy) was 0.84 [0.79, 0.89], and 0.95 [0.89, 1.00] for patients that underwent a resection only (Figure 1).

Comparable representativeness estimates were found among patients with esophageal cancer treated with palliative intent. In the largest group, patients treated with chemotherapy or target therapy, representativeness was 0.91 [0.88, 0.94]. Patients treated with chemoradiotherapy had a representativeness of 0.84 [0.76, 0.93]; 1.00 [0.97, 1.00] for radiotherapy; and 1.00 [0.78, 1.00] for palliative resection + chemo(radio)therapy or radiotherapy. Among patient with gastric cancer, only patients treated with chemotherapy or targeted therapy was sufficiently large to compute representativeness which was 0.93 [0.88, 0.97].

Across the entire POCOP cohort and the curative and palliative subgroups, the partial R-indicators showed that treatment contributed most to the degree of non-representativeness (Figure 2).

Table 1. Descriptive statistics of the NCR population database, the POCOP database and calibrated POCOP database. Calibrated POCOP refers to the POCOP database after Inverse Propensity Weighting calibration to the NCR population distributions.

	NCR (N=16,856)	POCOP (N=2,706)	Calibrated POCOP
Treatment intent			
Potentially curable	10,966 (65.1%)	2,197 (81.2%)	10,907.4 (65.0%)
Palliative	5,890 (34.9%)	509 (18.8%)	5,868.5 (35.0%)
Sex			
Male	11,883 (70.5%)	2,064 (76.3%)	11,626.0 (69.3%)
Female	4,973 (29.5%)	642 (23.7%)	5,149.9 (30.7%)
WHO performance status			
0	5,228 (31.0%)	1,330 (49.2%)	5,289.6 (31.5%)
1	5,700 (33.8%)	989 (36.5%)	5,507.7 (32.8%)
2	1,707 (10.1%)	126 (4.7%)	2,271.2 (13.5%)
>2	656 (3.9%)	21 (0.8%)	504.2 (3.0%)
Unknown	3,565 (21.1%)	240 (8.9%)	3,203.2 (19.1%)
Stage			
1	1,429 (8.5%)	139 (5.1%)	1,276.4 (7.6%)
2	2,771 (16.4%)	570 (21.1%)	2,771.9 (16.5%)
3	4,432 (26.3%)	1,145 (42.3%)	4,454.2 (26.6%)
4	6,742 (40.0%)	768 (28.4%)	6,905.7 (41.2%)
M/X	1,482 (8.8%)	84 (3.1%)	1,367.8 (8.2%)
Histology			
Adenocarcinoma	13,342 (79.2%)	2,217 (81.9%)	13,231.5 (78.9%)
Squamous cell carcinoma	2,919 (17.3%)	459 (17.0%)	2,884.0 (17.2%)
Other/not microscopically verified	595 (3.5%)	30 (1.1%)	660.4 (3.9%)
Age			
< 49	775 (4.6%)	119 (4.4%)	700.2 (4.2%)
50-59	2,199 (13.0%)	449 (16.6%)	2,245.3 (13.4%)
60-69	5,046 (29.9%)	1,069 (39.5%)	4,847.9 (28.%9)
70-79	5,880 (34.9%)	916 (33.9%)	6,129.7 (36.5%)
> 80	2,956 (17.5%)	153 (5.7%)	2,852.8 (17.0%)
Number of comorbidities			
0	8,085 (48.0%)	1,557 (57.5%)	8,181.4 (48.8)
≥1	8,771 (52.0%)	1,149 (42.5%)	8,594.5 (51.2)

Table 2. Observed frequencies of patients' treatments in the NCR and POCOP. CRT=neoadjuvant chemoradiotherapy, nCT =neoadjuvant chemotherapy, CT=chemotherapy, RT=radiotherapy.

Treatment intent	Treatment	Tumor location	NCR frequency	POCOP frequency
Potentially curable	Definitive CRT	Oesophagus	1276 (11.56%)	171 (6.7%)
	Endoscopic resection	Oesophagus	664 (6.02%)	26 (1.02%)
	nCRT + resection	Oesophagus	3018 (27.34%)	1145 (44.85%)
	nCRT no resection	Oesophagus	892 (8.08%)	313 (12.26%)
	nCT + resection (+- adjuvant CT)	Oesophagus	314 (2.84%)	78 (3.06%)
	nCT + resection (+- adjuvant CT)	Stomach	1250 (11.32%)	288 (11.28%)
	Other treatment	Oesophagus	1201 (10.88%)	35 (1.37%)
	Resection only	Stomach	687 (6.22%)	52 (2.04%)
Palliative	Chemoradiotherapy	Oesophagus	318 (2.88%)	56 (2.19%)
	Chemotherapy or targeted therapy	Oesophagus	2037 (18.45%)	249 (9.75%)
	Chemotherapy or targeted therapy	Stomach	876 (7.94%)	70 (2.74%)
	Palliative resection + C(R)T or RT	Oesophagus	97 (0.88%)	28 (1.1%)
	Radiotherapy only	Oesophagus	962 (8.71%)	42 (1.65%)

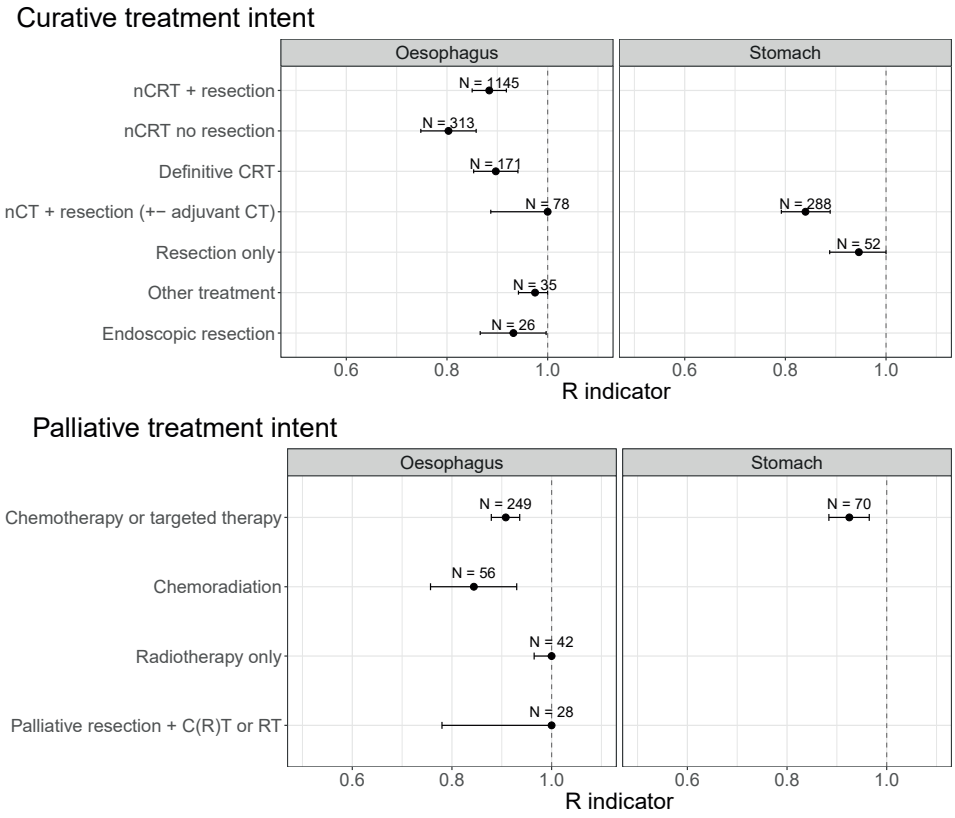


Figure 1. R-indicators of subgroups with curative and palliative treatment intent. Sample sizes of the patients in POCOP in specific subgroups are reported. The dashed line indicates perfect representativeness. Horizontal bars represent the 95% confidence intervals of the R-indicators. nCRT=neoadjuvant chemoradiotherapy, nCT=neoadjuvant chemotherapy, CT=chemotherapy, RT=radiotherapy.

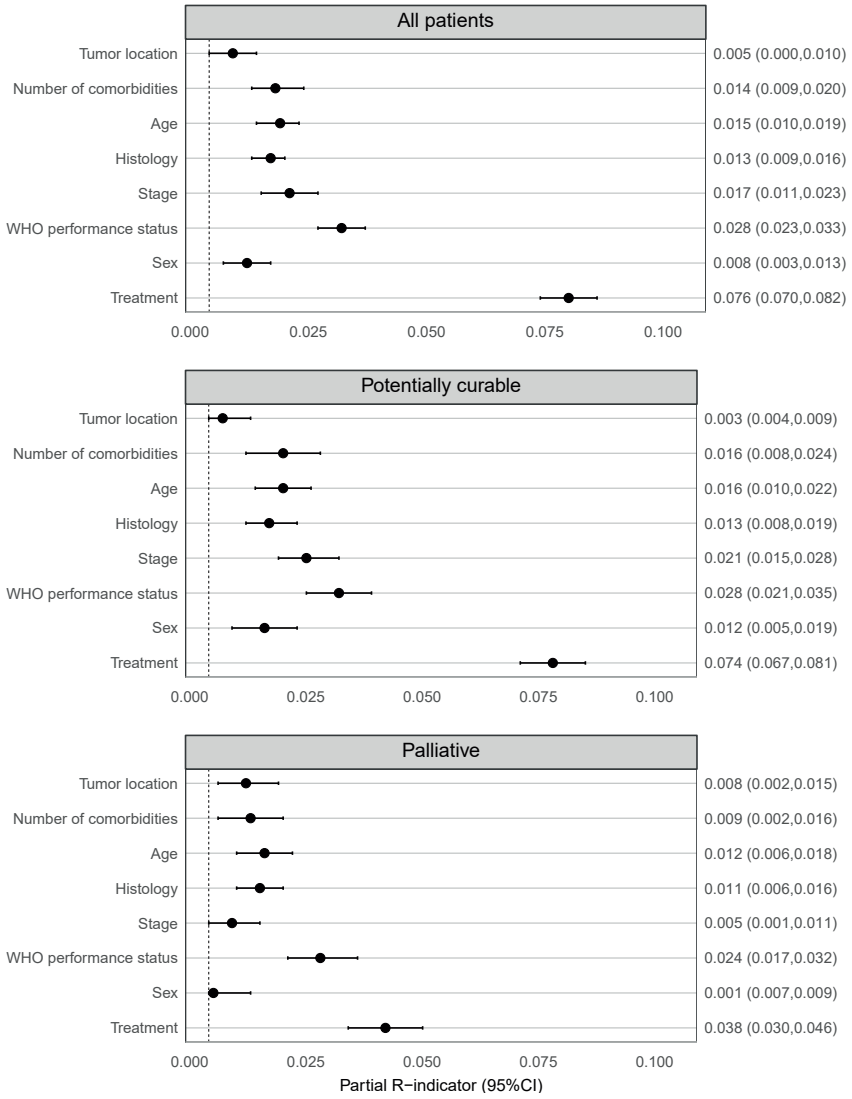


Figure 2. Partial R-indicators for all patients and stratified to treatment intent. Values should be interpreted relative to each other. Higher values of the partial R-indicator correspond to a larger contribution to non-representativeness due to that variable relative to the other variables.

Calibration and survival

The Kaplan-Meier curves of all patients and conditioned on treatment intent are displayed in Figure 3. Overall survival was higher for patients in POCOP compared to the NCR population data. After calibration, upon visual inspection survival of patients in POCOP as a whole was more alike the NCR population data. Survival curves of patients in the calibrated POCOP and NCR conditioned on type of treatment are shown in Figure 4-5.

Median survival of the NCR, POCOP and calibrated POCOP was 19 [18, 20], 32 [31, 36], and 23 [20, 25] months, respectively. For potentially curable patients, median survival of the NCR, POCOP and calibrated POCOP was 32 [30, 33], 43 [40, 47], and 36 [32, 42] months, respectively. For palliative patients, median survival of the NCR, POCOP and calibrated POCOP was 9 [9, 9], 13 [11, 14], and 11 [10, 12] months. Median survival of clinically relevant subgroups stratified to treatment can be found in Table 3.

The 5-year overall survival rates of patients in the NCR, POCOP and calibrated POCOP were 26%, 36%, and 27%, respectively. For potentially curable patients in the NCR, POCOP, and calibrated POCOP, the 5-year overall survival rates were 36%, 42% and 37%, respectively. For palliative treated patients in the NCR, POCOP, and calibrated POCOP, the 5-year overall survival rates were 4%, 8% and 6%, respectively. The 5-year overall survival of clinically relevant subgroups stratified to treatment can be found in Table 3.

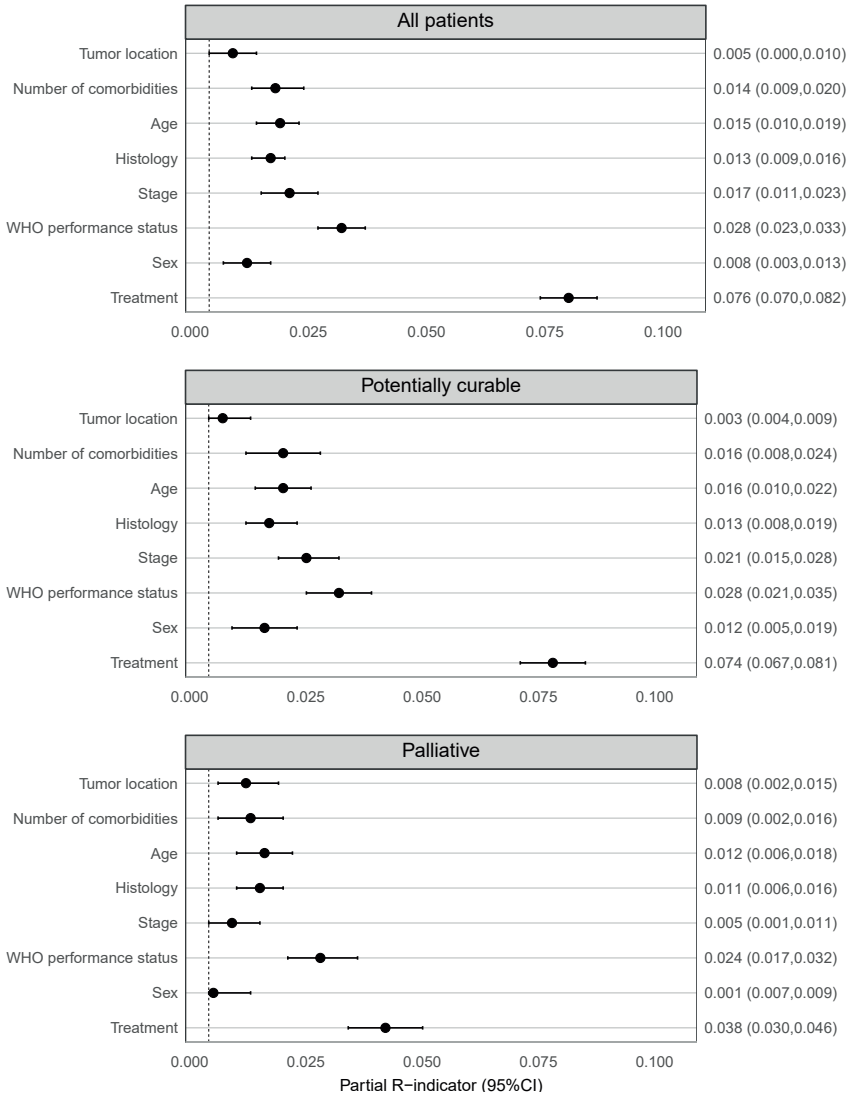


Figure 3. Kaplan-Meier curves of the NCR (population), POCOP and calibrated POCOP for all patients.

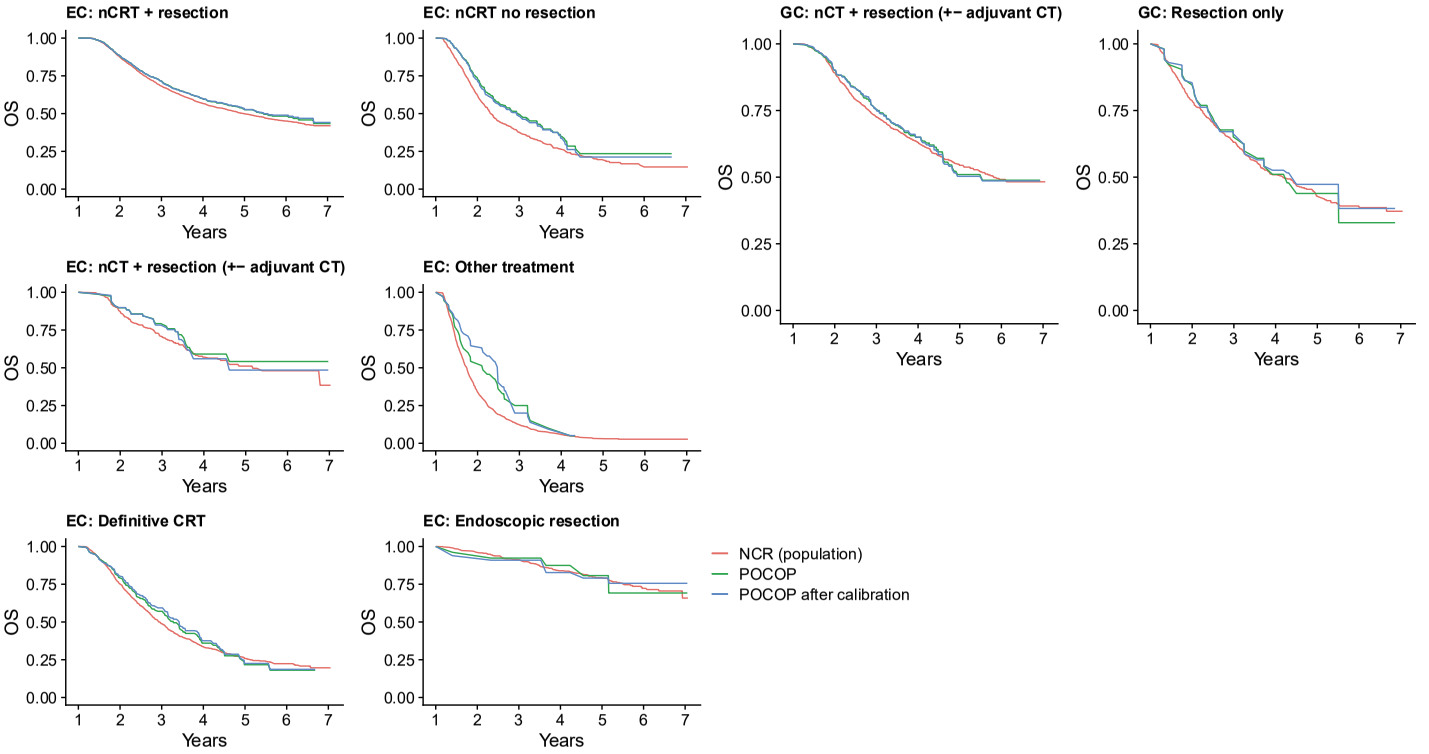


Figure 4. Kaplan-Meier curves of the NCR (population), POCOP and calibrated POCOP, of potentially curable patients conditioned on treatment.
EC = Esophageal cancer, GC: Gastric cancer, nCRT = Neoadjuvant chemoradiotherapy, CT = Chemotherapy, CRT = Chemoradiotherapy.

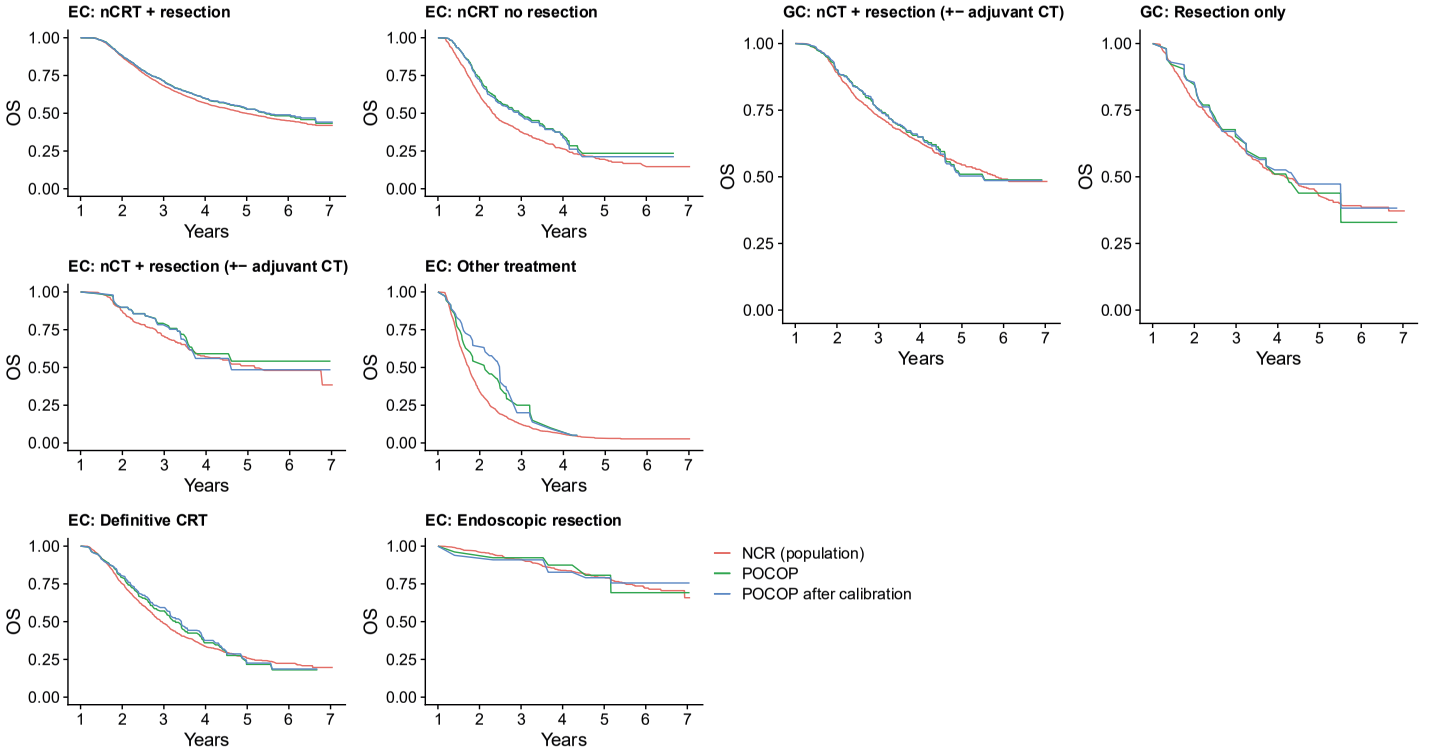


Figure 5. Kaplan-Meier curves of the NCR (population), POCOP and calibrated POCOP, of palliative patients conditioned on treatment. EC = Esophageal cancer, GC = Gastric cancer. CRT = Chemoradiotherapy, RT = Radiotherapy.

Table 3. Median overall survival (months) of patients in the NCR, POCOP and calibrated POCOP. OS = Overall survival. nC(R)T=neoadjuvant chemo(radio)therapy, nCT =neoadjuvant chemotherapy, CT=chemotherapy, RT=radiotherapy. - Not observed

	NCR		POCOP		POCOP calibrated	
	Median OS (months, 95%CI)	5-year OS (%)	Median OS (months, 95%CI)	5-year OS (%)	Median OS (months, 95%CI)	5-year OS (%)
Potentially curable						
Esophagus						
nCRT + resection	48 [44, 53]	45%	53 [48, NA]	48%	54 [48, NA]	49%
nCRT no resection	16 [15, 17]	15%	24 [20, 29]	23%	23 [18, 29]	22%
Other treatment	9 [8, 9]	3%	14 [8, 23]	-	18 [10, 23]	-
Definitive CRT	23 [22, 25]	22%	27 [23, 34]	18%	29 [25, 36]	19%
Endoscopic resection	-	72%	-	69%	-	77%
nCT + resection (+- adjuvant CT]	50 [40, NA]	48%	-	54%	-	51%
Stomach						
nCT + resection (+- adjuvant CT]	58 [51, NA]	49%	55 [43, NA]	49%	55 [43, NA]	49%
Resection only	38 [32, 45]	39%	39 [27, NA]	33%	42 [27, NA]	38%
Palliative						
Esophagus						
Radiotherapy only	5 [5, 6]	-	7 [5, 10]	-	6 [5, 10]	-
Palliative resection + C(R]T or RT	28 [23, 42]	30%	41 [24, NA]	38%	41 [26, NA]	39%
Chemoradiation	15 [12, 18]	12%	17 [13, 25]	10%	17 [14, 28]	12%
Chemotherapy or targeted therapy	10 [9, 10]	3%	12 [11, 14]	5%	12 [11, 14]	5%
Stomach						
Chemotherapy or targeted therapy	10 [9, 10]	3%	11 [9, 14]	-	11 [9, 14]	-

Discussion

In this study we evaluated the representativeness of a prospective registry cohort for real-world data in oncology, by applying a metric to quantify representativeness called R-indicators for the first time in medical oncology, and the extent to which differences between the patients included in the PROMs and population could be corrected to produce generalizable estimates.

We found that subgroups stratified to treatment generally had a higher real-world representativeness with respect to their respective target populations than the complete unstratified prospective registry cohort (POCOP). This implies that accounting for the variability of included treatments in the PROMs improved real-world representativeness of the prospective registry cohort. This pattern was also observed in survival analyses in which survival estimates from patients included in the prospective registry cohort were more similar to the target population in treatment-stratified samples. Therefore, although the variability of treatment may potentially introduce selection bias, the effects of selection bias can be mitigated by accounting for treatment using stratification. This is an important finding for other cancer types where treatment may also be a contributing factor in the selection mechanism that determines willingness to participate in PROMs studies.

Another important finding from stratification of the prospective registry cohort to clinically relevant groups was that lower inclusion rates of patients did not correspond with lower real-world representativeness. For example, the real-world representativeness of palliative patients was higher compared to potentially curable patients despite the inclusion rates being lower. This implied that, insofar included patients on average resemble patients from the real world, the absolute number of included patients does not matter. This confirms previous findings that lower inclusion rates do not necessarily cause biased samples and is is contradictory to what is speculated in published studies using similar PROMs data.[1,2,23,24]

In addition to the finding that stratification of the prospective registry cohort to clinically relevant subgroups improved representativeness, our study also showed that calibration weighting can be used as an alternative to stratification to obtain generalizable estimates from the PROMs with respect to the population to a large extend. Stratification in combination with calibration weighting only marginally improved the estimates. The advantage of using calibration weighting techniques is that it allows to control for multiple variables without the need to creating subgroups, in contrast to stratification where the number of strata increase exponentially as the number of variables to control for increase.[25] However, a disadvantage of weighting calibration is that is requires the population data in addition to the prospective cohort data to be able to perform the calibration of the PROMs data to the population, whereas stratification does not need these population data. Moreover, in some instances the calibrated survival curve did not perfectly overlap with the population curve, indicating that there still was some unobserved confounding for which we did not account. Known factors that may induce non-representativeness in health related quality of life studies are physical condition and comorbidities; patients with better physical condition and fewer comorbidities are more likely to participate.[26,27] However, the WHO performance status and the number of comorbidities were included in the propensity model and are therefore unlikely to explain the remaining non-correctable bias.

The R-indicator has shown potential to be able to estimate real-world representativeness of a prospective registry cohort. R-indicators can be intuitively interpreted as

it expresses representativeness on the same scale (0-1) regardless of the type of variables and the set of data that is used. It provides a good alternative to more classic methods such as statistically testing observed frequencies of sample and population characteristics for significance, which is influenced by samples sizes and does not provide an overall summary statistic for the samples' representativeness.[28]

Considerations

This study has a number of limitations. First, is the relatively small sample size of some of the clinically relevant subgroups that leads to large confidence intervals of the R-indicator. Larger samples of treatment groups are needed to be able to prove a more precise estimate of the representativeness. A second limitation was that we could only estimate bias and corrected bias of survival analyses rather than health related quality of life or other patient reported outcomes. Health related quality of life measures were by definition only available for patients actively participating in the POCOP registry and not for patients in the NCR. However, built into the weighting calibration is that it increases covariate balance between the NCR and POCOP patients by creating a pseudo population from the original POCOP with properties that resemble the population from which it borrowed information. By analyzing this pseudo population (or calibrated sample) generalizability to the total population is increased.[29,30] Therefore, an outcome such as survival can still provide information on whether calibration can correct existing bias.

What is more, in the computation of the R-indicators there is an additional bias correction because the R-indicator is inherently biased to be smaller than one since the variance of the propensity scores is rarely zero. This bias is corrected through a built-in bias correction of the software and explains why some R-indicators are exactly one.

Major strength of this study was that used data from a large prospective cohort study and data from the reliable nationwide Netherlands Cancer Registry which includes all diagnosed malignancies and is thus a very comprehensive population database, which enabled us to make comparisons between the prospective registry cohort and the Dutch population of patients with esophageal or gastric cancer.

Future perspectives

Given the advantages of R-indicators to express representativeness, future observational and clinical studies could be evaluated and managed more structurally and more uniformly. Additionally, real-world representativeness of PROMs registries could be improved by monitoring characteristics of included patients and adjusting inclusion strategies to reflect the total population. Adjusting inclusion strategies to target such specific patient groups could improve the real-world representativeness of PROMs registries. Finally, more research is needed in other healthcare research settings to further investigate the suitability of prospective registry cohorts as real-world data.

Conclusion

This study demonstrated the assessment of real-world representativeness of patients who participated in a prospective registry cohort and showed that real-world representativeness improved when the variability in treatment was accounted for. Moreover, this study demonstrated the utility of representativeness indicators to explore real-world representativeness of prospective registry cohort as well as calibration techniques and stratification to correct for differences between the prospective registry cohort and the population.

References

1. Dijksterhuis WPM, Latenstein AEJ, Van Kleef JJ, et al. Cachexia and dietetic interventions in patients with esophagogastric cancer: A multicenter cohort study. *JNCCN Journal of the National Comprehensive Cancer Network*. 2021;19(2):144-152. doi:10.6004/jnccn.2020.7615
2. van Kleef JJ, Dijksterhuis WPM, van den Boorn HG, et al. Prognostic value of patient-reported quality of life for survival in oesophagogastric cancer: analysis from the population-based POCOP study. *Gastric Cancer*. 2021;(0123456789). doi:10.1007/s10120-021-01209-1
3. Davie A, Carter GC, Rider A, et al. Real-world patient-reported outcomes of women receiving initial endocrine-based therapy for HR+/HER2- advanced breast cancer in five European countries. *BMC Cancer*. 2020;20(1):1-15. doi:10.1186/s12885-020-07294-2
4. Jakob A, Zahn MO, Nusch A, et al. Real-world patient-reported outcomes of breast cancer or prostate cancer patients receiving antiresorptive therapy for bone metastases: Final results of the PROBone registry study. *Journal of Bone Oncology*. 2022;33:100420. doi:10.1016/j.jbo.2022.100420
5. Gotto G, Drachenberg DE, Chin J, et al. Real-world evidence in patient-reported outcomes (PROs) of metastatic castrate-resistant prostate cancer (mCRPC) patients treated with abiraterone acetate + prednisone (AA+P) across Canada: Final results of COSMiC. *Canadian Urological Association Journal*. 2020;14(12):10-14. doi:10.5489/CUAJ.6388
6. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15(5):615-625. doi:10.1097/01.ede.0000135174.63482.43
7. Unruh M, Yan G, Radeva M, et al. Bias in assessment of health-related quality of life in a hemodialysis population: A comparison of self-administered and interviewer-administered surveys in the HEMO Study. *Journal of the American Society of Nephrology*. 2003;14(8):2132-2141. doi:10.1097/01.ASN.0000076076.88336.B1
8. Derksen JWG, Vink GR, Elferink MAG, et al. The Prospective Dutch Colorectal Cancer (PLCRC) cohort: real-world data facilitating research and clinical care. *Scientific Reports*. 2021;11(1):1-12. doi:10.1038/s41598-020-79890-y
9. Schouten B, Bethlehem J, Beullens K, et al. Evaluating, Comparing, Monitoring, and Improving Representativeness of Survey Response Through R-Indicators and Partial R-Indicators. *International Statistical Review*. 2012;80(3):382-399. doi:10.1111/j.1751-5823.2012.00189.x
10. Schouten B, Cobben F, Bethlehem J. Indicators for the representativeness of survey response. *Survey Methodology*. 2009;35(1):101-113.
11. Shlomo N, Skinner C, Schouten B, Bethlehem J, Zhang L chun. RISQ Statistical Properties of R-indicators. Published online 2009:1-52.
12. Bethlehem J, Cobben F, Schouten B. Handbook of Nonresponse in Household Surveys. *Handbook of Nonresponse in Household Surveys*. Published online 2011. doi:10.1002/9780470891056
13. Coebergh van den Braak RRJ, van Rijssen LB, van Kleef JJ, et al. Nationwide comprehensive gastro-intestinal cancer cohorts: the 3P initiative. *Acta Oncologica*. 2018;57(2):195-202. doi:10.1080/0284186X.2017.1346381
14. Coebergh van den Braak RRJ, van Rijssen LB, van Kleef JJ, et al. Nationwide comprehensive gastro-intestinal cancer cohorts: the 3P initiative. *Acta Oncologica*. 2018;57(2):195-202. doi:10.1080/0284186X.2017.1346381
15. Casparie M, Tiebosch ATMG, Burger G, et al. Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. *Cellular Oncology*. 2007;29(1):19-24. doi:10.1155/2007/971816
16. TNM Classification of Malignant Tumours, 8th Edition. *International Union Against Cancer (UICC)*. Published online 2014.
17. Hulshof MCCM, van Laarhoven HWM. Chemoradiotherapy in tumours of the oesophagus and gastro-oesophageal junction. *Best Practice and Research: Clinical Gastroenterology*. 2016;30(4):551-563. doi:10.1016/j.bpg.2016.06.002
18. Luiten A, Schouten B. Tailored fieldwork design to increase representative household survey

response: An experiment in the Survey of Consumer Satisfaction. *Journal of the Royal Statistical Society Series A: Statistics in Society*. 2013;176(1):169-189. doi:10.1111/j.1467-985X.2012.01080.x
19. Moore JC, Durrant GB, Smith PWF. Data set representativeness during data collection in three UK social surveys: generalizability and the effects of auxiliary covariate choice. *Journal of the Royal Statistical Society Series A: Statistics in Society*. 2018;181(1):229-248. doi:10.1111/rssa.12256
20. De Heij, V., Schouten, B., Sholomo N. RISQ manual 2.1 Tools in SAS and R for the computation of R-indicators, partial R-indicators and partial coefficients of variation. Published online 2015.
21. Cole SR, Hernán MA. Adjusted survival curves with inverse probability weights. *Computer Methods and Programs in Biomedicine*. 2004;75(1):45-49. doi:10.1016/j.cmpb.2003.10.004
22. Stekhoven DJ, Bühlmann P. Missforest-Non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28(1):112-118. doi:10.1093/bioinformatics/btr597
23. Burbach JPM, Kurk SA, Coebergh van den Braak RRJ, et al. Prospective Dutch colorectal cancer cohort: an infrastructure for long-term observational, prognostic, predictive and (randomized) intervention research. *Acta Oncologica*. 2016;55(11):1273-1280. doi:10.1080/0284186X.2016.1189094
24. Groves RM, Peytcheva E. The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*. 2008;72(2):167-189. doi:10.1093/poq/nfn011
25. Deb S, Austin PC, Tu J V., et al. A Review of Propensity-Score Methods and Their Use in Cardiovascular Research. *Canadian Journal of Cardiology*. 2016;32(2):259-265. doi:10.1016/j.cjca.2015.05.015
26. Addington-Hall J, Kalra L. Measuring quality of life: Who should measure quality of life? *Br Med J*. 2001;322(7299):1417-1420. doi:10.1136/bmj.322.7299.1417
27. Unruh M, Yan G, Radeva M, et al. Bias in assessment of health-related quality of life in a hemodialysis population: A comparison of self-administered and interviewer-administered surveys in the HEMO Study. *Journal of the American Society of Nephrology*. 2003;14(8):2132-2141. doi:10.1097/01.ASN.0000076076.88336.B1
28. Westgeest HM, Uyl-de Groot CA, van Moorselaar RJA, et al. Differences in Trial and Real-world Populations in the Dutch Castration-resistant Prostate Cancer Registry. *European Urology Focus*. 2018;4(5):694-701. doi:10.1016/j.euf.2016.09.008
29. Stuart EA, Cole SR, Bradshaw CP, Leaf PJ. The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2011;174(2):369-386. doi:10.1111/j.1467-985X.2010.00673.x
30. Nohr EA, Liew Z. How to investigate and adjust for selection bias in cohort studies. *Acta Obstetricia et Gynecologica Scandinavica*. 2018;97(4):407-416. doi:10.1111/aogs.13319

Chapter 5

Predicting health-related quality of life for patients with gastroesophageal cancer

Steven C. Kuijper, Irene Cara, Gijs Geleijnse, Marije Slingerland, Grard A.P. Nieuwenhuijzen, Sjoerd Lagarde, Bastiaan R. Klarenbeek, Ewout Kouwenhoven, Richard van Hillegersberg, Rob H.A. Verhoeven, Hanneke W.M. van Laarhoven

Manuscript under review at Quality of Life Research

Abstract

Background: Gastroesophageal cancer has a poor prognosis, and treatment significantly impacts health-related quality of life (HRQoL). Accurate prediction of HRQoL changes after treatment can support shared decision-making. This study aimed to develop and validate HRQoL prediction models for patients with gastroesophageal cancer using established risk-prediction models and a newly proposed sequential score model.
Methods: HRQoL data came from the Prospective Observational Cohort Study of Esophageal-Gastric Cancer Patients registry, linked to the Netherlands Cancer Registry. The EORTC-C30 functioning scales were used as outcomes. Risk-prediction models, based on logistic elastic-net regression, estimated the probability of meaningful HRQoL deterioration at 3, 6, and 12 months post-treatment. The sequential score model, using XGBoost regression, predicted the next HRQoL score at any time. Calibration curves and integrated calibration index (ICI) assessed predictive performance, with Brier scores for risk-prediction models and root mean squared error plus Out-of-Sample r^2 for sequential models.
Results: Risk-prediction models showed strong performance (ICI: 0.03-0.08; Brier score: 0.09-0.17) for predicting significant deterioration in Summary Score, Physical Functioning, and Fatigue, with good calibration. Sequential score models explained up to 40% of the variance in HRQoL scores.
Conclusion: Both models effectively predicted HRQoL in gastroesophageal cancer patients, demonstrating potential to enhance patient care and information sharing through accurate prediction of HRQoL outcomes.

Introduction

It is well established that treatment of gastroesophageal cancer can affect health-related quality of life (HRQoL) both positively as well as negatively.[1–3] HRQoL is therefore an important factor to consider when patients and caregivers make decisions about treatment strategies. Clinical prediction models have the potential to support caregivers in providing personalized information regarding treatment outcomes, which could help in the process of shared decision making.[4,5]

Successful efforts to predict HRQoL have been made in colorectal cancer, cervical cancer and breast cancer.[6–9] These models are based on classical statistical models as well as machine learning models to predict HRQoL at a fixed time horizon. However, no such models are currently available for HRQoL prediction for patients with gastroesophageal cancer.[10,11] Given the high variability of treatment outcomes in terms of survival across patients with this cancer type, a HRQoL prediction model with good predictive performance could be valuable tool in proving treatment-related information.[12,13]

The primary aim of this study was to develop models to predict post-treatment HRQoL at the time of primary diagnosis for patients with gastroesophageal cancer. To achieve this, we employed a well-established modeling methodology that has been successfully used to assess the risk of significant HRQoL deterioration in colorectal cancer patients. [7,8] Given its strong predictive performance in the colorectal cancer context, this method was deemed an appropriate choice for gastroesophageal cancer data as well. However, while this approach has demonstrated success, it is not without limitations. A key limitation of this approach is that it simplifies the prediction process by focusing exclusively on risk estimation of HRQoL deterioration, potentially overlooking important nuances in how HRQoL evolves over time.

To address this, we introduce a novel modeling methodology in this study—the sequential score model—which aims to predict HRQoL on a continuous time scale. Unlike existing models, this approach offers the potential for more granular and dynamic predictions of HRQoL progression over time. To our knowledge, this method has not yet been applied in this field, but it holds promise as a more flexible and comprehensive tool for estimating HRQoL after treatment onset.

Since HRQoL prediction in gastroesophageal cancer remains a relatively new field, we conducted a comprehensive study to apply and evaluate two different modeling approaches. Our aim is to explore the potential of these models to provide more accurate HRQoL predictions for patients. By offering a quantitative assessment, these models can help patients better understand how their quality of life may be impacted after treatment. This study follows the TRIPOD guidelines, ensuring transparent and rigorous reporting of multivariable prediction models for individual prognosis or diagnosis.[14]

Methods

Patient population

We used data from 3,305 patients with malignant gastroesophageal cancer who were diagnosed between 2015-2021 in the Netherlands and participated in the Prospective Observational Cohort Study of Esophageal-Gastric Cancer Patients (POCOP). The POCOP study was designed to register patients' self-reported HRQoL before, during and after treatment and was used for this study to obtain data on HRQoL. The complete methodology of POCOP is described elsewhere,[15] but in short, individuals involved in the project are directed to project investigators by a member of the medical team. Subsequently, POCOP investigators reach out to potential participants through phone calls and dispatch the questionnaire via mail or email. Once patients provided written informed consent, they are contacted every three months in the initial year, twice in the second year, and annually thereafter, through phone, mail, or email, for the purpose of gathering patient-reported outcome measures.

The POCOP database is routinely linked to the population-based Netherlands Cancer Registry (NCR) in order to obtain patients, tumor and treatment related information. The NCR is a nationwide database that contains records of all individuals diagnosed with a malignant form of cancer. Trained data managers routinely extract details such as diagnosis, tumor stage, and treatment from the electronic medical records of patients, and integrate this information into the NCR. The primary method of identification relies largely on notifications received from the nationwide network and registry of histopathology and cytopathology in the Netherlands (PALGA).

Patients with esophageal cancer (ICD-O-3 topography codes: C15.0–C15.9), gastroesophageal junction/cardia (C16.0), and gastric cancer (C16.1–C16.9) were included in this study. Gastroesophageal junction/cardia tumors were grouped together esophageal cancer.

Health-related Quality of Life outcomes

The primary outcome for prediction were outcome scales of the EORTC QLQ-C30 questionnaire: global health status (GHS), physical functioning, role functioning, emotional functioning, social functioning, nausea/vomiting, fatigue, pain, appetite loss, diarrhea, dyspnea, constipation, insomnia and financial difficulties, as well as the summary score of the QLQ-C30 which was calculated as average of the QLQ-C30 outcome scores, excluding global health status and financial difficulties .[16–18] All items in the C30 questionnaire are scored using a Likert scale, and linearly transformed to be scaled between 0-100. To promote readability of the manuscript, we presented the results the GHS and Summary Score in the main text given that these outcome reflect general HRQoL. Additionally, we present the best performing subscales in the main text. Results from all other scales can be found in Supplementary Figures 1-3.

Predictor variables

For the risk prediction model, we used a total of 78 candidate predictors known at baseline. We used 15 clinical variables from the NCR: sex, age, height, weight, cT, cN, cM, performance status, tumor differentiation grade, hemoglobin (HB), albumin, lactic acid dehydrogenase (LDH), creatinine, Lauren classification, tumor morphology and treatment (modeled as a

single categorical covariate). From POCOP, we used 63 variables: the baseline QLQ-C30 scale scores, baseline EORTC QLQ-OG25 symptom scores, smoking cigarette/cigar/pipe, tube feeding, drinking alcohol, education level, married, EuroQoL-5d scales, Worry of Progression Scale (WOPS) scores and Hospital Anxiety and Depression Scale (HADS) scale scores.

In addition to the abovementioned predictors, the responses to previous QLQ-C30 and QLQ-OG25 questionnaires, the time between questionnaires and the time between treatment onset and previous questionnaires were also selected for the sequential score model.

Model pipeline and development

Two types of models were developed to make HRQoL predictions after treatment onset. Although modelling and prediction of HRQoL was fundamentally different in each approach, the general pipeline and validation was similar. (Figure 1) The modelling details of each approach are described below, but the pipeline can be summarized as follows. First, candidate predictors with more than 50% of missing data were removed from the dataset, which resulted in the removal of tube feeding, smoking cigar, smoking pipe, drinking alcohol, OG25 hair loss symptom scale, and the Worry Of Progression Scale. Any further missing data were imputed using a random forest imputation or K-Nearest-Neighbors imputation, except for the categorical variables tumor morphology, differentiation grade and Lauren classification for which a category "unknown" was used as they were not assumed to be missing at random.[19,20] Models were trained on the complete dataset with feature selection and their performance was evaluated on the same data (complete model performance). Using 10-fold internal-external cross-validation, in which all previous steps that were used in model development on the complete were repeated in every fold to prohibit leakage and optimism between train and test set, we tested our methodologies for potential overfitting (internal-external performance). In Finally, both the apparent performance as well as the internal-external performance were evaluated for all models.

Model development: Risk prediction

The primary aim of the risk prediction models was predicting the probability of experiencing a clinically meaningful deterioration in each of the sixteen QLQ-C30 outcomes at six months and one year after treatment onset. To this end, we developed a total of 48 prediction models, one for each of the sixteen outcomes for 3 months, 6 months and 12 months after treatment onset.

First, for each patient, we determined the value of the outcome at three, six and twelve months after start of treatment. It was required for each patient that they minimally filled-in two questionnaires, one before start of treatment (with a margin of up to14 days post-treatment) and one questionnaire around the prediction horizon (with a margin of +/- 30 days before and after the time horizon). At these time horizons, we determined if patients experienced a clinically meaningful deteriorations compared to before start of treatment using established cut-off points for small deteriorations (Supplementary Table 1).[21]

Then, for each outcome and time horizon, we trained an elastic net logistic regression model on the complete dataset, which resulted in a total of 48 models. Elastic net regression is a method that shrinks non-informative model coefficients to zero and performs variable selection, while being able to handle high dimensional and colinear data. Given the

large number of candidate predictors (78) which may likely be highly colinear, elastic net regression was the most fitting solution for model estimation and variable selection. In training the model on all available data, the lambda parameter and mixing parameter required to perform the elastic net regression were optimized using a nested repeated cross-validation (Figure 1). All modeling steps (including handling of missing data) were repeated in the 10-fold cross-validation to obtain the internal-external validation performance.

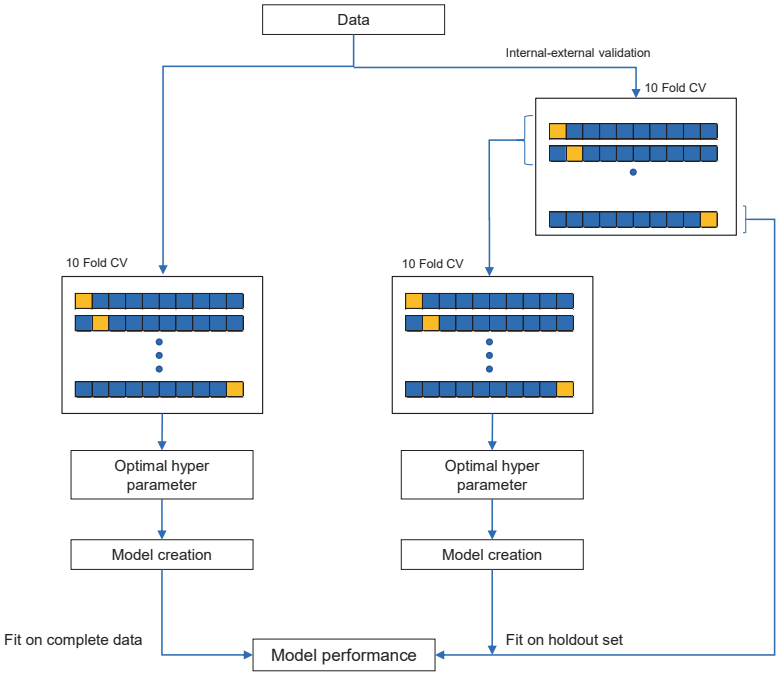


Figure 1. Modelling pipeline for the risk-prediction models and the sequential score prediction models

Model development: Sequential score model

The sequential score model is designed to predict the next QLQ-C30 outcome, at any time after treatment onset. A total of 16 models, one for each QLQ-C30 outcome, were developed using XGBoost, a tree boosting approach broadly known and used in the Data Science community for achieving state-of-the-art performance, thanks to the sequential fitting of trees on residuals.[22] Other key characteristics of XGBoost are the capability of handling high dimensional data with missing input and to capture non-linearities.

For each patient, the values of the outcomes at each questionnaire were determined. The time difference between questionnaires and treatment onset had to be in the range of 30 days to 1000 days. XGBoost regression models were trained, for each outcome, based on all available variables at the time of the previous questionnaire plus the time between questionnaires. Bayesian optimization was deployed to optimize the model parameters (column subsample ratio, training instance subsample ratio, gamma, learning rate, maximum tree depth, minimum child weight, number of estimators, loss function, L1 regularization term and L2 regularization term) in a nested repeated cross-validation.[23] In addition to XGBoost, an Elastic net regression model was trained for comparison. As Elastic net requires data imputation, K-Nearest-Neighbors imputation was used for this purpose, while this step was not required for XGBoost. The modeling strategy (Figure 1) was repeated in a 10-fold-cross-validation, where folds were selected based on patients rather than on questionnaires to avoid data spilling between the training and the validation sets.

Model performance

For both modelling approaches, the apparent predictive performance and the internal-externally validated performance was evaluated using calibration curves and the integrated calibration index (ICE). For the risk prediction models, calibration curves were constructed by directly plotting observed outcomes with predicted probabilities and fitting a smooth line using LOESS regression with 95% confidence interval.[24] For the sequential model, calibration curves were constructed by dividing the predicted score values into 10 quantiles and plotting the mean predicted values versus the mean observed scores. For the risk prediction model the Brier score was calculated and the Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and out of sample r2 was calculated for the sequential model.[25] Across all analyses, an alpha level of 0.05 was maintained for statistical significance. R version 4.3.1 and Python version 3.12.1 were used for all analyses.

Predictor importance

For the risk-prediction model, feature importance was determined by the absolute value of the standardized regression coefficient. Larger values corresponded with larger feature importance. Feature importance was determined for the complete models. For the sequential score model, the relative importance of each predictor can be determined by the XGBoost gain and this was evaluated on the complete model.

Results

Patient characteristics

In total, 3,305 patients were included in the dataset on which the prediction models were developed. Of all patients, 86% were diagnosed with esophageal cancer, of whom 89.2% were diagnosed with non-metastatic disease and 10.8% with metastatic disease (Table 1). For patients with gastric cancer, 85.6% were diagnosed with non-metastatic disease and 14.4% with metastatic disease.

Risk prediction model

In total, 48 prediction models were developed for risk prediction of a significant deterioration at 3, 6 and 12 months after treatment onset (Supplementary Figure 1, 2, 3) for all patients with gastroesophageal cancer. The risk prediction models for the GHS and Summary Score can be found in Figure 2. For the GHS models, the ICI varied between 0.04-0.07 and the Brier score 0.14-0.17. The internal-external cross-validation showed identical estimates compared to the complete model (Figure 2). The GHS baseline feature was the most important feature in predicting the risk of a significant deterioration across all time horizons. (Figure 4) For the 12-month model, the GHS baseline was the only feature that remained in the model after feature selection.

For the models predicting significant deteriorations in the Summary Score across all timepoints, the ICI of the complete model ranged between 0.05-0.08 and the Brier score between 0.14-0.17. Results from internal-external cross validations showed nearly identical estimates as the complete model. The most important features predicting the risk of a significant deterioration of the Summary Score were Social functioning, Constipation, and Appetite loss for the 3, 6 and 12 months models, respectively (Figure 4).

Based on visual inspection of the calibration curves, we identified two outcomes that showed particularly good calibration: physical functioning and fatigue across all time horizons (Figure 3). Results of all other outcomes are shown in Supplementary Figure 1-3. Generally, the other outcomes demonstrated suboptimal calibration and are likely not useable for clinical practice. For Physical functioning the ICI varied between 0.03-0.07 and the Brier score varied between 0.09-0.13. For Fatigue, the ICI varied between 0.03-0.06 and the Brier score ranges from 0.14-0.15 across all time horizons. For both Physical functioning and Fatigue the results from the internal-external cross validation were nearly identical to the complete model. Both for Physical functioning and Fatigue the most important features were the Physical functioning and Fatigue baseline questionnaires (Figure 5).

Table 1. Characteristics of included patients.

		Esophagus	Stomach
		2,799	506
Sex (%)	1	2207 (78.8)	305 (60.3)
	2	592 (21.2)	201 (39.7)
Age (mean (SD))		66.19 (8.76)	66.90 (10.42)
Weight (mean (SD))		81.69 (16.21)	75.68 (15.12)
WHO Performance status (%)	0	1417 (55.9)	261 (60.3)
	1	991 (39.1)	136 (31.4)
	2	105 (4.1)	26 (6.0)
	3	19 (0.7)	10 (2.3)
	4	2 (0.1)	0 (0.0)
Tumor morphology (%)	Adenocarcinoma	2239 (80.0)	504 (99.6)
	Squamous cell carcinoma	540 (19.3)	0 (0.0)
	Other or unknown	20 (0.7)	2 (0.4)
Differentiation grade (%)	G1	110 (3.9)	8 (1.6)
	G2	1117 (39.9)	113 (22.3)
	G3	958 (34.2)	260 (51.4)
	G4	5 (0.2)	1 (0.2)
	Unknown	609 (21.8)	124 (24.5)
Hemoglobin (mean (SD))		9.86 (17.98)	7.74 (1.41)
Creatinine (mean (SD))		85.11 (155.64)	80.43 (28.29)
Albumine (mean (SD))		41.07 (21.07)	39.42 (17.44)
LDH (mean (SD))		203.97 (128.74)	204.84 (111.48)
ct (%)	1	50 (1.8)	11 (2.2)
	2	701 (25.0)	103 (20.4)
	3	1814 (64.8)	261 (51.6)
	4A	47 (1.7)	43 (8.5)
	4B	44 (1.6)	27 (5.3)
	X	143 (5.1)	61 (12.1)
cn (%)	0	1023 (36.5)	277 (54.7)
	1	1015 (36.3)	117 (23.1)
	2	603 (21.5)	82 (16.2)
	3	127 (4.5)	17 (3.4)
	X	31 (1.1)	13 (2.6)

Table 1 (Continued). Characteristics of included patients.

cm (%)		0 (0.0)	0 (0.0)
	0	2352 (84.0)	393 (77.7)
	1	447 (16.0)	113 (22.3)
Treatment (%)	Neoadjuvant chemoradiotherapy followed by resection	1317 (47.1)	66 (13.0)
	Neoadjuvant chemoradiotherapy followed by resection followed by adjuvant nivolumab	74 (2.6)	0 (0.0)
	Neoadjuvant chemoradiotherapy not followed by resection	423 (15.1)	1 (0.2)
	Definitive chemoradiotherapy	245 (8.8)	6 (1.2)
	Neoadjuvant chemotherapy followed by resection	107 (3.8)	115 (22.7)
	Perioperative chemotherapy	95 (3.4)	124 (24.5)
	Endoscopic resection	23 (0.8)	5 (1.0)
	Resection	50 (1.8)	62 (12.3)
	Radiotherapy	94 (3.4)	6 (1.2)
	Systemic therapy	343 (12.3)	105 (20.8)
	Best supportive care	24 (0.9)	14 (2.8)
	Unknown or other	4 (0.1)	2 (0.4)

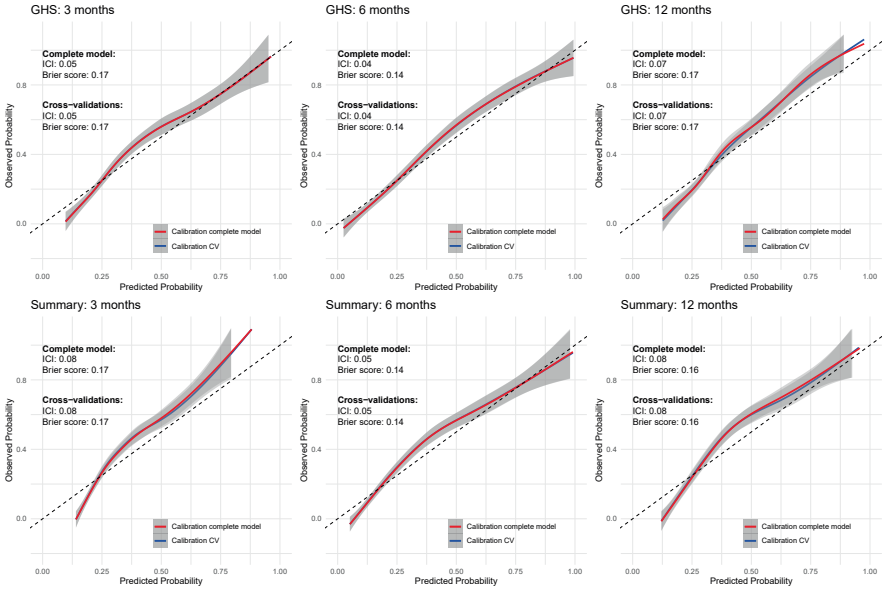


Figure 2. Calibration of the risk-prediction models for the Summary Score and GHS at 3, 6, and 12 months.

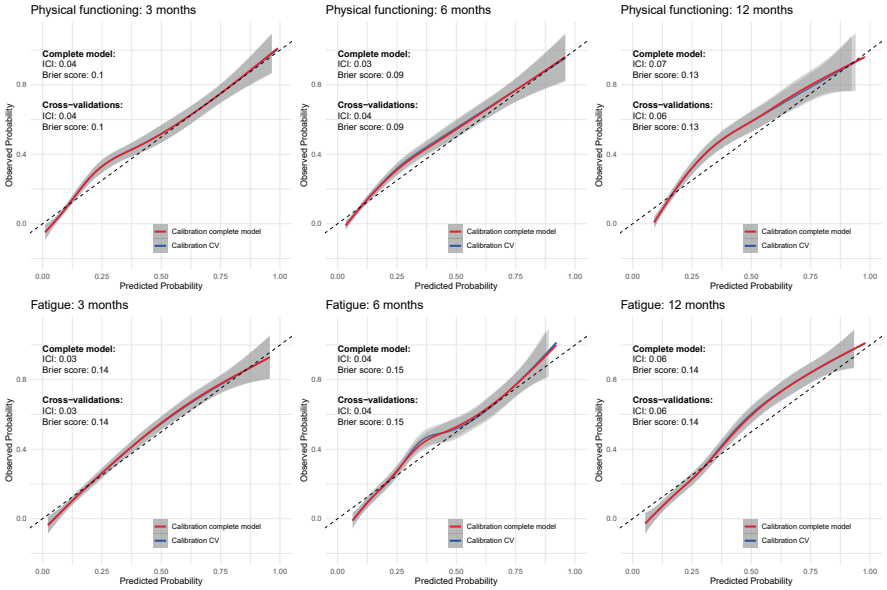


Figure 3. Calibration of the risk-prediction models for Fatigue and Physical functioning at 3, 6, and 12 months.

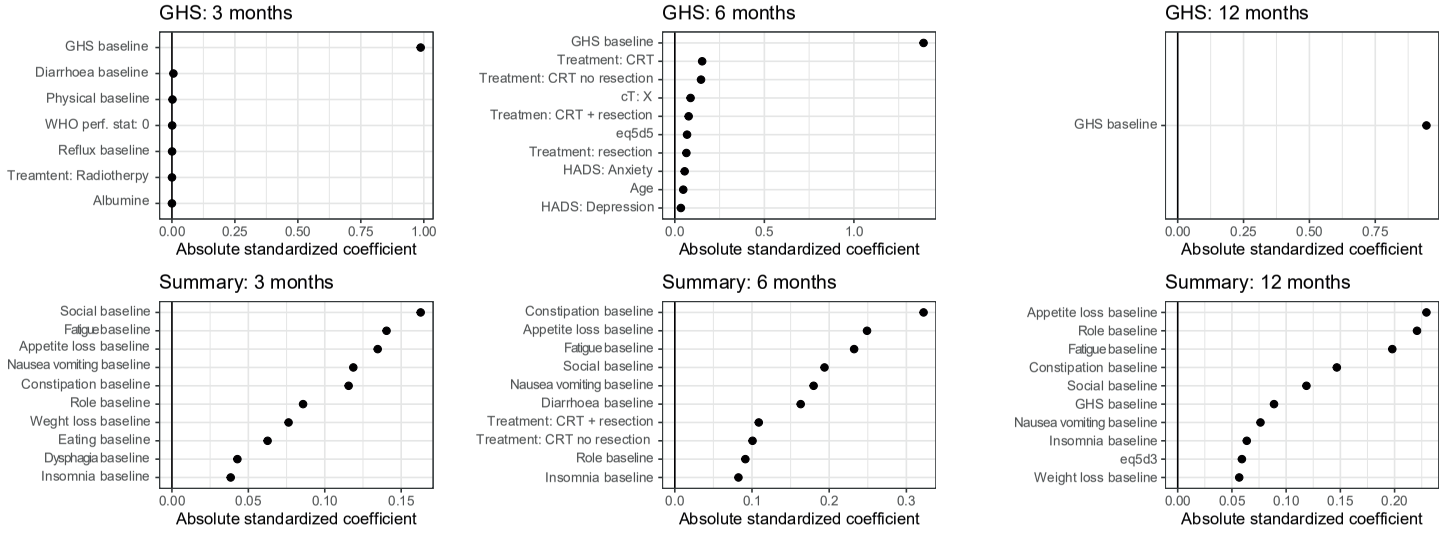


Figure 4. Feature importance of the GHS and Summary Score risk-prediction models. The list of displayed featured was truncated to show the top-10 most important features. Larger values of the absolute standardized coefficients correspond to higher feature importance.

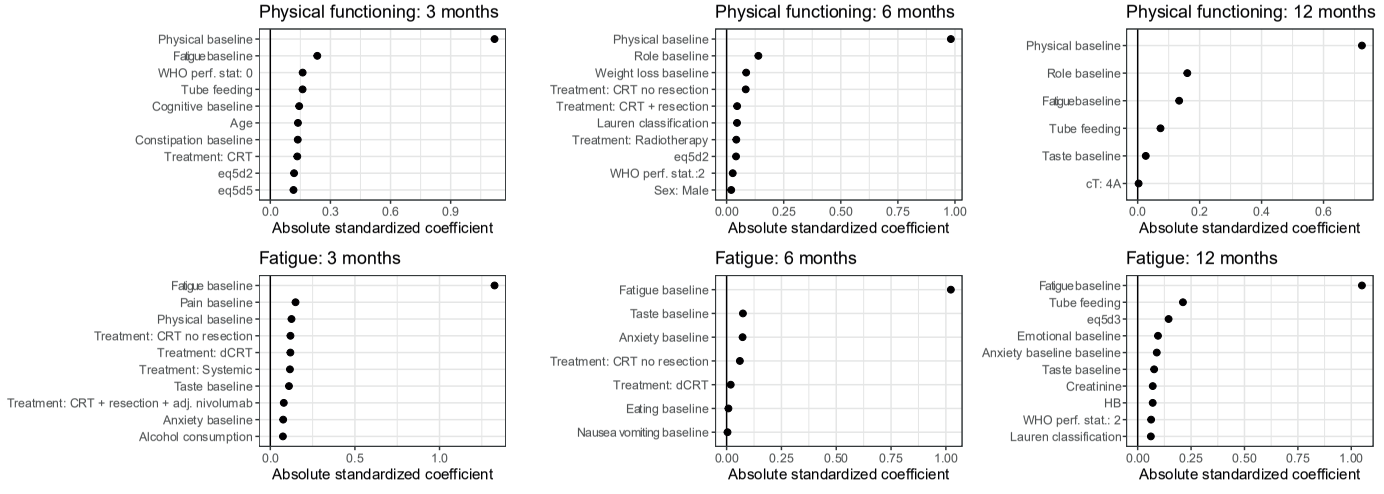


Figure 5. Feature importance of the Physical functioning and Fatigue risk-prediction models. The list of displayed featured was truncated to show the top-10 most important features. Larger values of the absolute standardized coefficients correspond to higher feature importance.

Sequential score model

The results of the internal external cross-validations and full models of Summary, GHS, Physical functioning and Fatigue are presented in Figure 6.

For the summary score, the RMSE was 9.9, the mean absolute error (MEA) was 7.1 and the r2 was 0.52. Across cross-validations the mean RMSE was 11.0, MAE was 7.9 and r2 was 0.41. The most important features in terms of relative importance were Summary score at the previous questionnaire, Fatigue and Physical Functioning at baseline (Figure 7). For the GHS, the RMSE was 14.0, the MAE was 10.3 and the r2 was 0.40. Across cross-validations the mean RMSE was 14.9, MAE was 11.0 and r2 was 0.35. GHS score at the previous questionnaire, EQ5D6 and the summary score were the most important features in terms of relative importance (Figure 7)

Based on the visual inspection of the calibration curves, it can also be observed that, in line with the risk-prediction model, physical functioning and fatigue demonstrated good calibration (Figure 6). For physical functioning, the RMSE was 12.8, the MAE was 8.9 and the r2 was 0.40. Across cross-validations the mean RMSE was 14.3, MAE was 9.9 and r2 was 0.40. The most important features in terms of relative importance were the Physical functioning score at the previous questionnaire, EQ5D1 and Physical functioning at baseline (Figure 7). For the Fatigue score, the RMSE was 17.2, the MAE was 13.1 and the r2 was 0.45. Across cross-validations the mean RMSE was 19.0, MAE was 14.4 and r2 was 0.35. Fatigue score at the previous questionnaire, Summary and Role functioning showed the largest relative feature importance (Figure 7).

In Supplementary Figure 4, the calibration plots for all 16 outcomes, based on XGboost trained on the full dataset, are shown. These plots reflect the results based on the RMSE, MAE and r2 reported at the beginning of the section. Summary, Cognitive, Emotional, Financial and Physical score models appear to be well calibrated, especially for higher scores where the largest part of the outcome distribution lies. Nausea-vomiting, Dyspnoea, Insomnia, Constipation and Appetite loss did not show a good calibration, confirming previous results based on cross-validations. The results of the internal-external cross-validation for each of the 16 outcomes can be seen in Supplementary Figure 5A-C. For the sequential score model, XGBoost outperformed Elastic net in terms of RMSE, MAE and r2 for all sixteen outcomes.

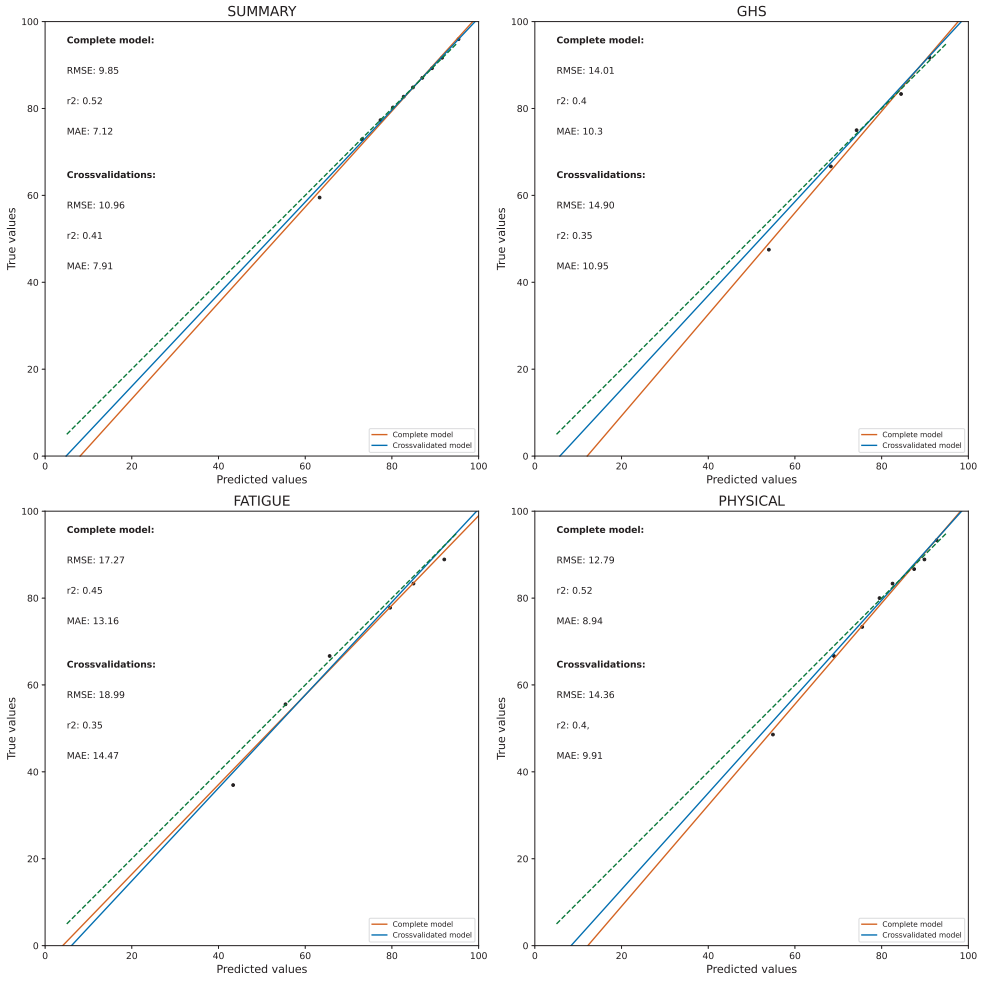


Figure 6. Calibration plots.

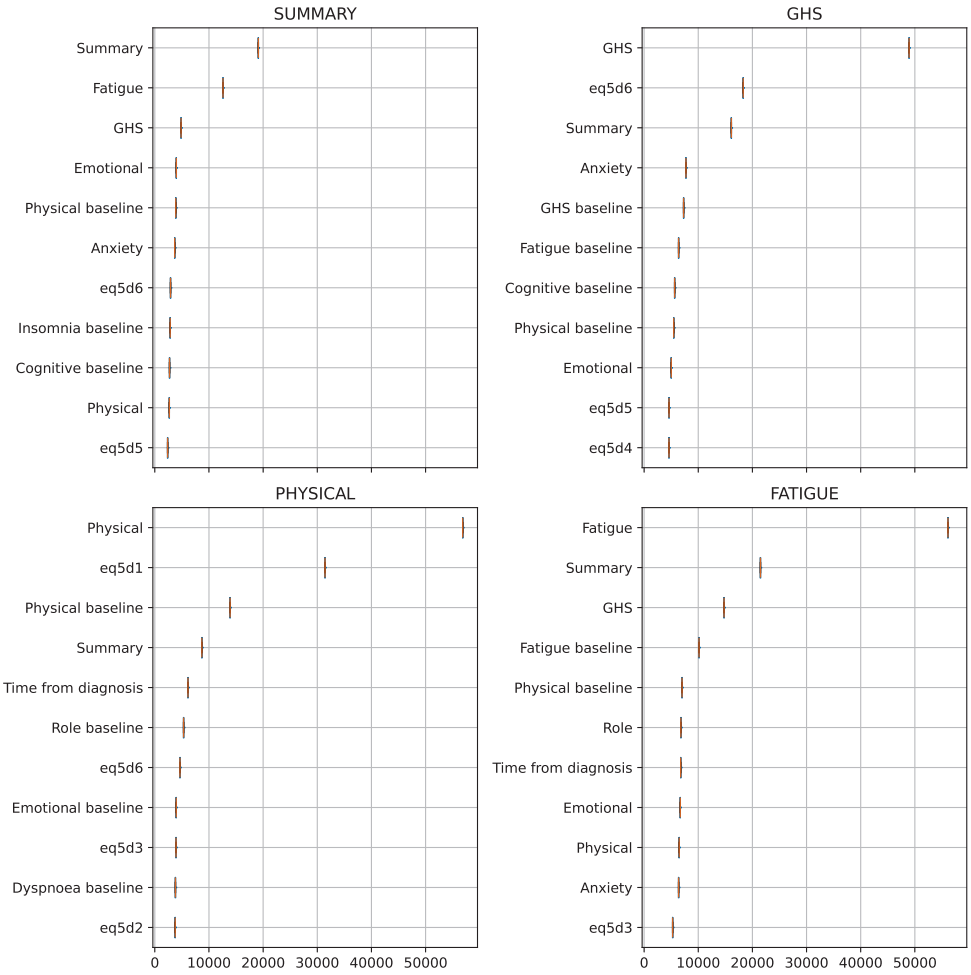


Figure 7. Feature importance of the sequential score models for the Summary Score, GHS, Physical function and Fatigue.

Discussion

In this study we have presented a large number of models to predict health-related quality of life after treatment at primary diagnosis. Using two methodically different approaches we showed that we could successfully predict global health-related quality of life, as represented by the GHS and Summary Score, as well as physical functioning and fatigue for patients with gastroesophageal cancer.

Model performance

Both the risk-prediction and sequential score models demonstrated good predictive performance for Summary Score, GHS, Physical Functioning, and Fatigue. Additionally, the sequential model showed strong performance for Cognitive, Emotional, and Financial difficulties. Brier scores indicated that the risk-prediction models outperformed random chance, with good calibration and agreement between observed and predicted probabilities. Prediction quality was consistent across time horizons (3 months to 1 year post-treatment).

The sequential model also showed good calibration, explaining up to 40% of the variance (r^2), but had a mean absolute error of 10-15 points, which corresponds to clinically significant changes in health-related quality of life. Despite good calibration, reducing this error may require modeling temporal variations throughout the patient journey, with future research focusing on integrating history from both questionnaires and clinical variables, and identifying additional predictive factors.

Results from internal-external validation of the risk-prediction model and the sequential score modeled indicated that there was relatively low overfitting of the models. The risk-prediction models in particular had model performance estimates of the complete model that were very similar to the estimates from internal-external cross validations. This is likely attributed to the elastic net modeling, which can handle a large number of (collinear) variables without overfitting.[26] For the sequential score model, the complete model showed a slightly better performance than the estimates of the internal-external cross validations. This difference, in the order of 5% for MAE, indicates that there might be some low overfitting, especially for the worst performing models. An explanation for this could lie on the tree-based nature of XGBoost, that is designed to improve incrementally performance.

Predictive features

Across all models we observed that the most common and predictive features were the baseline questionnaires or the previous questionnaires for the risk-prediction model and sequential score model, respectively. Other predictive factors that were observed using both methodologies included EORTC-C30 and OG25 baseline questionnaires, the Hospital Anxiety and Depression scale and the EQ5D questionnaire. However, there were also a notable number of differences in the selected features in the risk prediction models and the sequential score models. Clinical variables such as cT, cN, age, performance status, hemoglobin and creatinine and treatment, were selected in a number of risk-prediction models, but none of the clinical variables except for treatment were selected in the sequential score models. This may indicate that such clinical variables may be predictive of the risk of a significant deterioration in health-related quality of life, but do not predict the score itself.

A reason for this is that the clinical variables are measured at baseline only and not updated when the following questionnaires are filled in, making them less predictive in the short term for the sequential score.

Clinical implications

Since both global health-related quality of life and physical functioning and fatigue could be modelled and predicted with two different methodical approaches, this provides further evidence that the construct of health-related quality of life can be successfully predicted with a combination of clinical information such as personal, tumor and treatment related factors and baseline health-related quality of life factors. However, unlike results from health-related quality of life risk-prediction among colorectal cancer patients, our risk-prediction model did not find good predictive performance across all EORTC-C30 outcomes. It may be the case that for these outcomes, despite the comprehensive predictor set, we did not have the right predictors. The models predicting the Summary Score, GHS, Physical functioning, and Fatigue, seem to have adequate performance metrics to be used in clinical practice as the agreement between predicted and observed probabilities of a significant deterioration was very close.

We are the first to predict a sequence of health-related quality of life scores. While this needs improvement in terms of reducing prediction error, the methodology looks very promising for future application. A sequential prediction model such as the one we have developed exhibits several advantageous characteristics compared to conventional risk-prediction models. One primary advantage is its ability to incorporate time as a continuous predictor. Unlike the risk-prediction model that typically predicts risk at fixed time intervals (e.g. three, six and twelve months post start of treatment), our sequential model allows prediction at any desired time point. This flexibility enhances its utility by enabling predictions at multiple time points of interest. With our sequential model, we can effectively model the trajectory of health-related quality of life by predicting and plotting the predicted health-related quality of life score across various time intervals. This capability is particularly valuable for patients as it provides a personalized prediction of their expected health-related quality of life trajectory throughout and following treatment.

Considerations

It is important to consider some limitations of this study. First of all, due to the observational nature of the data, the models cannot be used to select a treatment which has the lowest impact on health-related quality of life. Such a prediction would be a counterfactual prediction and is currently impossible with the published models. Second, the data that was selected for fitting the models should be extended to exploit their full potential. The sequential score prediction model predicts the score of the next questionnaire based on the last questionnaires and the clinical variables at baseline, but it would be recommendable to collect clinical variables every time a new questionnaire is filled in. Finally, in the POCOP data, there was a disproportional amount of patients diagnosed with non-metastatic cancer as observed in the population. However, despite this discrepancy, an earlier study has shown that when accounting for treatment, as we have done in this study, it has very little effect on the real-world representativeness of findings obtained from POCOP data.[27]

These limitations notwithstanding, we have demonstrated two methodologically sound approaches which can be used to model and predict health-related quality of life and could also be used in different cancer settings. Testing two approaches allowed for the triangulation of evidence and given our findings that two approaches can both predict health-related quality of life, underscores the conclusion that health-related quality of life can in-fact be predicted. With this knowledge, the challenge for researchers working in this field is to find and refine predictive factors. Finally, in the future, it would be valuable if the presented models could be externally validated on new patient cohorts from the POPCOP study or on other external data from a similar population of patients.

Conclusion

This study successfully demonstrated that health-related quality of life for patients with esophageal and gastric cancer can be predicted using both risk-prediction and sequential score models. The models performed in predicting global health-related quality of life, Physical functioning and Fatigue. Our study highlights the promise of sequential prediction models, which allow for continuous predictions over time, offering personalized insights into patients' HRQoL trajectories during and after treatment. This dual-method approach underscores the robustness of HRQoL prediction and highlights the potential of HRQoL prediction for improving of patient information and care.

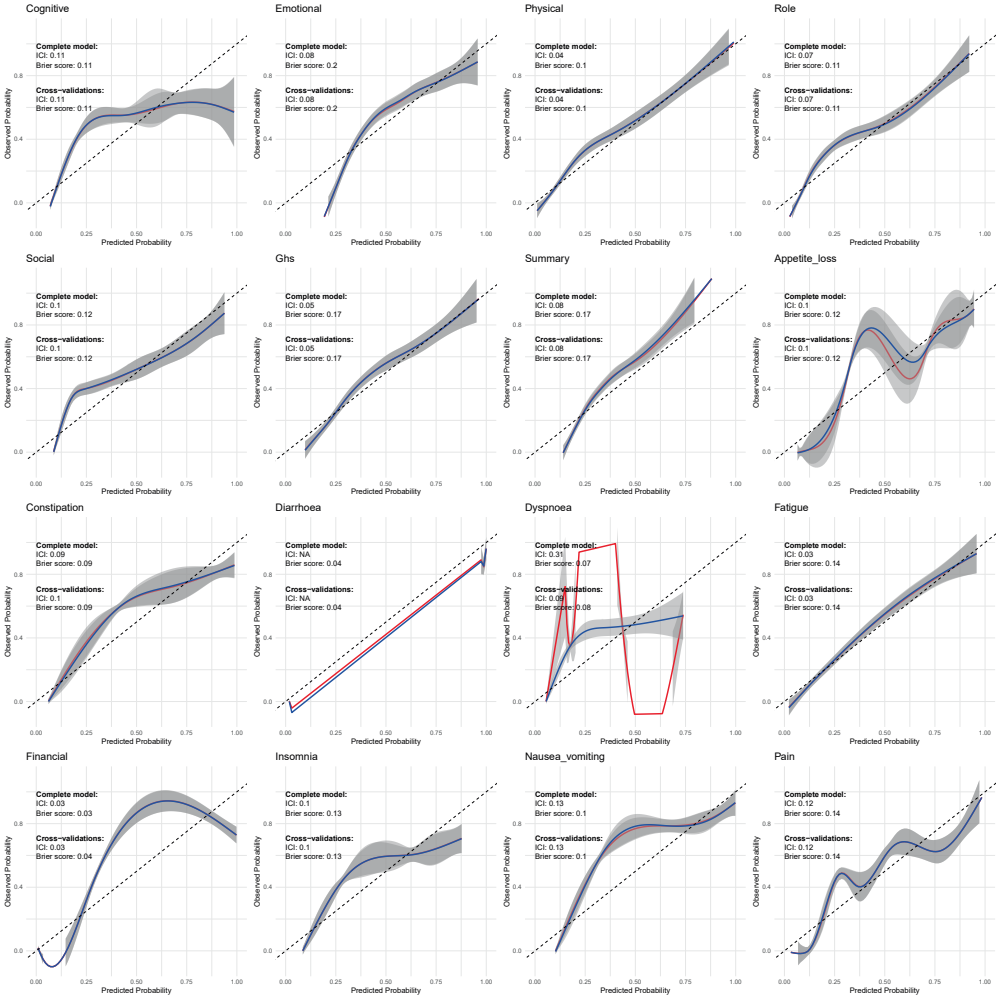
References

1. van den Boorn HG, Stroes CI, Zwinderman AH, Eshuis WJ, Hulshof MCCM, van Etten-Jamaludin FS, et al. Health-related quality of life in curatively-treated patients with esophageal or gastric cancer: A systematic review and meta-analysis. Crit Rev Oncol Hematol [Internet]. 2020;154:103069. Available from: https://doi.org/10.1016/j.critrevonc.2020.103069
2. van den Boorn HG, Stroes CI, Zwinderman AH, Eshuis WJ, Hulshof MCCM, van Etten-Jamaludin FS, et al. Health-related quality of life in curatively-treated patients with esophageal or gastric cancer: A systematic review and meta-analysis. Crit Rev Oncol Hematol [Internet]. 2020;154:103069. Available from: https://doi.org/10.1016/j.critrevonc.2020.103069
3. van Kleef JJ, ter Veer E, van den Boorn HG, Schokker S, Ngai LL, Prins MJ, et al. Quality of Life During Palliative Systemic Therapy for Gastroesophageal Cancer: Systematic Review and Meta-Analysis. JNCI: Journal of the National Cancer Institute [Internet]. 2020;112:12–29. Available from: https://doi.org/10.1093/jnci/djz133
4. van de Water LF, Kuijper SC, Henselmans I, van Alphen EN, Kooij ES, Calff MM, et al. Effect of a prediction tool and communication skills training on communication of treatment outcomes: a multicenter stepped wedge clinical trial (the SOURCE trial). EClinicalMedicine. 2023;64.
5. Van De Water LF, Van Den Boorn HG, Hoxha F, Henselmans I, Calff MM, Sprangers MAG, et al. Informing patients with gastroesophageal cancer about treatment outcomes by using a web-based tool and training: Development and evaluation study. J Med Internet Res. 2021;23.
6. Kumar S, Rana ML, Verma K, Singh N, Sharma AK, Maria AK, et al. PrediQt-Cx: Post treatment health related quality of life prediction model for cervical cancer patients. PLoS One. 2014;9.
7. Révész D, Van Kuijk SMJ, Mols F, Van Duijnhoven FJB, Winkels RM, Hoofs H, et al. Development and internal validation of prediction models for colorectal cancer survivors to estimate the 1-year risk of low health-related quality of life in multiple domains. BMC Med Inform Decis Mak. 2020;20.
8. Révész D, van Kuijk SMJ, Mols F, van Duijnhoven FJB, Winkels RM, Kant IjJ, et al. External validation and updating of prediction models for estimating the 1-year risk of low health-related quality of life in colorectal cancer survivors. J Clin Epidemiol. 2022;152:127–39.
9. Kang D, Kim H, Cho J, Kim Z, Chung M, Lee JE, et al. Prediction Model for Postoperative Quality of Life Among Breast Cancer Survivors Along the Survivorship Trajectory From Pretreatment to 5 Years: Machine Learning-Based Analysis. JMIR Public Health Surveill. 2023;9:e45212.
10. Kuijper SC, Pape M, Haj Mohammad N, van Voorthuizen T, Verhoeven RHA, van Laarhoven HWM. SOURCE beyond first-line: A survival prediction model for patients with metastatic gastroesophageal adenocarcinoma after failure of first-line palliative systemic therapy. Int J Cancer. 2023;152:1202–9.
11. van den Boorn HG, Abu-Hanna A, Haj Mohammad N, Hulshof MCCM, Gisbertz SS, Klarenbeek BR, et al. SOURCE: Prediction Models for Overall Survival in Patients With Metastatic and Potentially Curable Esophageal and Gastric Cancer. J Natl Compr Canc Netw. 2021;19:403–10.
12. Pape M, Kuijper SC, Vissers PAJ, Beerepoot L V., Creemers GJ, van Laarhoven HWM, et al. Beyond Median Overall Survival: Estimating Trends for Multiple Survival Scenarios in Patients With Metastatic Gastroesophageal Cancer. JNCCN Journal of the National Comprehensive Cancer Network. 2022;20:1321–9.
13. Kuijper SC, Pape M, Vissers PAJ, Jeene PM, Kouwenhoven EA, Haj Mohammad N, et al. Trends in best-case, typical and worst-case survival scenarios of patients with non-metastatic gastroesophageal cancer between 2006 and 2020: A population-based study. Int J Cancer. 2023;153:33–43.
14. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. BMC Med. 2015;13.
15. van Kleef JJ, Dijksterhuis WPM, van den Boorn HG, Prins M, Verhoeven RHA, Gisbertz SS, et al. Prognostic value of patient-reported quality of life for survival in ogastroesophageal cancer: ana-

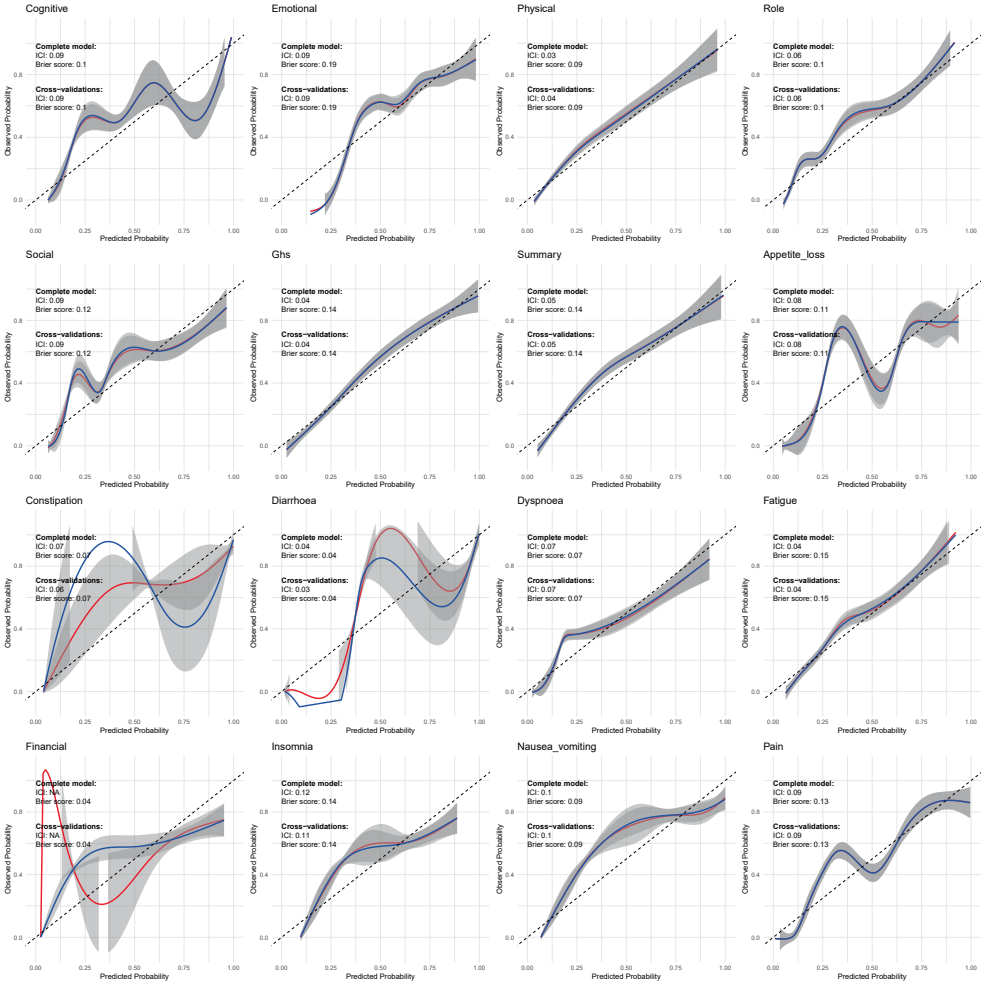
lysis from the population-based POCOP study. Gastric Cancer [Internet]. 2021;24:1203–12. Available from: https://doi.org/10.1007/s10120-021-01209-1
16. EORTC QLQ-C30 Scoring Manual The EORTC QLQ-C30. 2001.
17. Scott NW, Fayers PM, Aaronson NK, Others. EORTC QLQ-c30 Reference Values. Brussels, Belgium: EORTC Quality of Life. 2008 [cited 2019 Jan 21]; Available from: https://www.eortc.org/app/uploads/sites/2/2018/02/reference_values_manual2008.pdf
18. Giesinger JM, Kieffer JM, Fayers PM, Groenvold M, Petersen MA, Scott NW, et al. Replication and validation of higher order models demonstrated that a summary score for the EORTC QLQ-C30 is robust. J Clin Epidemiol [Internet]. 2016;69:79–88. Available from: https://www.sciencedirect.com/science/article/pii/S0895435615003832
19. Stekhoven DJ, Bühlmann P. Missforest-Non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;28:112–8.
20. Batista GEAPA, Monard MC. A Study of K-Nearest Neighbour as an Imputation Method.
21. Cocks K, King MT, Velikova G, de Castro G, Martyn St-James M, Fayers PM, et al. Evidence-based guidelines for interpreting change scores for the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. Eur J Cancer [Internet]. 2012;48:1713–21. Available from: https://www.sciencedirect.com/science/article/pii/S0959804912002110
22. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. 2016; Available from: http://arxiv.org/abs/1603.02754
23. Bergstra J, Komer B, Eliasmith C, Yamins D, Cox DD. Hyperopt: A Python library for model selection and hyperparameter optimization. Comput Sci Discov. 2015;8.
24. Austin PC, Harrell FE, van Klaveren D. Graphical calibration curves and the integrated calibration index (ICI) for survival models. Stat Med. 2020;39:2714–42.
25. Hawinkel S, Waegeman W, Maere S. The out-of-sample R^2: estimation and inference. 2023; Available from: http://arxiv.org/abs/2302.05131
26. Altelbany S. Evaluation of Ridge, Elastic Net and Lasso Regression Methods in Precedence of Multicollinearity Problem: A Simulation Study. Journal of Applied Economics and Business Studies. 2021;5:131–42.
27. Kuijper SC, Besseling J, Klausch T, Slingerland M, van der Zijden CJ, Kouwenhoven EA, et al. Assessing real-world representativeness of prospective registry cohorts in oncology: insights from patients with gastroesophageal cancer. J Clin Epidemiol [Internet]. 2023;164:65–75. Available from: https://linkinghub.elsevier.com/retrieve/pii/S089543562300269X

Supplementary Table 1. Cut-off scores used in the risk-prediction model to determine a clinically meaningful deterioration in health-related quality of life.

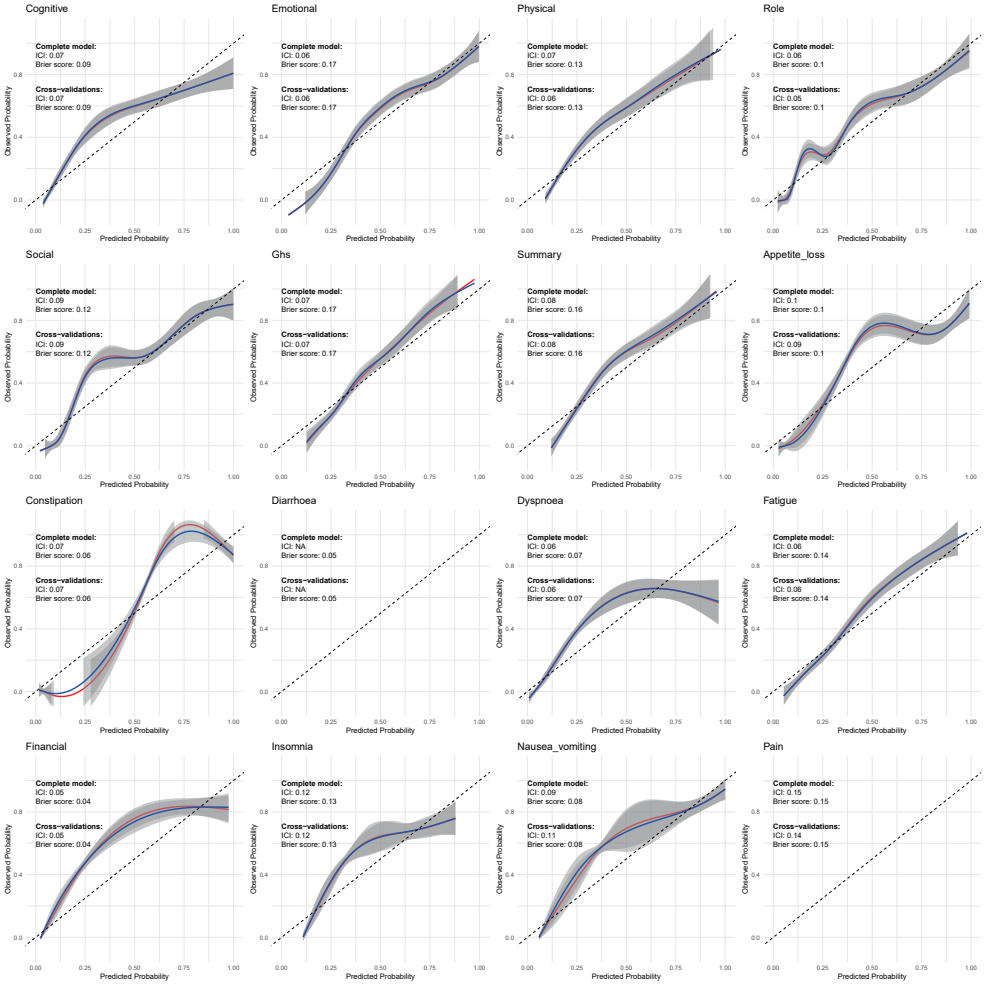
Scale	Cut-off
GHS	5
EORTC C30: Fatigue	5
EORTC C30: Nausea vomiting	5
EORTC C30: Pain	3
EORTC C30: Dyspnea	5
EORTC C30: Insomnia	2
EORTC C30: Appetite loss	2
EORTC C30: Constipation	5
EORTC C30: Diarrhoea	5
EORTC C30: Financial difficulty	2
EORTC C30: Physical functioning	5
EORTC C30: Role functioning	7
EORTC C30: Emotional functioning	3
EORTC C30: Cognitive functioning	1
EORTC C30: Social functioning	6
EORTC C30: Summary score	4



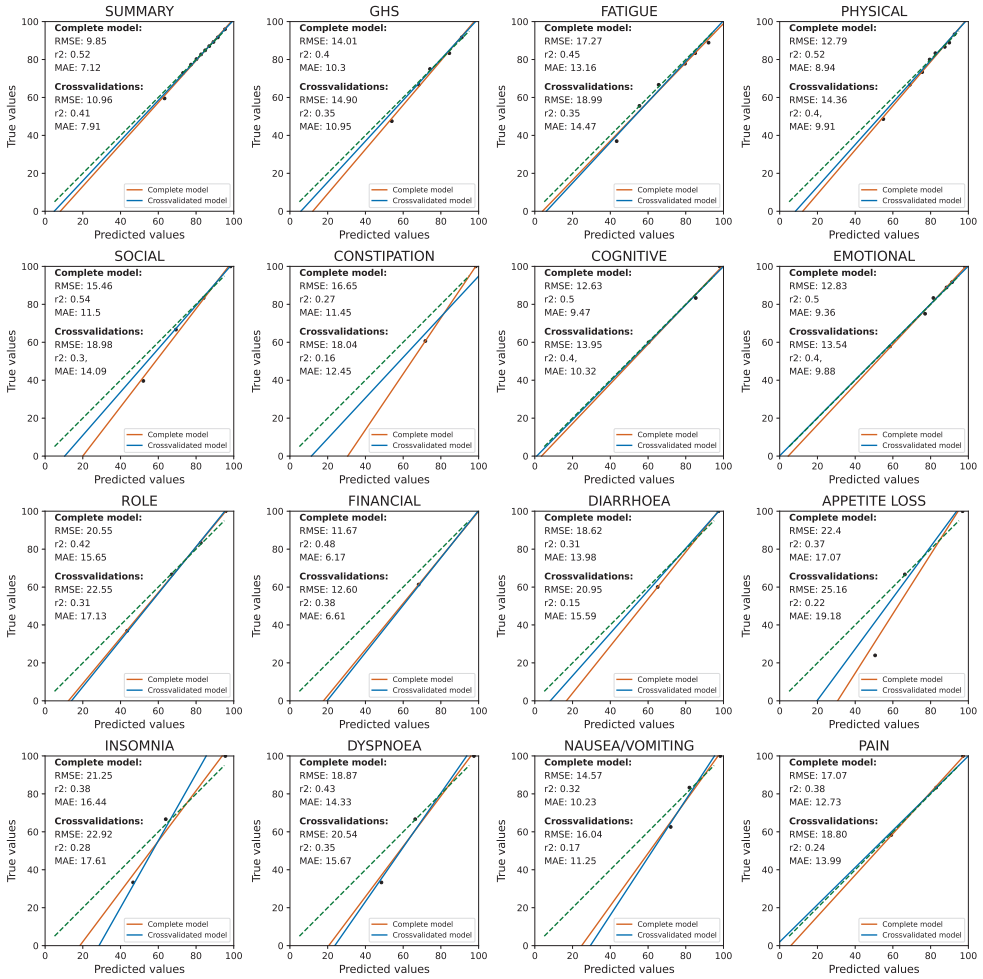
Supplementary Figure 1. Calibration of the risk-prediction models predicting the 3 months risk of a significant deterioration.



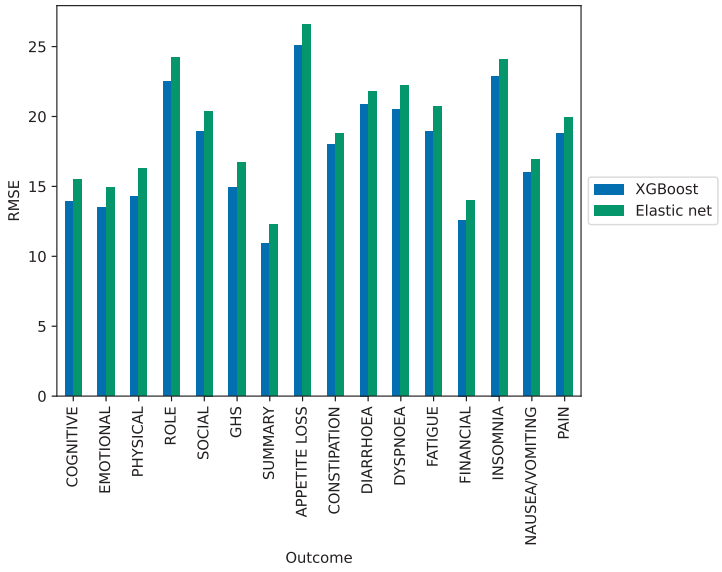
Supplementary Figure 2. Calibration of the risk-prediction models predicting the 6 months risk of a significant deterioration.



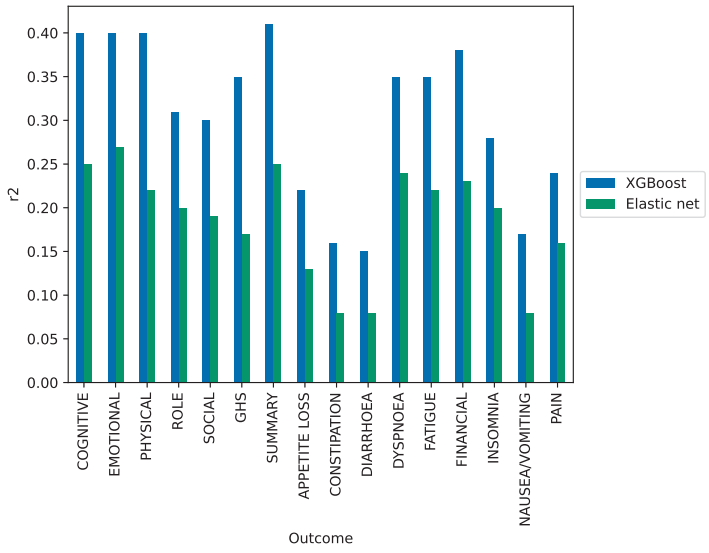
Supplementary Figure 3. Calibration of the risk-prediction models predicting the 12 months risk of a significant deterioration. Note: the model predicting Diarrhoea and Pain could not converge.



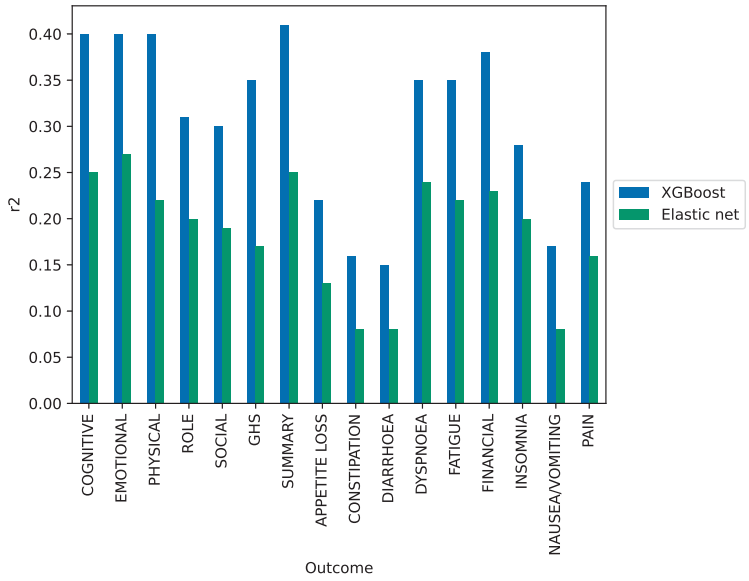
Supplementary Figure 4. Calibration of sequential score models.



Supplementary Figure 5A. Root mean squared error (RMSE) of Elastic net and XGBoost models.



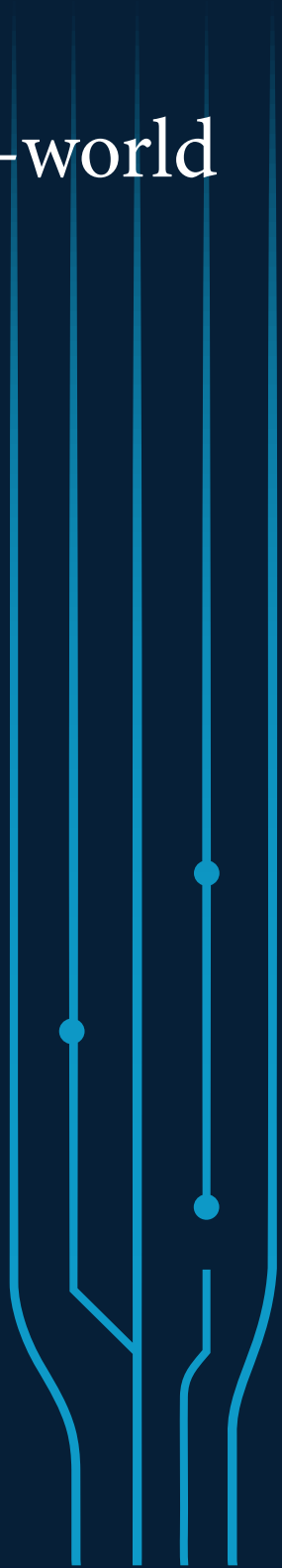
Supplementary Figure 5B. R2 of Elastic net and XGBoost models.



Supplementary Figure 5C. Mean absolute error (MEA) of Elastic net and XGBoost models.

Part III

Clinical trials and real-world data



Chapter 6

Trastuzumab Deruxtecan vs Ramucirumab-Paclitaxel as second-line therapy for patients with HER2-positive gastric or GEJ adenocarcinoma

Rob H.A. Verhoeven, Steven C. Kuijper, Florian Lordick, Marije Slingerland, Amy Qin, Hanneke W.M. van Laarhoven

Manuscript accepted for publication in Journal of the National Comprehensive Cancer Network (JNCCN)

Abstract

Background: The single arm phase 2 DESTINY-Gastric02 (DG-02) study investigated Trastuzumab Deruxtecan (T-DXd) as second-line therapy in Western patients with HER2-positive unresectable or metastatic gastric or gastro-esophageal junction (GEJ) adenocarcinoma who progressed during or after trastuzumab-based first-line therapy. The aim of the current study was to compare patients in the DG-02 study to a real-world reference group of patients who received Ramucirumab and Paclitaxel (Ram+Pac) as second-line therapy.
Methods: Patients from DG-02 who received trastuzumab in a metastatic setting were included (N=71/79). A comparable set of patients with HER2-positive esophageal, gastric or GEJ adenocarcinoma who received trastuzumab in first-line and Ram+Pac in second-line settings were identified from the Netherlands Cancer Registry (NCR) (*N*=120). Propensity score trimming and propensity score matching based on sex, age, WHO-performance status, primary tumor location, BMI, renal function, number of metastatic sites, presence of liver metastases, and duration of first-line therapy were used to select a reference group for the DG-02 from patients in the NCR.
Results: The propensity score trimming resulted in exclusion of *N*=12 DG-02 and *N*=33 NCR patients. Thereafter, propensity score matching resulted in a balanced group of patients from the NCR (*N*=78) and the DG-02 trial (*N*=58). Median overall survival was significantly longer among patients treated with T-DXd (11.6 months (95%CI: 9.0-20.5)) compared to the Ram+Pac reference group (6.2 months (95%CI: 4.5 – 10.0)) ($p<0.0001$).
Conclusions: Compared to patients with metastatic, trastuzumab-pretreated HER2-positive gastric or GEJ adenocarcinomas who received Ram+Pac, overall survival was better for patients who received T-DXd as second-line therapy.

Introduction

For patients with human epidermal growth factor receptor 2 (HER2) positive advanced gastric or gastro-esophageal adenocarcinoma trastuzumab in combination with chemotherapy is the advised first-line treatment according to European Society of Medical Oncology (ESMO) guidelines.[1–3] However, for second-line treatment, the strategy of continuation of trastuzumab beyond progression has failed.[4,5] Therefore, the ESMO guidelines advise ramucirumab plus paclitaxel (Ram+Pac), based on the RAINBOW study, as second-line systemic therapy for patients with gastric or gastro-esophageal junction (GEJ) adenocarcinomas, irrespective of HER2 status.[2,6] A subgroup analysis of the HER2-positive patients (N=39) of the RAINBOW trial revealed that the median OS was 11.4 months for patients treated with Ram+Pac compared to 7.0 months when treated with placebo plus paclitaxel.[4]

The randomized DESTINY-Gastric01 trial showed an improved median OS in South Korean and Japanese patients with HER2-positive gastric or GEJ adenocarcinoma. Patients treated with trastuzumab-deruxtecan (T-DXd) after at least two previous therapies - including trastuzumab – had a median overall survival of 12.5 months compared to 8.4 months when treated with chemotherapy.[7] Additionally, the single-arm phase two DESTINY-Gastric02 study investigated the efficacy and safety of T-DXd in Western patients and showed a median OS of 12.1 months, with 42% complete or partial repsonses.[8] The currently ongoing randomized global phase three DESTINY-Gastric04 study (NCT 04704934) compares T-DXd with Ram+Pac, but survival outcomes of this study are not yet available.

Based on the results of the DESTINY-Gastric02 trial, T-DXd was approved by the FDA for use in patients with locally advanced or metastatic HER2-positive gastric or GEJ adenocarcinoma who have received a prior trastuzumab-based regimen.[5,9] EMA approval followed in November 2022, after results of the DESTINY-Gastric02 study were presented at the ESMO conference of 2022.[8]

While awaiting the results of the DESTINY-Gastric04 study, the potential impact of T-DXd use in second line for Western patients may already be estimated via comparison to a similar cohort of patients treated with standard of care, using method of propensity score matching. Although comparison with a propensity score-matched cohort does not reach the scientific level of evidence of a prospective randomized study, propensity score matching can support the estimate of causal effects for an observational study by reducing confounding and bias. Therefore, the aim of this study was to compare OS between the patients included in the DESTINY-Gastric02 study and a propensity score matched cohort of Western patients who were treated with standard second-line systemic treatment of Ram+Pac.

Methods

Study design

In this study data from the Netherlands Cancer Registry (NCR) were used to create a selection of patients that could serve as a real-world reference group and thereafter be compared on overall survival with patients from the DESTINY-Gastric02 study. The NCR is a nation-wide database containing all patients with diagnosed malignancies and covers the entire Dutch population of almost 18 million people. Trained data managers routinely extract information on patient, tumor and treatment characteristics directly from patients' electronic medical records and include these in the NCR.

Selection of patients

A total of 79 patients with unresectable or metastatic gastric or GEJ adenocarcinomas who had confirmed HER2 positive tumors after progression of disease during or after treatment with a trastuzumab containing regimen were included in the single-arm DESTINY-Gastric02 study and received T-DXd at 6.4 mg/kg every 3 weeks. Except for a small group of DESTINY-Gastric02 patients who received a trastuzumab containing regimen in a non-metastatic setting (N=8), for which no similar group was available from the NCR, all patients from the DESTINY-Gastric02 study were included in the current study (N=71). Patients from the DESTINY-Gastric02 study were enrolled in the United States (43%) and Western European countries (57%, Great Britain, Belgium, Spain, Italy).

In the selection of patients from the NCR we kept as close as possible to the selection criteria of the DESTINY-Gastric02 study. Therefore, we included patients (N=120) with synchronous metastatic HER2-positive adenocarcinomas of the esophagus, GEJ, or stomach, who were diagnosed between 2014-2021 and received trastuzumab in first-line systemic therapy and Ram+Pac as second-line systemic therapy. Data on patients with metachronous metastatic HER2-positive adenocarcinomas was only limited available in the NCR and therefore not included in the current study. As HER2 testing is not routinely performed in daily practice after progression on trastuzumab in first line, HER2 status was almost always only known prior to start of first line treatment. To increase the number of patients from the NCR, we included patients with esophageal adenocarcinomas, as a recent publication of our group showed similar outcomes in patients with HER2 positive esophageal, GEJ/cardia and gastric adenocarcinomas.[10] Similar outcomes between these groups were found in other studies with HER2 negative patients or patients not tested on HER2 status.11–12 After imputation of missing data, patients with ECOG performance status ≥2 prior to start of Ram+Pac were excluded from the analysis (N=9). Figure 1 displays the complete patient inclusion flowchart. Data of the NCR have been previously been used as matched control comparisons in several other single arm studies.[13–16]

Variables and coding

To match the data of the DESTINY-Gastric02 study and the NCR we made use of the following variables: age, sex, ECOG performance status, primary tumor location, body mass index (BMI), number of metastatic sites, presence of liver metastases, renal function, and duration of first line therapy. Variables that could vary over time were collected prior to start of second line therapy.

In the NCR, serum creatinine was available, whereas creatinine clearance was used in the DESTINY-Gastric02 study. We used the Cockcroft-Gault equation to transform serum creatinine into creatinine clearance.[17] After transformation, creatinine clearance was classified into three renal function categories: normal (CLcr ≥90 mL/min), mild renal impairment (CLcr ≥60, <90 mL/min) and moderate renal impairment (≥30, <60 mL/min). As patients with primary esophageal cancer could formally not be included in the DESTINY-Gastric02 study but formed a considerable proportion of the NCR population (n=69), the primary tumor location of these patients was recoded into GEJ. To test the effect of this recoding a sensitivity analysis on the included patients from the NCR was performed, in which the overall survival of the patients that were recoded from esophageal to GEJ was compared to patients that were originally coded as GEJ or gastric cancer.

All other variables were equally coded between both datasets or could be recoded without adjustment of the content.

Statistical analyses

We used propensity score matching to construct a matched real-world reference group to compare OS between patients treated with T-DXd (DESTINY-Gastric02) and patients treated with Ram+Pac (NCR). Propensity score matching aims to select two groups of patients who received a different treatment, but have similar propensity of receiving the treatment conditional on a set of baseline characteristics. This increases covariate balance and thereby reduces the effect of confounding.[18] Prior to propensity score matching, missing data were imputed using random forest imputation from the R-package missRanger, separately for the DESTINY-Gastric02 and NCR data sets. After imputation, both data sets were merged and the propensity score was estimated using a multivariable logistic regression model which included sex, age, performance status (0 vs 1), primary tumor location (GEJ versus stomach), BMI (continuous), number of metastases (1 vs ≥2), presence of liver metastases (yes vs no), renal function (normal vs mild impairment vs moderate impairment), and duration of first line treatment (continuous). Trimming of the propensity scores that did not overlap between groups was performed, which resulted in exclusion of 12 patients from the DESTINY-Gastric02 group and 33 from the NCR group. Various matching methods (nearest neighbor matching (with and without replacement), optimal full matching and genetic matching) and calipers were iteratively tested to asses which would result in the best covariate balance.[18–20] The R-package MatchIt was used for propensity score matching. A standardized mean difference of <0.10 between the two groups was a priori determined to be a good balance. The dataset with the best covariate balance after propensity score matching was selected and only on this dataset OS analyses were performed.

Optimal full matching with a caliper of 0.16 times the logit of the propensity score resulted in the covariate balance with the lowest standardized mean differences. With optimal full matching, strata are created of either one patient from the DESTINY-Gastric02 and one or more patients in the NCR, or one patient from the NCR and one or more patients from the DESTINY-Gastric02 trial. Subsequently, weights are assigned to the patients in the control groups which are used to minimize remaining mean differences in strata between the control and treated group.[19] OS since start of second-line treatment of patients in the DESTINY-Gastric02 and the NCR was plotted in a Kaplan-Meier curve and tested for significance using the log-rank test. The R-package survival was used for the survival analyses. All analyses were performed in R version 4.1.0 using R studio. The study was performed according to the STROBE guidelines.[21]

Results

Propensity score matching

After propensity-score matching, we found matches for 58 patients from the DESTINY-Gastric02 study using N=78 patients from the NCR (Figure 1).

Standardized mean differences of sex, ECOG performance status, primary tumor location, BMI, the number of metastatic sites, renal function and the duration of first-line systemic therapy were all ≤0.10 and thus could be considered as well balanced (Table 2). Standardized mean differences of age (0.110) and presence of liver metastases (0.115) were slightly above 0.10. The weighted median age was 61 (IQR: 53-68) years in the DESTINY-Gastric02 trial and 55 (IQR: 52-66) years in the NCR, and the weighted percentage of the presence of liver metastases was 77.6% in the DESTINY-Gastric02 trial and 82.5% in the NCR. Additional tumor and disease characteristics that were not included in the matching procedure are presented in supplementary table 1.

Overall survival outcomes

Patients treated with T-DXd in the DESTINY-Gastric02 study had a significantly better OS compared to the patients from the NCR who were treated with Ram+Pac (p<0.0001) (Figure 2). Median OS of patients from the DESTINY-Gastric02 study was 11.6 (95%CI: 9.0-20.5) months and 6.2 (4.5 – 10.0) months for the patients from the NCR, with a hazard-ratio of 0.39 (95%CI: 0.26 –0.59, p<0.0001). One year OS was 49.5% (95% CI: 37.8% – 64.9%) for the T-DXd group and 21.5% (95% CI: 9.5% – 48.6%) for the patients treated with Ram+Pac.

Comparison of excluded and included patients from DESTINY-Gastric02 study

A comparison of patient and tumor characteristics of excluded and included patients from the DESTINY-Gastric02 study is presented in table 2. All characteristics were different (standardized mean difference of >0.10) between the excluded and included group. Patients that were excluded from the comparison were younger, more often female, more often had an ECOG performance status of 1, more often had a primary tumor located in the stomach, more often had a moderate renal impairment, less often had liver metastases and had a longer duration of first line therapy. In addition, patients excluded for receiving trastuzumab in a non-metastatic setting had a higher BMI and more often only 1 metastatic site, while patients excluded during the matching procedures had a lower BMI and a rather comparable proportion of ≥2 metastatic sites compared to patients from the DESTINY-Gastric02 study included in the comparison.

The OS did not differ between the excluded and included patients from the DESTINY-Gastric02 study (p=0.74, figure 3).

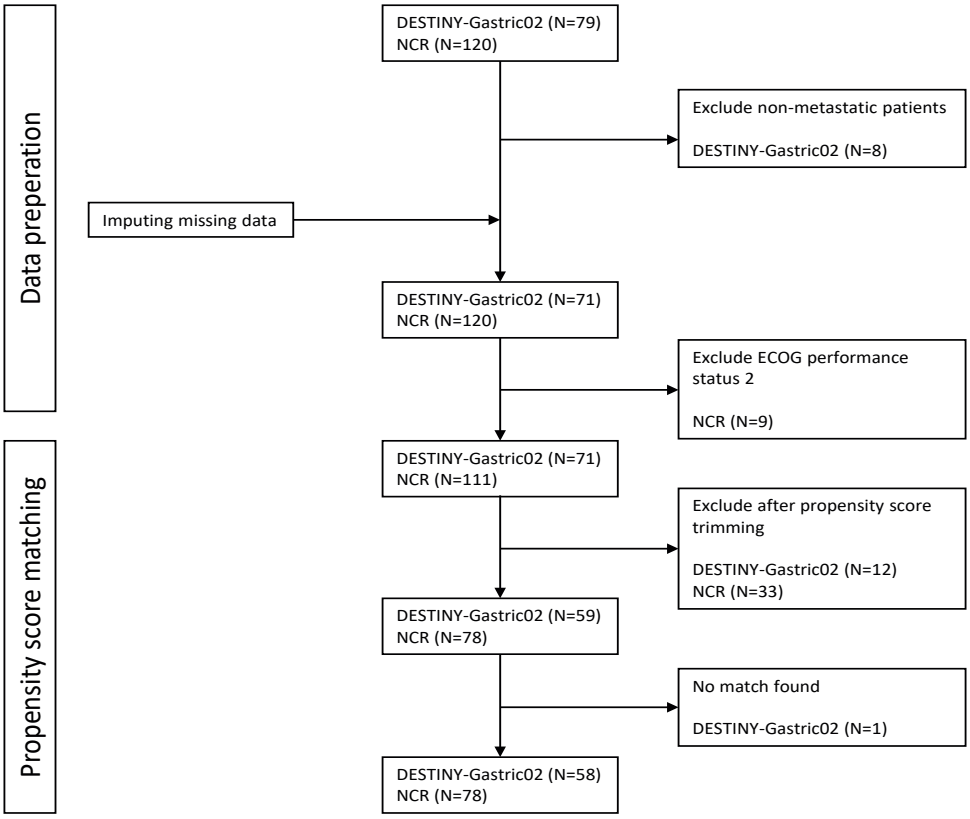


Figure 1. Flowchart of patients inclusion and propensity score matching.

Table 1. Overview of balance of patient, tumor and treatment characteristics of DESTINY-Gastric02 group (n=58) and NCR control group (n=78) after propensity-score matching.

	DESTINY-Gastric02 (N=58)	**NCR reference group** (N=78)	**Standardized mean difference**
Age *median (IQR)*	61 (53-68)	55 (52-66)	0.110
Sex			
- Male	81.0%	83.7%	0.068
- Female	19.0%	16.3%	0.068
ECOG performance status			
- 0	41.4%	41.2%	0.004
- 1	58.6%	58.8%	0.004
Primary tumor location			
- GEJ / cardia	75.9%	74.2%	0.040
- Stomach	24.1%	25.8%	0.040
BMI *median (IQR)*	25.2 (22.8-27.4)	25.3 (23.6-27.4)	0.090
Number of metastatic sites			
- 1	5.2%	4.6%	0.027
- ≥2	94.8%	95.4%	0.027
Presence of liver metastases			
- No	22.4%	17.5%	0.115
- Yes	77.6%	82.5%	0.115
Renal function			
- Normal	53.4%	55.6%	0.043
- Mild renal impairment	43.1%	42.5%	0.013
- Moderate renal impairment	3.4%	1.9%	0.083
Duration of first line therapy (months) *median (IQR)*	8.4 (5.0 – 12.8)	7.4 (4.0 – 13.1)	0.017

IQR = Interquartile range

Table 2. Baseline characteristics of all patients in the DESTINY-Gastric02 group

	DESTINY-Gastric02 Excluded during selection (N=8)	**DESTINY-Gastric02 Excluded during matching procedures** (N=13)	**DESTINY-Gastric02 Included** (N=58)	**Standardized mean difference**
Sex (%)				
Male	2 (25.0)	8 (61.5)	47 (81.0)	0.864
Female	6 (75.0)	5 (38.5)	11 (19.0)	
Age *median (IQR)*	63.5 (52.5 - 69.3)	52.0 (45.0 - 64.0)	62.0 (53.0 - 68.0)	0.380
ECOG performance status (%)				
0	2 (25.0)	3 (23.1)	24 (41.4)	0.266
1	6 (75.0)	10 (76.9)	34 (58.6)	
Primary tumor location (%)				
GEJ/cardia	5 (62.5)	3 (23.1)	44 (75.9)	0.801
Stomach	3 (37.5)	10 (76.9)	14 (24.1)	
BMI *median (IQR)*	27.3 (24.4 - 27.6)	23.5 (20.6 - 28.9)	25.2 (22.9 - 27.4)	0.218
Renal function (%)				
Normal	2 (25.0)	4 (30.8)	31 (53.4)	1.094
Mild renal impairment	3 (37.5)	3 (23.1)	25 (43.1)	
Moderate renal impairment	1 (12.5)	6 (46.2)	2 (3.4)	
Missing	2 (25.0)	0 (0.0)	0 (0.0)	
Number of metastatic sites (%)				
1	2 (25.0)	0 (0.0)	3 (5.2)	0.574
≥2	6 (75.0)	13 (100.0)	55 (94.8)	
Presence of liver metastases (%)				
No	6 (75.0)	10 (76.9)	13 (22.4)	0.861
Yes	2 (25.0)	3 (23.1)	45 (77.6)	
Duration of first line therapy (months) *median (IQR)*	15.2 (11.1 - 21.8)	17.3 (9.7 - 28.0)	8.5 (5.2 - 13.0)	0.551

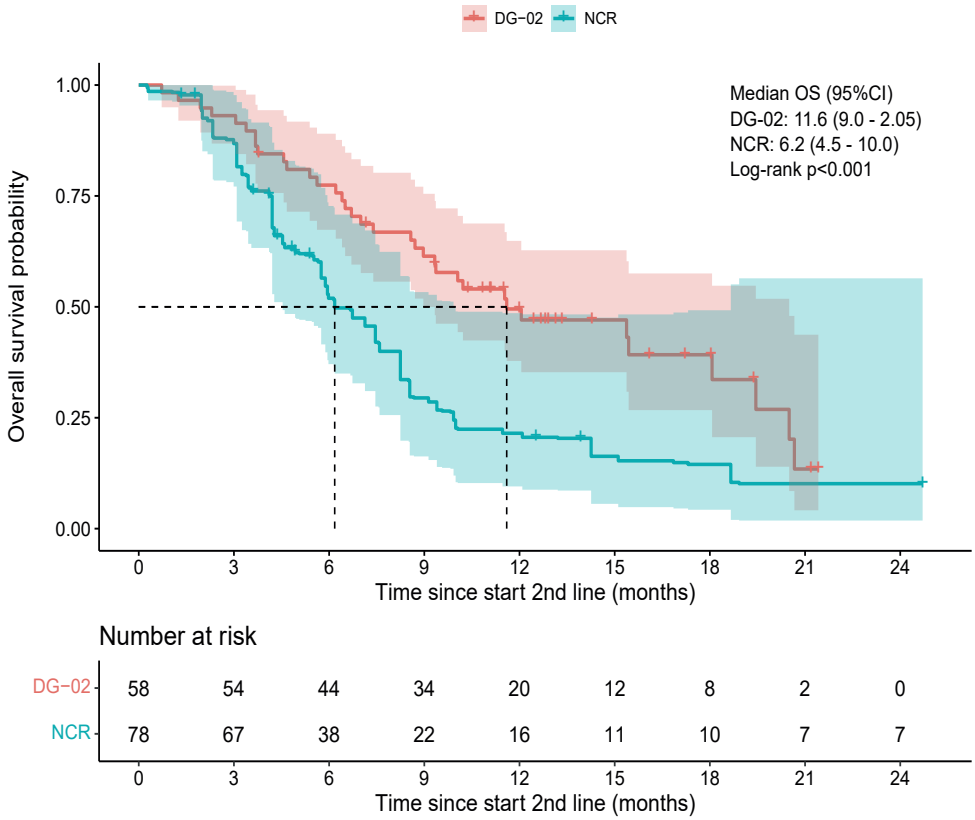


Figure 2. Kaplan-Meier curves on overall survival of the matched patients from the DESTINY-Gastric02 (DG-02) study and the NCR.

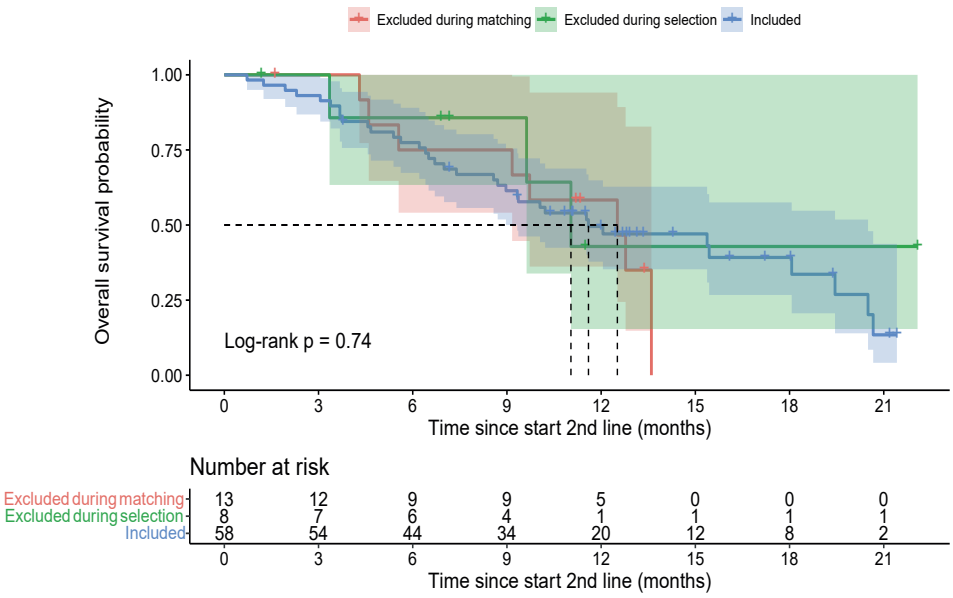


Figure 3. Kaplan-Meier curves on overall survival of the different DESTINY-Gastric02 study groups.

Sensitivity analysis on primary tumor location recoding

In the weighted analyses after matching the primary tumor location for the Ram+Pac patients was 74.2% GEJ (table 1), based on the original registration, this group consisted of 59.2% esophageal and 15.0% GEJ adenocarcinomas. The sensitivity analyses on the OS outcomes of the Ram+Pac patients that were originally coded as esophageal cancer compared to the Ram+Pac patients that were coded as GEJ/Gastric cancer showed no significant OS difference (p=0.50, supplemental figure 1).

Discussion

This study demonstrated a significantly longer survival of patients with metastatic, HER2-positive gastric and GEJ adenocarcinomas who received T-DXd in second-line in the DESTINY-Gastric02 study compared to a propensity score matched cohort of patients from the NCR who received Ram+Pac. These results are consistent with earlier findings in the DESTINY-Gastric01 trial, which showed that survival was higher among patients treated with T-DXd (12.5 months) compared to physicians' choice chemotherapy (8.4 months).[7]

Importantly, although the baseline characteristics were, as expected, different, the OS of the excluded and included patients from the DESTINY-Gastric02 did not differ. The OS among the patients of the DESTINY-Gastric02 trial included in the current study (11.6 months) was also quite similar compared to the earlier presented complete DESTINY-Gastric02 trial (12.1 months).[8] This indicates that the subset of patients of the DESTINY-Gastric02 trial that was included in this study for matching was representative of all patients in the DESTINY-Gastric02 trial in terms of OS.

The median OS of the matched NCR group (6.2 months, (95%CI: 4.5 – 10.0) was lower compared to all patients from the RAINBOW trial (9.6 months) as well as the small trastuzumab pre-treated group of the RAINBOW trial (11.4 months 95% CI: 7.0 – 17.9).[4,6] However, survival results were relatively similar to a recent Spanish real-world study which showed a median OS of 7.4 months (95%CI: 6.1 – 12.0) for HER2-positive patients treated with Ram+Pac in second line.[4,22] Although we had a study population that was matched on baseline characteristics to patients in the DESTINY-Gastric02 study, it was not matched to the original RAINBOW study. Thus, a comparison of OS between the trastuzumab pre-treated group of the RAINBOW study and the reference group treated with Ram+Pac in our current study should be made with great caution.

Evaluation of progression in clinical trials such as the DESTINY-Gastric02 study is highly standardized, with routinely planned scans which are evaluated using RECIST criteria. In daily clinical practice, follow-up with CT scans is more flexible and RECIST criteria are often not used. We have therefore deliberately chosen to not include a comparison of progression-free survival (PFS) between the patients from the DESTINY-Gastric02 and the NCR, but primarily focus on OS, which in the end remains the most important outcome parameter in this setting.

Some consideration regarding the propensity score matching to construct a reference group should be taken into account as possible limitations of the current study. First, the degree of HER2 expression/amplification (immunohistochemistry (IHC)3+ versus IHC2 and in situ hybridization (ISH)+) seems to be associated with response to T-DXd.[8] However, we could not match patients on the degree of HER2 expression as this information was not available in the NCR. Second, as HER2 testing prior to start of second line is not routine practice, we had no information on HER2 status prior to start of second line for patients from the NCR. It is known that a discordance of HER2 can occur between different specimens and/or over time.[23–25] We acknowledge the potential bias introduced by the post-progression HER2 testing in the DESTINY-Gastric02 study compared to the NCR cohort. However, it is important to note that HER2 is primarily considered a predictive biomarker for treatment response in relation to trastuzumab-based therapies. This means that while HER2 status is crucial for determining the likelihood of response to HER2-targeted therapies, the impact of HER2 status on overall prognosis independent of treatment is controversial.[25,26]

However, some impact on the outcomes of our study cannot be excluded. Third, in contrast to the DESTINY-Gastric02 study we also included patients with HER2 positive esophageal adenocarcinoma. However, our sensitivity analysis showed no difference in OS between the Dutch patients with a esophageal adenocarcioma compared to GEJ or gastric adenocarcinoma included in the current study. In addition, a recent publication of our group showed similar survival in HER2 positive esophageal, GEJ/cardia and gastric adenocarcinomas treated with a trastuzumab-based regimen in first-line and other studies reported similar survival between in HER2 negative esophageal, GEJ/Cardia and gastric adenocarcinomas in first-line treatment and also in patients not tested on HER2.[11,12,27] In addition, it was not possible to match on prior surgery, as none of the patients from the Ram+Pac group had undergone prior surgery. Information on prior surgery is in the patients treated with T-DXd presented in supplementary table 1. Last, although we constructed a reference group that was generally well-balanced, the standardized mean difference for age and presence of liver metastases was slightly above 0.10, indicating that these variables were not fully balanced. The median age of the NCR group was lower than that of the DESTINY-Gastric02 study (55 vs. 61 years), but the inter quartile ranges of both groups were quite similar (52-66 vs 53-68 years) indicating that the groups were still relatively well-balanced with regards to age. In the case that a lower age in the NCR group compared to the DESTINY-Gastric02 group would have affected the survival outcomes, one would expect that the lower age in the NCR group would result in a positive effect on the survival of the Ram+Pac group, which would make the actual difference between the two groups even larger than the currently presented survival difference. The presence of liver metastases was only slightly higher in the NCR control group (82.5% vs. 77.6%), which therefore most likely did not have a large effect on the survival outcomes.

A strength of this study was that multiple matching techniques and calipers were iteratively tested to determine the best method to maximize covariate balance, a necessary condition for causal inference. In our analysis the covariate balance was quite good as almost all standardized mean differences were below 0.10.

Conclusion

Compared to patients with metastatic or unresectable HER2-positive gastric and GEJ adenocarcinomas who received Ram+Pac in the second-line, OS was longer for patients who received T-DXd in second-line. Due to the non-randomized nature of the current comparison, the results should be interpreted with caution. However, while awaiting the results of the randomized phase III DESTINY-Gastric04 study, these results may aid in decision making regarding the use of T-DXd in clinical practice.

References

1. Bang YJ, Van Cutsem E, Feyereislova A, et al. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): A phase 3, open-label, randomised controlled trial. *The Lancet*. 2010;376(9742):687-697. doi:10.1016/S0140-6736(10)61121-X
2. Lordick F, Carneiro F, Cascinu S, et al. Gastric cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Annals of Oncology*. 2022;33(10):1005-1020. doi:10.1016/j.annonc.2022.07.004
3. ter Veer E, Creemers A, de Waal L, van Oijen MGH, van Laarhoven HWM. Comparing cytotoxic backbones for first-line trastuzumab-containing regimens in human epidermal growth factor receptor 2-positive advanced oesophagogastric cancer: A meta-analysis. *International Journal of Cancer*. 2018;143(2):438-448. doi:10.1002/ijc.31325
4. De Vita F, Borg C, Farina G, et al. Ramucirumab and paclitaxel in patients with gastric cancer and prior trastuzumab: Subgroup analysis from RAINBOW study. *Future Oncology*. 2019;15(23):2723-2731. doi:10.2217/fon-2019-0243
5. ter Veer E, van den Ende T, Creemers A, de Waal L, van Oijen MGH, van Laarhoven HWM. Continuation of trastuzumab beyond progression in HER2-positive advanced esophagogastric cancer: a meta-analysis. *Acta Oncologica*. 2018;57(12):1599-1604. doi:10.1080/0284186X.2018.1503421
6. Wilke H, Muro K, Van Cutsem E, et al. Ramucirumab plus paclitaxel versus placebo plus paclitaxel in patients with previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (RAINBOW): A double-blind, randomised phase 3 trial. *The Lancet Oncology*. 2014;15(11):1224-1235. doi:10.1016/S1470-2045(14)70420-6
7. Shitara K, Bang YJ, Iwasa S, et al. Trastuzumab Deruxtecan in Previously Treated HER2-Positive Gastric Cancer. *New England Journal of Medicine*. 2020;382(25):2419-2430. doi:10.1056/nejmoa2004413
8. Van Cutsem E, di Bartolomeo M, Smyth E, et al. Trastuzumab deruxtecan in patients in the USA and Europe with HER2-positive advanced gastric or gastroesophageal junction cancer with disease progression on or after a trastuzumab-containing regimen (DESTINY-Gastric02): primary and updated analyses from a. *The Lancet Oncology*. 2023;24(7):744-756. doi:10.1016/s1470-2045(23)00215-2
9. Makiyama A, Sukawa Y, Kashiwada T, et al. Randomized, Phase II Study of Trastuzumab beyond Progression in Patients with HER2-Positive Advanced Gastric or Gastroesophageal Junction Cancer: WJOG7112G (T-ACT Study). *Journal of Clinical Oncology*. 2020;38(17):1919-1927. doi:10.1200/JCO.19.03077
10. Pape M, Vissers PAJ, Dijksterhuis WPM, et al. Comparing treatment and outcomes in advanced esophageal, gastroesophageal junction, and gastric adenocarcinomas: a population-based study. *Therapeutic advances in medical oncology*. 2023;15:17588359231162576. doi:10.1177/17588359231162576
11. Chau I, Norman AR, Cunningham D, et al. The impact of primary tumour origins in patients with advanced oesophageal, oesophago-gastric junction and gastric adenocarcinoma - Individual patient data from 1775 patients in four randomised controlled trials. *Annals of Oncology*. 2009;20(5):885-891. doi:10.1093/annonc/mdn716
12. Shankaran V, Xiao H, Bertwistle D, et al. A Comparison of Real-World Treatment Patterns and Clinical Outcomes in Patients Receiving First-Line Therapy for Unresectable

Advanced Gastric or Gastroesophageal Junction Cancer Versus Esophageal Adenocarcinomas. *Advances in Therapy*. 2021;38(1):707-720. doi:10.1007/s12325-020-01567-9
13. Stroes CI, Schokker S, Khurshed M, et al. A phase Ib/II study of regorafenib and paclitaxel in patients with beyond first-line advanced esophagogastric carcinoma (REPEAT). *Therapeutic advances in medical oncology*. 2022;14:17588359221109196. doi:10.1177/17588359221109196
14. Stroes CI, Schokker S, Creemers A, et al. Phase II feasibility and biomarker study of neoadjuvant trastuzumab and pertuzumab with chemoradiotherapy for resectable human epidermal growth factor receptor 2-positive esophageal adenocarcinoma: Trap study. *Journal of Clinical Oncology*. 2020;38(5):462-471. doi:10.1200/JCO.19.01814
15. Stroes CI, Schokker S, Molenaar RJ, et al. A phase ii study demonstrates no feasibility of adjuvant treatment with six cycles of s-1 and oxaliplatin in resectable esophageal adenocarcinoma, with ercc1 as biomarker for response to sox. *Cancers*. 2021;13(4):1-15. doi:10.3390/cancers13040839
16. Van Den Ende T, De Clercq NC, Van Berge Henegouwen MI, et al. Neoadjuvant chemoradiotherapy combined with atezolizumab for resectable esophageal adenocarcinoma: A single-arm phase ii feasibility trial (PERFECT). *Clinical Cancer Research*. 2021;27(12):3351-3359. doi:10.1158/1078-0432.CCR-20-4443
17. Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. *Nephron*. 1976;16(1):31-41. doi:10.1159/000180580
18. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*. 2011;46(3):399-424. doi:10.1080/00273171.2011.568786
19. Austin PC, Stuart EA. Optimal full matching for survival outcomes: a method that merits more widespread use. *Statistics in Medicine*. 2015;34(30):3949-3967. doi:10.1002/sim.6602
20. Diamond A, Sekhon JS. Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics*. 2013;95(3):932-945. doi:10.1162/REST_a_00318
21. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Journal of Clinical Epidemiology*. 2008;61(4):344-349. doi:10.1016/j.jclinepi.2007.11.008
22. Valcarcel S, Gallego J, Jimenez-Fonseca P, et al. Does HER2 status influence in the benefit of ramucirumab and paclitaxel as second line treatment of advanced gastro-esophageal adenocarcinoma? Data from the AGAMENON-SEOM registry. *Journal of cancer research and clinical oncology*. Published online August 2022. doi:10.1007/s00432-022-04294-6
23. Creemers A, Ter Veer E, De Waal L, et al. Discordance in HER2 Status in Gastro-esophageal Adenocarcinomas: A Systematic Review and Meta-analysis. *Sci Rep*. 2017;7(1):1-10. doi:10.1038/s41598-017-03304-9
24. Creemers A, Ebbing EA, Hooijer GKJ, et al. The dynamics of HER2 status in esophageal adenocarcinoma. *Oncotarget*. 2018;9(42):26787-26799. doi:10.18632/oncotarget.25507
25. Sato Y, Okamoto K, Kawano Y, et al. Novel Biomarkers of Gastric Cancer: Current Research and Future Perspectives. *J Clin Med*. 2023;12(14). doi:10.3390/jcm12144646

26. Grabsch H, Sivakumar S, Gray S, Gabbert HE, Müller W. HER2 expression in gastric cancer: Rare, heterogeneous and of no prognostic value-conclusions from 924 cases of two independent series. *Cellular Oncology*. 2010;32(1-2):57-65. doi:10.3233/CLO-2009-0497
27. Pape M, Vissers PAJ, Dijksterhuis WPM, et al. Comparing treatment and outcomes in advanced esophageal, gastroesophageal junction, and gastric adenocarcinomas: a population-based study. *Therapeutic advances in medical oncology*. 2023;15:17588359231162576. doi:10.1177/17588359231162576

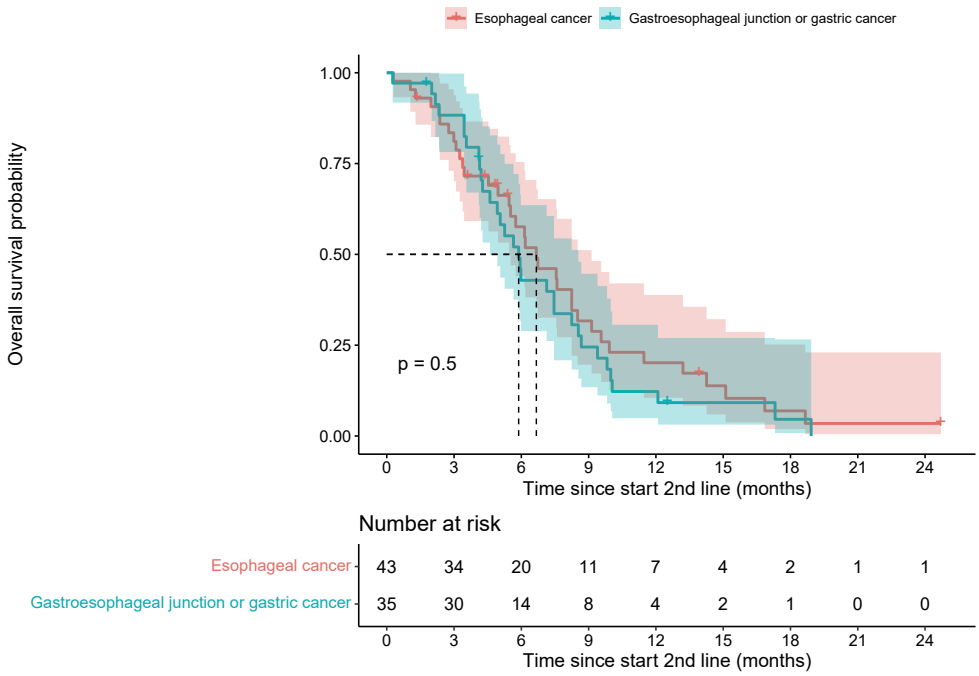
Supplementary table 1. additional disease and treatment characteristics after propensity score matching.

	DESTINY-Gastric 02 (n=58)		**NCR reference group** (n=78)	
	N	%	*N**	%
Differentiation grade				
- Well differentiated	1	1.7%	1	0.2%
- Moderately differentiated	31	53.4%	25.9	33.3%
- Poorly differentiated	10	17.2%	29.9	38.4%
- Unknown	16	27.6%	21.9	28.1%
Months between initial diagnosis and start study treatment (median, Inter-Quartile Range)	13.5 (9.8 – 19.8)		10.5 (7.6 – 15.1)	
Prior surgery of primary tumor				
- No	44	75.9%	78	100%
- Yes	14	24.1%	0	0%
Prior radiotherapy				
- Radiotherapy primary tumor	8	13.8%	20.6	26.4%
- Radiotherapy brain metastases	2	3.4%	10.3	13.2%
- Radiotherapy bone metastases	2	3.4%	10.4	13.3%
- Other prior radiotherapy	4	6.9%	1.3	1.7%
First-line systemic therapy				
- Trastuzumab	58	100%	78	100%
- Capecitabine	26	44.8%	68.3	87.5%
- Oxaliplatin	34	58.6%	64.4	82.5%
- Cisplatin	26	44.8%	13.6	17.5%
- Fluorouracil	38	65.5%	11.5	14.7%
- Other prior systemic therapy	28	48.3%	4.2	5.4%
First-line systemic therapy with immune-checkpoint inhibitor				
- No	53	91.4%	77.8	99.8%
- Yes	5	8.6%	0.2	0.2%

Supplementary table 1 (Continued). additional disease and treatment characteristics after propensity score matching.

Type of post study treatment				
- None	29	50.0%	45.6	58.5%
- Systemic therapy	29	50.%	17.1	21.9%
- Radiotherapy	2	3.4%	15.3	19.6%
- Surgery	1	1.7%	0	0%
Post study systemic treatment				
- Ramucirumab	17	29.3%	0	0%
- Irinotecan	5	8.6%	11.9	15.2%
- Trifluridine Tipiracil	1	1.7%	9.1	11.7%
- Pembrolizumab	6	10.3%	0	0%
- Tegafur Gimeracil Oteracil	0	0%	6.7	8.6%
- Fluorouracil	8	13.8%	2.7	3.4%
- Dacomitinib	0	0%	6.7	8.6%
- Carboplatin	1	1.7%	1.3	1.7%
- Other post study systemic treatment	19	32.8%	8.2	10.5%

** Number of patients for NCR reference group is based on weights applied after propensity score matching*



Supplementary figure 1. Kaplan-Meier curves comparing overall survival between the original primary tumor locations of the Ram+Pac patients

Chapter 7

Adjuvant nivolumab after chemoradiotherapy and resection for patients with esophageal cancer: a real-world matched comparison of overall survival

Rob H.A. Verhoeven*, Steven C. Kuijper*, Marije Slingerland, Bas Wijnhoven, Mark I. van Berge Henegouwen, Peter S.N. van Rossum, Sarah Derks, Bianca Mostert, Nadia Haj Mohammad, Hanneke W.M. van Laarhoven

* Shared first authorship

Manuscript submitted to the International Journal of Cancer

Abstract

Background. The Checkmate-577 trial showed a disease-free survival benefit for nivolumab compared to placebo in esophageal or gastroesophageal junction (GEJ) cancer patients with residual disease after neoadjuvant chemoradiotherapy (nCRT) and resection. Unfortunately, overall survival (OS) data have not yet been presented. The aim of this study was to evaluate OS of patients treated with or without adjuvant nivolumab in a nationwide real-world matched comparison.
Methods. For this study patients diagnosed with non-metastatic esophageal or GEJ cancer in 2020-2023 who had residual pathological disease after nCRT and resection were selected from the Netherlands Cancer Registry. 333 patients received treatment with adjuvant nivolumab. From the period before introduction of nivolumab, 486 patients were selected who received nCRT and resection alone. Propensity score trimming and nearest neighbor matching were used to create two well-balanced groups of 311 patients per treatment group.
Results. Median follow-up time was 24.4 months and 31.4 months for patients treated with and without adjuvant nivolumab, respectively. The 2-year OS was 66.8% (95% confidence interval [CI]: 61.6% - 72.44%) and 58.8% (95%CI: 53.5% - 64.5%) for the groups with and without nivolumab, respectively (log-rank p=0.024), hazard ratio: 0.75, 95%CI: 0.60-0.97 (p=0.024).
Conclusion. This matched real-world study showed an OS in favor of patients treated with nivolumab compared to patients without nivolumab. This represents the first report on an OS benefit in this setting. As follow-up and number of events are still limited, these analyses should be interpreted with caution and updated in the forthcoming years.

Introduction

Annually, more than 600,000 people are diagnosed with esophageal cancer worldwide, which is projected to increase to over 950,000 diagnoses in 2040.[1] Despite therapeutic improvements in the past decade, survival of patients with esophageal cancer remains low.[1, 2] After publication of the CROSS trial, neoadjuvant chemoradiotherapy (nCRT) followed by surgical resection has become a standard of care for patients with locally advanced resectable esophageal or gastroesophageal junction (GEJ) cancer in most Western countries.[3, 4] However, survival of patients with residual disease after treatment with nCRT is poor, underscoring the potential benefit of adjuvant treatment for these patients.[5, 6]

The randomized controlled CheckMate 577 trial investigated whether adjuvant immunotherapy with nivolumab for patients with esophageal or gastroesophageal junction (GEJ) cancer with pathologically confirmed residual disease after previous treatment with nCRT and resection would increase survival.[7] The trial showed a significantly longer median disease-free survival (DFS) of 22.4 months in the group who received adjuvant treatment with nivolumab compared to a median DFS of 11.0 months in patients who were treated with placebo (p<0.001).[7] Median distant metastasis-free survival (DMFS) was 28.3 and 17.6 months in the nivolumab and placebo groups, respectively.[7] No results on overall survival (OS) from the Checkmate 577 trial have been reported yet.

Although DFS may be a valid surrogate for OS in non-metastasized esophageal cancer,[8] significant differences in DFS do not always translate into significant and clinically meaningful OS differences. Also, findings from clinical trials may not consistently yield similar benefits in daily clinical practice.[9, 10] At this moment, it remains unclear whether adjuvant treatment with nivolumab will result in a clinically relevant OS benefit for patients with esophageal or GEJ cancer in routine clinical care, while nivolumab treatment can be associated with sometimes significant toxicity and costs.

Therefore, the aim of the current study was to conduct an initial assessment of OS among patients who received adjuvant nivolumab after nCRT and resection compared to patients who underwent nCRT and resection alone in a comprehensive nationwide real-world matched comparison.

Methods

Patient population

The Netherlands Cancer Registry (NCR) is a nationwide population-based database in which all Dutch inhabitants with a diagnosed malignant cancer are registered. On November 23, 2023 we extracted data of completely registered patients with non-metastatic esophageal, GEJ or cardia cancer diagnosed between 2020-2023 from the NCR. At that time, all Dutch patients with diagnosed esophageal, GEJ or cardia cancer in 2020 and 2021 and 72% and 3% of the patients diagnosed in 2022 and 2023, respectively, were completely registered in the NCR. Patients had to be treated with neoadjuvant chemoradiotherapy (nCRT) consisting of carboplatin and paclitaxel[11] and a surgical resection. In order to select a group of patients that would have been eligible for adjuvant nivolumab treatment, patients with complete pathological response (ypT0N0), distant metastases within 14 weeks after resection or patients deceased within 14 weeks after resection were excluded. As the EMA indication, in contrast to the CheckMate 577 trial, does not exclude patients with irradical resection (R1/R2), we have also not excluded this group from the current study.

Nivolumab was initially reimbursed for adjuvant treatment in esophageal or GEJ cancer patients in the Netherlands starting from January 1, 2022. However, reimbursement ceased as of January 1, 2024, following adjustments in Dutch guidelines related to reimbursement based on DFS outcomes. Therefore, all patients from the previously described group who started with adjuvant nivolumab treatment since January 1, 2022 were classified into the nivolumab group. To select a group of patients that would have been eligible for adjuvant treatment with nivolumab had it been available at the time, we selected all patients fulfilling the previously described criteria and who were treated with nCRT and a surgical resection prior to the November 1, 2021 but were not treated with nivolumab. Patients who underwent a surgical resection on or later than the November 1, 2021 were considered potential candidates for treatment with nivolumab, given that adjuvant nivolumab treatment typically commences more than 2 months after resection.

Information on vital status was obtained through annual linkage with the Dutch Personal Records Database and updated until January 31, 2025.

Statistical analyses

Propensity score matching was used to create two comparable and balanced groups of patients. Three medical oncologists (MS, NHM and HL) and two scientific researchers (RV and SK) selected a set of 18 patient (age, sex, Body Mass Index (BMI), Charlson comorbidity index, performance status prior to nCRT, and other malignancies), tumor (primary tumor location, histology, differentiation grade, ypT, and ypN) and treatment-related variables (number of chemotherapy cycles, radiotherapy dose, time between nCRT and resection, type of surgery, annual resection volume of hospital, surgical radicality, and post-operative complications) based on expert knowledge about factors that might influence the choice for whether or not starting adjuvant treatment with nivolumab, but also factors that might influence OS. These variables were used to estimate the propensity of receiving adjuvant nivolumab using a multivariable logistic regression model. Propensity score trimming was used to remove any non-overlapping propensity scores. After trimming, we iteratively tested five different matching methods (nearest neighbor, optimal matching, full matching, generic

matching and coarsened exact matching) and calipers (ranging from 0.01-0.20 times the standard deviation of the logit of the propensity score) without replacement, to evaluate which method resulted in the largest number of matched patients while maintaining a maximum allowable univariable covariate balance. We defined a standardized mean difference (SMD) of ≤0.1 as an appropriate univariable covariate balance.

To make the overall survival (OS) analyses as comparable as possible as the survival analyses of the CheckMate 577 trial, OS was calculated since start of adjuvant treatment. For patients who did not receive adjuvant treatment, this start date was obviously not existent. To ensure similar starting points to make a valid OS comparison, we based the starting point for patients not treated with adjuvant therapy on the number of days between resection and adjuvant treatment of the patient to whom they were matched in the propensity score matching. Kaplan-Meier curves and the univariable hazard ratio from Cox regression analysis were tested for significance using two-sided tests with an alpha level of 0.05. Subgroup analyses with univariable hazard ratios were conducted for histology, ypN status and surgical radicality, as the distribution of patients across histology and ypN status were considerably different in the current study compared to the CheckMate 577 study and as patients with R1/R2 surgical radicality may be treated in daily clinical practice, but were not included in the CheckMate 577 study. To investigate the potential differential effect for adjuvant nivolumab within these subgroups, we estimated and tested the hazard ratio of the interaction effect. Missing data was imputed with a random forest imputation using the missRanger package and the propensity score matching was performed using the MatchIt package.[12, 13] A sensitivity analyses was performed to assess the robustness of the treatment effect for unobserved confounding, for which the E-value was used.[14] The E-value reflects the magnitude that an unobserved confounder would need to have to negate the estimated treatment effect. The E-value in this analysis can be interpreted similarly to a hazard ratio. All analyses were performed by Steven Kuijper in R version 4.3.

Results

Prior to matching, 333 patients were identified who received adjuvant nivolumab after nCRT and resection and 485 patients were identified who received nCRT and resection alone (Table 1). Propensity score trimming resulted in the exclusion of 10 patients (1 with and 9 without nivolumab). Nearest neighbor matching with a caliper of 0.19 resulted in the largest number of matched patients (308 patients in each arm) while all univariable SMDs were ≤0.1 (Table 2). All patient, tumor, and treatment-related variables were well balanced in the matched groups.

The included patient population was predominantly male (83%) and was on average 65 years old. The primary tumor was located in the distal (or unknown) part of the esophagus in the majority of patients (84%) and most often was an adenocarcinoma (90%) (Table 2). With regards to the nCRT and surgery, 86% of the patients completed all 5 cycles of carboplatin and paclitaxel and 41.4Gy radiotherapy, while 71% of the patients underwent surgery in a hospital with ≥40 annual esophagectomies, 87% had a R0 resection and 87% had no grade 3 or 4 post-operative complication.

Median follow-up time was 24.4 (Interquartile range (IQR): 17.3 – 30.9) months and 31.4 (IQR: 15.3 – 43.4) months for patients treated with and without adjuvant nivolumab, respectively. The median survival times for group treated with and without adjuvant nivolumab was 36.2 months (95%CI: 33-NA) and 31.4 months (95%CI: 25.5-37.4), respectively. The 2-year OS was 66.8% (95%CI: 61.6%-72.4%) for the nivolumab group and 58.8% (95%CI: 53.5%-64.5%) for the group without nivolumab (log-rank p=0.024) (Figure 1, Table 3). The univariable hazard ratio for OS with nivolumab versus without nivolumab treatment was 0.75, 95%CI: 0.60-0.97, demonstrating favorable survival for patients treated with adjuvant nivolumab.

Subgroup analyses showed a significant OS difference for patient with adenocarcinoma treated with nivolumab with a hazard ratio of 0.76 (95%CI: 0.59-0.98, p=0.04) and patients who had irradical (R1/R2) surgical resection and treated with nivolumab with a hazard ratio of 0.46 (95%CI: 0.27-0.81, p=0.007). No significant difference was observed in any of the other subgroups (Figure 2). All interaction effects between adjuvant nivolumab and histology, surgical radicality and primary tumor location, were not significant. (Supplementary Table 1).

Sensitivity analysis of the observed treatment effect revealed that the E-value was 1.74.

Table 1. Patient characteristics of both treatment arms prior to imputation and propensity score matching.

	No Adjuvant nivolumab	Adjuvant nivolumab	SMD
n	485	333	
Age at resection (mean (SD))	65.80 (8.57)	65.32 (8.27)	0.058
Sex (%)			
Male	381 (78.6)	279 (83.8)	0.134
Female	104 (21.4)	54 (16.2)	
Body mass index (median (IQR))	25.85 (4.13)	26.65 (3.98)	0.197
Charlson Comorbidity Index (median (IQR))	0.0 (0.0-1.0)	0.0 (0.0-1.0)	0.118
WHO performance status prior to nCRT (%)			
0	274 (56.5)	212 (63.7)	0.202
1	160 (33.0)	96 (28.8)	
2	16 (3.3)	5 (1.5)	
3	1 (0.2)	0 (0.0)	
4	0 (0.0)	1 (0.3)	
Missing	34 (7.0)	19 (5.7)	
Other malignancies[a] (%)			
No	450 (92.8)	314 (94.3)	0.061
Yes	35 (7.2)	19 (5.7)	
Primary tumor location (%)			
Proximal or mid thoracic esophagus	38 (7.8)	24 (7.2)	0.100
Distal or unknown esophagus	391 (80.6)	280 (84.1)	
Junction or cardia	56 (11.5)	29 (8.7)	
Histology (%)			
Adenocarcinoma (intestinal type)	323 (66.6)	230 (69.1)	0.110
Adenocarcinoma (diffuse type)	47 (9.7)	35 (10.5)	
Adenocarcinoma (other/unknown type)	54 (11.1)	32 (9.6)	
Squamous cell carcinoma	59 (12.2)	33 (9.9)	
Other/Not otherwise specified	2 (0.4)	3 (0.9)	

Table 1 (Continued). Patient characteristics of both treatment arms prior to imputation and propensity score matching.

Differentiation grade (%)			
1-2	263 (54.2)	173 (52.0)	0.103
3-4	189 (39.0)	128 (38.4)	
Unknown	33 (6.8)	32 (9.6)	
ypT (%)			
ypT0	15 (3.1)	22 (6.6)	0.249
ypT 1	104 (21.4)	74 (22.2)	
ypT 2	103 (21.2)	80 (24.0)	
ypT 3	249 (51.3)	151 (45.3)	
ypT 4	5 (1.0)	5 (1.5)	
Missing	9 (1.9)	1 (0.3)	
ypN (%)			
ypN0	233 (48.0)	148 (44.4)	0.138
ypN 1	154 (31.8)	108 (32.4)	
ypN 2	61 (12.6)	53 (15.9)	
ypN 3	35 (7.2)	24 (7.2)	
Missing	2 (0.4)	0 (0.0)	
Number of cycles carboplatin and paclitaxel (median (IQR))	5 (5-5)	5 (5-5)	0.127
Radiotherapy dose (median (IQR))	41.4 (41.4-41.4)	41.4 (41.4-41.4)	0.047
Days between nCRT and resection (median (IQR))	75.0 (62.0-92.3)	77.0 (64.0-91.3)	0.045
Type of surgery (%)			
Transhiatal esophagectomy	49 (10.1)	31 (9.3)	0.216
Ivor-Lewis esophagectomy	320 (66.0)	229 (68.8)	
McKeown esophagectomy	90 (18.6)	55 (16.5)	
Other or unknown type of esophagectomy	15 (3.1)	17 (5.1)	
Gastrectomy	6 (1.2)	1 (0.3)	
Other	5 (1.0)	0 (0.0)	

Table 1 (Continued). Patient characteristics of both treatment arms prior to imputation and propensity score matching.

Annual resection volume of hospital (%)			
<20	25 (5.2)	14 (4.2)	0.169
20-29	59 (12.2)	34 (10.2)	
30-39	58 (12.0)	44 (13.2)	
>40	338 (69.7)	241 (72.4)	
Missing	5 (1.0)	0 (0.0)	
Surgical radicality (%)			
R0	394 (81.2)	271 (81.4)	0.052
R1/2	68 (14.0)	43 (12.9)	
Missing	23 (4.7)	19 (5.7)	
Post-operative complications (%)			
No grade 3/4 complication	377 (77.7)	276 (82.9)	0.188
Grade 3 complication	62 (12.8)	30 (9.0)	
Grade 4 complication	35 (7.2)	15 (4.5)	
Missing	11 (2.3)	12 (3.6)	

SMD = Standardized Mean Difference, IQR = Interquartile range

[a] Other malignancies 12 months prior to or 12 months after diagnosis esophageal or gastroesophageal junction cancer.

Table 2. Patient characteristics of both treatment arms after propensity score matching.

	No Adjuvant nivolumab	Adjuvant nivolumab	SMD
N	308	308	
Age at resection (mean (SD))	65.70 (8.60)	65.62 (8.09)	0.010
Sex (%)			
Male	257 (83.4)	255 (82.8)	0.017
Female	51 (16.6)	53 (17.2)	
Body mass index (median (IQR))	26.26 [23.98, 28.66]	26.03 [24.29, 28.53]	0.031
Charlson Comorbidity Index (median (IQR))	0.0 (0.0-1.0)	0.0 (0.0-1.0)	0.005
WHO performance status prior to nCRT (%)			
0	210 (68.2)	207 (67.2)	0.021
1	93 (30.2)	96 (31.2)	
2	5 (1.6)	5 (1.6)	
Other malignancies[a] (%)			
No	292 (94.8)	290 (94.2)	0.028
Yes	16 (5.2)	18 (5.8)	
Primary tumor location (%)			
Proximal or mid thoracic esophagus	17 (5.5)	20 (6.5)	0.066
Distal or unknown esophagus	258 (83.8)	260 (84.4)	
Junction or cardia	33 (10.7)	28 (9.1)	
Histology (%)			
Adenocarcinoma (intestinal type)	213 (69.2)	215 (69.8)	0.026
Adenocarcinoma (diffuse type)	33 (10.7)	31 (10.1)	
Adenocarcinoma (other/unknown type)	31 (10.1)	30 (9.7)	
Squamous cell carcinoma	29 (9.4)	30 (9.7)	
Other/Not otherwise specified	2 (0.6)	2 (0.6)	
Differentiation grade (%)			
1-2	167 (54.2)	165 (53.6)	0.024
3-4	117 (38.0)	117 (38.0)	
Unknown	24 (7.8)	26 (8.4)	

Table 2 (Continued). Patient characteristics of both treatment arms after propensity score matching.

ypT (%)			
yp0	12 (3.9)	12 (3.9)	0.058
yp1	70 (22.7)	70 (22.7)	
yp2	68 (22.1)	76 (24.7)	
yp3	156 (50.6)	147 (47.7)	
yp4	2 (0.6)	3 (1.0)	
ypN (%)			
yp0	147 (47.7)	143 (46.4)	0.058
yp1	91 (29.5)	98 (31.8)	
yp2	47 (15.3)	47 (15.3)	
yp3	23 (7.5)	20 (6.5)	
Number of cycles carboplatin and paclitaxel (median (IQR))	5 (5-5)	5 (5-5)	0.010
Radiotherapy dose (median (IQR))	41.4 (41.4-41.4)	41.4 (41.4-41.4)	0.005
Days between nCRT and resection (median (IQR))	76.00 [64.75, 94.25]	77.00 [64.00, 92.00]	0.014
Type of surgery (%)			
Transhiatal esophagectomy	31 (10.1)	30 (9.7)	0.094
Ivor-Lewis esophagectomy	215 (69.8)	211 (68.5)	
McKeown esophagectomy	48 (15.6)	52 (16.9)	
Other or unknown type of esophagectomy	13 (4.2)	15 (4.9)	
Annual resection volume of hospital (%)			
<20	17 (5.5)	13 (4.2)	0.069
20-29	31 (10.1)	32 (10.4)	
30-39	44 (14.3)	41 (13.3)	
>40	216 (70.1)	222 (72.1)	
Surgical radicallity (%)			
R0	275 (89.3)	267 (86.7)	0.080
R1/2	33 (10.7)	41 (13.3)	
Post-operative complications (%)			
No grade 3/4 complication	272 (88.3)	266 (86.4)	0.072
Grade 3 complication	22 (7.1)	28 (9.1)	
Grade 4 complication	14 (4.5)	14 (4.5)	

SMD = Standardized Mean Difference, IQR = Interquartile range

[a] Other malignancies 12 months prior to or 12 months after diagnosis esophageal or gastroesophageal junction cancer.

Table 3. Overall survival at 6, 12 and 18 months.

Time since start adjuvant nivolumab	No adjuvant nivolumab		Adjuvant nivolumab	
	Number at risk	OS (95% CI)	Number at risk	OS (95% CI)
6 months	283	91.9 (88.9 - 95.0)	286	92.9 (90.0 - 95.8)
12 months	250	81.2 (76.9 - 85.7)	263	85.4 (81.5 - 89.4)
18 months	218	70.8 (65.8 - 76.0)	224	77.2 (72.6 - 82.0)
24 months	181	58.8 (53.5 - 64.5)	162	66.8 (61.6 - 72.4)
30 months	157	51.0 (45.7 - 56.9)	88	61.7 (56.1 - 70.0)

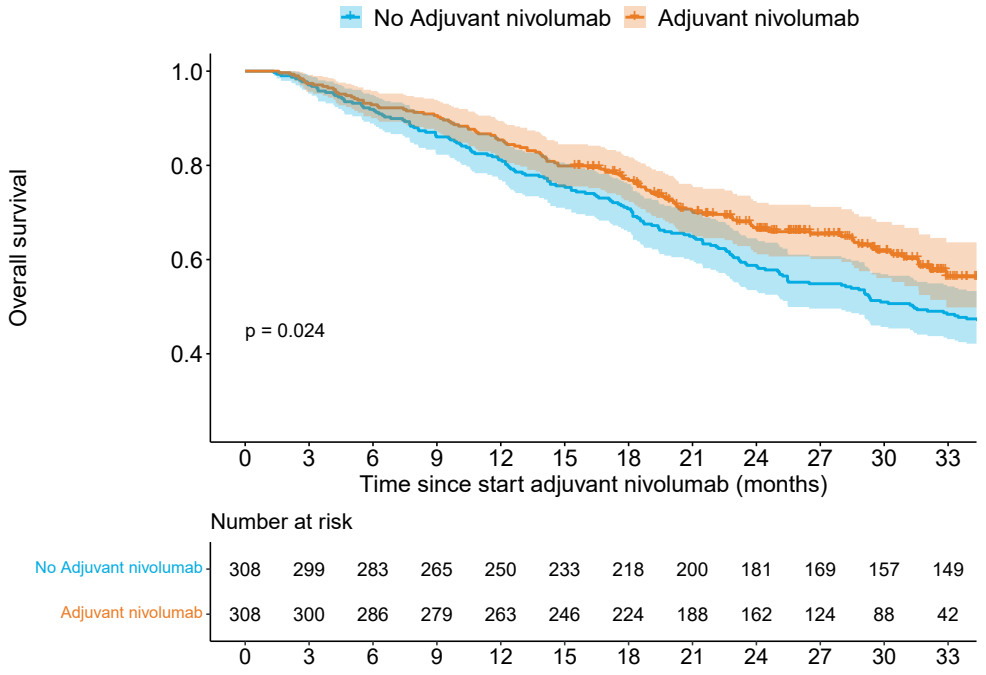


Figure 1. Overall survival analyses on adjuvant nivolumab versus no adjuvant nivolumab groups

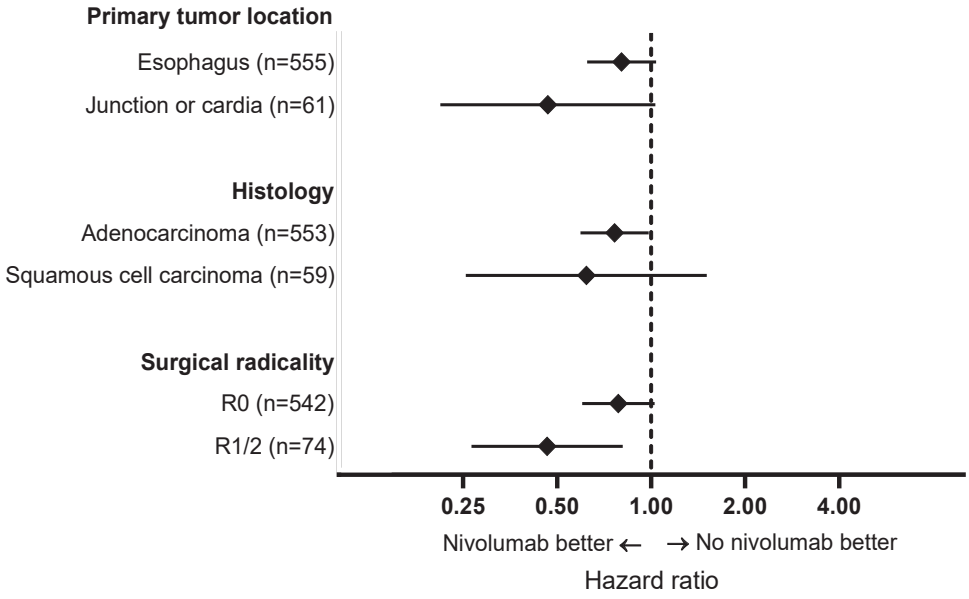


Figure 2. Subgroup analyses on overall survival of adjuvant nivolumab versus no adjuvant nivolumab

Discussion

In this nationwide real-world propensity score matched study with a median follow-up of 24 months of the nivolumab group, we found a significantly higher overall survival for patients with non-metastatic esophageal or gastroesophageal cancer who received adjuvant nivolumab compared to patients who did not receive adjuvant nivolumab after an incomplete pathologic response to nCRT.

For both the DFS and DMFS the curves of the patients treated with nivolumab and placebo in the CheckMate 577 trial started to separate at approximately 6 months, after which the difference between both groups remained relatively stable.[7] Although separation of the survival curve in the current study began around 6 months, the most significant divergence was noted around 15 to 18 months. Thus, in our current study, we observed a more gradual separation of the overall survival (OS) curve compared to the divergence seen in DFS/DFMS curves in the CheckMate 577 trial. This can primarily be attributed to the inclusion of post-recurrence survival times in the OS analysis, which are not accounted for in DFS/DMFS analyses.

While a direct comparison of the absolute values between the DFS in the CheckMate 577 trial and the OS in our present study is not feasible, due to the absence of routine imaging in the follow-up of the patients not treated with nivolumab and the absence of published OS data from the CheckMate 577 trial. There does exist a noteworthy discrepancy between the DFS in the CheckMate 577 trial and OS observed in the current study. The reported 1-year DFS in CheckMate 577 was roughly 60% in the adjuvant nivolumab arm and roughly 50% in the control arm. While the absolute difference in the 1-year OS of the current study was considerably smaller, as we observed an absolute difference of 4.2% between patients treated with nivolumab (85.4%) and not treated with nivolumab (81.2%). As collection of data on DFS requires additional manual data registration by data managers of the NCR, we currently lack comprehensive data on DFS for the patients in our study. Conversely, vital status information can be seamlessly obtained through an automated linkage with the Dutch Personal Records Database. Unfortunately, this means we do not have DFS data available for comparison with the DFS data from the CheckMate 577 trial at this time. However, we previously compared DFS between Dutch patients not treated with nivolumab and the placebo group of the CheckMate 577 trial[15] and showed that the median DFS of Dutch patients was considerably higher than the placebo group of the trial (19.7 versus 11.0 months, respectively). Both in our previous and the present study, it thus appears that the survival rate among patients treated with nCRT and surgical resection in the Netherlands is considerably higher than that observed in the placebo group of the CheckMate 577 trial. The explanation for the difference of the DFS could be in the absence of routine imaging in the follow-up of the patients not treated with nivolumab in Dutch daily clinical practice. Another possible explanation could be that the notably elevated survival rates in the Netherlands among patients treated with nCRT and resection may be attributed to the rigorous adherence to the CROSS protocol for nCRT (86% of the current study population completed all 5 cycles of carboplatin and paclitaxel and 41.4 Gy radiotherapy, while nCRT treatment was considerably more heterogenous in the CheckMate 577 trial, with only 71% receiving carboplatin and paclitaxel, number of chemotherapy cycles not reported and 63% of patients receiving a radiotherapy dose between 41.4 and 50.4 Gy) and centralized surgery in hospitals meeting minimum annual resection volumes (71% of current study population

underwent surgery in a hospital with ≥40 annual esophagectomies per year, annual hospital resection volumes were not reported for CheckMate 577 trial). This may suggest that the high survival baseline in the Netherlands could potentially reduce the impact of adjuvant nivolumab on survival. On the contrary, in other countries, the effectiveness of less-than-ideal nCRT and surgical procedures could potentially have been enhanced through adjuvant nivolumab treatment. Unfortunately, from Checkmate 577, no data are available on completion rates of neoadjuvant chemoradiotherapy treatment and quality of surgical procedures.

The Checkmate 577 highlighted a greater DFS benefit in the squamous cell carcinomas compared to the adenocarcinomas. Conversely, patients with GEJ cancer exhibited a lower DFS benefit of adjuvant treatment with nivolumab compared to patients with the primary tumor located in the esophagus. This is in contrast to our subgroup analyses in which we found a significant higher survival for patients with adenocarcinoma but not for patients with a squamous cell carcinoma and no indication that patients with a GEJ would have a lower benefit compared to patients with the primary tumor located in the esophagus. It is important to note that some groups within our study were relatively small, especially the squamous cell carcinoma group, and when combined with limited follow-up, this resulted in constrained statistical power. Consequently, the findings from the subgroup analyses should be interpreted with considerable caution.

In general we do not know whether treating earlier means treating better. Adjuvant treatment has the goal to cure patients with undetectable micro metastases. Monotherapy checkpoint inhibition in first line for metastatic gastroesophageal cancer failed to show survival gain.[16] In combination with chemotherapy checkpoint inhibition has shown improved survival outcomes in patients with PD-L1 positive tumors.[17, 18] At this time it is unknown whether adjuvant monotherapy with nivolumab to treat undetected micro metastases is beneficial in term of OS gain over starting treatment at time of diagnosis of metachronous metastatic disease. This study cannot answer that question, as patients in our study who were not treated with adjuvant nivolumab, most likely only received chemotherapy upon disease recurrence, as immune checkpoint inhibition was not reimbursed yet in the metastatic setting.

Several limitations of our study should be discussed. First, adjuvant treatment nivolumab was only available since 2022 in the Netherlands, which resulted in short follow-up compared to patients who received nCRT and resection. Potential long-term outcomes of patients treated with nivolumab in adjuvant setting in our real-world cohort could not yet be investigated. Although patients with esophageal cancer worldwide are increasingly being treated with adjuvant nivolumab based on the DFS outcomes from the CheckMate-577 trial, it is crucial to note that DFS serves as a surrogate endpoint for OS, which is the ultimate outcome of paramount importance to patients. Despite the significant improvement in DFS observed in the CheckMate-577 trial, there has been no published data to date on the potential overall survival benefit of adjuvant nivolumab treatment. Given this gap in the literature, even with our study's limited follow-up period, we believe it is essential to share our findings with the medical community. Providing insights into the possible overall survival benefits of adjuvant nivolumab treatment can help clinicians and patients make more informed decisions and may contribute to the ongoing evaluation and optimization of treatment strategies for esophageal cancer. Therefore, publishing our data, despite its limitations, is highly relevant and necessary to advance the understanding and management of esophageal cancer

Second, despite construction of two comparable treatment groups based on factors which are known to be related to survival outcomes and treatment allocation, patients were obtained from real-world observational data. Thus, despite our efforts, there might be some unmeasured residual confounding and results should be interpreted with caution. However, sensitivity analyses of the robustness of the treatment effects revealed that a potential unobserved confounder would need to have a HR of 1.74 or larger to negate the treatment effect. Given that we have accounted for most of the known clinical confounders, it is unlikely that there is an unobserved confounder with a HR of 1.74 or higher that would thereby entirely negate the treatment effect. Third, there was no information available on treatment of recurrent disease, patients treated without adjuvant nivolumab largely had their disease relapse in an era in which ICIs in the metastatic setting were not yet reimbursed, in contrast to those treated with adjuvant nivolumab. As the addition of ICI to chemotherapy increased survival in the biomarker-positive subgroups in the registration trials[17, 18], this may have positively impacted the post-recurrence survival of the adjuvant nivolumab group, and thus lead to imbalances in the two groups

A strength of our study is the thorough identification and inclusion of most of the pertinent measurable factors related to survival and treatment allocation, achieved through the utilization of expert knowledge from physicians and healthcare professionals. This approach allowed us to successfully balance treatment groups, thereby minimizing the risk of confounding as much as possible. Another strength of our study lays in the iterative comparisons between matching method to determine the matching method that yielded the highest number of matched patients and achieved optimal covariate balance.

In conclusion, our study revealed a significantly higher OS in patients with esophageal or GEJ cancer treated with adjuvant nivolumab after nCRT and resection compared to patients without adjuvant treatment in a real-world matched comparison after a median of 24 months of follow-up for the nivolumab group. Given the current limitations in follow-up duration, the relatively low number of events, the possible residual confounding and the possible differences in post-recurrence treatment, these findings should be approached with caution and warrant reevaluation in the coming years as additional data accrue.

References

1. Morgan E, Soerjomataram I, Rumgay H, Coleman HG, Thrift AP, Vignat J, Laversanne M, Ferlay J, Arnold M. The Global Landscape of Esophageal Squamous Cell Carcinoma and Esophageal Adenocarcinoma Incidence and Mortality in 2020 and Projections to 2040: New Estimates From GLOBOCAN 2020. Gastroenterology 2022;163:649-58 e2.
2. van Putten M, de Vos-Geelen J, Nieuwenhuijzen GAP, Siersema PD, Lemmens V, Rosman C, van der Sangen MJC, Verhoeven RHA. Long-term survival improvement in oesophageal cancer in the Netherlands. Eur J Cancer 2018;94:138-47.
3. Shapiro J, van Lanschot JJB, Hulshof M, van Hagen P, van Berge Henegouwen MI, Wijnhoven BPL, van Laarhoven HWM, Nieuwenhuijzen GAP, Hospers GAP, Bonenkamp JJ, Cuesta MA, Blaisse RJB, et al. Neoadjuvant chemoradiotherapy plus surgery versus surgery alone for oesophageal or junctional cancer (CROSS): long-term results of a randomised controlled trial. Lancet Oncol 2015;16:1090-98.
4. Obermannova R, Alsina M, Cervantes A, Leong T, Lordick F, Nilsson M, van Grieken NCT, Vogel A, Smyth EC, clinicalguidelines@esmo.org EGCEa. Oesophageal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol 2022;33:992-1004.
5. Al-Kaabi A, van der Post RS, van der Werf LR, Wijnhoven BPL, Rosman C, Hulshof M, van Laarhoven HWM, Verhoeven RHA, Siersema PD. Impact of pathological tumor response after CROSS neoadjuvant chemoradiotherapy followed by surgery on long-term outcome of esophageal cancer: a population-based study. Acta Oncol 2021;60:497-504.
6. Blum Murphy M, Xiao L, Patel VR, Maru DM, Correa AM, F GA, Liao Z, Komaki R, Lin SH, Skinner HD, Vaporciyan A, Walsh GL, et al. Pathological complete response in patients with esophageal cancer after the trimodality approach: The association with baseline variables and survival-The University of Texas MD Anderson Cancer Center experience. Cancer 2017;123:4106-13.
7. Kelly RJ, Ajani JA, Kuzdzal J, Zander T, Van Cutsem E, Piessen G, Mendez G, Feliciano J, Motoyama S, Lievre A, Uronis H, Elimova E, et al. Adjuvant Nivolumab in Resected Esophageal or Gastroesophageal Junction Cancer. N Engl J Med 2021;384:1191-203.
8. Ajani JA, D'Amico TA, Almhanna K, Bentrem DJ, Besh S, Chao J, Das P, Denlinger C, Fanta P, Fuchs CS, Gerdes H, Glasgow RE, et al. Esophageal and esophagogastric junction cancers, version 1.2015. J Natl Compr Canc Netw 2015;13:194-227.
9. Templeton AJ, Booth CM, Tannock IF. Informing Patients About Expected Outcomes: The Efficacy-Effectiveness Gap. J Clin Oncol 2020;38:1651-54.
10 . Kuijper SC, Pape M, Vissers PAJ, Jeene PM, Kouwenhoven EA, Haj Mohammad N, Ruurda JP, Sosef MN, Verhoeven RHA, van Laarhoven HWM. Trends in best-case, typical and worst-case survival scenarios of patients with non-metastatic esophagogastric cancer between 2006 and 2020: A population-based study. Int J Cancer 2023;153:33-43.
11. van Hagen P, Hulshof MC, van Lanschot JJ, Steyerberg EW, van Berge Henegouwen MI, Wijnhoven BP, Richel DJ, Nieuwenhuijzen GA, Hospers GA, Bonenkamp JJ, Cuesta MA, Blaisse RJ, et al. Preoperative chemoradiotherapy for esophageal or junctional cancer. N Engl J Med 2012;366:2074-84.
12. Ho D, Imai K, King G, Stuart E. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. Journal of Statistical Software 2011;42:1-28.
13. M M. missRanger: Fast Imputation of Missing Values. R package version 2.4.0, https://github.com/mayer79/missRanger., 2024.
14. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. Ann Intern Med 2017;167:268-74.
15. Pape M, Vissers PAJ, Beerepoot LV, van Berge Henegouwen MI, Lagarde SM, Mook S, Moehler M, van Laarhoven HWM, Verhoeven RHA. A population-based study in resected esophageal or gastroesophageal junction cancer aligned with CheckMate 577. Ther Adv Med Oncol 2022;14:17588359221075495.

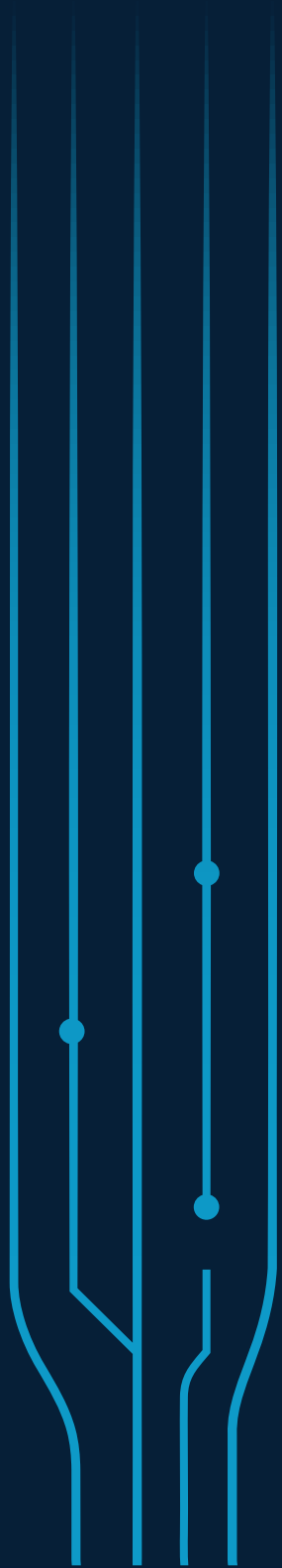
16. Shitara K, Van Cutsem E, Bang YJ, Fuchs C, Wyrwicz L, Lee KW, Kudaba I, Garrido M, Chung HC, Lee J, Castro HR, Mansoor W, et al. Efficacy and Safety of Pembrolizumab or Pembrolizumab Plus Chemotherapy vs Chemotherapy Alone for Patients With First-line, Advanced Gastric Cancer: The KEYNOTE-062 Phase 3 Randomized Clinical Trial. JAMA Oncol 2020;6:1571-80.
17. Sun JM, Shen L, Shah MA, Enzinger P, Adenis A, Doi T, Kojima T, Metges JP, Li Z, Kim SB, Cho BC, Mansoor W, et al. Pembrolizumab plus chemotherapy versus chemotherapy alone for first-line treatment of advanced oesophageal cancer (KEYNOTE-590): a randomised, placebo-controlled, phase 3 study. Lancet 2021;398:759-71.
18. Janjigian YY, Shitara K, Moehler M, Garrido M, Salman P, Shen L, Wyrwicz L, Yamaguchi K, Skoczylas T, Campos Bragagnoli A, Liu T, Schenker M, et al. First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): a randomised, open-label, phase 3 trial. Lancet 2021;398:27-40.

Supplementary table 1

Model	Parameter		HR (95%CI)	P value
Surgical radicality				
	Adjuvant nivolumab	No (REF)	-	-
		Yes	0.78 (0.62-1.02)	0.072
	Surgical radicality	0 (REF)	-	-
		1	2.58 (1.71-3.82)	<0.001
	Adjuvant nivolumab * Surgical radicality		0.65 (0.35-1.21)	0.173
Primary tumor location				
	Adjuvant nivolumab	No (REF)	-	-
		Yes	0.79 (0.54-1.02)	0.071
	Primary tumor location	Junction or cardia (REF)	-	-
		Esophagus	1.13 (0.71-1.80)	0.609
	Adjuvant nivolumab * Primary tumor location		0.63 (0.28-1.45)	0.280
Histology				
	Adjuvant nivolumab	No (REF)	-	-
		Yes	0.76 (0.60-0.97)	0.029
	Histology	Adenocarcinoma (REF)	-	-
		Squamous cell carcinoma	0.77 (0.44-1.33)	0.355
	Adjuvant nivolumab * Histology		0.90 (0.37-2.22)	0.825

Part IV

Prediction of survival outcomes



Chapter 8

Improving survival prediction of esophageal cancer patients treated with external beam radiotherapy for dysphagia

P.M. Jeene*, S.C. Kuijper*, H.G. van den Boorn, S.Y. El Sharouni, P.M Braam, V. Oppedijk, R.H.A.Verhoeven, M.C.C.M Hulshof$, H.W.M. van Laarhoven$

* shared first authorship
$ shared last authorship

Based on:
Jeene PM, Kuijper SC, van den Boorn HG, et al. Improving survival prediction of oesophageal cancer patients treated with external beam radiotherapy for dysphagia. Acta Oncol. 2022

Abstract

Introduction: The recent POLDER trial investigated the effects of external beam radiotherapy (EBRT) on dysphagia caused by incurable esophageal cancer. An estimated life expectancy of minimally three months was required for inclusion. However, nearly one-third of the included patients died within three months. The aim of this study was to investigate if the use of prediction models could have improved the physician's estimation of the patient's survival.
Methods: Data from the POLDER trial (N=110) were linked to the Netherlands Cancer Registry to retrieve patient, tumor and treatment characteristics. Two published prediction models (the SOURCE model and Steyerberg model) were used to predict 3-month survival for all patients included in the POLDER trial. Predicted survival probabilities were dichotomized and the accuracy, sensitivity, specificity and the area under the curve (AUC) were used to evaluate the predictive performance.
Results. The SOURCE and Steyerberg model had an accuracy of 79% and 64%, and an AUC of 0.76 and 0.60 (p = .017), respectively. The SOURCE model had higher specificity across survival cut-off probabilities, the Steyerberg model had a higher sensitivity beyond the survival probability cut-off of 0.7. Using optimal cut-off probabilities, SOURCE would have wrongfully included 16/110 patients into the POLDER and Steyerberg 34/110.
Conclusion: The SOURCE model was found to be a more useful decision aid than the Steyerberg model. Results showed that the SOURCE model could be used for three-month survival predictions for patients that are considered for palliative treatment of dysphagia caused by esophageal cancer in addition to clinicians' judgement.

Introduction

Esophageal cancer is the seventh most prevalent cancer in men and the thirteenth most commonly occurring cancer in women worldwide.[1]"ISSN":"1542-4863","P-MID":"30207593","abstract":"This article provides a status report on the global burden of cancer worldwide using the GLOBOCAN 2018 estimates of cancer incidence and mortality produced by the International Agency for Research on Cancer, with a focus on geographic variability across 20 world regions. There will be an estimated 18.1 million new cancer cases (17.0 million excluding nonmelanoma skin cancer Roughly a third of patients with esophageal cancer have a metastatic disease at primary diagnosis and the median overall survival (OS) ranges between eleven to fourteen months.[2–4]this study evaluated trends in stage distribution, treatment and survival of oesophageal cancer patients in the last 26 years in the Netherlands. Patients and methods: Patients with oesophageal cancer diagnosed in the period 1989–2014 were selected from the Netherlands Cancer Registry. Patients were divided into two groups: non-metastatic (M0 Around 80–90 % of esophageal cancer patients report dysphagia during their clinical course.[5,6]12 women In the recently published POLDER trial, it was shown that short course external beam radiotherapy (EBRT) was preferable over brachytherapy for palliation of dysphagia.[7]both between the original cohorts and between 1:1 propensity score–matched cohorts. The primary end point was an improvement of dysphagia at 3 months without reintervention. The secondary end points included toxicity and time-to-effect. Results: A total of 115 patients treated with EBRT and 93 patients who underwent brachytherapy were eligible for analysis. In the original cohorts, dysphagia improved after EBRT in 79% of patients compared with 64% after brachytherapy (p = 0.058

In the POLDER study an estimated life expectancy of minimally three months was required for inclusion. However, about one-third of patients survived shorter. Survival estimates were based on clinical judgement of the treating physician. To aid in predicting survival for esophageal cancer patients, various prediction models are available.[8,9]however, have mostly been developed for survival prediction after surgery (ie, when treatment has already been completed In the SIREC trial published in 2004, a total of 209 patients with dysphagia caused by incurable esophageal cancer were randomized between intraluminal brachytherapy and stent placement. Based on these patients, a prediction tool for survival was developed by Steyerberg and colleagues.[9]however, only occurred after a relatively long survival. The objective is to develop a model that distinguishes patients with a poor prognosis from those with a relatively good prognosis. Methods: Survival was analyzed with Cox regression analysis. Dysphagia-adjusted survival (alive with no or mild dysphagia More recently, the SOURCE prediction model was published based on 3271 metastatic esophageal cancer patients.[8] Results show that the SOURCE model for metastatic esophageal cancer patients demonstrates fair discrimination and good calibration. Although the SOURCE model is more recent and based on more patients, the Steyerberg model is based on patients treated for dysphagia only, and thus perhaps a better representative for this specific group.

The aim of the current study was to evaluate if the use of prediction tools would have improved survival prediction compared to clinical judgement in patients treated in the POLDER trial. In addition, the model's performances of predicting survival at three months will be used to determine which model is more suitable as a tool to determine which patients are eligible for EBRT treatment.

Methods

Study sample

This study is performed according to the TRIPOD checklist for the validation of prediction models[10] to inform their decision making. However, the overwhelming evidence shows that the quality of reporting of prediction model studies is poor. Only with full and clear reporting of information on all aspects of a prediction model can risk of bias and potential usefulness of prediction models be adequately assessed. The Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD The data used in the study originated from the POLDER study, a Dutch multicenter prospective cohort study of patients with metastasized or otherwise incurable esophageal cancer requiring palliation of dysphagia between 2016 and 2019.[7]both between the original cohorts and between 1:1 propensity score–matched cohorts. The primary end point was an improvement of dysphagia at 3 months without reintervention. The secondary end points included toxicity and time-to-effect. Results: A total of 115 patients treated with EBRT and 93 patients who underwent brachytherapy were eligible for analysis. In the original cohorts, dysphagia improved after EBRT in 79% of patients compared with 64% after brachytherapy (p = 0.058 The sample consisted of (N=115) patients with incurable esophageal $T_{2, 3, 4A, 4B, IS,X}$ $N_{0-3}M_{0-1}$ that were treated with EBRT in five fractions of 4 Gy. Patients with non-metastatic disease in poor condition and for whom treatment with curative intent was not deemed feasible were also included in the POLDER trial. The data was linked to the Netherlands Cancer Registry (NCR), a nation-wide database containing tumor, patient and treatment characteristics of patients diagnosed with cancer. Data from the NCR were used to retrieve the characteristics that were required for the prediction models but were not recorded in the POLDER study.

One patient was excluded from the analyses because this patient's T-stage was in situ. Four patients were excluded because the date of the start of their treatment was missing, thus leaving 110 patients for the analyses.

Furthermore, only the weight, but not the height of patients could be obtained due to practical constraints. To approximate patients' BMI, we used the average height of Dutch men and women as reported by Statistics Netherlands (CBS). For men this was 180.8 cm and for women this was 167.7 cm.[11] Additional sensitivity analyses were performed to investigate the effect of patient with a height of -10 cm and +10 cm.

Prediction models

The published SOURCE and Steyerberg prediction models were retrospectively used to predict three month survival probabilities of patients treated in the POLDER trial. The SOURCE prediction model was recently developed for patients suffering from metastatic or potentially curable esophageal or stomach cancer.[8] Since most patients treated in the POLDER trial had metastatic esophageal cancer (87%), for the current study, the model for patients with metastatic esophageal cancer was used. For the remaining 13% of patients without distant metastases, the general condition was considered too poor for curative or more radical treatment. Therefore, the model for metastatic patients was also used to predict survival for these 13% of patients. The predictors in the SOURCE model included the following patient characteristics: age, sex, body mass index, performance status, Albumin, LDH, Creatinine, type of treatment and the following tumor characteristics: cT and cN stage, diffe-

rentiation grade, HER2 status, only distant lymph node metastases, peritoneal metastases, and number of metastatic sites.

The Steyerberg prediction model has been developed prior to the SOURCE prediction model and was intended to predict survival for esophageal patients treated for dysphagia.[9]however, only occurred after a relatively long survival. The objective is to develop a model that distinguishes patients with a poor prognosis from those with a relatively good prognosis. Methods: Survival was analyzed with Cox regression analysis. Dysphagia-adjusted survival (alive with no or mild dysphagia The predictors in the Steyerberg model differ from the SOURCE model. These predictors include the following patient characteristics: sex, age (per ten years), WHO performance status and tumor length. In this analysis, we fitted the cox regression model with the reported model's coefficients to the data. As the baseline hazard function was not reported by Steyerberg and colleagues, we estimated the baseline hazard on the POLDER data on the assumption that patients in the POLDER study had similar characteristics as patients in the SIREC trial on which the Steyerberg model was developed.

As the primary aim of this study was to investigate to what extent both models would perform better than the clinician's survival predictions, the main focus was on predicting survival at three months. Furthermore, threshold probabilities were used to evaluate if a patient was predicted to be deceased or alive at three months: the survival cut-off probability. Since the choice of such a cut-off probability is arbitrary and was unknown at the time of patient inclusion, we used multiple cut-off probabilities to evaluate the models. For example, if the cut-off probability was at 0.7, we assumed that patients with lower and higher values than 0.7 were predicted to be deceased and alive, respectively.

Statistical Analyses

Three-month survival probabilities were computed with the published model coefficients of the SOURCE and Steyerberg models using the Prediction Error Curves for Risk Prediction Models in Survival Analysis (PEC) package for R.[12] For each model, the area under the curve (AUC) was calculated and the difference of the AUC between the SOURCE and Steyerberg model was tested for significance with a two-sided DeLong test with an alpha level of 0.05. The accuracy (the percentage of correct decisions), the sensitivity and specificity were calculated to evaluate the models' predictions. To estimate the optimal cut-off survival probability Youden's-index was used.[13] This method is developed to determine the optimal balance between sensitivity and specificity. Furthermore, the sensitivity and specificity for all cut-off scores between 0.5 and 1.0 were plotted and smoothed using locally estimated scatterplot smoothing (LOESS). All analyses were performed in R version 4.0.3.[14]

Robustness

Missing data on the variables in the dataset were imputed via random forest imputation using the missForest package in R.[15] Missing forest imputation with missForest can handle missing values in data with different types of variables, complex interactions between variables and has been found to outperform other imputation methods such as multivariate imputation by chained equations in biological and medical datasets.[15] In addition, the missForest algorithm also provides an out of bag error estimate to evaluate the imputation error.

This error is estimated by iteratively training the algorithm on a bootstrapped sample and testing on a number of complete cases that are not in the bootstrapped sample. The difference between observed and expected is defined as the out of bag error estimate. Furthermore, to evaluate optimism of estimating the optimal cut-off and testing the model on the same data, we used 20 repeated five-fold cross validations.[16] This emulates the procedure of validating the cut-off probability with new data. For each repetition the data were randomly partitioned into five folds. Four folds were used for determining the optimal cut-off probability an one fold was used for testing. This was repeated five times so that every patient was in the training and test fold at least one. The process of five-fold cross validation was repeated 20 times to increase stability of the estimates. The mean accuracy with 95% confidence interval was evaluated.

Results

Characteristics of patients treated in the POLDER study are shown in Table 1. Three months after the onset of treatment, 35 patients were deceased. The AUC of the SOURCE and Steyerberg models were 0.76 and 0.60, respectively (p = .017). Based on Youden's index, the optimal survival probability cut-off was 0.70 and 0.87 for the SOURCE and Steyerberg model respectively. Using 0.70 as a cut-off the accuracy of the SOURCE model was 79%, the sensitivity was 93% and the specificity was 54%. Using 0.87 as a cut-off the accuracy of the Steyerberg model was 64%, the sensitivity was 67% and the specificity was 51%.

Table 2 shows how many patients would have been justly and unjustly included if the decision was only based on predicted survival using ideal cut-off probabilities. Retrospectively, SOURCE would have wrongfully included in total 16 patients as opposed to 35 patients that were wrongfully included in the POLDER trial. Steyerberg would have wrongfully included 34 patients.

Extending beyond the optimal survival cut-offs, the general trend was that the sensitivity of the SOURCE model was lower compared to the Steyerberg model (Figure 1) across cut-off probabilities higher than 0.7. The SOURCE model's specificity was higher than the Steyerberg model across all cut-off probabilities. Additionally, Figure 1 can be used to investigate the sensitivity and specificity given a different cut-off probability.

A nomogram of the SOURCE model (Figure 2) can be used to obtain the three month survival probability.

Robustness

After 20 five-fold cross validations, the accuracy of the SOURCE model was 0.74 (0.54-0.94) and 0.53 (0.31-0.75) for the Steyerberg model. Thus the optimism of retrospectively estimating the cut-off probability of both models were 5% and 11% for SOURCE and Steyerberg, respectively.

Furthermore, the imputation error and the effect of varying the average patient height for the BMI calculation was separately tested. The normalized root mean squared error, which reflects the imputation error of continuous variables, was $2.52*10^{-7}$. The proportion of falsely classified entries, which reflects the imputation error of categorical variables, was 0.15. Values close to zero indicate low imputation error whereas values near one indicate high imputation error. Thus, the overall imputation error was low.

In additional sensitivity analyses, varying patients' average height with -10 cm and +10 cm had no effect on overall results (Supplementary Table 1). Therefore, for all analyses heights of 180.8 cm for men and 167.7 cm for women were used to calculate BMI.

Table 1. Descriptive statistics of the included patients.

	Overall
N	110
Tumor length > 10 cm (%)	6 (5.5)
Peritoneal metastases (%)	
No	87 (79.1)
Yes	2 (1.8)
Missing	21 (19.1)
Age (mean (SD))	71.36 (9.36)
Sex = Female (%)	25 (22.7)
BMI (mean (SD))	24.11 (4.66)
WHO performance status (%)	
0	20 (18.2)
1	37 (33.6)
2	20 (18.2)
3+	6 (5.5)
Missing	27 (24.5)
Albumine (mean (SD))	36.70 (5.27)
LDH (mean (SD))	227.91 (121.32)
Creatinine (mean (SD))	86.22 (27.58)
Clinical M-stage = 1 (%)	89 (80.9)
Clinical T-stage (%)	
2	37 (33.6)
3	48 (43.6)
4	9 (8.2)
X	16 (14.5)
Clinical N-stage (%)	
0	12 (10.9)
1	38 (34.5)
2	47 (42.7)
3	13 (11.8)

Table 1 (Continued). Descriptive statistics of the included patients.

Differentation grade (%)	
G1	3 (2.7)
G2	24 (21.8)
G3	43 (39.1)
Missing	40 (36.4)
HER2 status (%)	
Negative	43 (39.1)
Positive	10 (9.1)
Missing	57 (51.8)
Only lymph node metastases (%)	
No	69 (62.7)
Yes	20 (18.2)
Missing	21 (19.1)
Number of metastases (%)	
0	21 (19.1)
1	45 (40.9)
2	27 (24.5)
3	17 (15.5)
First line treatment (%)	
Chemoradiation	31 (28.2)
Chemotherapy	4 (3.6)
Other	1 (0.9)
Radiotherapy metastases	1 (0.9)
Radiotherapy of primary tumor	73 (66.4)

Table 2. Correct and incorrect in- and exclusions based on optimal cut-off probabilities.

	POLDER trial (N=110)	SOURCE (cut-off probability = 0.70)	Steyerberg (cut-off probability = 0.87)
Wrongfully included	35	16	34
Wrongfully excluded	-	7	6
Correctly included	75	68	69
Correctly excluded	-	19	1

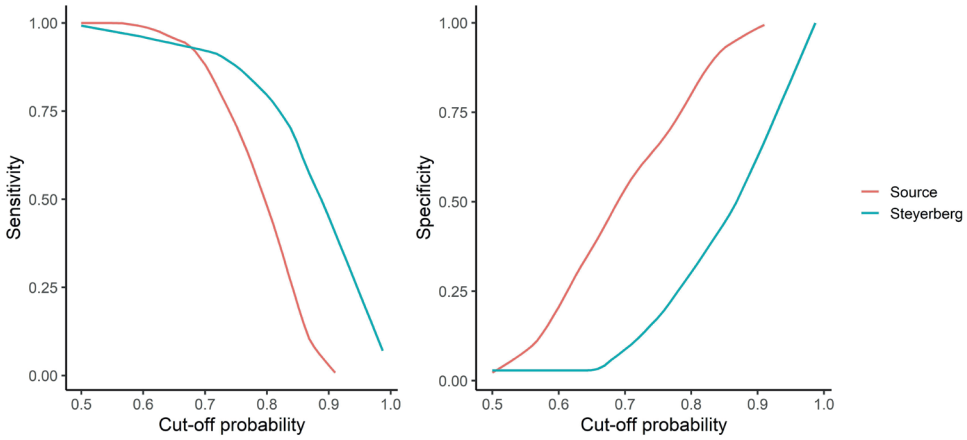


Figure 1. Sensitivity and specificity of the SOURCE and Steyerberg model as function of cut-off survival probability

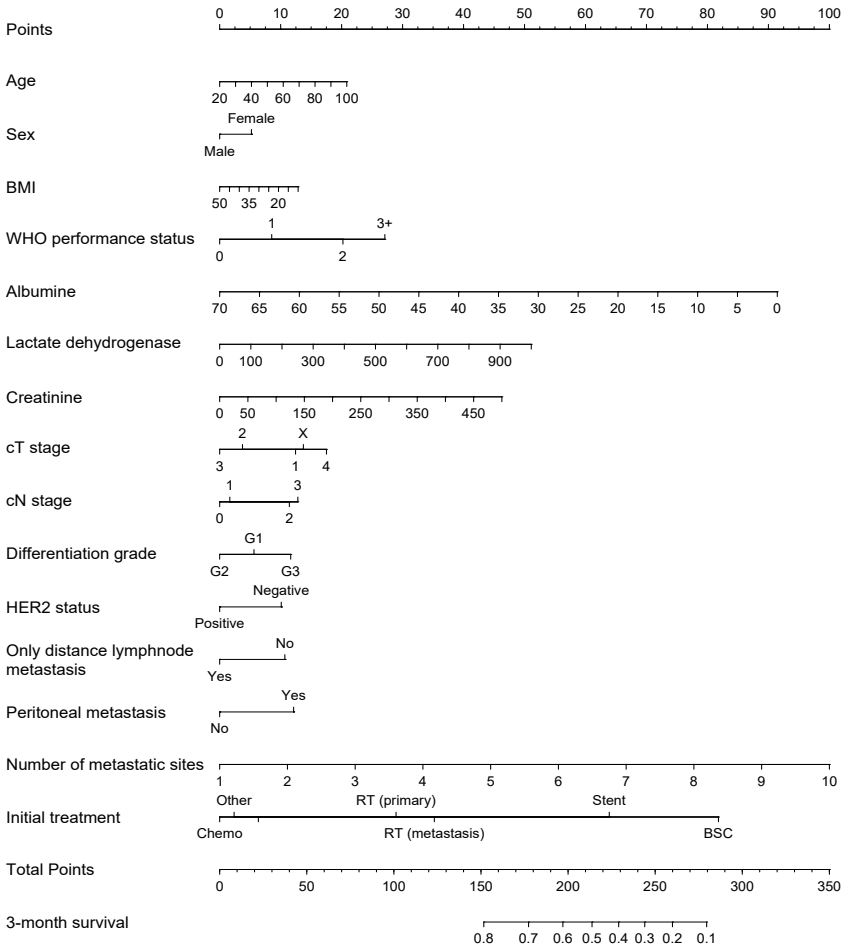


Figure 2. Nomogram for 3-month survival of the SOURCE prediction model for metastatic patients. The SOURCE prediction model for patients with metastatic esophageal cancer was developed on 3271 patients.[8]

Discussion

The POLDER trial investigated the effects of EBRT on dysphagia caused by incurable esophageal cancer.[7]both between the original cohorts and between 1:1 propensity score–matched cohorts. The primary end point was an improvement of dysphagia at 3 months without reintervention. The secondary end points included toxicity and time-to-effect. Results: A total of 115 patients treated with EBRT and 93 patients who underwent brachytherapy were eligible for analysis. In the original cohorts, dysphagia improved after EBRT in 79% of patients compared with 64% after brachytherapy (p = 0.058 Both the SOURCE and Steyerberg prediction models might have improved survival predictions for these patients in addition to the clinicians' judgement, albeit with different predictive characteristics.

Overall, the SOURCE model displayed a higher accuracy than the Steyerberg model. Furthermore, the SOURCE model was a more specific prediction model whereas the Steyerberg model was more sensitive. This implies that when the SOURCE model would have been used as a decision aid, less patients in the POLDER trial would have been included that did not meet the criteria of surviving three months. On the other hand, this also implies that if the SOURCE model was used, some patients would not have been included but did survive three months. Based on the retrospectively estimated survival-cutoff scores, the SOURCE model outperformed the Steyerberg model because fewer patients would have been incorrectly included. Based on the prediction models only, SOURCE would have incorrectly included 16 patients and Steyerberg 34 patients.

Clinical implications

There is considerable treatment variation for patients with esophageal cancer in the palliative setting [3,17,18] but experience with its administration may be limited and vary among hospitals. In a population-based study, we analysed the association between hospital systemic treatment volume and administration of beyond first-line treatment in oesophagogastric adenocarcinoma, as well as the effect on overall survival (OS For example, a significant hospital variation in treating patients with either EBRT or stent placement has been observed.[19] In daily practice, when the patient is considered for stent placement to relieve dysphagia, the SOURCE model can be used to determine whether EBRT treatment would be a good alternative. SOURCE outperforms the Steyerberg model in filtering patients that are likely to survive three months and as such identify patients for whom EBRT would be a good treatment option. In this scenario, the Steyerberg model would incorrectly select more patients for EBRT treatment.

For relieving dysphagia, treating patients with EBRT when they will not survive three months is undesirable, since the effect of EBRT on dysphagia relief is not immediate and patients will thus potentially not experience its effect.[7]both between the original cohorts and between 1:1 propensity score–matched cohorts. The primary end point was an improvement of dysphagia at 3 months without reintervention. The secondary end points included toxicity and time-to-effect. Results: A total of 115 patients treated with EBRT and 93 patients who underwent brachytherapy were eligible for analysis. In the original cohorts, dysphagia improved after EBRT in 79% of patients compared with 64% after brachytherapy (p = 0.058 These patients will likely benefit more from stent placement, which relieves dysphagia more

rapidly.[20] Therefore, for patients that are likely to die soon or patients for whom it is unclear whether they will survive three months, stent placement is potentially a better option.

Inherent to SOURCE's conservative survival predictions, some patients will not receive EBRT treatment when they are alive after three months. This is the cost of using conservative survival predictions. However, making this error has less severe consequences for patients since these patients may have experienced rapid dysphagia relief and retreatment with stent replacement can be performed when necessary.[21] Alternatively, stent removal and subsequent EBRT can be considered. Nevertheless, dysphagia recurrence after stent placement is high (31%) and possibly negatively impacts quality of life.[22] For practical application of the SOURCE model, the 3-month survival nomogram for patients with metastatic esophageal cancer (Figure 2) can be used.

For clinical application of the SOURCE and Steyerberg prediction models, the optimal cut-off probability can be used as this maximizes the model accuracy, However, Figure 1 can also be used to visually inspect and select a different cut-off probability given desired sensitivities and specificities, as an alternative to the cut-off point based on the Youden index.

Strengths and limitations

This study has a number of strengths. First, it concerns a specific patient group in which research is rarely performed. Also, the data of this study were based on recent patient data. Moreover, multiple steps were undertaken to evaluate robustness of results. We conducted a repeated cross-validation to evaluate the optimism of estimating the cut-off probability and testing the model on the same data with that cut-off. To improve stability of the estimates we repeated 5-fold cross-validation 20 times, which showed that overfitting of the cut-off probability was fairly low. Furthermore, even though we imputed missing data and calculated BMI using average heights' of men and women, analyses showed that these missing data methods did not affect our conclusions.

A limitation of this study was that only treated patients were included in the POLDER study and thus the analysis. Unfortunately, data of excluded patients were not available. A second limitation was that we could not use the baseline survival hazard of the Steyerberg model because this was not reported. Alternatively, we used the baseline survival hazard of the patients of the POLDER trial. Patients in the POLDER trial were similar to patients in the SIREC trial on which the Steyerberg model was developed as the inclusions criteria were the same. [23] We therefore assumed similarity of their baseline survival hazard. Furthermore, patients in the POLDER trial were registered in the NCR and as such used to develop the SOURCE model. Overfitting was a potential hazard, however the patients in the POLDER trial were only 3% of all patients used for fitting the SOURCE model. Thus, the risk of overfitting was relatively low.

Conclusion

Both the SOURCE and Steyerberg models could have improved three-month survival predictions in addition to clinical judgement alone for patients with incurable esophageal cancer experiencing dysphagia. The SOURCE model was found to be a more useful decision aid than the Steyerberg model as it was more accurate, albeit slightly more conservative. Results showed that the SOURCE model could be used for patients that are considered for palliative treatment of dysphagia caused by esophageal cancer.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394-424. doi:10.3322/caac.21492
2. van Putten M, de Vos-Geelen J, Nieuwenhuijzen GAP, et al. Long-term survival improvement in oesophageal cancer in the Netherlands. *Eur J Cancer*. 2018;94:138-147. doi:10.1016/j.ejca.2018.02.025
3. Dijksterhuis WPM, Verhoeven RHA, Slingerland M, et al. Heterogeneity of first-line palliative systemic treatment in synchronous metastatic esophagogastric cancer patients: A real-world evidence study. *Int J Cancer*. 2020;146(7):1889-1901. doi:10.1002/ijc.32580
4. Janjigian YY, Shitara K, Moehler M, et al. First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): a randomised, open-label, phase 3 trial. *Lancet*. 2021;398(10294):27-40. doi:10.1016/S0140-6736(21)00797-2
5. Watkinson AF, Ellul J, Entwisle K, Mason RC, Adam A. Esophageal carcinoma: Initial results of palliative treatment with covered self-expanding endoprostheses. *Radiology*. 1995;195(3):821-827. doi:10.1148/radiology.195.3.7538682
6. Dai Y, Li C, Xie Y, et al. Interventions for dysphagia in oesophageal cancer. *Cochrane Database Syst Rev*. 2014;2014(10). doi:10.1002/14651858.CD005048.pub4
7. Jeene PM, Vermeulen BD, Rozema T, et al. Short-Course External Beam Radiotherapy Versus Brachytherapy for Palliation of Dysphagia in Esophageal Cancer: A Matched Comparison of Two Prospective Trials. *J Thorac Oncol*. 2020;15(8):1361-1368. doi:10.1016/j.jtho.2020.04.032
8. van den Boorn HG, Abu-Hanna A, Haj Mohammad N, et al. SOURCE: Prediction Models for Overall Survival in Patients With Metastatic and Potentially Curable Esophageal and Gastric Cancer. *J Natl Compr Canc Netw*. 2021;19(4):403-410. doi:10.6004/jnccn.2020.7631
9. Steyerberg EW, Homs MYV, Stokvis A, Essink-Bot ML, Siersema PD. Stent placement or brachytherapy for palliation of dysphagia from esophageal cancer: A prognostic model to guide treatment selection. *Gastrointest Endosc*. 2005;62(3):333-340. doi:10.1016/S0016-5107(05)01587-7
10. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *BMC Med*. 2015;13(1). doi:10.1186/s12916-014-0241-z
11. CBS. Lengte en gewicht van personen, ondergewicht en overgewicht; vanaf 1981. Centraal Bureau voor de Statistiek. https://www.cbs.nl/nl-nl/cijfers/detail/81565NED?dl=35805. Published 2021.
12. Gerds TA. Prediction Error Curves for Risk Prediction Models in Survival. 2022. https://cran.r-project.org/web/packages/pec/.
13. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3(1):32-35. doi:10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3
14. R Core Team. R: A Language and Environment for Statistical Computing. 2021. https://www.r-project.org/.
15. Stekhoven DJ, Bühlmann P. Missforest-Non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28(1):112-118. doi:10.1093/bioinformatics/btr597
16. Kim JH. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Comput Stat Data Anal*. 2009;53(11):3735-3745. doi:10.1016/j.csda.2009.04.009
17. Dijksterhuis WPM, Verhoeven RHA, Pape M, et al. Hospital volume and beyond first-line palliative systemic treatment in metastatic oesophagogastric adenocarcinoma: A population-based study. *Eur J Cancer*. 2020;139:107-118. doi:10.1016/j.ejca.2020.08.010
18. Dijksterhuis WPM, Verhoeven RHA, Meijer SL, et al. Increased assessment of HER2 in metastatic gastroesophageal cancer patients: a nationwide population-based cohort study. *Gastric Cancer*. 2020;23(4):579-590. doi:10.1007/s10120-020-01039-7

19. Opstelten JL, de Wijkerslooth LRH, Leenders M, et al. Variation in palliative care of esophageal cancer in clinical practice: Factors associated with treatment decisions. *Dis Esophagus*. 2017;30(2). doi:10.1111/dote.12478
20. van der Bogt RD, Vermeulen BD, Reijm AN, Siersema PD, Spaander MCW. Palliation of dysphagia. *Best Pract Res Clin Gastroenterol*. 2018;36-37:97-103. doi:10.1016/j.bpg.2018.11.010
21. Homs MYV, Steyerberg EW, Kuipers EJ, et al. Causes and treatment of recurrent dysphagia after self-expanding metal stent placement for palliation of esophageal carcinoma. *Endoscopy*. 2004;36(10):880-886. doi:10.1055/s-2004-825855
22. Reijm AN, Didden P, Schelling SJC, Siersema PD, Bruno MJ, Spaander MCW. Self-expandable metal stent placement for malignant esophageal strictures - Changes in clinical outcomes over time. *Endoscopy*. 2018;51(1):18-29. doi:10.1055/a-0644-2495
23. Homs MYV, Steyerberg EW, Eijkenboom WMH, et al. Single-dose brachytherapy versus metal stent placement for the palliation of dysphagia from oesophageal cancer: Multicentre randomised trial. *Lancet*. 2004;364(9444):1497-1504. doi:10.1016/S0140-6736(04)17272-3

Supplementary Table 1. Sensitivity analysis for different average heights for males and females. Average height was used to calculate BMI which is used in the SOURCE model.

		Cut-off probability	Youden-index	Sensitivity	Specificity	AUC
Average height -10cm	SOUCE	0.707	0.478	0.907	0.571	0.764
Average height +10cm	SOUCE	0.699	0.478	0.907	0.571	0.762

Chapter 9

SOURCE beyond first-line: A survival prediction model for patients with metastatic esophagogastric adenocarcinoma after failure of first-line palliative systemic therapy

Steven C. Kuijper, Marieke Pape, Nadia Haj Mohammad, Theo van Voorthuizen, Rob H.A. Verhoeven, Hanneke W.M. van Laarhoven

Based on:
Kuijper SC, Pape M, Haj Mohammad N, van Voorthuizen T, Verhoeven RHA, van Laarhoven HWM. SOURCE beyond first-line: A survival prediction model for patients with metastatic esophagogastric adenocarcinoma after failure of first-line palliative systemic therapy. *Int J Cancer*. 2023

Abstract

Background. Prior models have been developed to predict survival for patients with esophagogastric cancer undergoing curative treatment or first-line chemotherapy (SOURCE models). Comprehensive clinical prediction models for patients with esophagogastric cancer who will receive second-line chemotherapy or best supportive care are currently lacking. The aim of this study was to develop and internally validate a new clinical prediction model, called SOURCE beyond first-line, for survival of patients with metastatic esophagogastric adenocarcinoma after failure of first-line palliative systemic therapy.
Methods. Patients with unresectable or metastatic esophageal or gastric adenocarcinoma (2015-2017) who received first-line systemic therapy (N=1067) were selected from the Netherlands Cancer Registry. Patient, tumor and treatment characteristics at primary diagnosis and at progression of disease were used to develop the model. A Cox proportional hazards regression model was developed through forward and backward selection using Akaike's Information Criterion. The model was internally validated through 10-fold cross-validations to assess performance. Model discrimination (C-index) and calibration (slope and intercept) were used to evaluate performance of the complete and cross-validated models.
Results. The final model consisted of 11 patient tumor and treatment characteristics. The C-index was 0.75 (0.73 to 0.78), calibration slope 1.01 (1.00 to 1.01) and calibration intercept 0.01 (0.01 to 0.02). Internal cross-validation of the model showed that the model performed adequately on unseen data: C-index was 0.79 (0.77 to 0.82), calibration slope 0.93 (0.85 to 1.01) and calibration intercept 0.02 (-0.01 to 0.06).
Conclusion. The SOURCE beyond first-line model predicted survival with fair discriminatory ability and good calibration.

Introduction

Survival of patients with metastatic esophagogastric cancer is poor.[1,2] First-line palliative systemic treatment for patients with metastatic esophagogastric cancer has the potential to extend survival, and to improve or sustain quality of life.[3–5] After failure of first-line systemic treatment, patients have the option to continue with second-line palliative systemic therapy or best-supportive care. Second-line treatment with paclitaxel and ramucirumab is considered standard of care for patients with esophagogastric adenocarcinoma.[6–8] In clinical practice, roughly a quarter of patients that received first-line systemic therapy continue with second-line systemic therapy and have a median overall survival (OS) of 5.4 months since start of second-line treatment.[9]

The emergence of prediction models have enabled physicians to improve communication of individualized information regarding life expectancy and can aid in shared decision making.[10] Recently, the SOURCE and SOURCE-PANC prediction models for patients with curable or incurable esophagogastric cancer and incurable pancreatic cancer, respectively, have shown good predictive performances and are important in informing patients about treatment outcomes.[11–13]

Currently, two prediction models exist for the survival after failure of first-line systemic treatment for patients with gastric cancer.[14,15] The first consisted of a prognostic model for patients with gastric cancer who received second-line chemotherapy who were treated with second-line chemotherapy. [14] However, this model lacks internal and external validation, and thus predictive performance on novel data cannot be assessed. Furthermore, at the time of publication second-line therapy ramucirumab and paclitaxel was not available and therefore not included in the model. Since second-line therapy with ramucirumab and paclitaxel has improved survival in recent years, the existing model has become less relevant for current clinical practice.[8] The second prediction model did internally and externally validated the model, but the model was trained on relatively small number of patients and only included patients that received second-line chemotherapy.[15] Patients that received best supportive care were not included. Finally, both models were developed for patients with gastric cancer only, and cannot be used for patients with esophageal cancer.

The aim of this study was to develop and internally validate a survival prediction model using nationwide population-based data of patients with esophagogastric adenocarcinoma after failure of first-line palliative systemic treatment for use in clinical practice with patient, tumor and treatment characteristics.

Methods

Data collection

This manuscript is written according to the Transparent Reporting of a Multivariable Prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement. Patients with synchronous metastatic adenocarcinoma (2015-2017) of the esophagus (C15.0-C15.9), gastroesophageal junction (GEJ)/cardia (C16.0) or stomach (C16.1-C16.9) and patients with a metachronous metastatic disease initially treated with curative intent (2015-2016) for a non-metastatic esophageal, GEJ/cardia or gastric adenocarcinoma, who received first-line palliative systemic treatment in the Netherlands were selected from the Netherlands Cancer Registry (NCR). The NCR is a nationwide based registry that covers the total Dutch population of more than 17 million people. The NCR is linked to the pathology archive in the Netherlands (PALGA), which contains information from all newly diagnosed malignancies. Trained data managers routinely extract patient and treatment information from electronic medical records.

Follow-up information on tumour and treatment characteristics (including metachronous metastatic disease) was collected in the second half of 2019, with the exception of two hospitals due to logistical constraints. Data on vital status were obtained through annual linkage to the Dutch Personal Records Database and updated until February 2021. Metachronous metastatic disease was defined as diagnosis of metastases at least five days after end of treatment with curative intent for primary non-metastatic disease for patients diagnosed in 2015-2016. Treatment with curative intent was defined as endoscopic resection, surgical resection or definitive chemoradiotherapy (chemotherapy with concurrent radiotherapy consisting of ≥28 fractions or total radiation dose of ≥50 Gy). For patients who developed metachronous metastases within six months after end of neoadjuvant chemotherapy, the neoadjuvant chemotherapy was considered as first-line systemic therapy. First-line palliative systemic therapy was defined as all chemotherapy or targeted agents that started within three days of each other, as described in more detail in a previous publication.[16] Second-line treatment was considered when a new agent of a different drug group was started that was not administered in first-line.[9] Patients with first-line treatment failure for other reasons than disease progression were excluded. Furthermore, patients were also excluded if patients first-line therapy despite progression or if a restart of the first-line was initiated after disease progression A comprehensive overview of included patients is available in Supplementary Figure 1.

Model development and validation

Characteristics of patients included in this study were summarized with mean and standard deviation for continuous variables, frequencies for categorical variables, and median for overall survival estimates. An initial variable selection was performed to select predictors that were available for at least 50% of patients. After the selection, all variables that were available at primary diagnosis and at progression after the first-line treatment were used for the modelling procedure. Potential predictors included patient, tumor and treatment characteristics. Type of treatment (including best supportive care) after first-line systemic

therapy was a mandatory variable and was forced to be in the model, since the model's primary aim is predicting treatment effects.

Next, a multivariable Cox proportional hazard model was fitted. Through back- and forward variable selection, the final set of predictors was determined and fitted as the final model. Predictor selection was performed based on Akaike's Information Criterion (AIC). Multiple imputations by chained equations (MICE) with 10 iterations was used to handle missing data, with the exception of cN-stage, cT-stage, differentiation grade and HER2-status.[17] These variables could not be assumed to be missing at random and the fact that they are missing was likely to have predictive information. We therefore included a separate category "unknown" in these variables, which was attributed to patients whose data was missing on that variable.

The predictive performance of the final model was evaluated with the concordance index (C-index), calibration slope and calibration intercept. The C-index is a measure for model discrimination and ranges from 0.5 (random chance) to 1.0 (perfect discrimination).[18] C-indices of 0.60-0.69 are typically interpreted as poor discrimination, 0.70-0.79 fair discrimination, 0.80-0.89 good discrimination and 0.90-1.0 excellent discrimination.[19] The calibration slope and intercept refer to the accordance between predicted and observed survival outcomes.[20] For each prediction, we calculated the partial chi-squared statistic minus the predictor degree of freedom which quantified the relative importance of each variable.[21] Higher values correspond with higher relative variable importance.

To assess model performance on unseen data, 10-fold cross-validation was performed.[20] With this method, the data is randomly shuffled and split into 10 equal parts called folds. The model was then trained in 9 folds and tested in the remaining fold. This process is repeated ten times so that every patient is included in the train and test fold at least once. The C-index, calibration slope and intercept across cross-validations were evaluated with a meta-analysis to obtain pooled performance estimates similarly to previously published SOURCE and SOURCE-PANC models.[11,12]

Results

Predictors

We identified 1067 patients with metastatic esophagogastric adenocarcinoma with failure on first-line palliative systemic treatment (Table 1). Median OS of all patients since progression was 3.6 (95%CI: 3.2 to 3.8) months (Table 1; Supplementary Figure 2). After back and forward predictor selection, the final model contained 11 patient tumor and treatment characteristics (Table 2). A significant predictor at primary diagnosis was cN-stage. Although tumor differentiation grade remained in the final model, its hazard ratios were not significant compared to the reference. Significant predictors after progression on first line systemic therapy were WHO performance status, albumin (g/L), lactate dehydrogenase (LDH) (U/L), neutrophils count (10^9/L), human epidermal growth factor receptor 2 (HER2) status, duration of first-line systemic therapy (months), type of metastatic disease (synchronous or metachronous), number of metastatic sites and type of treatment after failure of first-line therapy (including best supportive care).

At primary diagnosis, a cN1 and cN2 were associated with higher OS compared to cN0. At progression, poorer WHO performance status, higher LDH concentrations, higher neutrophils count, a higher number of metastatic sites was associated with lower OS. Patients with HER2 positive tumors, higher albumin concentrations, synchronous metastatic disease, and a longer duration of first-line therapy were associated with higher OS.

Model performance

The final model had a C-index of 0.75 (0.73 to 0.78), calibration slope of 1.01 (1.00 to 1.01) and calibration intercept of 0.01 (0.01 to 0.02) (Figure 1). 10-Fold cross-validation showed similar point estimates, C-index of 0.79 (0.77 to 0.82), calibration slope 0.93 (0.85 to 1.01) and calibration intercept of 0.02 (-0.01 to 0.06) (Figure 1). In the final model, the type of treatment after first-line systemic therapy was the most predictive for survival, followed by the number of metastatic sites and duration of the first-line therapy (Figure 2).

A nomogram of the model predicting six month and one-year survival is available in Supplementary Figure 3. Predictions can be made by adding the points of each variable, and finding the corresponding probability to the total amount of points.

Table 1. Patient, disease and treatment characteristics at progression of disease after failure of first-line.

	All patients (N=1067)
Median survival (95% CI), months	3.55 (3.29-3.84)
Variables at primary diagnosis	
Sex	
Male	835 (78.3%)
Female	232 (21.7%)
cT	
1	2 (0.2%)
1A	2 (0.2%)
1B	1 (0.1%)
2	386 (36.2%)
3	355 (33.3%)
4A	36 (3.4%)
4B	48 (4.5%)
X	237 (22.2%)
cN	
0	205 (19.2%)
1	359 (33.6%)
2	359 (33.6%)
3	91 (8.5%)
X	53 (5.0%)
Primary tumor location	
Oesophagus	563 (52.8%)
Stomach	317 (29.7%)
GE-junction/Cardia	187 (17.5%)
Tumor differentiation	
Well	21 (2.0%)
Moderate	256 (24.0%)
Poorly	417 (39.1%)
Unknown	373 (35.0%)

Table 1 (Continued). Patient, disease and treatment characteristics at progression of disease after failure of first-line.

Variables at progression of disease	
Age	
Mean (SD)	63.38 (10.00)
Albumin (g/l)	
Mean (SD)	35.43 (7.13)
Missing (N)	390
LDH (U/L)	
Mean (SD)	412.04 (695.25)
Missing (N)	162
Neutrophile count (x10^9/L)	
Mean (SD)	5.76 (3.786)
Missing (N)	384
HER2 status	
Negative	678 (63.5%)
Positive	192 (18.0%)
Unknown	197 (18.5%)
WHO performance status	
0	116 (10.9%)
1	310 (29.1%)
2	110 (10.3%)
>2	97 (9.1%)
Missing (N)	431 (40.5%)
Duration first line therapy (monhts)	
Mean (SD)	8.80 (8.223)
Type of metastatic disease	
Metachronous metastases	185 (17.3%)
Synchronous metastases	882 (82.7%)
Number of metastatic sites	
Mean (SD)	2.57 (1.39)

Table 1 (Continued). Patient, disease and treatment characteristics at progression of disease after failure of first-line.

First-line therapy	
Monotherapy	51 (4.8%)
Doublet therapy	627 (58.8%)
Triplet therapy	212 (19.9%)
Trastuzumab-containing regimen	164 (15.4%)
Non-trastuzumab targeted therapy-containing regimen	13 (1.2%)
Type of second-line treatment	
Paclitaxel and ramucirumab	232 (21.7%)
Monochemotherapy	114 (10.7%)
Doublet or triplet chemotherapy	81 (7.6%)
Best supportive care	640 (60.0%)

Table 2. Hazard ratios (HR) of overall survival and 95% confidence intervals of predictors in the model.

	HR (95% CI)	p-value
Variables at primary diagnosis		
cN		
0	Reference	
1	0.75 (0.63-0.90)	0.002
2	0.81 (0.68-0.97)	0.022
3	1.05 (0.81-1.36)	0.734
X	1.06 (0.78-1.45)	0.703
Tumor differentiation grade		
Well	Reference	
Moderate	1.19 (0.76-1.88)	0.449
Poorly	1.41 (0.90-2.21)	0.129
Unknown	1.16 (0.74-1.82)	0.520
Variables at progression of the disease		
WHO performance status		
0	Reference	
1	1.33 (1.11-1.59)	0.002
2	1.38 (1.10-1.73)	0.006
>2	3.06 (2.41-3.87)	<.001
Albumin (g/l)	0.99 (0.98-1.00)	0.014
LDH (U/l)	1.0002 (1.0001-1.0003)	<.001
Neutrophile count (10^9/L)	1.04 (1.02-1.06)	<.001
HER2 status		
Negative	Reference	
Positive	0.80 (0.67-0.94)	0.008
Unknown	0.96 (0.82-1.14)	0.650
Duration first-line systemic therapy (months)	0.96 (0.95-0.97)	<.001
Type of metastatic disease		
Metachronous	Reference	
Synchronous	0.65 (0.53-0.81)	<.001
Number of metastatic sites after first-line therapy	1.21 (1.16-1.26)	<.001

Table 2 (Continued). Hazard ratios (HR) of overall survival and 95% confidence intervals of predictors in the model.

Treatment after first-line therapy		
Paclitaxel and Ramucirumab	Reference	
Monochemotherapy	1.23 (0.97-1.54)	0.082
Doublet or triplet chemotherapy	1.21 (0.93-1.56)	0.149
Best supportive care	2.65 (2.24-3.14)	<.001

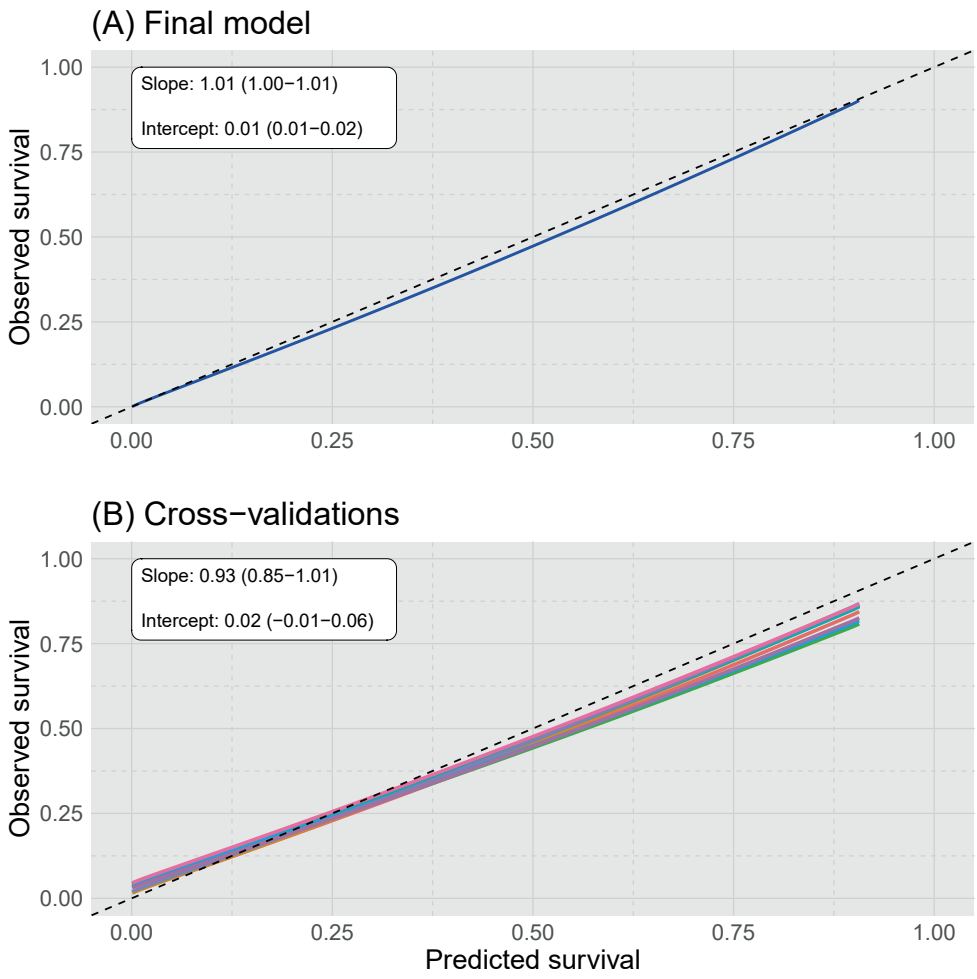


Figure 1. Calibration plot of complete model and cross-validations. This shows the accordance between the observed survival and the predicted survival of the final model (A) and across 10-fold cross-validations (B). The coloured lines represent 10 different validation folds on which the trained model was tested. Perfect values are a slope of 1 and an intercept of 0.

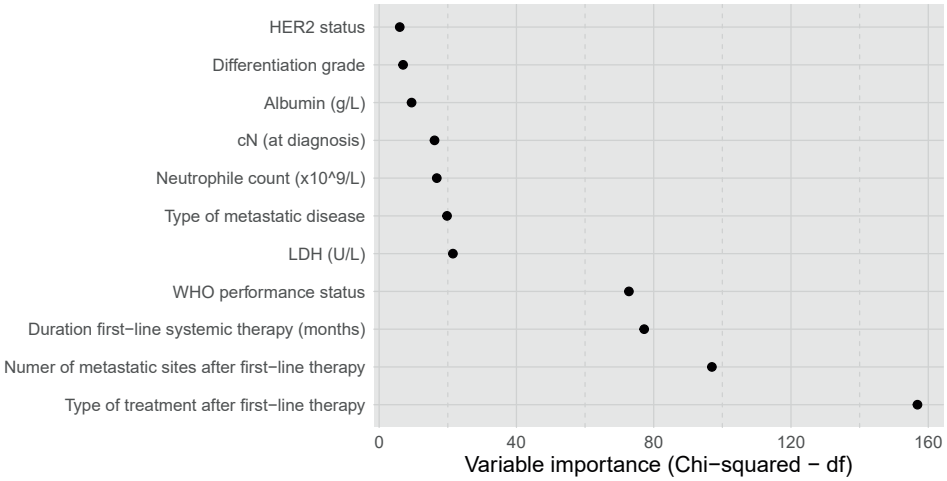


Figure 2. Relative variable importance. Variables with higher values correspond with a larger relative importance in predicting survival in the final Cox-regression model. Relatively, type of treatment after first-line systemic therapy had the most predictive capability. LDH=Lactate dehydrogenase, WHO= World Health Organization, df = degrees of freedom.

Discussion

This study developed the first population-based prediction model for survival of patients with metastatic esophagogastric adenocarcinoma after failure of first-line palliative systemic therapy. The model showed fair discrimination (0.75) and good accordance between predicted and observed overall survival. This indicates that the SOURCE beyond first-line model can be valuable for shared decision making between patient and physician when considering second-line palliative systemic therapy or best-supportive care.

The predictive performance of the prediction model was fairly similar to the previously developed SOURCE models for esophagogastric cancer where the C-indexes ranged from 0.73 to 0.78, and calibration estimates were alike.[11,12] In line with the previous SOURCE model for patients with metastatic esophagogastric cancer, WHO performance status, albumin, LDH, HER2 status, cN stage and number of metastatic sites were predictive for overall survival. In the final model, poorer WHO performance status, and higher LDH concentrations were predictive of lower OS. Higher albumin concentrations, cN1 and cN2 compared to cN0 were predictive of higher OS. Unlike previous SOURCE models for esophagogastric cancer, patient characteristics such as age, sex and body mass index were not predictive for survival after failure of first-line systemic therapy due to progression.[11] Since survival of these patients is generally poor (median survival of around 4 months in this study), variables that reflect patients' fitness may be more predictive than general patient characteristics.[22]

Novel predictor variables included the neutrophils count, duration of the first-line systemic therapy and whether a patient had synchronous or metachronous metastatic disease. Higher neutrophils count were predictive for lower OS, which is consistent with earlier findings.[23,24] It is suggested that increased numbers of neutrophils can reduce anticancer activity and increase tumor growth.[25–27] Furthermore, longer first-line therapy was predictive for higher OS, which showed that patients that respond well to first-line chemotherapy have a better OS. Synchronous metastases were predictive for higher OS compared to patients with metachronous metastases. Finally, compared to paclitaxel and ramucirumab best supportive care was predictive of a lower OS. Monochemotherapy and doublet or triplet chemotherapy were not predictive of a different OS compared to the reference treatment paclitaxel and ramucirumab.

Furthermore, although WHO performance status after failure of the first-line due to progression was predictive for survival, it was not the most predictive variable as this was the type of treatment after first-line therapy. Performance status should be accounted for in the decision to start or forgo second-line treatment, however our results show that variability of survival among patients cannot be solely accounted for by patients' performance status.[28,29] It should be noted, that 40% of performance status scores were missing.

The robustness and generalizability of the models was assessed and tested with an internal-external 10-fold cross validation scheme. With this method it can be assessed how the model performs on data that was not used for training the model. In development of prior SOURCE models, a temporal cross-validation scheme was employed which mimics real-world practice of testing the model on a new sample of patients.[11] However, follow-up of patients diagnosed between 2015-2017 was obtained in 2019. Hence, the follow-up time for patients diagnosed in 2017 was shorter compared to patients diagnosed in 2015. Survival estimates from these cohorts may therefore be different. To counter this potential source of

bias, we created the folds using a random patient sample rather than consecutive cohorts of patients. Additionally, missing data of continuous variables were handled through multiple imputation with chained equations (MICE), which reduces bias due to missing data and is preferred over complete case analysis.[30] Missing differentiation grade, cN, cT, and HER2-status variables were handled by including 'unknown' as a separate category. The combination of missing indicators with multiple imputation has been found to be a valid method to handle missing data that are not missing at random. [31] This is also useful in clinical practice since not all variables can always be known for some patients.

This study has several limitations. First, this prediction model is only developed using population-based data from the Netherlands which could affect the generalizability to other populations of patient with esophagogastric cancer. Second, health related quality of life is an important prognostic factor for survival in patients with metastatic esophagogastric cancer, but was not available to use in this study.[32]

A strength of this study is that the prediction model was developed on data from the population-based Netherlands cancer registry which is directly linked to the national pathology archive. Additionally, steps were taken to increase the robustness and generalizability of the results. Finally, this is the first model predicting survival for patients with esophagogastric cancer after failure of first-line treatment due to progression which includes paclitaxel and ramucirumab as second-line therapy and can be used as a treatment decision aid.

Conclusion

This study presented a prediction model for patients with esophagogastric adenocarcinoma that receive second-line systemic therapy or best-supportive care after failure of the first-line due to progression. The SOURCE beyond first-line model predicted survival with fair discriminatory ability and good calibration. In the future this model will be integrated in an online decision support tool to be used in clinical practice.

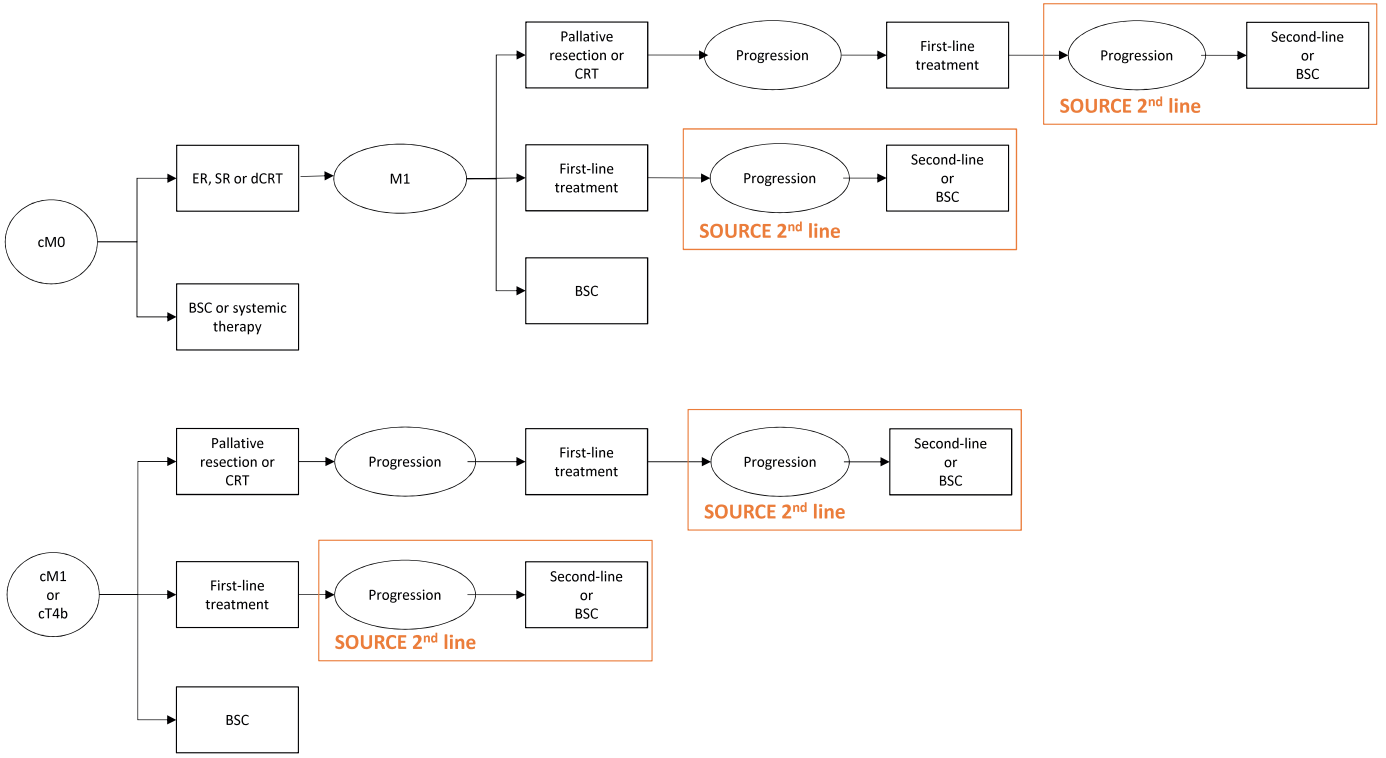
References

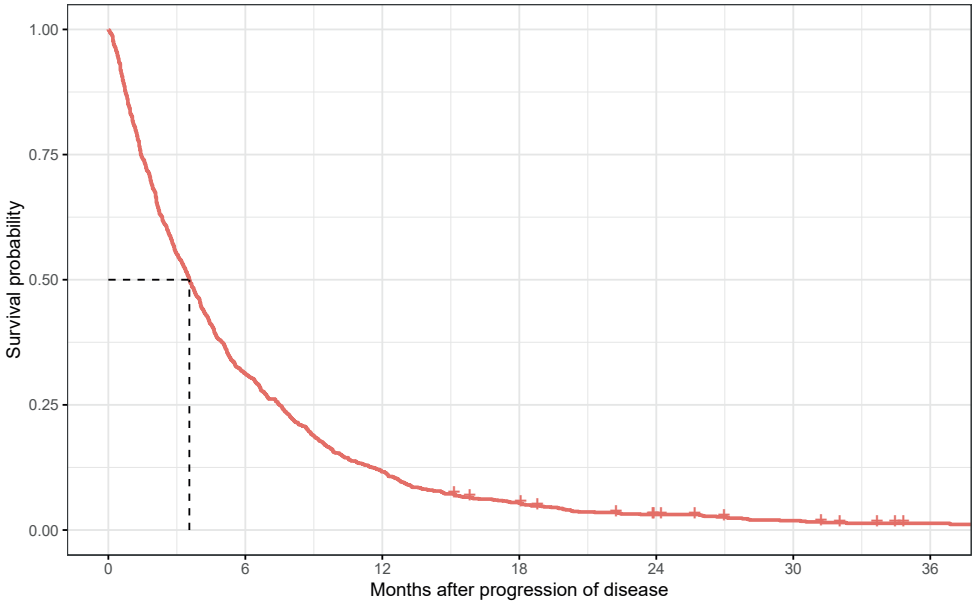
1. van Putten M, de Vos-Geelen J, Nieuwenhuijzen GAP, Siersema PD, Lemmens VEPP, Rosman C, van der Sangen MJC, Verhoeven RHA. Long-term survival improvement in oesophageal cancer in the Netherlands. *Eur J Cancer*. 2018;94:138-147. doi:10.1016/j.ejca.2018.02.025
2. Bernards N, Creemers GJ, Nieuwenhuijzen GAP, Bosscha K, Pruijt JFM, Lemmens VEPP. No improvement in median survival for patients with metastatic gastric cancer despite increased use of chemotherapy. *Ann Oncol*. 2013;24(12):3056-3060. doi:10.1093/annonc/mdt401
3. Al-Batran SE, Ajani JA. Impact of chemotherapy on quality of life in patients with metastatic esophagogastric cancer. *Cancer*. 2010;116(11):2511-2518. doi:10.1002/cncr.25064
4. Veer E Ter, Mohammad NH, Van Valkenhoef G, Ngai LL, Mali RMA, Anderegg MC, Van Oijen MGH, Van Laarhoven HWM. The Efficacy and Safety of First-line Chemotherapy in Advanced Esophagogastric Cancer: A Network Meta-analysis. *J Natl Cancer Inst*. 2016;108(10):1-13. doi:10.1093/jnci/djw166
5. Janmaat VT, Steyerberg EW, van der Gaast A, Mathijssen RHJ, Bruno MJ, Peppelenbosch MP, Kuipers EJ, Spaander MCW. Palliative chemotherapy and targeted therapies for esophageal and gastroesophageal junction cancer. *Cochrane Database Syst Rev*. 2017;2017(11). doi:10.1002/14651858.CD004063.pub4
6. Lordick F, Mariette C, Haustermans K, Obermannová R, Arnold D, on behalf of the ESMO Guidelines Committee clinicalguidelines@esmo org. Oesophageal cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2016;27:v50-v57. doi:10.1093/annonc/mdw329
7. Smyth EC, Verheij M, Allum W, Cunningham D, Cervantes A, Arnold D, on behalf of the ESMO Guidelines Committee clinicalguidelines@esmo org. Gastric cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2016;27:v38-v49. doi:10.1093/annonc/mdw350
8. Wilke H, Muro K, Van Cutsem E, Oh SC, Bodoky G, Shimada Y, Hironaka S, Sugimoto N, Lipatov O, Kim TY, Cunningham D, Rougier P, Komatsu Y, Ajani J, Emig M, Carlesi R, Ferry D, Chandrawansa K, Schwartz JD, Ohtsu A. Ramucirumab plus paclitaxel versus placebo plus paclitaxel in patients with previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (RAINBOW): A double-blind, randomised phase 3 trial. *Lancet Oncol*. 2014;15(11):1224-1235. doi:10.1016/S1470-2045(14)70420-6
9. Dijksterhuis WPM, Verhoeven RHA, Pape M, Slingerland M, Haj Mohammad N, de Vos-Geelen J, Beerepoot L V., van Voorthuizen T, Creemers GJ, Lemmens VEPP, van Oijen MGH, van Laarhoven HWM. Hospital volume and beyond first-line palliative systemic treatment in metastatic oesophagogastric adenocarcinoma: A population-based study. *Eur J Cancer*. 2020;139:107-118. doi:10.1016/j.ejca.2020.08.010
10. Vickers AJ. Prediction models in cancer care. *CA Cancer J Clin*. 2011;61(5):n/a-n/a. doi:10.3322/caac.20118
11. van den Boorn HG, Abu-Hanna A, Haj Mohammad N, Hulshof MCCM, Gisbertz SS, Klarenbeek BR, Slingerland M, Beerepoot L V., Rozema T, Sprangers MAG, Verhoeven RHA, van Oijen MGH, Zwinderman KH, van Laarhoven HWM. SOURCE: Prediction Models for Overall Survival in Patients With Metastatic and Potentially Curable Esophageal and Gastric Cancer. *J Natl Compr Canc Netw*. 2021;19(4):403-410. doi:10.6004/jnccn.2020.7631
12. van den Boorn HG, Dijksterhuis WPM, van der Geest LGM, de Vos-Geelen J, Besselink MG, Wilmink JW, van Oijen MGH, van Laarhoven HWM. SOURCE-PANC: A prediction model for patients with metastatic pancreatic ductal adenocarcinoma based on nationwide population-based data. *JNCCN J Natl Compr Cancer Netw*. 2021;19(9):1045-1053. doi:10.6004/jnccn.2020.7669
13. Van De Water LF, Van Den Boorn HG, Hoxha F, Henselmans I, Calff MM, Sprangers MAG, Abu-Hanna A, Smets EMA, Van Laarhoven HWM. Informing patients with esophagogastric cancer about treatment outcomes by using a web-based tool and training: Development and evaluation study.

J Med Internet Res. 2021;23(8). doi:10.2196/27824
14. Kanagavel D, Pokataev IA, Fedyanin MY, Tryakin AA, Bazin IS, Narimanov MN, Yakovleva ES, Garin AM, Tjulandin SA. A prognostic model in patients treated for metastatic gastric cancer with second-line chemotherapy. *Ann Oncol*. 2010;21(9):1779-1785. doi:10.1093/annonc/mdq032
15. Pietrantonio F, Barretta F, Fanotto V, Park SH, Morano F, Fucà G, Niger M, Prisciandaro M, Silvestris N, Bergamo F, Fornaro L, Bordonaro R, Rimassa L, Santini D, Tomasello G, Antonuzzo L, Noventa S, Avallone A, Leone F, Faloppi L, Donato S Di, De Braud F, Lee J, Vita F De, Bartolomeo M Di, Miceli R, Aprile G. Estimating survival probabilities of advanced gastric cancer patients in the second-line setting: The gastric life nomogram. *Oncol*. 2018;95(6):344-352. doi:10.1159/000491753
16. Dijksterhuis WPM, Verhoeven RHA, Slingerland M, Haj Mohammad N, de Vos-Geelen J, Beerepoot L V., van Voorthuizen T, Creemers GJ, van Oijen MGH, van Laarhoven HWM. Heterogeneity of first-line palliative systemic treatment in synchronous metastatic esophagogastric cancer patients: A real-world evidence study. *Int J Cancer*. 2020;146(7):1889-1901. doi:10.1002/ijc.32580
17. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45(3):1-67. doi:10.18637/jss.v045.i03
18. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128-138. doi:10.1097/EDE.0b013e3181c30fb2
19. van den Boorn HG, Engelhardt EG, van Kleef J, Sprangers MAG, van Oijen MGH, Abu-Hanna A, Zwinderman AH, Coupé VMH, van Laarhoven HWM. Prediction models for patients with esophageal or gastric cancer: A systematic review and meta-analysis. *PLoS One*. 2018;13(2):1-20. doi:10.1371/journal.pone.0192310
20. Steyerberg EW. *Clinical Prediction Models*. 2nd ed. Springer Nature Switzerland; 2019.
21. Harrell FE. *Regression Modeling Strategies*. Vol 45.; 2003. doi:10.1198/tech.2003.s158
22. Abraham P, Gricar J, Zhang Y, Shankaran V. Real-World Treatment Patterns and Outcomes in Patients Receiving Second-Line Therapy for Advanced/Metastatic Esophageal Squamous Cell Carcinoma. *Adv Ther*. 2020;37(7):3392-3403. doi:10.1007/s12325-020-01394-y
23. Fuchs CS, Muro K, Tomasek J, Van Cutsem E, Cho JY, Oh SC, Safran H, Bodoky G, Chau I, Shimada Y, Al-Batran SE, Passalacqua R, Ohtsu A, Emig M, Ferry D, Chandrawansa K, Hsu Y, Sashegyi A, Liepa AM, Wilke H. Prognostic factor analysis of overall survival in gastric cancer from two phase iii studies of second-line ramucirumab (REGARD and RAINBOW) using pooled patient data. *J Gastric Cancer*. 2017;17(2):132-144. doi:10.5230/jgc.2017.17.e16
24. Xu J, Li Y, Fan Q, Shu Y, Yang L, Cui T, Gu K, Tao M, Wang X, Cui C, Xu N, Xiao J, Gao Q, Liu Y, Zhang T, Bai Y, Li W, Zhang Y, Dai G, Ma D, Zhang J, Bai C, Huang Y, Liao W, Wu L, Chen X, Yang Y, Wang J, Ji S, Zhou H, Wang Y, Ma Z, Wang Y, Peng B, Sun J, Mancao C. Clinical and biomarker analyses of sintilimab versus chemotherapy as second-line therapy for advanced or metastatic esophageal squamous cell carcinoma: a randomized, open-label phase 2 study (ORIENT-2). *Nat Commun*. 2022;13(1):1-12. doi:10.1038/s41467-022-28408-3
25. Shau HY, Kim A. Suppression of lymphokine-activated killer induction by neutrophils. *J Immunol*. 1988;141(12):4395-4402.
26. Gregory AD, Houghton AMG. Tumor-associated neutrophils: New targets for cancer therapy. *Cancer Res*. 2011;71(7):2411-2416. doi:10.1158/0008-5472.CAN-10-2583
27. An X, Ding PR, Li YH, Wang FH, Shi YX, Wang ZQ, He YJ, Xu RH, Jiang WQ. Elevated neutrophil to lymphocyte ratio predicts survival in advanced pancreatic cancer. *Biomarkers*. 2010;15(6):516-522. doi:10.3109/1354750X.2010.491557
28. Orditura M, Galizia G, Sforza V, Gambardella V, Fabozzi A, Laterza MM, Andreozzi F, Ventriglia J, Savastano B, Mabilia A, Lieto E, Ciardiello F, De Vita F. Treatment of gastric cancer. *World J Gastroenterol*. 2014;20(7):1635-1649. doi:10.3748/wjg.v20.i7.1635
29. Wesolowski R, Lee C, Kim R. Is there a role for second-line chemotherapy in advanced gastric cancer? *Lancet Oncol*. 2009;10(9):903-912. doi:10.1016/S1470-2045(09)70136-6
30. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpen-

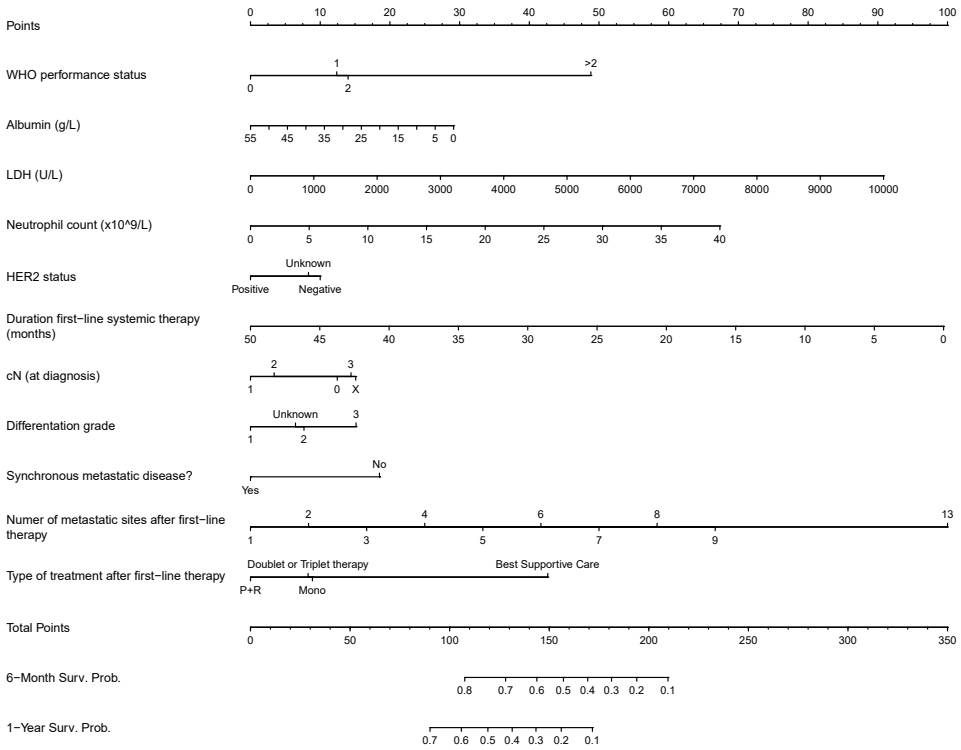
ter JR. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338:b2393. doi:10.1136/bmj.b2393
31. Sperrin M, Martin GP. Multiple imputation with missing indicators as proxies for unmeasured variables: Simulation study. *BMC Med Res Methodol*. 2020;20(1):1-11. doi:10.1186/s12874-020-01068-x
32. van Kleef JJ, Dijksterhuis WPM, van den Boorn HG, Prins M, Verhoeven RHA, Gisbertz SS, Slingerland M, Mohammad NH, Creemers GJ, Neelis KJ, Heisterkamp J, Rosman C, Ruurda JP, Kouwenhoven EA, van de Poll-Franse L V., van Oijen MGH, Sprangers MAG, van Laarhoven HWM. Prognostic value of patient-reported quality of life for survival in oesophagogastric cancer: analysis from the population-based POCOP study. *Gastric Cancer*. 2021;(0123456789). doi:10.1007/s10120-021-01209-1

Supplementary Figure 1. Flowchart of which patients were selection from the Netherland Cancer Registry. Patients included in the final dataset for the development of the SOURCE 2nd line model are highlighted in orange. ER= Endoscopic resection, SR= Surgical resection, dCRT= Definitive chemoradiotherapy (chemotherapy with concurrent radiotherapy consisting of ≥28 fractions or total radiation dose of ≥50 Gy), BSC= Best Supportive Care. Neoadjuvant chemotherapy for patients with cM1 and T4b was also seen as first line treatment. Neoadjuvant chemotherapy for patients that had metachronous metastases within six months was seen as first-line therapy. Progression during first-line treatment was not considered an event for treatment failure of





Supplementary Figure 2. Kaplan-Meier curve of the sample of patients with esophagogastric cancer on which the model was developed. The median survival time (3.6 months) is shown with the dashed line.



Supplementary Figure 3. Nomogram of prediction model for 6-month and 1-year overall survival.
P+R = Paclitaxel and ramucirumab, Mono = Monochemotherapy.

Chapter 10

Integrating Clinical Variables, Radiomics, and Tumor-derived Cell-free DNA for Enhanced Prediction of Resectable Esophageal Adenocarcinoma Outcomes

Tom van den Ende*, Steven C. Kuijper*, Yousif Widaatalla, Wyanne A. Noortman, Floris H. P. van Velden, Henry C. Woodruff, Ymke van der Pol, Norbert Moldovan, D. Michiel Pegtel, Sarah Derks, Maarten F. Bijlsma, Florent Mouliere, Lioe-Fee de Geus-Oei, Philippe Lambin, Hanneke W.M. van Laarhoven

* Shared first authorship

Based on:

Ende TVD, Kuijper SC, Widaatalla Y, et al. Integrating Clinical Variables, Radiomics, and Tumor-derived Cell-Free DNA for Enhanced Prediction of Resectable Esophageal Adenocarcinoma Outcomes. *Int J Radiat Oncol Biol Phys*. 2025

Abstract

Background: The value of integrating clinical variables, radiomics, and tumor-derived cell-free DNA (cfDNA) for the prediction of survival and response to chemoradiation of resectable esophageal adenocarcinoma (rEAC) patients is not yet known. Our aim was to investigate if radiomics and cfDNA metrics combined with clinical variables can improve personalized predictions.
Methods: A cohort of 111 rEAC patients from two centers treated with neoadjuvant chemoradiotherapy was used for exploratory retrospective analyses. Models combining the clinical variables of the SOURCE survival model with radiomic features and cfDNA, were built using elastic net regression and internally validated using 5-fold cross validation. Model performance for overall survival (OS) and time to progression (TTP) were evaluated with the C-index and the area under the curve (AUC) for pathological complete response (pCR)
Results: The best performing baseline models for OS and TTP were based on the combination of SOURCE-cfDNA which reached a C-index of 0.55 and 0.59 compared to 0.44-0.45 with SOURCE alone. The addition of re-staging PET radiomics to SOURCE was the most promising addition for predicting OS (C-index: 0.65) and TTP (C-index: 0.60). Baseline risk-stratification was achieved for OS and TTP by combining SOURCE with radiomics or cfDNA, log-rank $p<0.01$. The best performing combination model for the prediction of pCR reached an AUC of 0.61 compared to 0.47 with SOURCE variables alone.
Conclusions: The addition of radiomics and cfDNA can improve the performance of an established survival model. External validity needs to be further assessed in future studies together with the optimization of radiomic pipelines.

Introduction

Currently, personalized outcome predictions to guide treatment for resectable esophageal adenocarcinoma (rEAC) are lacking.[1] One of the primary treatment modalities for rEAC involves neoadjuvant carboplatin and paclitaxel based chemoradiotherapy (nCRT) according to CROSS and adjuvant nivolumab for incomplete responders.[2,3] This treatment regimen has shown improved survival compared to surgery alone, but locoregional and systematic relapses negatively impacts long-term outcome.[3,4] An established alternative for treating rEAC is perioperative chemotherapy according to the FLOT protocol.[5] The ESOPEC trial presented at ASCO 2024 compared FLOT with CROSS and found superior survival results for FLOT.[6] However, potentially the CROSS arm underperformed with a lower complete response rate and less patients able to finish the full protocol compared to historical data.[2] Moreover, perioperative chemotherapy is associated with more neutropenia and diarrhoea based on data from the Neo-AEGIS trial.[7] Identifying patients who are likely to benefit from either CROSS or FLOT through prediction of survival or response could aid in selecting the most suitable candidates. Such an approach would not only benefit treated patients, but will also reduce healthcare costs and protect individuals from potential complications or side effects associated with ineffective treatments. To fulfil this purpose, models that integrate clinical characteristics, radiologic, and nuclear imaging as well as other biomarker data can be employed.

Routine medical imaging provides qualitative information, but it fails to capture the wealth of hidden information that is invisible to the human eye. Radiomics, on the other hand, involves extracting multiple additional features or combinations thereof, such as intensity, shape, and texture features, from the image voxels of both tumors and healthy tissue.[8-10] By leveraging radiomics alongside clinical data and other biomarkers, we can enhance our understanding of tumor biology, and develop prediction models to inform clinical decision making.[8] Several studies have found that handcrafted manual delineations and deep learning derived radiomic feature extraction can be used to predict survival, therapy response and adverse events, alone or in combination with biomarker and clinical data.[11-16] In head and neck squamous cell carcinoma and diffuse large B-cell lymphoma adding computed tomography (CT) or 2-deoxy-2-[^{18}F]fluoro-D-glucose ([^{18}F]FDG) positron emission tomography (PET) radiomic features to clinical or biological variables improved overall survival (OS) and two-year time to progression model estimates.[11,14] A recently conducted study in resectable esophageal cancer found an improvement in response prediction after adding data relating to the expression of the immunohistochemistry tumor markers Human Epidermal growth factor Receptor 2 (HER2) and CD44 to clinico-radiomic models.[13] Other promising biomarkers such as tumor-derived cell-free DNA (cfDNA) from liquid biopsy can provide information on cancer detection, prognosis, treatment response and targeted therapy.[17,18] In rEAC patients it has already been shown cfDNA tumor fraction quantification and mutation detection is prognostic for survival but if it is a useful addition in clinico-radiomic models remains unknown.[19-22]

In our previous work, we developed the externally validated prediction model SOURCE for overall survival, utilizing clinical variables from 13,080 patients obtained from the Netherlands Cancer Registry (NCR).[23,24] In this model, relevant baseline clinical parameters were identified from electronic health records specific for esophageal cancer patients treated with curative intent. To further enhance personalized treatment predictions, we aim

to explore the potential improvement in the SOURCE model's performance by integrating liquid biopsy and radiomics data. Therefore, in this exploratory retrospective study, we investigate whether the addition of radiomic and/or cfDNA features can enhance the predictive power of the SOURCE model for survival and response prediction.

Patients and methods

In total 111 stage II-III (M0) resectable esophageal or gastroesophageal junction adenocarcinoma patients were included in this study treated consecutively between 2014 and 2019 in the Amsterdam UMC (n=104) or UMC Utrecht (n=7). Neoadjuvant treatment consisted in 40 patients of neoadjuvant chemoradiotherapy according to CROSS (nCRT) combined with PD-L1 immune checkpoint inhibition (ICI) and in 71 patients of nCRT only. The ICI treated patients were part of a prospective phase II non-randomized feasibility trial (PERFECT).[25] The patients whom were enrolled in the PERFECT trial received neoadjuvant immunotherapy intravenous atezolizumab (1,200 mg) concurrent with chemoradiotherapy in week 1 and week 4. Atezolizumab monotherapy was administered after neoadjuvant chemoradiotherapy in week 7, 10, and 13 before surgery. The nCRT only patients were included from the prospectively collected BIOES Amsterdam UMC biobank. All patients provided written, informed consent prior to study participation. This study was conducted in accordance with the Declaration of Helsinki and the international standards of good clinical practice.

SOURCE model clinical variables

For downstream analyses, the original linear predictor of the SOURCE survival prediction model for potentially curable esophageal cancer was used as reported earlier.[23] The linear predictor refers to the weighted sum of the covariates for each patient in the data, where the weights are the regression coefficients. The following baseline clinical variables on which the SOURCE model was based were extracted for the whole cohort: body mass index (BMI), albumin, hemoglobin, lactate dehydrogenase (LDH), age, clinical tumor stage (cT), clinical nodal stage (cN), tumor topography and differentiation grade.[23]

Image acquisition and reconstruction from CT and [^{18}F]FDG-PET

From the complete cohort, baseline CT images used for diagnostic or radiotherapy treatment planning with comparable quality were available from 111 patients. Additionally, re-staging CT scans after neoadjuvant treatment were available from 109 patients. Post-treatment imaging was performed with a median interval of 27 days (min 18 days to max 77 days) measured from the end of chemoradiation. The EANM Research limited (EARL)-compliant [^{18}F]FDG-PET scans were available from 61 patients at baseline and from 105 at re-staging.[26] The EARL guidelines help to ensure uniform imaging standards and enhance reproducibility. The scans were acquired and reconstructed according to standard operating procedures (SOPs) at the respective centres for diagnostic imaging.[26] The [^{18}F]FDG-PET instructions for patients included fasting for at least 6 hours before scanning. Serum glucose levels were measured and were in the range of 4 mmol/l and 11 mmol/l. Acquisition of the PET scan was scheduled 60 ± 5 minutes after administration of an intravenous [18F]FDG bolus of approximately 3 MBq/kg. Alongside the [18F]FDG-PET, diagnostic CTs were made

(5 mm slices, 3 mm reconstruction) with intravenous and oral contrast for the Amsterdam UMC patients and 5 mm reconstruction for patients treated in the UMC Utrecht. Manufacturers and respective convolution kernel reconstruction was from either Philips or Siemens for CT and PET. Except for the baseline CT scans of 10 patients which were acquired on a Toshiba (n=5) or GE HealthCare scanner (n=5). Additional information regarding patient and imaging characteristics according to the image biomarker standardisation initiative reporting guidelines is given in **Table S1**.

Radiomic feature extraction and harmonization

For the CT images manual delineation of the primary gross tumor volume (GTV) in MIM Maestro (MIM software, Cleveland, Ohio) was performed by TvdE and peer-reviewed by YW. The GTV delineation was adjusted according to the radiology or endoscopic ultrasound report. In cases with available [^{18}F]FDG-PET scans they were used in conjunction to guide the delineation. In total 105 radiomic features were extracted from the GTV using the PyRadiomics python package version 3.01.[27] The extracted quantitative metrics were 14 shape features from the region of interest, 18 intensity features and 73 texture features (22 derived from the grey level co-occurrence matrix [glcm], 16 from the grey level run length matrix [glrlm], 16 from the grey level size zone matrix [glszm], 14 from the grey level dependence matrix [gldm] and five from the neighbouring grey tone difference matrix [ngtdm]). The radiomic feature 'original_shape_VoxelVolume' was regarded as a proxy marker for tumor volume. Additionally, the [^{18}F]FDG PET images were used for the same radiomic feature extraction based on the primary esophageal tumor volume of interest (VOI). The VOI delineation was performed in 3DSlicer version 4.11 (www.slicer.org) and in-house built software implemented in Python 3.7.2 (Python Software Foundation).[28] Boxing was applied to exclude surrounding [^{18}F]FDG-avid tissues. Esophageal primary tumor location was delineated using an isocontour that applies an adaptive threshold of 50% of the SUVpeak, obtained using a sphere of 12 mm diameter, corrected for local background.[29,30] At baseline there were less than 100 PET scans available (n=61) and were therefore not used for radiomic analysis.[31] Only the re-staging PET-scans made after neoadjuvant therapy were used (n=105). For the radiomic feature extraction in PyRadiomics a fixed bin size of 0.5 g/mL was used for PET and 25 HU for CT. The interpolator used for resampling was sitkBSpline. The pixel spacing was set to ($4 \times 4 \times 4$ mm^3) for PET and ($1 \times 1 \times 1$ mm^3) for CT in PyRadiomics. The surrogate variable analysis R package (version 3.38.0) was used for ComBat post-processing harmonization to correct for the two main sources of data variability namely: convolution kernel reflected by manufacturer (Philips, Siemens, Toshiba, GE healthcare) and slice thickness (3 mm or less vs. 4-5 mm).[32-34] This post-processing using ComBat ensured that the variability of the convolution kernel and slice thickness due to different manufacturers was accounted for and removed any unwanted variability that was not due to tumor-related differences. This was done for CT baseline, CT post-treatment and PET post-treatment.

Liquid biopsy cfDNA metrics

Details on blood plasma collection, DNA isolation, library preparation and sequencing have previously been published.[35,36] For this study we only used baseline cfDNA features and tumor agnostic mutation data. In short, blood samples were collected into EDTA tubes and processed with double-centrifugation (1600g for 10 minutes and 16000g for 10 minutes) before storage at -80°C. Plasma cfDNA was extracted using QIAGEN kits. Baseline

liquid biopsy data was derived from shallow whole genome sequencing (sWGS) <1× depth of coverage on a NovaSeq 6000 and an EAC ion-torrent amplicon targeted gene panel (23 genes).[18,21,37] Library preparation of the amplicon panel was performed according to the standard operating procedure of the Ion AmpliSeq HD Library Kit and the Ion AmpliSeq HD Dual Barcode kit with 5 ng of input per pool. Sequencing of libraries was done on the Ion S5 NGS system. The mean base coverage depth for cfDNA samples was 10,524× and for white blood cell samples 9,647×. The following cfDNA metrics were used to estimate tumor fraction from sWGS: short fragments (P20-150), somatic copy number aberrations (IchorCNA), and fragment end sequence score (FrEIA) as previously described.[35,36] The amplicon sequencing results corrected for white blood cell variants were used for the detection of mutations.

Clinical outcomes

The following outcome measures were used: overall survival (OS), time-to-progression (TTP), and pathological complete response (pCR; ypT0N0). OS was defined as the days elapsed from start of treatment until death or censored at the end of follow-up, and for TTP until disease progression, recurrence, or censored at the end of follow-up. Data cut-off for OS and TTP was 14-01-2022. pCR was a binary outcome comparing complete responders to patients with residual disease or pre-surgery progression. The SOURCE model, radiomic features, and cfDNA metrics served as input for the different prediction models constructed for these three outcomes (OS, TTP, pCR), **Fig 1A.** Models were first constructed to predict each outcome based on the single input of SOURCE, baseline or re-staging radiomics or cfDNA. Thereafter, double parameter models were constructed by combining SOURCE with radiomics or cfDNA. Finally, all three parameters were combined and compared with the different radiomic combinations (baseline CT, re-staging CT, re-staging PET).

Statistical analysis

Prior to predictive modelling, for each time point (baseline, follow-up) and scan-type (CT, PET) an initial feature selection of the 105 radiomic features was performed to remove highly correlating features (≥0.75) using the redundancy filter algorithm based on a Spearman correlation matrix within the FMradio (Factor Modelling for Radiomics Data) R-package (version 1.1.1).[38] After removing the highly correlated features between 28-34 were left depending on time-point and scan-type. Next, elastic net regularization from the glmnet R package (version 4.1-7) was used to develop the prediction models for OS, TTP and pCR using clinical, cDNA and radiomic features as predictors.[39] Elastic net is a linear regression regularization technique that combines Lasso (L1) and Ridge (L2) penalties. It balances sparsity (zero coefficients) and shrinkage (small coefficients) using two parameters, lambda and alpha, offering a flexible approach to improve model performance and handle correlated features. Consequently, elastic net also performs feature selection. OS and TTP were modelled using an elastic net Cox regression and pCR was modelled using a logistic elastic net regression. The lambda parameter and alpha penalty mixing parameter were optimized using a 10-fold cross-validation scheme, **Fig. 1B**. Across all models, feature selection was handled via elastic net modelling.

Predictors in each model included the redundancy filtered radiomic features, a predefined set of four cfDNA metrics reflective of tumor fraction: short fragments P20-150 (dichotomous; threshold 0.2), ichorCNA (dichotomous; threshold 0.3), FrEIA (continuous

parameter) and mutation (dichotomous; threshold VAF 1%), the linear predictor or clinical variables from the SOURCE prediction model, and all respective combinations. Missing data on clinical variables were handled and imputed with a random forest imputation using the missRanger package in R (version 2.2.1).[40] For the time-to-event models (OS, TTP) the linear predictor of the SOURCE prediction model was used as a single predictor to reflect the clinical variables. Using the linear predictor, the original covariances of the original SOURCE model remained intact. For logistic models (pCR) all clinical variables from the SOURCE prediction model were used in the modelling procedure as the linear prediction was developed in the context of a Cox regression model as opposed to a logistic regression model. Additionally, an exploratory analysis was performed for pCR by using the delta radiomic values from matched baseline and re-staging CT scans (re-staging value – baseline value).

For time-to-event models, the concordance (C-index) was used to assess predictive performance and the AUC-ROC curve to evaluate performance of the logistic models. In the context for survival models the C-index expresses whether the models can distinguish between patients with low-risk and high-risk of an event.[41] A C-index of 0.5 indicates poor discrimination and 1.0 indicates perfect discrimination. For logistic models, the AUC expresses whether models can discriminate between patients with and without a pathologically complete response.[42] An AUC of 0.5 indicates poor discrimination and 1.0 indicates perfect discrimination. Both the C-index and AUC provide an indication of the model' ability to discriminate between patients with short and long survival or detect a higher likelihood of pathological complete response. Given the explorative nature of this study these metrics were chosen as the primary metric to evaluate all models on.

Bootstrapping (500 iterations) was used to empirically estimate 95% confidence intervals for the performance metrics. Furthermore, for every fitted prediction model, we performed an internal-external five-fold cross-validation to assess performance of the model on data on which the model was not trained (80% training and 20% validation). In each fold, the entire modelling pipeline was repeated. Finally, a random permutation test of the outcomes was performed to investigate overfitting, **Table S1**.[43] The values obtained by cross-validation was the primary outcome of this study, as these would better reflect the external reproducibility.

Based on the fitted models we constructed Kaplan-Meier curves of the time-to-event models and ROC curves of the logistic models. For the Kaplan-Meier curves, patients were classified into low-risk and high-risk groups based on the linear predictor values of each patient from the fitted model. The optimal cut point for the determination of high-risk and low-risk was optimized by finding the maximum rank statistic (prognostic index).[44] A log-rank test was used to statistically test the two arms of the curve. For this study the TRIPOD statement (transparent reporting of multivariable prediction model for individual prognosis or diagnosis, version 1 October 2020) can be found in **Table S2**.[45]

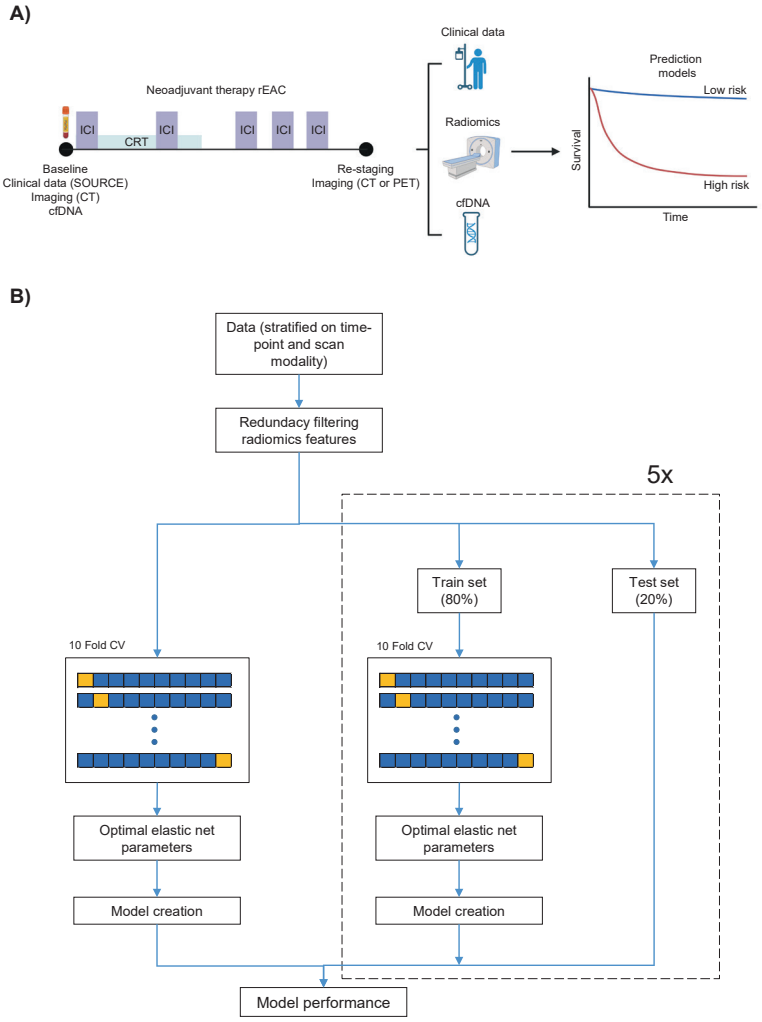


Figure 1. Study method for developing predictions for survival and pathological response.
A) Treatment schedule of patients included in this study. Patients were treated with neoadjuvant chemoradiotherapy (nCRT) and a subset of patients also received neoadjuvant immune checkpoint inhibition (ICI). The predictions for survival and response are based on radiomics, SOURCE and cfDNA.
B) Elastic-net regression analysis with cross-validation method.

Results

Prediction models for survival and disease progression

Table 1 displays all clinical characteristics of patients included in this study. Survival models were constructed based on three different types of data: the linear predictor of SOURCE, radiomics (CT or PET) and cfDNA data. The apparent C-indices derived from Cox regression analysis on the full data-set (apparent) and five-fold cross-validated estimates are given in **Fig. 2**. Below we report the cross-validated results as these are most representative for external performance. The C-index of the clinical model (SOURCE) for OS was 0.45, **Fig. 2A**. The addition of baseline CT radiomics or cfDNA to the clinical model improved the C-indices to 0.54 and 0.55, respectively. Combining all three baseline metrics did not lead to any additional improvement, as SOURCE, CT radiomics and cfDNA reached a C-index of 0.54. Baseline features selected in the CT radiomic models were two features namely: gldm_SmallDependenceEmphasis and ngtdm_Strength. Additionally, we assessed if the re-staging CT or PET derived radiomic features could be used for OS prediction together with the clinical model and/or cfDNA. The addition of re-staging PET radiomics to the clinical model improved the C-index for OS to 0.65 which was better than the addition of re-staging CT radiomics (C-index: 0.48). The addition of cfDNA to the clinico-radiomic re-staging PET model did not improve the C-index (0.62), **Fig. 2A**. The features selected in the PET-models were three features namely: firstorder_Skewness, gldm_DependenceNonUniformity and ngtdm_Contrast. The complete overview of selected features for each model and coefficients are given in **Table S3**. These findings suggest the addition of radiomic features or cfDNA can improve the performance of an established OS model such as SOURCE.

Next, we investigated the performance of different models for the prediction of TTP, **Fig. 2B**. Below we report the five-fold cross-validated estimates. The C-index for the clinical model (SOURCE) was 0.44, **Fig. 2B**. The addition of baseline CT radiomics or cfDNA to the clinical model improved the C-indices to 0.55 and 0.59, respectively. Combining all three baseline metrics SOURCE, CT radiomics and cfDNA reached a C-index of 0.56 which was not better than the SOURCE-cfDNA model, **Fig. 2B**. Baseline CT features selected were glrlm_RunEntropy, gldm_SmallDependenceEmphasis, ngtdm_Coarseness and ngtdm_Strength. Additionally, we assessed if the re-staging CT or PET derived radiomic features could be used for TTP prediction together with the clinical model and/or cfDNA. The addition of re-staging PET radiomics to the clinical model improved the C-index for TTP to 0.60 which was better than the addition of re-staging CT radiomics (C-index: 0.45). The addition of cfDNA to the clinico-radiomic re-staging PET model did not improve the C-index (0.59), **Fig. 2B**. The features selected in the PET models were firstorder_Skewness, glcm_Idmn, gldm_DependenceNonUniformity and ngtdm_Contrast. The complete overview of selected features for each model and coefficients are given in **Table S3**. The prediction of TTP could thus be enhanced by adding radiomics or cfDNA to the clinical SOURCE model.

To investigate if these models could be used for baseline risk stratification we performed an exploratory analysis by Kaplan–Meier analysis of the prognostic index by only using baseline metrics (SOURCE, baseline CT radiomics, cfDNA). For both OS and TTP it was possible to identify a high and low risk group after determining the optimal cut off point of the prognostic index for each model, **Fig. 3**. Similar stratification for OS was achie-

ved by the SOURCE-radiomics (**Fig. 3B**) and SOURCE-radiomics-cfDNA model, log-rank p=0.0017, **Fig. 3D**. For TTP the SOURCE-radiomics-cfDNA model was able to provide the best separation of the Kaplan-Meier curve, log-rank p=0.0001, **Fig. 3H**. Risk-stratification could thus be achieved by combining SOURCE with radiomics or cfDNA metrics.

Prediction models for treatment response

To predict pCR after neoadjuvant therapy we compared the performance of AUC classification based on SOURCE, cfDNA and radiomics from baseline or re-staging imaging, **Fig. 4**. Below we report the five-fold cross-validated estimates. First, we investigated if baseline CT derived radiomics can be used in combination with other metrics to predict pCR. The classification of response with only the SOURCE variables reached an AUC of 0.47, **Fig. 4**. The combination model of SOURCE with baseline CT radiomics improved the AUC to 0.61 while this was not seen with the addition of cfDNA to SOURCE, AUC: 0.48. Combining all three baseline metrics SOURCE, CT radiomics and cfDNA reached an AUC of 0.61. Baseline features in the CT radiomic models were among others: glrlm_RunEntropy, glszm_SizeZoneNonUniformity, gldm_DependenceEntropy, gldm_SmallDependenceEmphasis and ngtdm_Busyness. Next, we assessed if the re-staging CT or PET derived radiomic features could be used for pCR prediction together with the clinical variables and/or cfDNA. The addition of re-staging CT or PET radiomics to the SOURCE variables did not lead to any meaningful improvement with AUCs of 0.49 and 0.42, respectively, **Fig. 4**. The combination of all three: SOURCE, re-staging CT or PET and cfDNA was also of no additional value with AUCs of 0.50 and 0.42. An exploratory analysis of delta CT radiomics alone or together with the SOURCE variables and/or cfDNA reached AUCs after cross validation between 0.37 and 0.44, **Fig. 4**. The only features selected in the delta radiomic models was glszm_ZoneEntropy. The complete overview of selected features for each model and coefficients are given in **Table S3**. In conclusion, the prediction of pCR improves after combining the SOURCE variables, baseline CT radiomics and baseline cfDNA data compared to clinical variables alone.

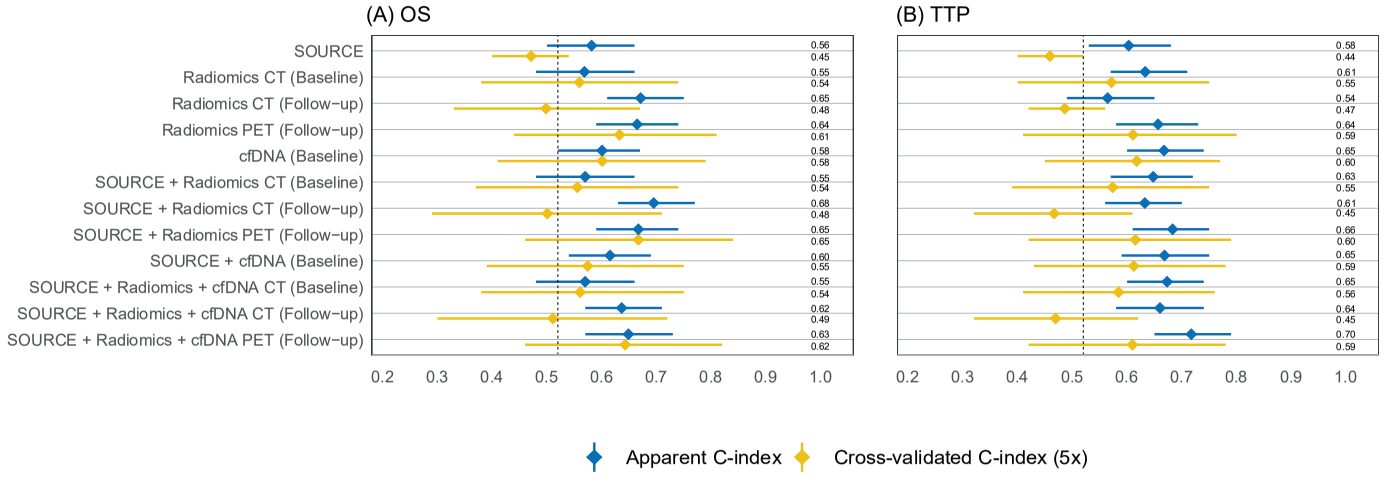


Figure 2. Prediction models for OS and TTP with in blue the apparent C-indices and in yellow the cross-validated estimates.

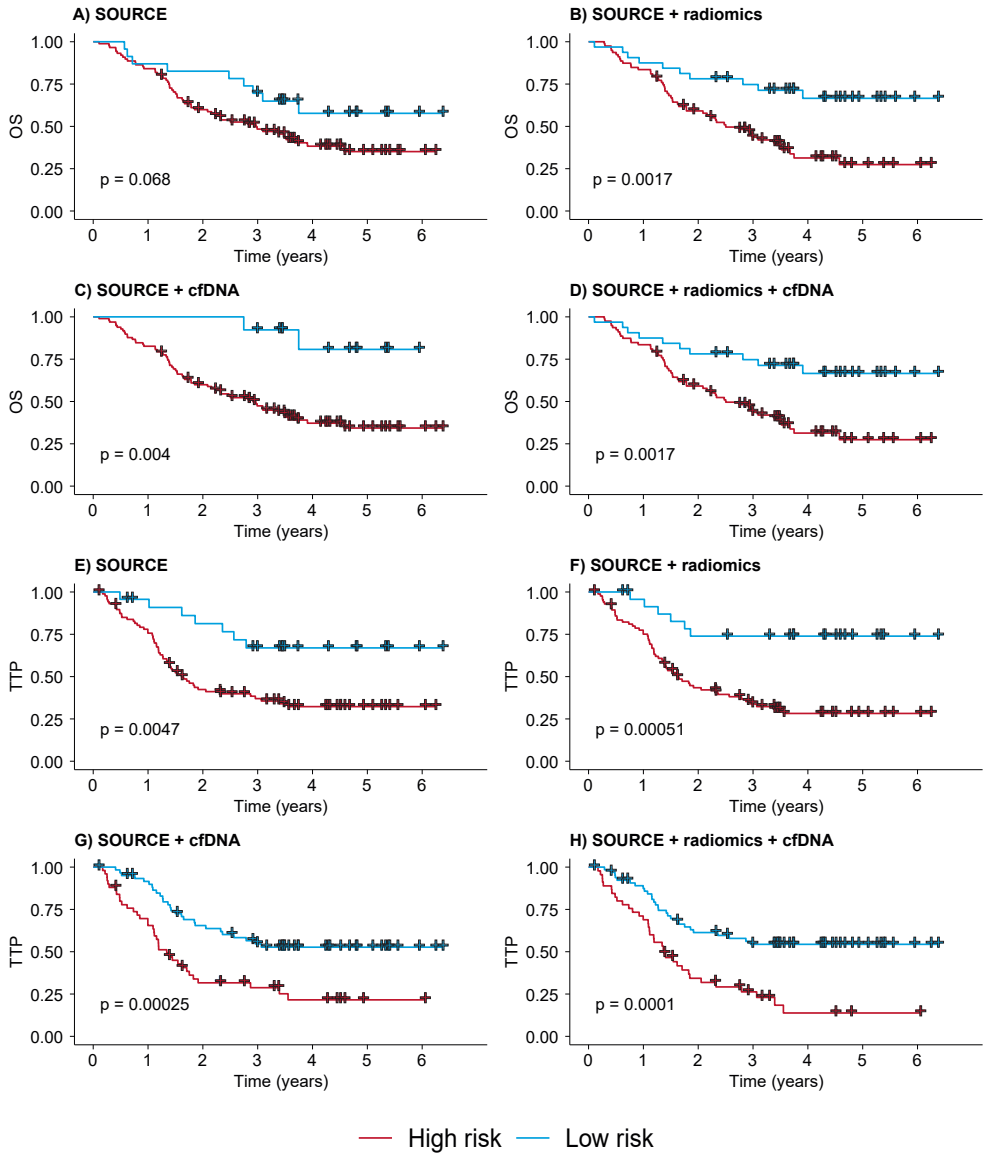


Figure 3. Baseline risk stratification by Kaplan-Meier for OS (A-D) and TTP (E-H).

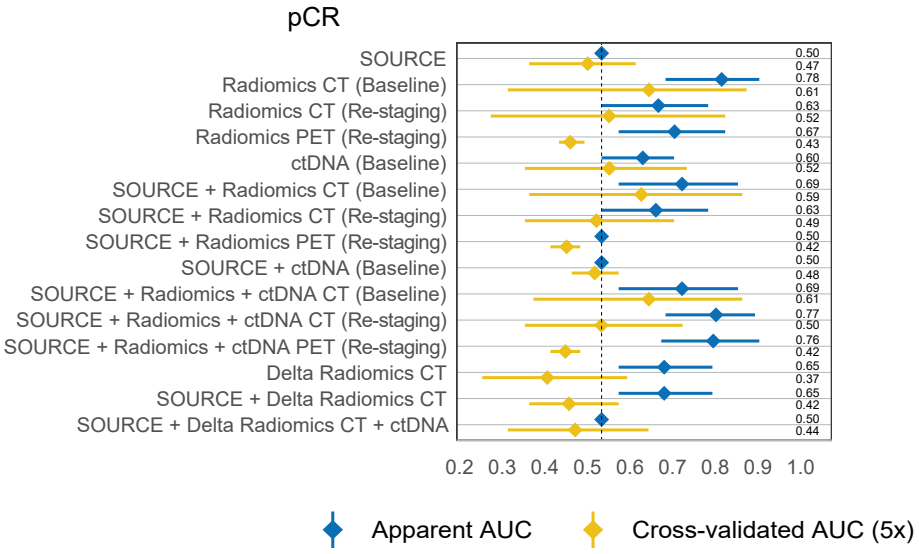


Figure 4. Classification of pCR with in blue the apparent AUCs and in yellow the cross-validated estimates.

Discussion

This study evaluated whether integrating clinical variables, radiomics, and tumor-derived cfDNA data can improve survival and response prediction for rEAC patients treated in the neoadjuvant setting compared to the clinical SOURCE model alone. The addition of cfDNA or CT radiomics to the clinical SOURCE model was able to improve baseline prediction models for OS and TTP. However, the cross-validated estimates were relatively low. Baseline risk stratification for survival was possible based on the prognostic index of the regression models. Re-staging PET radiomics was the most promising addition to the clinical SOURCE model. The addition of baseline CT radiomics and cfDNA to clinical variables was able to marginally improve the prediction of pCR compared to clinical variables alone.

Several studies in different cancer types have found that radiomics can improve upon conventional staging or clinical variables.[10,14,46,47] Most studies on radiomics use a select number of clinical variables including among others the American Joint Committee on Cancer (AJCC) staging system. In our proof of concept study we were able to improve the performance of the clinical SOURCE model by adding cfDNA or handcrafted radiomic features. Moreover, baseline regression models were able to classify patients into a high and low survival group. It must however be noted that despite the observed improvement the cross-validated C-indices were relatively low ≤ 0.65 and are thus not yet ready for clinical implementation. As a next step, external validation and optimization in larger cohorts will be necessary. Unfortunately, analyzing large-data sets in radiomics is a time-consuming process partly due to the manual delineation step. A potential solution to this problem is automatic segmentation which can help speed up delineation, improve accuracy and provide better risk stratification. A recently conducted study in lung cancer showed faster and better reproducible delineations by an automated pipeline.[48] Moreover, in the majority of cases the radiologist or radiation oncologist preferred the automated segmentation compared to the manually delineated volume.[48] Other techniques that can improve radiomics performance is combining handcrafted radiomics and deep learning algorithms with ensemble learning or consensus algorithms.[49] This has already led to improvements in the classification of idiopathic pulmonary fibrosis and prediction of adverse radiation effects in patients with brain metastases.[16,50] Our study provides support for the continued investigation of these techniques for radiomic feature extraction, and data analysis alongside integration into gastroesophageal prognostic models such as SOURCE.[23] Future studies should consider addressing certain aspects, such as enhancing preprocessing techniques and incorporating additional harmonization methods like ComBat, as utilized in our study.[51]

This study observed that baseline cfDNA biomarker data was of additive value for the prediction of OS or TTP together with SOURCE and/or radiomics. Previous studies were also able to establish the additional value of combining genomic or pathology biomarkers with clinical data and radiomics.[13,52-54] In a non-small cell lung cancer study the addition of cfDNA data to a clinico-radiomic model improved the prediction of survival models in metastatic patients treated with epidermal growth factor receptor targeted therapy.[53] The combination of baseline cfDNA metrics and SOURCE showed better capacity to inform on TTP (C-index: 0.59) than OS (C-index: 0.55). Other studies also found that cfDNA was a marker for progression at baseline or at later time-points in rEAC.[19,20,55,56] Current results do not support the integration of cfDNA into rEAC specific survival models as it is not yet sufficiently informative regarding OS prediction and TTP. Also, cfDNA analysis is relatively

expensive while radiomics uses readily available diagnostic imaging. In our models we only used baseline cfDNA data and used a tumor-agnostic approach. To further improve the utility of cfDNA profiling repeated sampling or sequencing for methylation changes may improve prognostication.[20,57] Several other biomarkers in esophageal cancer detected in blood or tissue could be included in predictive models.[58,59] For example, the presence of tumor-associated immune cells in the tumor microenvironment was predictive of nCRT response.[60] Molecular characterization by subtyping esophageal cancer could help stratify patients for certain treatments such as immunotherapy for the microsatellite instable tumors.[61,62] Moving forward these biomarkers that can technically be implemented in patient care need to be further evaluated if they can improve clinical prediction models such as SOURCE.

In our study the re-staging PET radiomics were the most promising addition to the clinical SOURCE model. From a clinical perspective baseline models would be preferred to select patients upfront for certain treatments and this can be combined with for example baseline immune profiling of the TME to predict response to chemoradiation.[63] Incorporating non-baseline measurements into these decision systems may nevertheless be of value to select patients for surgery after re-staging imaging in rEAC patients. Putting our results into context of current literature, longitudinal radiomic features were able to predict patient outcome in several tumor types.[64-67] A recently conducted systematic review highlights its advantages but also its limitations, especially regarding the heterogeneous methods used within each paper.[65] Based on our study further exploration of adding longitudinal imaging radiomics for personalized outcome prediction seems more promising than only baseline measurements, although the difference was small between baseline and re-staging models. Features from the best performing re-staging PET model for OS were related to voxel skewness, local intensity variation between neighboring voxels and gray level homogeneity. Interestingly, a previous study in non-small cell lung cancer also found skewness, as a PET radiomic feature, to be associated with a higher chance of progression after immunotherapy.[68] These features could be related to the degree of tissue heterogeneity and need further biological validation.[69] In conclusion our results support the further development of longitudinal radiomic models in decision support systems.[70,71]

In this study, there were several methodological limitations worth considering. Due to the relatively small sample size of the cohort the risk of overfitting was present. Cross-validation was used to investigate this and revealed that the cross-validated point estimates were generally significantly lower than the apparent estimates. This suggests further external validation will be necessary. Another limitation is the use of the linear predictor from SOURCE for TTP while it was originally developed for OS. This was based on the assumption that the covariance structure of the SOURCE model would also be valid for modelling TTP. However, as SOURCE was never validated for TTP we could not test this assumption. In this study we were also not able to look separately at patients treated with or without immunotherapy as the numbers were too small for separate analyses.

Conclusion

Clinical survival prediction models such as SOURCE could be improved by integrating radiomics or cfDNA measurements. Moreover, the addition of re-staging radiomic PET features to a clinical model were promising to predict OS and TTP. The prediction of pCR improved after adding baseline CT radiomic features and cfDNA to the SOURCE variables. However, currently developed models are not yet sufficient for clinical implementation due to possibly poor external validity. Future studies should explore the optimization of radiomic pipelines e.g. integrating handcrafted and deep radiomics, in personalized prognostic predictions or treatment stratification for rEAC.

References

1. Obermannova R, Alsina M, Cervantes A, et al. Oesophageal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Ann Oncol* 2022; **33**(10): 992-1004.
2. van Hagen P, Hulshof MC, van Lanschot JJ, et al. Preoperative chemoradiotherapy for esophageal or junctional cancer. *N Engl J Med* 2012; **366**(22): 2074-84.
3. Kelly RJ, Ajani JA, Kuzdzal J, et al. Adjuvant Nivolumab in Resected Esophageal or Gastroesophageal Junction Cancer. *N Engl J Med* 2021; **384**(13): 1191-203.
4. Shapiro J, van Lanschot JJB, Hulshof M, et al. Neoadjuvant chemoradiotherapy plus surgery versus surgery alone for oesophageal or junctional cancer (CROSS): long-term results of a randomised controlled trial. *Lancet Oncol* 2015; **16**(9): 1090-8.
5. Al-Batran SE, Homann N, Pauligk C, 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* 2019; **393**(10184): 1948-57.
6. Hoeppner J, Brunner T, Lordick F, et al. Prospective randomized multicenter phase III trial comparing perioperative chemotherapy (FLOT protocol) to neoadjuvant chemoradiation (CROSS protocol) in patients with adenocarcinoma of the esophagus (ESOPEC trial). *Journal of Clinical Oncology* 2024; **42**(17_suppl): LBA1-LBA.
7. Reynolds JV, Preston SR, O'Neill B, et al. Trimodality therapy versus perioperative chemotherapy in the management of locally advanced adenocarcinoma of the oesophagus and oesophagogastric junction (Neo-AEGIS): an open-label, randomised, phase 3 trial. *Lancet Gastroenterol Hepatol* 2023; **8**(11): 1015-27.
8. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017; **14**(12): 749-62.
9. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012; **48**(4): 441-6.
10. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014; **5**: 4006.
11. Eertink JJ, van de Brug T, Wiegers SE, et al. (18)F-FDG PET baseline radiomics features improve the prediction of treatment outcome in diffuse large B-cell lymphoma. *Eur J Nucl Med Mol Imaging* 2022; **49**(3): 932-42.
12. Beukinga RJ, Poelmann FB, Kats-Ugurlu G, et al. Prediction of Non-Response to Neoadjuvant Chemoradiotherapy in Esophageal Cancer Patients with (18)F-FDG PET Radiomics Based Machine Learning Classification. *Diagnostics (Basel)* 2022; **12**(5).
13. Beukinga RJ, Wang D, Karrenbeld A, et al. Addition of HER2 and CD44 to (18)F-FDG PET-based clinico-radiomic models enhances prediction of neoadjuvant chemoradiotherapy response in esophageal cancer. *Eur Radiol* 2021; **31**(5): 3306-14.
14. Keek SA, Wesseling FWR, Woodruff HC, et al. A Prospectively Validated Prognostic Model for Patients with Locally Advanced Squamous Cell Carcinoma of the Head and Neck Based on Radiomics of Computed Tomography Images. *Cancers (Basel)* 2021; **13**(13).
15. Salahuddin Z, Chen Y, Zhong X, Rad NM, Woodruff HC, Lambin P. HNT-AI: An Automatic Segmentation Framework for Head and Neck Primary Tumors and Lymph Nodes in FDG- PET/CT Images. 2023; Cham: Springer Nature Switzerland; 2023. p. 212-20.
16. Keek SA, Beuque M, Primakov S, et al. Predicting Adverse Radiation Effects in Brain Tumors After Stereotactic Radiotherapy With Deep Learning and Handcrafted Radiomics. *Front Oncol* 2022; **12**: 920393.
17. Dang DK, Park BH. Circulating tumor DNA: current challenges for clinical utility. *J Clin Invest* 2022; **132**(12).
18. Mouliere F, Chandrananda D, Piskorz AM, et al. Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med* 2018; **10**(466).
19. Ococks E, Frankell AM, Masque Soler N, et al. Longitudinal tracking of 97 esophageal aden-

ocarcinomas using liquid biopsy sampling. *Ann Oncol* 2021; **32**(4): 522-32.
20. Ococks E, Sharma S, Ng AWT, Aleshin A, Fitzgerald RC, Smyth E. Serial Circulating Tumor DNA Detection Using a Personalized, Tumor-Informed Assay in Esophageal Adenocarcinoma Patients Following Resection. *Gastroenterology* 2021; **161**(5): 1705-8 e2.
21. Moldovan N, van der Pol Y, van den Ende T, et al. Genome-wide cell-free DNA termini in patients with cancer. *medRxiv* 2021: 2021.09.30.21264176.
22. Wallander K, Eisfeldt J, Lindblad M, et al. Cell-free tumour DNA analysis detects copy number alterations in gastro-oesophageal cancer patients. *PLoS One* 2021; **16**(2): e0245488.
23. van den Boorn HG, Abu-Hanna A, Haj Mohammad N, et al. SOURCE: Prediction Models for Overall Survival in Patients With Metastatic and Potentially Curable Esophageal and Gastric Cancer. *J Natl Compr Canc Netw* 2021; **19**(4): 403-10.
24. van Kleef JJ, van den Boorn HG, Verhoeven RHA, et al. External Validation of the Dutch SOURCE Survival Prediction Model in Belgian Metastatic Oesophageal and Gastric Cancer Patients. *Cancers (Basel)* 2020; **12**(4).
25. van den Ende T, de Clercq NC, van Berge Henegouwen MI, et al. Neoadjuvant Chemoradiotherapy Combined with Atezolizumab for Resectable Esophageal Adenocarcinoma: A Single-arm Phase II Feasibility Trial (PERFECT). *Clin Cancer Res* 2021; **27**(12): 3351-9.
26. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging* 2015; **42**(2): 328-54.
27. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* 2017; **77**(21): e104-e7.
28. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging* 2012; **30**(9): 1323-41.
29. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *J Nucl Med* 2009; **50 Suppl 1**(Suppl 1): 122S-50S.
30. Frings V, van Velden FH, Velasquez LM, et al. Repeatability of metabolically active tumor volume measurements with FDG PET/CT in advanced gastrointestinal malignancies: a multicenter study. *Radiology* 2014; **273**(2): 539-48.
31. Orlhac F, Nioche C, Klyuzhin I, Rahmim A, Buvat I. Radiomics in PET Imaging:: A Practical Guide for Newcomers. *PET Clin* 2021; **16**(4): 597-612.
32. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012; **28**(6): 882-3.
33. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007; **8**(1): 118-27.
34. Ligero M, Jordi-Ollero O, Bernatowicz K, et al. Minimizing acquisition-related radiomics variability by image resampling and batch effect correction to allow for large-scale data analysis. *Eur Radiol* 2021; **31**(3): 1460-70.
35. Moldovan N, van der Pol Y, van den Ende T, et al. Multi-modal cell-free DNA genomic and fragmentomic patterns enhance cancer survival and recurrence analysis. *Cell Rep Med* 2024; **5**(1): 101349.
36. van den Ende T, van der Pol Y, Creemers A, et al. Genome-wide and panel-based cell-free DNA characterization of patients with resectable esophageal adenocarcinoma. *J Pathol* 2023; **261**(3): 286-97.
37. van der Pol Y, Moldovan N, Verkuijlen S, et al. The Effect of Preanalytical and Physiological Variables on Cell-Free DNA Fragmentation. *Clin Chem* 2022; **68**(6): 803-13.
38. Peeters CFW ÜC, Mes SW. Stable prediction with radiomics data. *ArXiv* 2019: 2019;abs/1903.11696.
39. Friedman JH, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* 2010; **33**(1): 1 - 22.
40. Stekhoven DJ, Buhlmann P. MissForest--non-parametric missing value imputation for

mixed-type data. *Bioinformatics* 2012; **28**(1): 112-8.
41. Longato E, Vettoretti M, Di Camillo B. A practical perspective on the concordance index for the evaluation and selection of prognostic time-to-event models. *J Biomed Inform* 2020; **108**: 103496.
42. de Hond AAH, Steyerberg EW, van Calster B. Interpreting area under the receiver operating characteristic curve. *Lancet Digit Health* 2022; **4**(12): e853-e5.
43. Buvat I, Orlhac F. The Dark Side of Radiomics: On the Paramount Importance of Publishing Negative Results. *J Nucl Med* 2019; **60**(11): 1543-4.
44. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol* 2013; **13**: 33.
45. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015; **350**: g7594.
46. Mossinelli C, Tagliabue M, Ruju F, et al. The role of radiomics in tongue cancer: A new tool for prognosis prediction. *Head Neck* 2023.
47. Li W, Zhang L, Tian C, et al. Prognostic value of computed tomography radiomics features in patients with gastric cancer following curative resection. *Eur Radiol* 2019; **29**(6): 3079-89.
48. Primakov SP, Ibrahim A, van Timmeren JE, et al. Automated detection and segmentation of non-small cell lung cancer computed tomography images. *Nat Commun* 2022; **13**(1): 3423.
49. Rogers W, Thulasi Seetha S, Refaee TAG, et al. Radiomics: from qualitative to quantitative imaging. *Br J Radiol* 2020; **93**(1108): 20190948.
50. Refaee T, Salahuddin Z, Frix AN, et al. Diagnosis of Idiopathic Pulmonary Fibrosis in High-Resolution Computed Tomography Scans Using a Combination of Handcrafted Radiomics and Deep Learning. *Front Med (Lausanne)* 2022; **9**: 915243.
51. Mali SA, Ibrahim A, Woodruff HC, et al. Making Radiomics More Reproducible across Scanner and Imaging Protocol Variations: A Review of Harmonization Methods. *J Pers Med* 2021; **11**(9).
52. Fathi Kazerooni A, Saxena S, Toorens E, et al. Clinical measures, radiomics, and genomics offer synergistic value in AI-based prediction of overall survival in patients with glioblastoma. *Sci Rep* 2022; **12**(1): 8784.
53. Yousefi B, LaRiviere MJ, Cohen EA, et al. Combining radiomic phenotypes of non-small cell lung cancer with liquid biopsy data may improve prediction of response to EGFR inhibitors. *Sci Rep* 2021; **11**(1): 9984.
54. Feng L, Liu Z, Li C, et al. Development and validation of a radiopathomics model to predict pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a multicentre observational study. *Lancet Digit Health* 2022; **4**(1): e8-e17.
55. Bonazzi VF, Aoude LG, Brosda S, et al. ctDNA as a biomarker of progression in oesophageal adenocarcinoma. *ESMO Open* 2022; **7**(3): 100452.
56. Hofste LSM, Geerlings MJ, von Rhein D, et al. Circulating Tumor DNA-Based Disease Monitoring of Patients with Locally Advanced Esophageal Cancer. *Cancers (Basel)* 2022; **14**(18).
57. Luo H, Wei W, Ye Z, Zheng J, Xu RH. Liquid Biopsy of Methylation Biomarkers in Cell-Free DNA. *Trends Mol Med* 2021; **27**(5): 482-500.
58. McClurg DP, Sanghera C, Mukherjee S, Fitzgerald RC, Jones CM. A systematic review of circulating predictive and prognostic biomarkers to aid the personalised use of radiotherapy in the radical treatment of patients with oesophageal cancer. *Radiother Oncol* 2024; **195**: 110224.
59. Booth ME, Smyth EC. Immunotherapy in Gastro-Oesophageal Cancer: Current Practice and the Future of Personalised Therapy. *BioDrugs* 2022; **36**(4): 473-85.
60. Soeratram TT, Creemers A, Meijer SL, et al. Tumor-immune landscape patterns before and after chemoradiation in resectable esophageal adenocarcinomas. *J Pathol* 2022; **256**(3): 282-96.
61. van Velzen MJM, Derks S, van Grieken NCT, Haj Mohammad N, van Laarhoven HWM. MSI as a predictive factor for treatment outcome of gastroesophageal adenocarcinoma. *Cancer Treat Rev* 2020; **86**: 102024.

62. Cancer Genome Atlas Research N, Analysis Working Group: Asan U, Agency BCC, et al. Integrated genomic characterization of oesophageal carcinoma. *Nature* 2017; **541**(7636): 169-75.
63. Goedegebuure RSA, Harrasser M, de Klerk LK, et al. Pre-treatment tumor-infiltrating T cells influence response to neoadjuvant chemoradiotherapy in esophageal adenocarcinoma. *Oncoimmunology* 2021; **10**(1): 1954807.
64. Guo L, Du S, Gao S, et al. Delta-Radiomics Based on Dynamic Contrast-Enhanced MRI Predicts Pathologic Complete Response in Breast Cancer Patients Treated with Neoadjuvant Chemotherapy. *Cancers (Basel)* 2022; **14**(14).
65. Nardone V, Reginelli A, Grassi R, et al. Delta radiomics: a systematic review. *Radiol Med* 2021; **126**(12): 1571-83.
66. Xie D, Xu F, Zhu W, et al. Delta radiomics model for the prediction of progression-free survival time in advanced non-small-cell lung cancer patients after immunotherapy. *Front Oncol* 2022; **12**: 990608.
67. Peeken JC, Asadpour R, Specht K, et al. MRI-based delta-radiomics predicts pathologic complete response in high-grade soft-tissue sarcoma patients treated with neoadjuvant therapy. *Radiother Oncol* 2021; **164**: 73-82.
68. Polverari G, Ceci F, Bertaglia V, et al. (18)F-FDG Pet Parameters and Radiomics Features Analysis in Advanced Nsclc Treated with Immunotherapy as Predictors of Therapy Response and Survival. *Cancers (Basel)* 2020; **12**(5).
69. Tomaszewski MR, Gillies RJ. The Biological Meaning of Radiomic Features. *Radiology* 2021; **298**(3): 505-16.
70. Fournier L, Costaridou L, Bidaut L, et al. Incorporating radiomics into clinical trials: expert consensus endorsed by the European Society of Radiology on considerations for data-driven compared to biologically driven quantitative biomarkers. *Eur Radiol* 2021; **31**(8): 6001-12.
71. Lambin P, van Stiphout RG, Starmans MH, et al. Predicting outcomes in radiation oncology--multifactorial decision support systems. *Nat Rev Clin Oncol* 2013; **10**(1): 27-40.

Supplementary Table 1. Random permutation test for the three different outcomes

	OS (C-index)		TTP (C-index)		pCR (AUC)	
Baseline	Apparent	Cross-validated	Apparent	Cross-validated	Apparent	Cross-validated
Clinical (SOURCE)	0.54	0.45	0.55	0.45	0.50	0.56
Radiomics CT (Baseline)	0.50	0.50	0.55	0.52	0.50	0.45
Radiomics CT (Re-staging)	0.59	0.53	0.66	0.56	0.62	0.48
Radiomics PET (Re-staging)	0.50	0.50	0.61	0.56	0.74	0.67
cfDNA (Baseline)	0.56	0.54	0.50	0.52	0.50	0.48
SOURCE+Radiomics CT (Baseline)	0.55	0.51	0.50	0.47	0.50	0.44
SOURCE+Radiomics CT (Re-staging)	0.50	0.47	0.66	0.52	0.62	0.51
SOURCE+Radiomics PET (Re-staging)	0.64	0.48	0.59	0.57	0.70	0.59
SOURCE+cfDNA (Baseline)	0.56	0.53	0.58	0.53	0.50	0.54
SOURCE+Radiomics CT+cfDNA (Baseline)	0.57	0.50	0.50	0.47	0.50	0.44
SOURCE+Radiomics CT+cfDNA (Re-staging)	0.50	0.46	0.64	0.49	0.67	0.52
SOURCE+Radiomics PET+cfDNA (Re-staging)	0.64	0.50	0.63	0.56	0.79	0.58
Delta Radiomics CT	NA	NA	NA	NA	0.69	0.59
SOURCE+Delta radiomics CT	NA	NA	NA	NA	0.79	0.67
SOURCE+Delta radiomics CT+cfDNA	NA	NA	NA	NA	0.79	0.66

Supplementary Table 2. Image Biomarker Standardisation Initiative (IBSI) Reporting Guidelines.

Patient	
Volumes of interest	Baseline: primary gross tumor volume in the esophagus and gastro-esophageal junction Re-staging: remaining and former primary tumor volume
Patient preparation • Patient instructions • Drugs • Equipment	Patients were advised to fast for at least 6 hours before [^{18}F]FDG-PET imaging. Additional instructions included to drink 2L of water 24 hours before the PET-scan and 1L on the day of the PET-scan. No heavy physical exercise was allowed 24 hours before the PET-scan.
Radioactive tracer • Tracer • Administration method • Injected activity • Uptake time prior to acquisition • Competing substances	PET acquisition was started 60 (55-75) minutes after intravenous administration of [^{18}F]FDG (3 MBq/kg of body weight) in a peripheral vein. Serum glucose levels were in the range of 4 mmol/l and 11 mmol/l.
Contrast agent	Intravenous (Ultravist 300) and oral (Telebrix 5%).
Comorbidities	Patients with diabetes mellitus were allowed as long as serum glucose was below 11.1 mmol/L. Separate instructions were given to patients with diabetes mellitus to ensure glucose levels were within the range of 4 mmol/l and 11 mmol/l.
Acquisition	
Acquisition protocol	For patients with both a CT and PET-scan: first, a diagnostic CT scan was performed, followed by a static PET-scan from the skull base up until the thighs of the patient approximately 60 minutes post-injection. In case only a diagnostic CT-scan was performed this was done from the neck up until the abdomen.
Scanner type	Scans were acquired on the following devices: Philips: iCT 256, Brilliance 64, Gemini-GXL 16, Gemini TF 16 Siemens: SOMATOM Definition AS, Edge or Force, Sensation 64, Biograph mCT 40, Biograph mCT 128 GE medical: Discovery CT750 HD Toshiba: Aquilion PRIME
Imaging modality	PET-CT or CT
Static/dynamic scans	Static
Scanner calibration	Cross-calibration was performed accord to the EARL-accreditation program and the guidelines for quality control of the Dutch Association of Nuclear Medicine (NVNG).
Patient instructions	Free-breathing PET and CT scans were acquired with the patient instructed not to move and positioned in restraining/supportive devices.
Anatomical motion correction	No anatomical motion correction was performed.
Scan duration	1.5-5 min per bed position
Tube voltage CT	Range 100-140 kVp Median: 120 kVp
Tube current CT	Range 36-701 mA Median: 217
Time-of-flight	Re-staging (105/105) 100%
Reconstruction	
In plane resolution	PET: 2.04×2.04 – 4.11×4.11 mm^2 CT: 0.45×0.45 – 1.52×1.52 mm^2
Image slice thickness	PET: 2-5 mm CT: 2-5 mm
Image slice spacing	PET: 2-5 mm CT: 2-5 mm
Convolution kernel	Philips: A or B Siemens: B40f, B60f, I30f;3, I31f;3 GE medical: Standard Toshiba: FC07, FC12
Reconstruction method	PSF+TOF 2i21s and PSF+TOF 4i21s BLOB-OS-TF
Point spread function modelling	Re-staging (52/105) 49.5%
Image corrections • Attenuation correction • Other corrections	Attenuation correction based on CT. Correction for scatter, randoms, normalization, dead time and physical decay was applied.

Image processing – data conversion	
SUV normalisation	Body weight (in grams)
Other data conversions	NA
Image processing – post acquisition processing	
Anti-aliasing	NA
Noise suppression	NA
Post-reconstruction smoothing filter	5-7.5 mm FWHM Gaussian kernel
Intensity normalisation	NA
Other post-acquisition processing methods	NA
Segmentation	
Segmentation method • Method • Number of experts, expertise, consensus strategies • Settings • Images	The PET VOIs were delineated semi-automatically using 3DSlicer (version 4.11; www.slicer.org) and in-house built software implemented in Python 3.7.2 (Python Software Foundation, Wilmington, Delaware). VOI were delineated on the [^{18}F]FDG-PET scans using an isocontour that applies a threshold of 50% of the peak standardized uptake value (SUVpeak), obtained using a sphere 1 cm^3, corrected for local background. Boxing was applied to exclude surrounding [^{18}F]FDG-avid tissues. For the CT images manual delineation of the primary gross tumor volume (GTV) in MIM Maestro (MIM software, Cleveland, Ohio) was performed by TvdE and peer-reviewed by YW. The GTV delineation was adjusted according to the radiology or endoscopic ultrasound report. In cases with available [^{18}F]FDG-PET scans they were used in conjunction to guide the delineation.
Conversion to mask	NA
Image processing – image interpolation	
Interpolation algorithm • Algorithm • Interpolation grid • Dimensions • Extrapolation	Images were interpolated to isotropic voxels using B-spline interpolation, with grids aligned by the input origin and only covering the VOI (PyRadiomics default).
Interpolated voxel dimensions	PET: 4×4×4 mm^3 CT: 1×1×1 mm^3
Image processing – ROI interpolation and re-segmentation	
Interpolation algorithm	NA
Partially masked voxels	NA
Re-segmentation methods	NA
Image processing – discretisation	
Discretisation method • Method • Number of bins/ bin size • Lowest intensity first bin	Discretisation using a fixed bin size. Bin edges were equally spaced from 0 (e.g., 0-0.5, 0.5-1, etc) and the lowest grey value was discretized into the first bin. PET: 0.5 g/mL CT: 25 HU
Image processing – image transformation	
Image filter	NA
Image biomarker computation	

Biomarker set (PyRadiomics nomenclature, if IBSI nomenclature differed, it was added in brackets)	PET and CT • First Order Statistics (18 features): 10th Percentile, 90th Percentile, Energy, Entropy (Intensity Histogram Entropy), Interquartile Range, Kurtosis, Maximum, Mean Absolute Deviation, Mean, Median, Minimum, Range, Robust Mean Absolute Deviation, Root Mean Squared, Skewness, Total Energy (not present in IBSI definitions), Uniformity (Intensity histogram uniformity), Variance • Shape based (14 features): Elongation, Flatness, Least Axis Length, Major Axis Length, Maximum 2D Diameter Column, Maximum 2D Diameter Row, Maximum 2D Diameter Slice, Maximum 3D Diameter, Mesh Volume (Volume), Minor Axis Length, Sphericity, Surface Area, Surface Volume Ratio, Voxel Volume (Approximate Volume) • Grey Level Co-occurrence Matrix (GLCM; 22 features): Autocorrelation, Joint Average, Cluster Prominence, Cluster Shade, Cluster Tendency, Contrast, Correlation, Difference Average, Difference Entropy, Difference Variance, Joint Energy (Angular Second Moment), Joint Entropy, Informational Measure of Correlation 1, Informational Measure of Correlation 2, Inverse Difference Moment, Inverse Difference Moment Normalized, Inverse Difference, Inverse Difference Normalized, Inverse Variance, Maximum Probability (Joint Maximum), Sum Entropy, Sum of Squares (Joint Variance) • Grey Level Run Length Matrix (GLRLM; 16 features): Short Run Emphasis, Long Run Emphasis, Grey Level Non-Uniformity, Grey Level Non-Uniformity Normalized, Run Length Non-Uniformity, Run Length Non-Uniformity Normalized, Run Percentage, Grey Level Variance, Run Variance, Run Entropy, Low Grey Level Run Emphasis, High Grey Level Run Emphasis, Short Run Low Grey Level Emphasis, Short Run High Grey Level Emphasis, Long Run Low Grey Level Emphasis, Long Run High Grey Level Emphasis • Grey Level Size Zone Matrix (GLSZM; 16 features): Small Area Emphasis (Small Zone Emphasis), Large Area Emphasis (Large Zone Emphasis), Grey Level Non-Uniformity, Grey Level Non-Uniformity Normalized, Size-Zone Non-Uniformity (Zone Size Non-Uniformity), Size-Zone Non-Uniformity Normalized (Zone Size Non-Uniformity Normalized), Zone Percentage, Grey Level Variance, Zone Variance (Zone Size Variance), Zone Entropy (Zone Size Entropy), Low Grey Level Zone Emphasis, High Grey Level Zone Emphasis, Small Area Low Grey Level Emphasis (Small Zone Low Grey Level Emphasis), Small Area High Grey Level Emphasis (Small Zone High Grey Level Emphasis), Large Area Low Grey Area Emphasis (Large Zone Low Grey Level Emphasis), Large Area High Grey Level Emphasis (Large Zone High Grey Level Emphasis) • Grey Level Dependence Matrix (GLDM; 14 features): Small Dependence Emphasis (Low Dependence Emphasis), Large Dependence Emphasis (High Dependence Emphasis), Grey Level Non-Uniformity, Dependence Non-Uniformity (Dependence Count Non-Uniformity), Dependence Non-Uniformity Normalized (Dependence Count Non-Uniformity Normalized), Grey Level Variance, Dependence Variance (Dependence Count Variance), Dependence Entropy (Dependence Count Entropy), Low Grey Level Emphasis (Low Grey Level Count Emphasis), High Grey Level Emphasis (High Grey Level Count Emphasis), Small Dependence Low Grey Level Emphasis (Low Small Dependence Low Grey Level Emphasis), Small Dependence High Grey Level Emphasis (Low Dependence High Grey Level Emphasis), Large Dependence Low Grey Level Emphasis (High Dependence Low Grey Level Emphasis), Large Dependence High Grey Level Emphasis (High Dependence Low Grey Level Emphasis) • Neighbouring Grey Tone Difference Matrix (NGTDM; 5 features): Coarseness, Contrast, Busyness, Complexity, Strength
IBSI compliance	Yes
Robustness	Not assessed
Software availability	PyRadiomics 3.01 in Python 3.7.2 (Python Software Foundation, Wilmington, Delaware)
Image biomarker computation – texture parameters	
Texture matrix aggregation	GLCM and GLRLM: 3D: average; GLSZM, GLDM and NGTDM: 3D
Distance weighting	No weighting
Cooccurrence matrix symmetry	Symmetric
Cooccurrence matrix distance	Chebyshev distance of 1
Size zone matrix linkage distance	Chebyshev distance of 1
Distance zone matrix linkage distance	NA
Distance zone matrix distance norm	NA
Neighbouring grey tone difference matrix distance	Chebyshev distance of 1

Grey level dependence matrix distance	Chebyshev distance of 1
Grey level dependence matrix coarseness	0
Machine learning and radiomic analysis	
Diagnostic and prognostic modelling	Documented in the section below using the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD, version October 1 2020) Checklist: Prediction Model Development and Validation.
Comparison with known factors	SOURCE esophageal cancer survival prediction model
Multicollinearity	Prior to using the radiomic features as input for the prediction models an initial feature selection of the 105 features was performed to remove highly correlating features (≥0.75) using the redundancy filter algorithm based on a Spearman correlation matrix within the FMradio (Factor Modelling for Radiomics Data) R-package (version 1.1.1). This was done for for each time point (baseline, follow-up) and scan-type (CT, PET).
Model availability	The models in this study are available from the corresponding author on reasonable request.
Data availability	The datasets in this study are available from the corresponding author on reasonable request.
TRIPOD checklist	
Objectives	We investigated whether the addition of radiomic and/or cfDNA features can enhance the predictive power of the SOURCE model for survival and response prediction.
Source of data	Patients from the prospectively collected BIOES Amsterdam UMC biobank (METC 2013_241) or patients included in a phase II non-randomized feasibility trial ((NCT03087864; PERFECT).
Participants	In total 111 stage II-III (M0) resectable esophageal or gastroesophageal junction adenocarcinoma patients were included in this study treated consecutively between 2014 and 2019 in the Amsterdam UMC (n=104) or UMC Utrecht (n=7). Neoadjuvant treatment consisted in 40 patients of neoadjuvant chemoradiotherapy according to CROSS (nCRT) combined with PD-L1 immune checkpoint inhibition (ICI) and in 71 patients of nCRT only. Patient characteristics are described in table 1 of the manuscript.
Outcome	Improvement in survival or response prediction for resectable esophageal adenocarcinoma patients by combining a clinical model with radiomics and cfDNA data.
Predictors	Clinical: linear predictor of SOURCE (OS or TTP), clinical variables from SOURCE (pCR) Radiomic features as specified above from baseline CT, re-staging CT or re-staging PET. cfDNA: short fragments (P20-150), somatic copy number aberrations (IchorCNA), fragment end sequence score (FrEIA) and detection of mutation with a VAF>1%.
Statistical analysis	ComBat post-processing harmonization to correct for the two main sources of data variability namely: convolution kernel reflected by manufacturer (Philips, Siemens, Toshiba, GE healthcare) and slice thickness (3 mm or less vs. 4-5 mm). Prior to predictive modelling, for each time point (baseline, follow-up) and scan-type (CT, PET) an initial feature selection of the 105 radiomic features was performed to remove highly correlating features (≥0.75) using the redundancy filter algorithm based on a Spearman correlation matrix within the FMradio (Factor Modelling for Radiomics Data) R-package (version 1.1.1). OS and TTP were modelled using an elastic net Cox regression and pCR was modelled using a logistic elastic net regression. The lambda parameter and alpha penalty mixing parameter were optimized using a 10-fold cross-validation scheme,. Additionally, internal-external five-fold cross-validation was performed to assess performance of the model on data on which the model was not trained (80% training and 20% validation).
Model performance	C-index for OS (Figure 2A) and TTP (Figure 2B) AUC for pCR (Figure 4)

Supplementary Table 3. Features selected for each model

OS models	
Clinical (SOURCE)	Linear predictor
Radiomics CT (Baseline)	2 features; gldm: SmallDependenceEmphasis, ngtdm: Strength
Radiomics CT (Re-staging)	13 radiomic features; 3 shape, 1 first order, 1 glcm, 3 glszm, 4 gldm, 1 ngtdm
Radiomics PET (Re-staging)	3 features; firstorder: Skewness, gldm: DependenceNonUniformity, ngtdm: Contrast
cfDNA (Baseline)	Freia baseline
SOURCE+Radiomics CT (Baseline)	1 feature; gldm: SmallDependenceEmphasis
SOURCE+Radiomics CT (Re-staging)	Linear predictor, 14 radiomic features; 3 shape, 1 first order, 1 glcm, 3 glszm, 5 gldm, 1 ngtdm
SOURCE+Radiomics PET (Re-staging)	Linear predictor, 3 features; firstorder: Skewness, gldm: DependenceNonUniformity, ngtdm: Contrast
SOURCE+cfDNA (Baseline)	Linear predictor, cfDNA: baseline VAF > 1pct, Freia baseline
SOURCE+Radiomics CT+cfDNA (Baseline)	1 feature; Gldm: SmallDependenceEmphasis
SOURCE+Radiomics CT+cfDNA (Re-staging)	Linear predictor, cfDNA: Freia baseline, 2 features; shape: LeastAxisLength, glcm: ClusterProminence
SOURCE+Radiomics PET+cfDNA (Re-staging)	Linear predictor, 2 features; firstorder: Skewness, ngtdm: Contrast
TTP models	
Clinical (SOURCE)	Linear predictor
Radiomics CT (Baseline)	4 features; glrlm: RunEntropy, gldm: SmallDependenceEmphasis, ngtdm: Coarseness, ngtdm: Strength
Radiomics CT (Re-staging)	1 feature; ngtdm: Coarseness
Radiomics PET (Re-staging)	3 features; firstorder: Skewness, gldm: DependenceNonUniformity, ngtdm: Contrast
cfDNA (Baseline)	cfDNA: P20-150, ichorCNA, baseline VAF > 1pct, Freia baseline
SOURCE+Radiomics CT (Baseline)	Linear predictor, 4 features; glrlm: RunEntropy, gldm: SmallDependenceEmphasis, ngtdm: Coarseness, ngtdm: Strength
SOURCE+Radiomics CT (Re-staging)	Linear predictor, 5 features; shape: LeastAxisLength, shape: Maximum3DDiameter, shape: Sphericity, glcm: ClusterProminence, ngtdm: Coarseness
SOURCE+Radiomics PET (Re-staging)	Linear predictor, 4 features; firstorder: Skewness, glcm: Ldmn, gldm: DependenceNonUniformity, ngtdm: Contrast
SOURCE+cfDNA (Baseline)	Linear predictor, cfDNA: P20-150, ichorCNA, baseline VAF > 1pct, Freia baseline
SOURCE+Radiomics CT+cfDNA (Baseline)	Linear predictor, baseline VAF > 1pct, Freia baseline, 4 features; glrlm: RunEntropy, gldm: SmallDependenceEmphasis, ngtdm: Coarseness, ngtdm: Strength
SOURCE+Radiomics CT+cfDNA (Re-staging)	Linear predictor, cfDNA: baseline VAF > 1pct, Freia baseline, 3 features; shape: LeastAxisLength, shape: Maximum3DDiameter, ngtdm: Coarseness

SOURCE+Radiomics PET+cfDNA (Re-staging)	Linear predictor, cfDNA: baseline VAF > 1pct, Freia baseline, 4 features; firstorder: Skewness, glcm: Ldmn, gldm: DependenceNonUniformity, ngtdm: Contrast
pCR models	
Clinical (SOURCE)	Intercept (no variables selected)
Radiomics CT (Baseline)	9 features: 1 shape, 2 glcm, 1 glrlm, 1 glszm, 2 gldm, 2 ngtdm
Radiomics CT (Re-staging)	2 features: glszm: ZonePercentage, gldm: DependenceVariance
Radiomics PET (Re-staging)	8 features: shape: SurfaceVolumeRatio, firstorder: Skewness, glcm: Ldmn, glszm: GrayLevelNonUniformity, glszm: SizeZoneNonUniformity, gldm: SmallDependenceLowGrayLevelEmphasis, ngtdm: Busyness, ngtdm: Contrast
cfDNA (Baseline)	cfDNA: P20-150
SOURCE+Radiomics CT (Baseline)	Intercept, 4 features; glrlm: RunEntropy, glszm: SizeZoneNonUniformity, gldm: DependenceEntropy, gldm: SmallDependenceEmphasis
SOURCE+Radiomics CT (Re-staging)	glszm: ZonePercentage
SOURCE+Radiomics PET (Re-staging)	Intercept (no variables selected)
SOURCE+cfDNA (Baseline)	Intercept (no variables selected)
SOURCE+Radiomics CT+cfDNA (Baseline)	Intercept, 4 features: glrlm: RunEntropy, glszm: SizeZoneNonUniformity, gldm: DependenceEntropy, gldm: SmallDependenceEmphasis
SOURCE+Radiomics CT+cfDNA (Re-staging)	BMI, Albumin, Hb, LDH, age, *c*T, cN, tumor location, differentiation grade, cfDNA: P20-150, ichorCNA, baseline VAF > 1pct, Freia baseline, 30 features; 6 shape, 2 first order, 5 glcm, 6 glszm, 6 gldm, 5 ngtdm
SOURCE+Radiomics PET+cfDNA (Re-staging)	*c*T, cfDNA: P20-150, shape: SurfaceVolumeRatio, firstorder: Skewness, glcm: Ldmn, glszm: GrayLevelNonUniformity, glszm: SizeZoneNonUniformity, ngtdm: Contrast
Delta Radiomics CT	1 feature; glszm: ZoneEntropy
SOURCE+Delta radiomics CT	1 feature; glszm: ZoneEntropy
SOURCE+Delta radiomics CT+cfDNA	Intercept (no variables selected)

Chapter 11

Neoadjuvant chemoradiotherapy versus definitive chemoradiotherapy for patients with esophageal cancer: Development and validation of a counterfactual prediction model on real-world data

Steven C. Kuijper*, Thomas Klausch*, Bastiaan R. Klarenbeek, Peter S.N. van Rossum, Rob H.A. Verhoeven, Hanneke W.M. van Laarhoven

* Shared first authorship

Abstract

Background. Counterfactual prediction modeling is a method that estimates outcomes under alternative treatment strategies. Unlike traditional clinical models based on observational data, which are inherently unusable for optimal treatment selection, counterfactual models can be used for treatment optimization through predicting individualized outcomes for hypothetical treatment scenarios. In this proof-of-concept study we developed and validated counterfactual prediction models for esophageal cancer, comparing neoadjuvant chemoradiotherapy and definitive chemoradiotherapy.
Methods. We used data from the Netherlands Cancer Registry to develop and validate a counterfactual prediction model for esophageal cancer treatments. Data from 4,388 patients (2015-2019) was used for training, and 1,693 patients (2020-2021) for validation. The focus was on non-metastatic esophageal cancer treated with neoadjuvant (nCRT) or definitive chemoradiotherapy (dCRT) for which we used the target trial framework to emulate this trial with observational data. Inverse propensity score weighting to control for confounding, and parametric survival and random survival forest models were developed and validated using counterfactual calibration and survival curves. Finally, the counterfactual performance of the existing SOURCE prediction model was tested.
Results. Inverse propensity score weighting ensured acceptable covariate balance in the training and validation data. External validation revealed that the random forest model demonstrated good counterfactual calibration for nCRT (intercept: 0.12 95%CI(-0.01-0.26) ; slope: 0.79 95%CI (0.53-1.05)) and dCRT (intercept: 0.24 (0.09-0.39); slope: 0.42 (0.14-0.71)) and aligned more closely to the true counterfactual survival curve up until 2.5 years post diagnosis. The SOURCE model demonstrated good calibration in counterfactual predictions despite not being specifically designed for counterfactual analysis or individualized treatment effects.
Conclusion. This study's development and validation of counterfactual prediction models for esophageal cancer treatments underscore their potential to improve personalized decision-making. By enabling patients and clinicians to evaluate individualized survival probabilities up to 2.5 years under different treatment strategies, these models can facilitate treatment choices, ultimately enhancing patient outcomes in clinical settings where multiple viable options are available.

Introduction

In every day oncological care, physicians and patients face the challenge of making the right choice. A choice that not only reflects medical effectiveness but also aligns with the patient's personal needs and preferences. In this context of shared decision making, clinical prediction models have been developed to support the decision making process. For example, by predicting expected survival for a particular treatment. [1–3] Ideally, such models can be used to predict outcomes for a number of different treatment options, which would help patients and physicians in choosing the right treatment.

The main challenge of clinical prediction models, however, is that most models are inherently not capable of making predictions under different outcomes.[4] While these models can be used to predict outcomes once a treatment has been chosen, they do not provide insights into what might happen if a different treatment strategy would be applied, i.e. the counterfactual outcome. Another common shortcoming of prediction models is failure to include treatment effect heterogeneity. This assumes that the effect of a particular treatment on the outcome is equal for all patients, when in reality it is highly likely that the effect of treatment is different for each patient. To address these issues, counterfactual predictive models are needed.[5,6] Counterfactual prediction models estimate the outcome, or risk for an event like death, for a patient under different treatment strategies while taking causal aspects into account during model development.[4,7] This is crucial for personalized medicine, as it can support treatment decision making.[8]

The key reason why most predictive models are inherently not suitable for causal predictions, is due to the type of data they are usually trained on. An important source of data used for the development of predictive models, for example, are registries based on data from electronic health records. These data sets are observational, as treatment outcomes are observed under the treatments that have been assigned to a patient in clinical practice. This can be considered a limitation, since it only reflects what has already occurred, which means the data do not offer information about what might have happened if a patient had received a different treatment (the counterfactual). This problem, known as the fundamental problem of causal inference, makes it challenging to estimate effects of different treatments on an individual level.

A second challenge in observational data is confounding, which denotes the situation when prognostic patient characteristics differ systematically between treatment groups. Confounding is a consequence of non-random treatment assignment in observational data and biases model predictions if not all relevant confounders are controlled for during model development. Furthermore, since confounding leads to differences in patient mix across treatment groups, a predictive model developed on a group of patients receiving a specific treatment may perform adequately in this group but not in the population as a whole. Consequently, the goal of counterfactual prediction modeling is to develop predictive models that work well for the population as a whole under any treatment option considered.

To ou knowledge, such counterfactual prediction models are not available within the field of oncology. Therefore, the aim of this proof-of-concept study was to develop and validate counterfactual prediction models that can estimate individualized survival probabilities under two different treatment strategies, based on real-world, observational data using statistical and machine learning models. We used data from patients diagnosed with esophageal cancer who underwent either neoadjuvant chemoradiotherapy followed by surgery

(nCRT) or observation according to the SANO protocol,[9] or definitive chemoradiotherapy (dCRT). Both treatments are viable options for potentially curable esophageal cancer, but the choice between these strategies in practice can depend on various factors such as patient age, physical fitness, and personal preferences.[10] Given that there are no models existent that can help to assess the optimal treatment for a patient, the proposed counterfactual prediction models could be highly valuable tool in clinical daily practice.

Methods

We used the target trial framework to emulate a trial.[11] This framework facilitates preparing the observational data in a way that they closely resemble data that would have been obtained if during a clinical randomized controlled comparison of neoadjuvant chemoradiotherapy to definitive chemoradiotherapy was carried out (Table 1).

Target trial emulation

Data for this study was obtained from the Netherlands Cancer Registry (NCR). The NCR is a nationwide population-based database that systematically collects data on all cancer diagnoses in the Netherlands. Trained data managers regularly retrieve details about diagnosis, tumor stage, and treatment from patients' electronic medical records and incorporate this information into the NCR. Identification of cases primarily relies on notifications from PALGA, the nationwide network and registry for histopathology and cytopathology in the Netherlands. A total of 4,388 patients diagnosed between 2015-2019 with esophageal cancer were retrieved from the NCR which were used for training the models.

We included patients with primary non-metastatic (cT1-4a, cN0-3, cM0) esophageal and gastroesophageal junction adenocarcinoma or squamous cell carcinoma who either were treated with neoadjuvant chemoradiotherapy according to nCRT or dCRT. Neoadjuvant chemoradiotherapy consists of five cycles chemotherapy (carboplatin-paclitaxel) combined with radiotherapy with a radiation dose of 41.4 Gy in 23 fractions followed by a resection of the tumor or watchful waiting according to the SANO protocol.[12] Definitive chemoradiotherapy consists of chemotherapy and radiotherapy without surgery, but with a higher dose of radiation (50.4 Gy in 28 fractions) and six cycles of chemotherapy.[13]

Patients were not randomly assigned but were sourced from observational real-world data. To control for unobserved confounding, we included 15 predetermined treatment outcome-related variables: sex, age, histological subtype,, WHO performance status, cT, cN, hemoglobin, BMI, comorbidities, creatinine, LDH, albumin, tumor differentiation grade, and tumor sub-localization which were chosen based on expert knowledge and as well as known to be predictive features.[14] The follow-up period started at the beginning of treatment and ended at death or loss to follow-up, with a maximum follow-up time of 9 years. The primary outcome was overall survival, employing an intention-to-treat analysis to assess treatment-specific counterfactual risks, with the survival risk difference serving as our primary causal contrast.

Validation data

A cohort of patients (n=1,693) diagnosed between 2020-2021 with non-metastatic esophageal cancer that had followed either neoadjuvant chemoradiotherapy of definitive chemoradiotherapy was identified in the NCR and was used as model validation cohort.

Table 1. Target trial emulation of the data underlying this study.

	Emulated target trial
Eligibility	Patients with primary non-metastatic (cM0) esophageal or gastro-esophageal junction cancer (cT1-4a, cN0-3) who were diagnosed between 2015-2023 and are registered in the Netherlands Cancer Registry
Treatment	Neoadjuvant chemoradiotherapy, which consists of: 5 Cycles of carboplatin-paclitaxel combined with a radiation dose of 41.4 Gy in 23 fractions followed by a resection of the tumour or watchful waiting according to the SANO protocol. Definitive chemoradiotherapy, which consists of 6 Cycles of carboplatin-paclitaxel combined to a radiation dose of 50.4 Gy in 28 fractions
Assignment procedure	Participants were not randomly assigned, but obtained from observational real-world data. To control for confounding we will include the following pre-determined treatment outcome-related variables: Sex, age, histological subtype, performance status, cN, cT, haemoglobin, BMI, comorbidities, creatinine, LDH, albumin, tumour differentiation grade, tumour sub-localization.
Follow-up	Starts at start of treatment and ends at death or loss to follow-up. Maximum follow-up time was 9 years.
Outcome	Overall survival
Causal contrast	Intention to treat analysis of the treatment-specific counterfactual risks with survival risk difference as primary causal contrast

Model development

Our aim was to develop models that predict survival probabilities under two different treatments. To this end, the training and validation data were required to be suitable for causal inference by ensuring the data satisfied the conditions for causal inference before modelling and validation.

Inverse propensity score weighting

We employed inverse propensity score weighting to control for confounding in the training data and the validation data.[15] These aforementioned variables were used in a multivariable (logistic) regression model to estimate the propensity of being treated.

To ensure that the data satisfied the requirement for causal inference, we checked for covariate overlap and covariate balance before and after weighting. Covariate balance implies that all levels of the covariates are present in both treatment arms. Multivariate covariate overlap was assessed via visual inspection of the linearized propensity scores. Propensity score trimming of the 98th percentiles was applied in both the train and validation data to improve multivariate overlap, after which we removed patients with linearized propensity scores <-3. Univariate balance was assessed through calculation of the standardized mean difference (SMD) between covariates in both treatment arms. A SMD≤0.1 was considered good balance. Interaction effects and higher-order polynomials were added in the logistic model until all covariates had an SMD≤0.1.

Model training

Two types of models, a classical statistical model and a machine learning model, were trained on the data to test which type of model was most suitable for the counterfactual model: a multivariable parametric survival model and a random survival forest model. Each model was trained separately on each of the two treatment arms, resulting in four different models. By training the models on treatment-specific data, we modelled treatment effect heterogeneity (i.e. individual treatment effects) implicitly without the need to include treatment-predictor interactions. Specifically, we can obtain one prediction of survival probability for each treatment from each model, and compare the difference in predicted probabilities (individualized causal risk difference) for an assessment of treatment effectiveness. For both models we used the same set of confounders as predictors that were used in the calculation of the propensity score. During training of the models, the estimated weights obtained from the inverse propensity score weighting were applied.

Parametric survival models require specification of the baseline hazard function. We tested six different base functions (i.e. exponential, Weibull, gamma, lognormal, log-logistic, and the generalized gamma) and evaluated model fit (using log-likelihood).

Random survival forests require two hyper-parameters: the number of trees and the number of variables to randomly sample as candidates at each split (mtry). With a 5-fold cross-validation of a range of values for the hyperparameter we tested which combination of hyperparameters produced the lowest prediction error across cross-validations. Propensity score weights were added in model training by letting the selection probability of bootstrapped cases depend on the relative size of the weights, as implemented in the R package 'ranger'.[16]

Existing model: SOURCE prediction model

In addition to the development of counterfactual models, we also tested the SOURCE prediction model to estimate counterfactual probabilities on the validation data. The SOURCE prediction model was published in 2021, is aimed at predicting survival probabilities for patients with esophageal and gastric cancer, and demonstrated adequate external validation performance.[14,17] The SOURCE model is a Cox regression model that models treatment as a main effect. We applied the SOURCE model to investigate to what extent existing prediction models—that were not optimized for counterfactual estimation—can in-fact be used for counterfactual prediction. All model parameters (e.g. hazard ratios), including a Breslow estimate of the baseline hazard function were available to produce predictions on the validation data.

Counterfactual validation

To validate the models, we employed both internal and external counterfactual validation techniques. Counterfactual validation involves assessing the model's ability to predict outcomes for a treatment that was not actually received by the patient, as opposed to merely validating against observed outcomes. For instance, if a patient received nCRT and we have observed their survival outcome under this treatment, we aim to predict and validate what their outcome would have been under dCRT. Although the actual outcome under dCRT is unobservable for this patient, we can estimate this counterfactual outcome using inverse propensity score weighting. [6]

When we apply inverse propensity score weighting to one treatment arm, the distribution of covariates in the treatment arm mirrors the distribution of covariates that would have existed if the whole population had been treated with that treatment.[15] This weighted sample then serves as a pseudopopulation that reflects what the outcomes would look like if all patients, regardless of their actual treatment, had received that particular treatment. By applying these weights, we can estimate what the outcomes would be for patients under the alternative treatment scenario, thereby validating the model's predictions in a counterfactual context.

All counterfactual validations were performed for on the training data itself (internal counterfactual validation) and on the external sample (external counterfactual validation).

Counterfactual calibration

We evaluated the models' counterfactual validation using counterfactual calibration.[6] Calibration refers to the agreement between predicted survival probabilities by the model and estimated counterfactual survival probabilities.[18] To obtain counterfactual calibration, we obtained survival probabilities at the median overall survival for all patients and split them into 10 equally sized bins. In each bin, the mean survival probability served as the predicted survival probability. To obtain the counterfactual survival probability, we fitted a weighted Kaplan-Meier model (i.e. survival of the pseudopopulation) and evaluated the survival probability at the median over survival time that was used to obtain the model predictions.[6] By plotting the predicted and counterfactual survival probabilities, we obtained the calibration curves. Calibration slope and intercept were estimated and evaluated.

Counterfactual survival curves

We also evaluated the counterfactual predictions in terms of survival curves. To this end, we plotted an estimate of the counterfactual survival by plotting a weighted Kaplan-Meier curve and plotted the predicted survival curves by our counterfactual models.

Counterfactual prediction demonstration

To demonstrate what the counterfactual prediction would look like, we applied the model that performed best on the validation data to ten randomly selected patients, for whom we used the model to predict counterfactual survival curves.

Missing data

All missing data in train and validation data were considered missing at random, and imputed with a random forest imputation using the 'missForest' implementation for R.[19] Separate imputations were performed for each treatment arm and for the train and validation data, resulting in 4 different imputation rounds to prevent any type of dependency between treatment arms and data sources.

Results

Training data

In total, 5,170 patients were selected for the training data in the NCR before any exclusions. Next, we excluded a total of 149 patients with WHO performance status 4 or cT0 to meet the requirement of covariate balance (Table 2). A total of 5,021 patients, of whom 3,845 patients underwent neoadjuvant chemoradiotherapy (nCRT) and 1,176 patients underwent definitive chemoradiotherapy (dCRT). After removal of the 98th percentiles of the propensity score and removal of patients with a linearized propensity score of <-3, a total of 4,230 patients remained in the training data, of whom 3,300 patients underwent nCRT and 903 patients underwent dCRT.

After inverse propensity score weighting using the final model (Formula S1), we obtained good multivariate (Figure S1) and univariate balance (Table 2). All standardized mean differences after weighting were considered acceptable (SMD<0.1).

Model training and internal counterfactual validation

Two models were trained on the training data. The parametric survival model with a generalized gamma baseline hazard function was fitted on the data, as this baseline hazard function demonstrated the best fit (model parameters are reported in Table S1-S2). For the random survival forest model a grid search was performed to find the optimal hyperparameters for the number of trees and mtry. After 5-fold cross-validation, the random survival forest was fitted to the data with mtry=1 and number of trees=50.

Internal validation demonstrated (i.e. fitting the model on the training data) that the parametric survival model had a calibration intercept and slope close to the ideal values in both the nCRT counterfactual treatment arm (intercept: 0.14 95%CI(0.08-0.20)) ; slope: 0.78 95%CI (0.66-0.90)) and in the dCRT counterfactual treatment arm (intercept: 0.14 95%CI (-0.08-0.36)) ; slope: 0.75 95%CI (0.35-1.14)) (Figure 1A-B). The random forest model showed a calibration slope and intercept that were more distant of the ideal diagonal line for both nCRT (intercept: 0.33 95%CI (0.26-0.40) ; slope: 0.34 95%CI (0.24-0.45)) and dCRT treatment (intercept: 0.34 95%CI (0.27-0.42); slope: 0.34 95%CI (0.23-0.45)). This is likely caused by a large number of extreme values of exactly zero and one in the observed survival probability, and could indicate overfitting on the training data.

In the counterfactual survival plots (Figure 1C-D) it can be seen that for both nCRT and dCRT treatments the shape of the counterfactual Kaplan-Meier survival curve was approximated well by both models until roughly 2.5 years post-diagnosis for the nCRT model and 2 years post-diagnosis for the dCRT model.

Table 2. Patient characteristics in the training data after inverse propensity score weighting

		Before weighting			After weighting		
		Definitive chemoradiotherapy	Neoadjuvant chemoradiotherapy	SMD	Definitive chemoradiotherapy	Neoadjuvant chemoradiotherapy	SMD
n		991	3397		4093.5	4305.4	
Sex (%)	1	661 (66.7)	2678 (78.8)	0.275	3067.5 (74.9)	3286.6 (76.3)	0.032
	2	330 (33.3)	719 (21.2)	-	1026.0 (25.1)	1018.8 (23.7)	-
Age (mean (SD))		71.77 (8.44)	65.60 (8.66)	-0,722	66.74 (8.78)	67.47 (8.54)	0,089
Histology (%)	Adenocarcinoma	458 (46.2)	2754 (81.1)	0.777	2954.6 (72.2)	3118.4 (72.4)	0.006
	Squamous cell	528 (53.3)	618 (18.2)	-0.787	1099.8 (26.9)	1159.1 (26.9)	0.001
	Other	5 (0.5)	25 (0.7)	0.029	39.0 (1.0)	27.9 (0.6)	-0.038
WHO performance status (%)	0	285 (28.8)	1876 (55.2)	0.557	2019.8 (49.3)	2070.1 (48.1)	-0.026
	1	520 (52.5)	1402 (41.3)	-0.226	1778.2 (43.4)	1886.6 (43.8)	0.008
	2	170 (17.2)	114 (3.4)	-0.467	275.7 (6.7)	311.6 (7.2)	0.017
	3	16 (1.6)	5 (0.1)	-0.158	19.8 (0.5)	37.1 (0.9)	0.040
cN (%)	0	394 (39.8)	1360 (40.0)	0.006	1513.9 (37.0)	1727.0 (40.1)	0.064
	1	404 (40.8)	1304 (38.4)	-0.049	1606.0 (39.2)	1664.4 (38.7)	-0.012
	2	166 (16.8)	644 (19.0)	0.058	861.8 (21.1)	799.6 (18.6)	-0.065
	3	27 (2.7)	89 (2.6)	-0.006	111.8 (2.7)	114.4 (2.7)	-0.004
cT (%)	1	9 (0.9)	29 (0.9)	-0.006	24.1 (0.6)	38.1 (0.9)	0.032
	2	330 (33.3)	971 (28.6)	-0.102	1147.1 (28.0)	1290.8 (30.0)	0.042
	3	627 (63.3)	2357 (69.4)	0.130	2859.3 (69.9)	2894.5 (67.2)	-0.056
	4	25 (2.5)	40 (1.2)	-0.100	63.0 (1.5)	82.0 (1.9)	0.028
Hemoglobin (mean (SD))		8.37 (1.14)	8.76 (1.06)	0,348	8.68 (1.05)	8.68 (1.06)	0,001
BMI (mean (SD))		24.77 (4.30)	26.19 (4.09)	0,339	25.55 (4.08)	25.80 (4.17)	0,059

Table 2 (Continued). Patient characteristics in the training data after inverse propensity score weighting

Number of comorbidities (%)	0	305 (30.8)	1809 (53.3)	0.468	1955.3 (47.8)	2022.2 (47.0)	-0.017
	1	381 (38.4)	1090 (32.1)	-0.133	1451.2 (35.5)	1493.0 (34.7)	-0.016
	2	305 (30.8)	498 (14.7)	-0.392	687.0 (16.8)	790.2 (18.4)	0.039
Creatinine (mean (SD))		81.09 (28.23)	82.24 (20.27)	0,047	81.30 (22.86)	81.85 (22.01)	0,022
LDH (mean (SD))		194.64 (51.19)	185.51 (40.12)	-0,198	186.15 (43.74)	188.81 (45.01)	0,059
Albumin (mean (SD))		39.50 (5.02)	40.61 (4.05)	0,245	40.26 (4.29)	40.29 (4.15)	0,008
Differentiation Grade (%)	G1	57 (5.8)	126 (3.7)	-0.096	146.7 (3.6)	187.3 (4.3)	0.036
	G2	628 (63.4)	1797 (52.9)	-0.213	2093.7 (51.1)	2371.2 (55.1)	0.080
	G3	306 (30.9)	1474 (43.4)	0.261	1853.1 (45.3)	1747.0 (40.6)	-0.097
Tumor sublocation (%)	Upper thoracic	175 (17.7)	25 (0.7)	-0.612	114.3 (2.8)	126.9 (2.9)	0.007
	Mid-thoracic	215 (21.7)	358 (10.5)	-0.307	663.7 (16.2)	616.2 (14.3)	-0.051
	Lower thoracic	496 (50.1)	2645 (77.9)	0.605	2925.2 (71.5)	3068.8 (71.3)	-0.004
	Junction	49 (4.9)	292 (8.6)	0.146	271.9 (6.6)	340.7 (7.9)	0.050
	Overlapping lesion	21 (2.1)	31 (0.9)	-0.099	50.4 (1.2)	56.4 (1.3)	0.006
	Esophagus NOS	35 (3.5)	46 (1.4)	-0.141	67.9 (1.7)	96.4 (2.2)	0.037

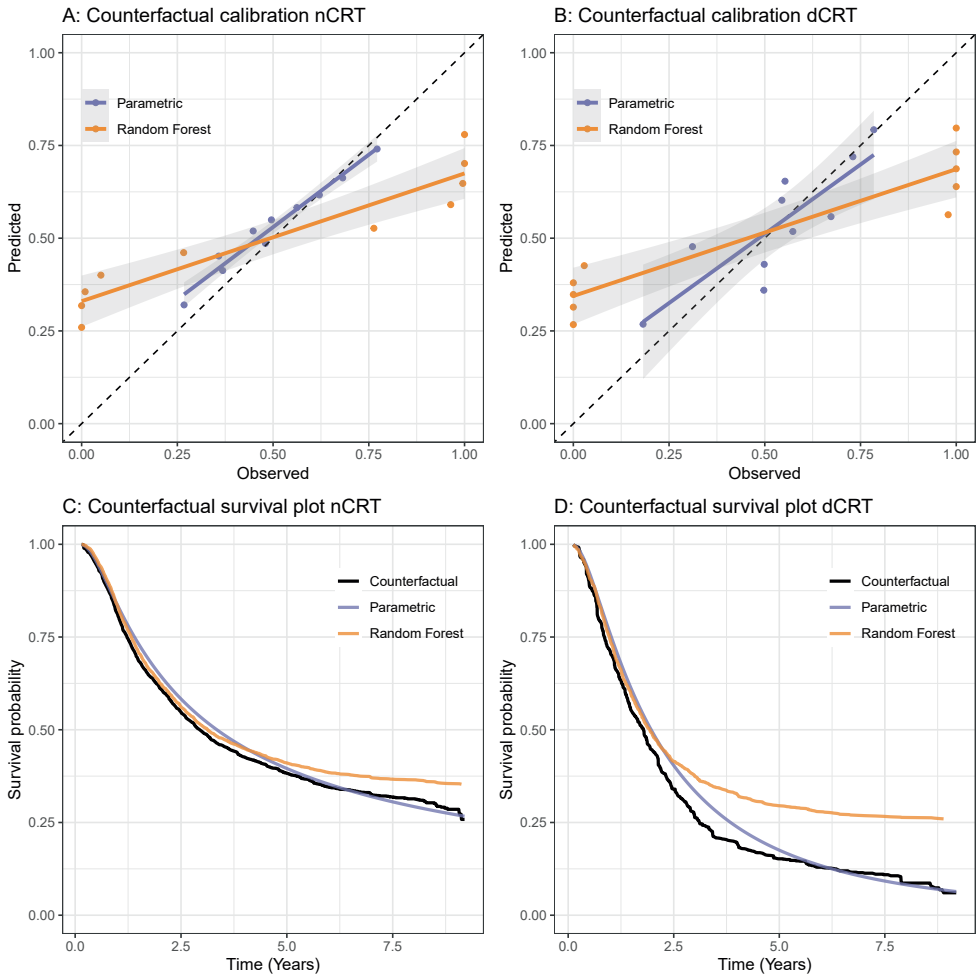


Figure 1. Calibration and survival plots of the trained models on the training data. dCRT = Definitive chemoradiotherapy, nCRT = Neoadjuvant chemoradiotherapy

External counterfactual validation

In total, 1,693 patients were used as external-validation cohort, of whom 1,221 underwent nCRT and 472 underwent dCRT. After removal of the 98th percentiles of the propensity score and removal of patients with a linearized propensity score of <-3 and >4, a total of 1,536 patients remained in the testing data, of whom 1,145 patients underwent nCRT and 418 patients underwent dCRT. After inverse propensity score weighting using the final model (Formula S2), we obtained good multivariate (Figure S2) and univariate covariate balance (Table 3).

External validation of the nCRT model showed calibration relatively close to the ideal line for the parametric model (intercept: 0.09 95%CI (-0.04-0.23); slope: 0.91 95%CI (0.64-1.17)), random forest (intercept: 0.12 95%CI (-0.01-0.26) ; slope: 0.79 95%CI (0.53-1.05)) and SOURCE model (intercept: 0.21 95%CI (0.09-0.33) ; slope: 0.72 95%CI (0.49-0.95)). For the dCRT model, the estimates for the parametric model (intercept: 0.23 95%CI (0.00-0.46); slope: 0.53 95%CI (0.10-0.97)), random forest (intercept: 0.24 95%CI (0.09-0.39) ; slope: 0.42 95%CI (0.14-0.71)) and SOURCE model (intercept: 0.33 95%CI (0.18-0.49); slope: 0.31 95%CI (0.01-0.62)) were lower than for the nCRT model.

Based on visual inspection of the counterfactual survival curves, the random forest model seemed to overlap closest to the counterfactual Kaplan-Meier survival curve for the nCRT model. For the dCRT model, all models performed relatively similar up until 2.5 years post-diagnosis after which the parametric survival model underestimated survival.

Demonstration of individualized counterfactual predictions

The random forest model was used to apply ten randomly selected patients from the validation data (Figure 3). The patient's characteristics are displayed in the plots. The plots demonstrate predicted survival probabilities up to 2.5 years post diagnosis. For eight out of ten randomly selected patients nCRT was the superior treatment according to the model. It can be observed that for two patients (71 years old, T3N1M0, WHO performance status 1; 75 years old, T3N1M0, WHO performance status 1) definitive chemoradiotherapy predicted similar survival outcomes than for neoadjuvant chemoradiotherapy.

Table 3. Patient characteristics in the validation data after inverse propensity score weighting. [a] Level of covariate was removed after propensity score trimming.

		Before weighting			After weighting		
		Definitive chemoradiotherapy	Neoadjuvant chemoradiotherapy	SMD	Definitive chemoradiotherapy	Neoadjuvant chemoradiotherapy	SMD
n		472	1221		1576.6	1642.9	
Sex (%)	1	316 (66.9)	918 (75.2)	0,182	1218.3 (77.3)	1224.7 (74.5)	-0,061
	2	156 (33.1)	303 (24.8)	-	358.3 (22.7)	418.2 (25.5)	-
Age (mean (SD))		66.74 (8.78)	72.32 (8.25)	-0,623	68.02 (9.01)	68.42 (9.20)	0,047
Histology (%)	Adenocarcinoma	221 (46.8)	993 (81.3)	0,771	1138.4 (72.2)	1160.8 (70.7)	-0,035
	Squamous cell	246 (52.1)	219 (17.9)	-0,767	429.2 (27.2)	470.8 (28.7)	0,032
	Other	5 (1.1)	9 (0.7)	-0,034	9.1 (0.6)	11.3 (0.7)	0,012
WHO performance status (%)	0	127 (26.9)	666 (54.5)	0,586	737.6 (46.8)	730.7 (44.5)	-0,048
	1	244 (51.7)	494 (40.5)	-0,227	688.2 (43.6)	729.5 (44.4)	0,015
	2	93 (19.7)	57 (4.7)	-0,472	141.0 (8.9)	170.9 (10.4)	0,046
	3	8 (1.7)	4 (0.3)	-0,137	9.8 (0.6)	11.8 (0.7)	0,010
cN (%)	0	191 (40.5)	502 (41.1)	0,013	634.2 (40.2)	655.1 (39.9)	-0,007
	1	167 (35.4)	458 (37.5)	0,044	593.4 (37.6)	614.6 (37.4)	-0,005
	2	98 (20.8)	233 (19.1)	-0,042	312.1 (19.8)	331.7 (20.2)	0,010
	3	16 (3.4)	28 (2.3)	-0,066	36.9 (2.3)	41.5 (2.5)	0,012
cT (%)	1	7 (1.5)	3 (0.2)	-0,134	8.9 (0.6)	18.7 (1.1)	0,063
	2	108 (22.9)	287 (23.5)	0,015	416.9 (26.4)	370.0 (22.5)	-0,091
	3	346 (73.3)	911 (74.6)	0,030	1126.4 (71.4)	1227.0 (74.7)	0,073
	4	11 (2.3)	20 (1.6)	-0,050	24.4 (1.6)	27.2 (1.7)	0,007
Hemoglobin (mean (SD))		8.68 (1.05)	8.29 (1.08)	0,264	8.43 (1.05)	8.49 (1.17)	0,050
BMI (mean (SD))		25.55 (4.08)	24.87 (4.75)	0,219	25.11 (4.48)	25.52 (4.63)	0,087

Table 3 (Continued). Patient characteristics in the validation data after inverse propensity score weighting. [a] Level of covariate was removed after propensity score trimming.

Number of comorbidities (%)	0	152 (32.2)	582 (47.7)	0,320	620.9 (39.4)	690.5 (42.0)	0,054
	1	169 (35.8)	411 (33.7)	-0,045	618.9 (39.3)	622.6 (37.9)	-0,028
	2	151 (32.0)	228 (18.7)	-0,310	336.8 (21.4)	329.8 (20.1)	-0,031
Creatinine (mean (SD))		81.30 (22.86)	78.64 (26.24)	0,119	82.40 (24.56)	80.74 (22.85)	-0,071
LDH (mean (SD))		186.15 (43.74)	200.16 (51.85)	-0,229	192.68 (37.20)	190.72 (42.61)	-0,047
Albumin (mean (SD))		40.26 (4.29)	39.19 (4.39)	0,303	40.19 (4.51)	40.29 (4.48)	0,023
Differentiation Grade (%)	G1	25 (5.3)	40 (3.3)	-0,100	72.6 (4.6)	63.9 (3.9)	-0,036
	G2	322 (68.2)	696 (57.0)	-0,233	943.2 (59.8)	1009.1 (61.4)	0,033
	G3	125 (26.5)	485 (39.7)	0,284	560.8 (35.6)	569.9 (34.7)	-0,019
Tumor sublocation (%)	Upper thoracic	81 (17.2)	9 (0.7)	-0,601	67.4 (4.3)	94.6 (5.8)	0,058
	Mid-thoracic	113 (23.9)	139 (11.4)	-0,334	254.2 (16.1)	251.9 (15.3)	-0,021
	Lower thoracic	236 (50.0)	943 (77.2)	0,590	1097.6 (69.6)	1126.1 (68.5)	-0,023
	Junction	30 (6.4)	123 (10.1)	0,136	145.2 (9.2)	146.6 (8.9)	-0,010
	Overlapping lesion	2 (0.4)	1 (0.1)	-0,068	a	a	a
	Esophagus NOS	10 (2.1)	6 (0.5)	-0,144	12.2 (0.8)	23.7 (1.4)	0,065

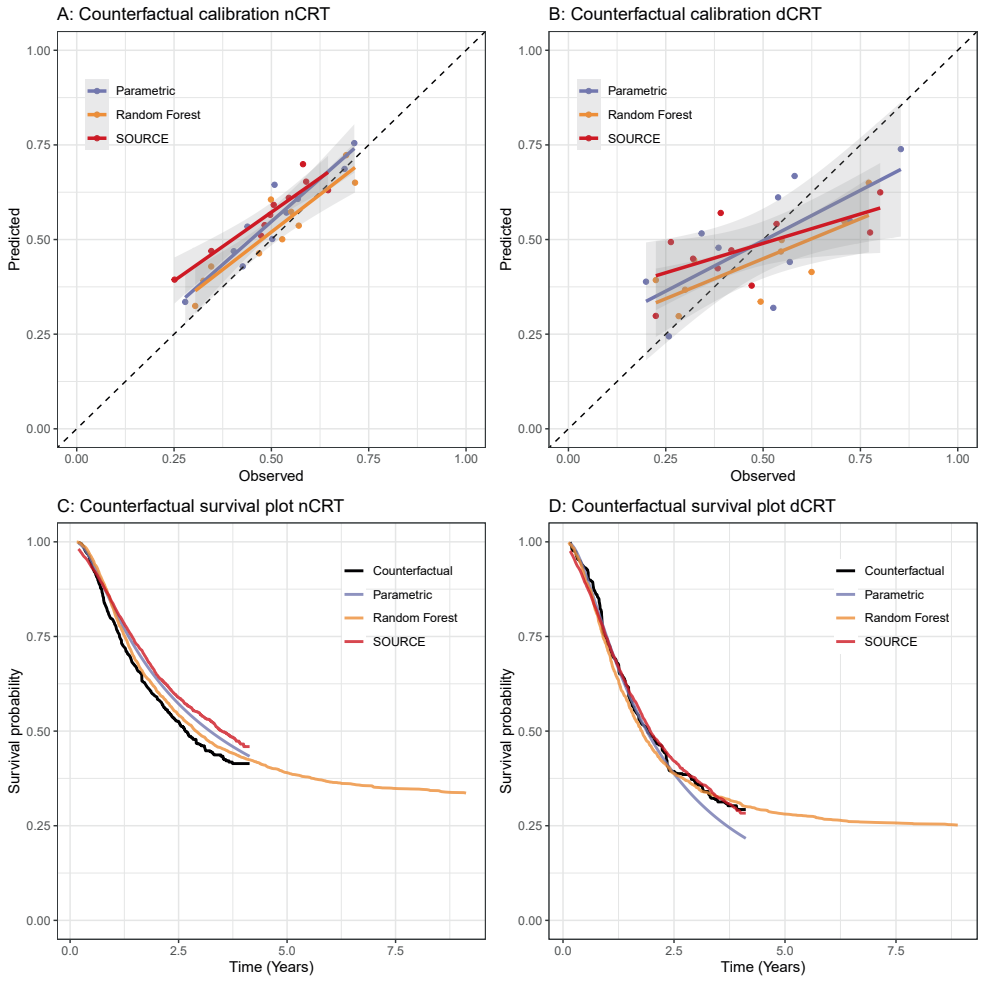


Figure 2. Calibration and survival plots of the trained models on the validation data. dCRT = Definitive chemoradiotherapy, nCRT = Neoadjuvant chemoradiotherapy

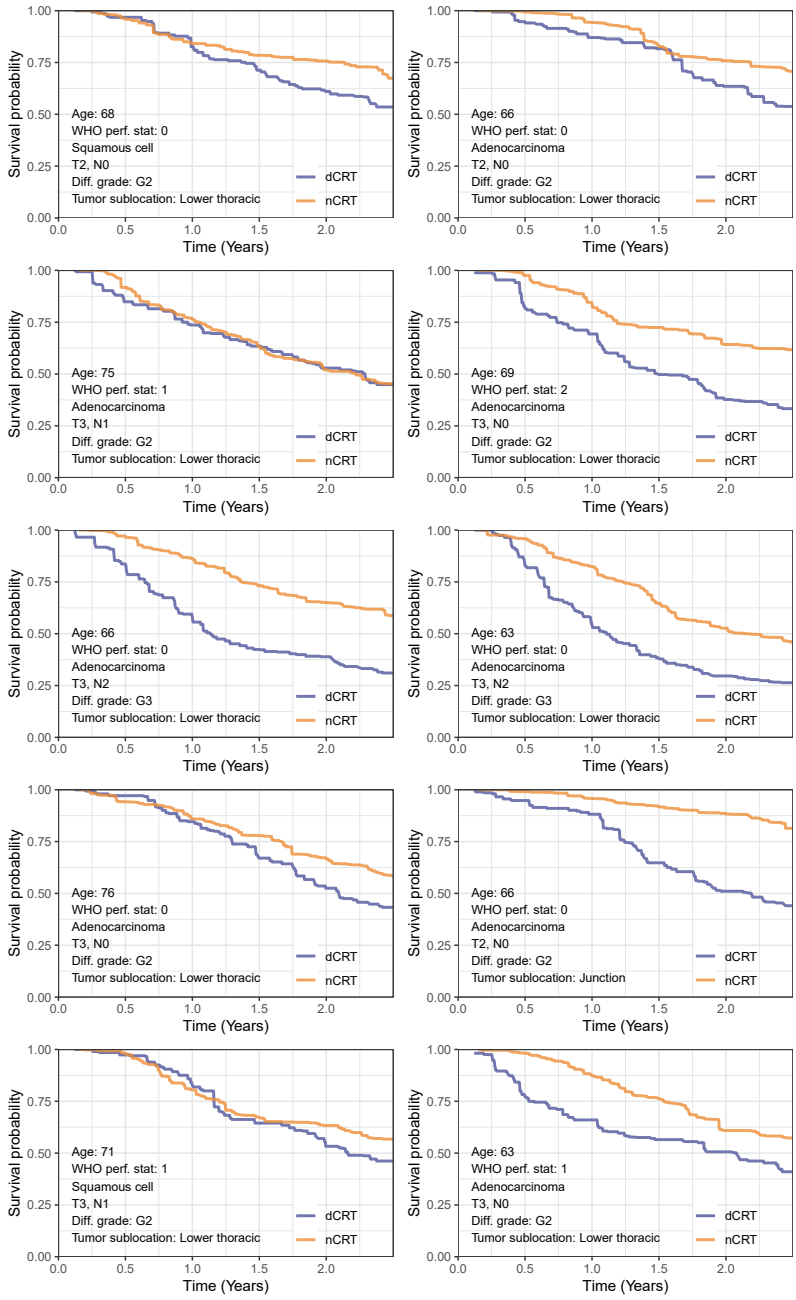


Figure 3. Individual counterfactual predictions using the Random Forest model for ten randomly selected patients from the validation data. Patients' clinical characteristics (Age, WHO performance status, cT, cN, Differentiation grade, and tumor sublocation) are reported in each panel.

Discussion

This study demonstrated the development and validation of a counterfactual prediction model that is clinically useful for assisting in the choice between neoadjuvant chemoradiotherapy and definitive chemoradiotherapy among patients with esophageal cancer. To our knowledge these are the first reported counterfactual prediction models in oncology.

Generally, the nCRT models exhibited better calibration, as indicated by intercepts close to zero and slopes approaching one across the parametric, random forest, and SOURCE models. The random forest model demonstrated the strongest alignment with the counterfactual survival curves, suggesting that it provides more accurate long-term survival predictions in this cohort. The high predictive performance of the random forest model may be attributable to the model's ability to capture complex, non-linear interactions between covariates that were not modelled in the parametric and SOURCE models.[20] In contrast, the dCRT models showed less reliable calibration, particularly in long-term survival predictions. While model performance was relatively consistent up to 2.5 years post-diagnosis, beyond this point, the parametric survival model underestimated survival rates. Therefore, we suggest that both models could be employed for predicting treatment-specific survival until this point in time.

It is important to note the discrepancy between the predictive performance of the random forest model validated on the training data and on the validation data. The random forest model seemed to under-perform on the training data compared to the validation data in terms of calibration. However, we hypothesize that this could be explained by overfitting of the random forest model on the training data. The calibration method we used involved binning the predicted survival probabilities into ten equally sized bins (from lowest, to highest predicted probabilities), after which in each bin the median survival was estimated which served as the observed survival. We believe that the random forest model's predictions stratified patients so efficiently to patients with low and high survival probabilities, that their observed survival in corresponding bins was exactly zero or one. In the validation data, we did not observe this effect, and the model performed well.

A surprising finding was that the SOURCE prediction model showed relatively good calibration and alignment with the counterfactual survival plots, compared to the models that were specifically designed for counterfactual prediction. The SOURCE model is a Cox proportional hazards model published in 2021 and developed using real-world observational data. Unlike the counterfactual parametric and random forest models presented in this study, the development of the SOURCE model was focused on predicting the outcome under the factual treatment and did not apply any additional causal inference methodology to extrapolate predictions and model validation to the whole population. This was achieved in the present study through inverse propensity score weighting. Furthermore, the SOURCE model does not model individualized treatment effects (e.g. through including treatment-covariate interactions) and instead includes the treatment as a main effect only.[21] As a Cox proportional hazards model it also made the assumption that the baseline hazard does not depend on treatment. However, despite these limitations, SOURCE demonstrated relatively good calibration and counterfactual survival estimation in the dCRT treatment arm. This relative good predictive performance may have been facilitated by the absence of strong treatment effect heterogeneity across covariates and baseline hazards It is important to emphasize that SOURCE was developed on a different developed on a different patient

population and we applied the baseline hazard estimated in the SOURCE study during model validation. The baseline hazard was not re-estimated on the validation data and hence there is no risk of our results being overly optimistic.

Taken together, these findings underscore that counterfactual prediction for this patient population under these treatment strategies is possible until 2.5 years after treatment. The presented counterfactual models could help to inform patients and physicians on which potentially curable treatment to opt for. These models are clinically useful as both neoadjuvant chemoradiotherapy and definitive chemoradiotherapy can be viable potentially curable treatments strategies, but the choice for which treatment to undergo is not strictly protocolized.[22,23] In such clinical cases, counterfactual prediction models are highly relevant because there is an actual choice in treatment, which leads to sufficient covariate overlap between treatment groups that is needed for valid causal inference. For highly protocolized treatment strategies, where there is no real choice between treatment options, counterfactual prediction models are less suitable.

In conclusion, this study provides evidence supporting the development and validation of counterfactual prediction models to aid in clinical decision-making for esophageal cancer treatment. By presenting the first models to guide treatment decisions between nCRT and dCRT, we offer a novel tool that could significantly impact personalized patient care. Interestingly, the SOURCE model—despite its inherent limitations—demonstrated better-than-expected calibration, suggesting that even conventional models may have value in counterfactual prediction. In wider view, our study underscore the importance of using counterfactual prediction models and validation. Importantly, our models may be used in shared decision-making, particularly in scenarios where multiple curative treatment options are available. Future work should focus on validating these models in more diverse populations and refining methods to address overfitting and interpretability challenges. Overall, this study advocates for more personalized, data-driven treatment choices, which could improve decision-making processes and ultimately lead to better patient outcomes in esophageal cancer care.

References

1. Shen A, Wei X, Zhu F, et al. Risk prediction models for breast cancer-related lymphedema: A systematic review and meta-analysis. *European Journal of Oncology Nursing*. 2023;64. doi:10.1016/j.ejon.2023.102326
2. van den Boorn HG, Engelhardt EG, van Kleef J, et al. Prediction models for patients with esophageal or gastric cancer: A systematic review and meta-analysis. *PLoS One*. 2018;13(2). doi:10.1371/journal.pone.0192310
3. Jha AK, Mithun S, Sherkhane UB, et al. Systematic review and meta-analysis of prediction models used in cervical cancer. *Artif Intell Med*. 2023;139. doi:10.1016/j.artmed.2023.102549
4. van Amsterdam WAC, de Jong PA, Verhoeff JJC, Leiner T, Ranganath R. From algorithms to action: improving patient care requires causality. *BMC Med Inform Decis Mak*. 2024;24(1). doi:10.1186/s12911-024-02513-3
5. Feuerriegel S, Frauen D, Melnychuk V, et al. Causal machine learning for predicting treatment outcomes. *Nat Med*. 2024;30(4):958-968. doi:10.1038/s41591-024-02902-1
6. Keogh RH, Van Geloven N. Prediction under Interventions: Evaluation of Counterfactual Performance Using Longitudinal Observational Data. *Epidemiology*. 2024;35(3):329-339. doi:10.1097/EDE.0000000000001713
7. Dickerman BA, Hernán MA. Counterfactual prediction is not only for causal inference. *Eur J Epidemiol*. 2020;35(7):615-617. doi:10.1007/s10654-020-00659-8
8. Liang Y, Yew PY, Loth M, et al. Personalized statin treatment plan using counterfactual approach with multi-objective optimization over benefits and risks. *Inform Med Unlocked*. 2023;42. doi:10.1016/j.imu.2023.101362
9. Noordman BJ, Wijnhoven BPL, Lagarde SM, et al. Neoadjuvant chemoradiotherapy plus surgery versus active surveillance for oesophageal cancer: A stepped-wedge cluster randomised trial. *BMC Cancer*. 2018;18(1). doi:10.1186/s12885-018-4034-1
10. Obermannová R, Alsina M, Cervantes A, et al. Oesophageal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Annals of Oncology*. 2022;33(10):992-1004. doi:10.1016/j.annonc.2022.07.003
11. Hernán MA, Wang W, Leaf DE. Target Trial Emulation: A Framework for Causal Inference From Observational Data. *JAMA*. 2022;328(24):2446-2447. doi:10.1001/jama.2022.21383
12. van Hagen P, Hulshof MCCM, van Lanschot JJB, et al. Preoperative Chemoradiotherapy for Esophageal or Junctional Cancer. *New England Journal of Medicine*. 2012;366(22):2074-2084. doi:10.1056/nejmoa1112088
13. Hulshof MCCM, Geijsen ED, Rozema T, et al. Randomized Study on Dose Escalation in Definitive Chemoradiation for Patients With Locally Advanced Esophageal Cancer (ARTDECO Study). *J Clin Oncol*. 2021;39(25):2816-2824. doi:10.1200/JCO.20.03697
14. van den Boorn HG, Abu-Hanna A, Haj Mohammad N, et al. SOURCE: Prediction Models for Overall Survival in Patients With Metastatic and Potentially Curable Esophageal and Gastric Cancer. *J Natl Compr Canc Netw*. 2021;19(4):403-410. doi:10.6004/jnccn.2020.7631
15. Cole SR, Hernán MA. Adjusted survival curves with inverse probability weights. *Comput Methods Programs Biomed*. 2004;75(1):45-49. doi:10.1016/j.cmpb.2003.10.004
16. Wright MN, Ziegler A. Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *J Stat Softw*. 2017;77(1). doi:10.18637/jss.v077.i01
17. van Kleef JJ, van den Boorn HG, Verhoeven RHA, et al. External validation of the dutch SOURCE survival prediction model in belgian metastatic oesophageal and gastric cancer patients. *Cancers (Basel)*. 2020;12(4). doi:10.3390/cancers12040834
18. Austin PC, Harrell FE, van Klaveren D. Graphical calibration curves and the integrated calibration index (ICI) for survival models. *Stat Med*. 2020;39(21):2714-2742. doi:10.1002/sim.8570
19. Mayer M. missRanger: Fast Imputation of Missing Values. Published online 2024.

20. Fife DA, D'Onofrio J. Common, uncommon, and novel applications of random forest in psychological research. *Behav Res Methods*. 2023;55(5):2447-2466. doi:10.3758/s13428-022-01901-9
21. Mueller S, Pearl J. Personalized decision making – A conceptual introduction. *J Causal Inference*. 2023;11(1). doi:10.1515/jci-2022-0050
22. Koëter M, van Putten M, Verhoeven RHA, Lemmens VEPP, Nieuwenhuijzen GAP. Definitive chemoradiation or surgery in elderly patients with potentially curable esophageal cancer in the Netherlands: a nationwide population-based study on patterns of care and survival. *Acta Oncol (Madr)*. 2018;57(9):1192-1200. doi:10.1080/0284186X.2018.1450521
23. Hulshof MCCM, van Laarhoven HWM. Chemoradiotherapy in tumours of the oesophagus and gastro-oesophageal junction. *Best Pract Res Clin Gastroenterol*. 2016;30(4):551-563. doi:10.1016/j.bpg.2016.06.002

*Propensity score ~ sex+age^2+creatinine^3+albumin^2+BMI^2+tumor histology+WHO performance status+cN+cT+number of comorbidities+HB+LDH+differentiation grade+tumor sublocation+tumorhistology*tumor sublocation*

Formula S1. Formula used to estimate the propensity score in the training data.

*Propensity score ~ sex+age^2+creatinine^3+albumin^2+BMI^2+tumor histology+WHO performance status+cN+cT+number of comorbidities+HB+LDH+differentiation grade+tumor sublocation+tumorhistology*tumor sublocation+number of comborbidities*age+cT*age+tumor sublocation*age*

Formula S2. Formula used to estimate the propensity score in the validation data.

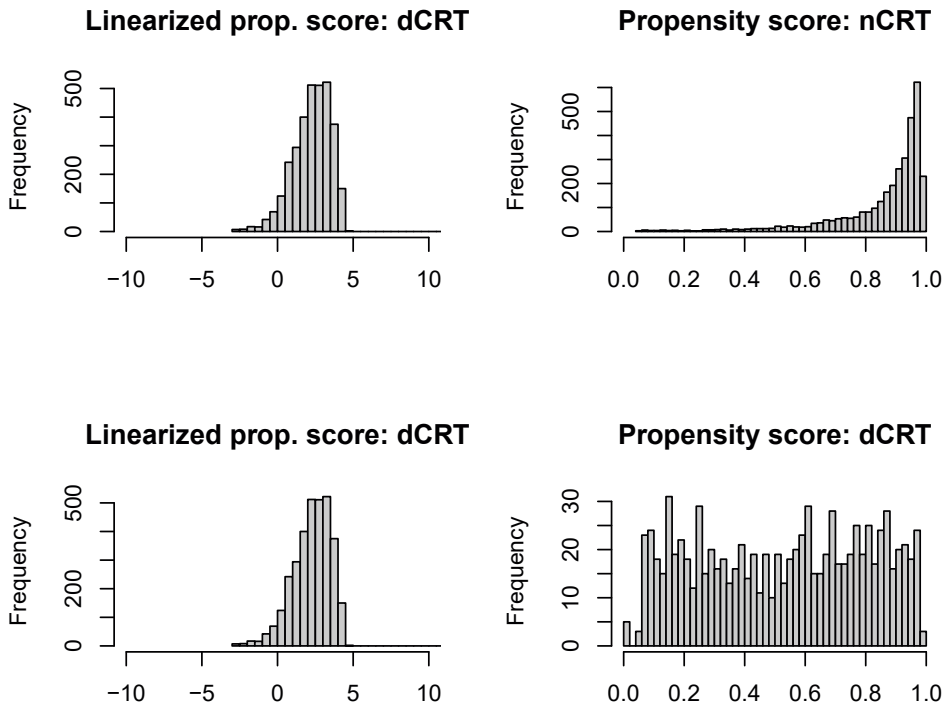


Figure S1. Histograms of the linearized and non-linearized propensity scores after weighting of the training data.

Table S1. Model parameters of the parametric survival model for the neoadjuvant chemoradiotherapy treatment arm.

	Parameter	Coefficient (95%CI)
Baseline hazard function	mu	7.83 (7.12-8.55)
	sigma	0.15 (0.11-0.18)
	Q	-1.14 (-1.29--0.99)
Sex	Male	REF
	Female	0.12 (0.03-0.22)
Age	leeft	-0.01 (-0.02--0.01)
Histology	Adenocarcinoma	REF
	Squamous cell carcinoma	0.24 (0.13-0.35)
	Other	-0.07 (-0.51-0.37)
WHO performance status	0	REF
	1	-0.23 (-0.31--0.16)
	2	-0.35 (-0.5--0.2)
	3	-1.59 (-1.99--1.2)
cN	0	REF
	1	-0.18 (-0.26--0.1)
	2	-0.32 (-0.42--0.22)
	3	-0.58 (-0.8--0.35)
cT	1	REF
	2	-0.27 (-0.65-0.11)
	3	-0.38 (-0.76-0)
	4	-0.51 (-0.97--0.05)
HB		-0.02 (-0.06-0.02)
BMI		0.02 (0.01-0.02)
Number of comorbidities	0	REF
	1	-0.13 (-0.21--0.05)
	Two or more	-0.16 (-0.26--0.06)
Creatinine		0.00 (0.00-0.00)
LDH		0.00 (0.00-0.00)
Albumin		0.03 (0.02-0.03)
Differentiation grade	G1	REF
	G2	-0.17 (-0.35-0)
	G3	-0.59 (-0.77--0.41)
Tumor sublocation	Lower thoracic	REF
	Junction	0 (-0.13-0.13)
	Mid-thoracic	-0.41 (-0.53--0.28)
	Overlapping lesion	-0.73 (-1.06--0.4)
	Upper thoracic	-0.51 (-0.74--0.28)
	Esophagus NOS	0.14 (-0.1-0.37)

Table S2. Model parameters of the parametric survival model for the definitive chemoradiotherapy treatment arm.

	Parameter	Coefficient (95%CI)
Baseline hazard function	mu	7.44 (6.82-8.06)
	sigma	-0.1 (-0.13--0.08)
	Q	-0.15 (-0.25--0.04)
Sex	Male	REF
	Female	0.01 (-0.09-0.06)
Age	leeft	0.00 (-0.01-0)
Histology	Adenocarcinoma	REF
	Squamous cell carcinoma	0.20 (0.12-0.28)
	Other	-0.13 (-0.43-0.17)
WHO performance status	0	REF
	1	-0.02 (-0.09-0.04)
	2	-0.11 (-0.23-0.01)
	3	-0.21 (-0.62-0.2)
cN	0	REF
	1	-0.19 (-0.26--0.12)
	2	-0.39 (-0.48--0.31)
	3	-0.13 (-0.31-0.05)
cT	1	REF
	2	-0.49 (-0.86--0.11)
	3	-0.91 (-1.28--0.54)
	4	-0.69 (-1.13--0.25)
HB		-0.01 (-0.04-0.02)
BMI		0.01 (0-0.01)
Number of comorbidities	0	REF
	1	0 (-0.07-0.07)
	Two or more	-0.09 (-0.18-0)
Creatinine		0.00 (0.00-0.00)
LDH		0.00 (0.00-0.00)
Albumin		0.00 (-0.01-0.01)
Differentiation grade	G1	REF
	G2	0.00 (-0.15-0.16)
	G3	-0.41 (-0.57--0.25)
Tumor sublocation	Lower thoracic	REF
	Junction	0.25 (0.12-0.37)
	Mid-thoracic	-0.08 (-0.17-0.01)
	Upper thoracic	0.2 (0.01-0.39)
	Esophagus NOS	-0.02 (-0.24-0.21)

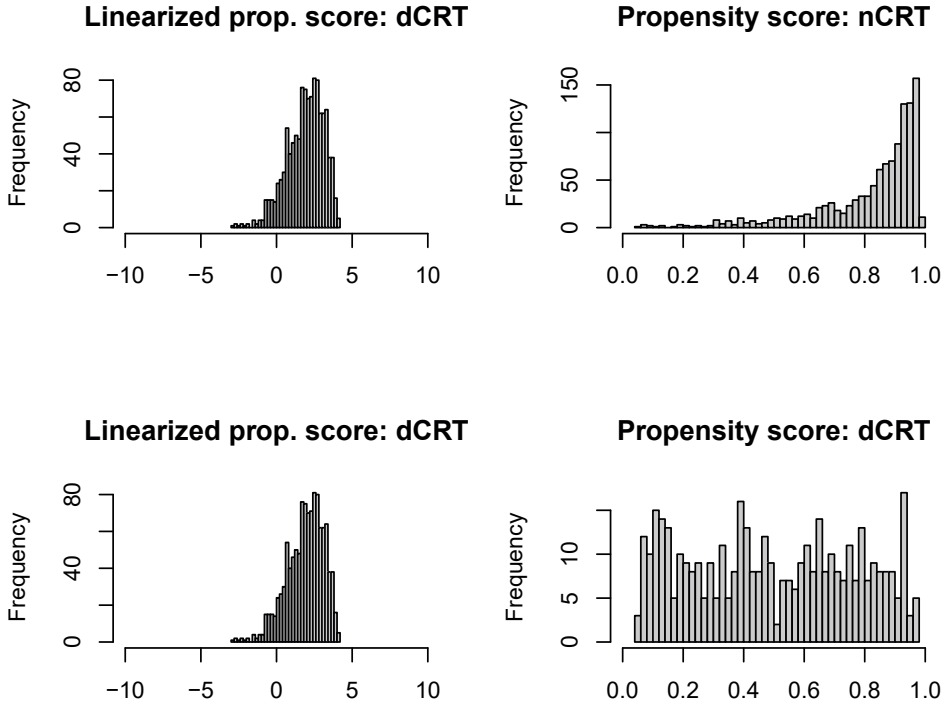
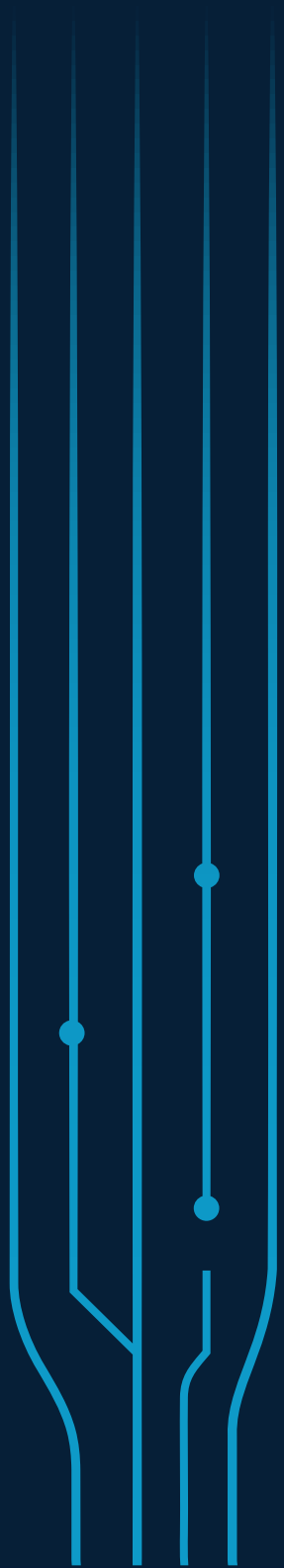


Figure S2. Histograms of the linearized and non-linearized propensity scores after weighting of the validation data.

General discussion



General discussion

The general aim of the research presented in this thesis was to apply statistical and machine learning methods to real-world data from patients with esophageal and gastric cancer, to advance understanding on the application of these methods to real-world data and to advance understanding of treatment outcomes for this patient population. In this discussion, we will discuss findings presented in this thesis in broader context with a focus on the future.

Good Models and Representative Data in Health-Related Quality of Life Research

Health-related quality of life research is a growing field with significant potential to make a difference for patients. Models can help identify risk factors of health-related quality of life and prediction models can, in turn, use these risk factors to develop health-related quality of life prediction models. In Chapter 5, we presented the first health-related quality of life prediction models for patients with esophageal and gastric cancer, demonstrating that it is indeed possible to predict health-related quality of life using two distinct types of models.

Despite this positive conclusion, predictive performance of the models could generally be improved in the coming years. One area for improvement is the inclusion of additional predictive variables. With a subjective construct such as health-related quality of life, it could be hypothesized that patients' expectations about the potential effects of treatment may influence the relationship between treatment and health-related quality of life. Different studies have suggested that this relationship exists, generally finding that positive expectations about anticancer treatment correlate with higher post-treatment health-related quality of life and, conversely, negative expectations correlate with poorer health-related quality of life.[1–4] Given this relationship, it may be highly informative to have data on treatment expectations as this is likely an important baseline factor in predicting post-treatment health-related quality of life. A future consideration for the POCOP questionnaire may be the inclusion of an item that measures patients' expectations about the effects of treatment.

Alternatively, electronic health records (EHRs) could be used to this extend and is potentially an existing avenue to be explored. Large language models, which have increased in popularity recently, could be used to extract valuable insights from written physician notes, capturing variables like patient expectations or longitudinal well-being data not traditionally included in standardized datasets.[5–8] This integration could potentially enhance the precision and scope of health-related quality of life predictions, and could further our understanding of how health-related quality of life and treatment are intertwined.

In addition, our novel approach to quantifying the representativeness of longitudinal cohort samples, as demonstrated in Chapter 4, ensures that future health-related quality of life models are built on data reflective of the real-world patient population. These methods can also be extended to clinical trials, enabling the adjustment of trial results to better represent real-world populations through calibration techniques.

Future Perspectives on Real-World Data

While trial data from randomized controlled trials (RCTs) remains the gold standard for assessing drug safety and efficacy, this thesis has demonstrated the growing potential of real-world data in further our understanding of therapeutic outcomes in real-world clinical populations. In Chapter 6, we utilized real-world data from the Netherlands Cancer Registry to create a control group that could be matched to the single-arm DESTINY-Gastric02 trial. This approach exemplifies how real-world data can generate comparative analyses in situations where traditional RCTs may be limited or infeasible.

Such analyses not only provide unique clinical insights that may be more reflective of everyday patient experiences, but they also hold the potential to inform drug regulatory policies and decision-making processes. Both the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA) recognize the value of real-world evidence and have begun to incorporate real-world data to complement evidence obtained from clinical trials, in areas such as label expansions and post-market safety.[9–11] A recent systematic review of FDA approval documents from January 2019 to June 2021 found that 116 approvals incorporated real-world evidence. Of these, 88 approvals used RWE for safety or effectiveness, influencing the FDA's decisions in 65 cases and being included in 38 product labels, demonstrating the clear shift to the use of real-world evidence.[9]

With this shift, the reliability and credibility of real-world data studies must be safeguarded. Despite the attractiveness of real-world data, it is also much more susceptible to different types of bias. Therefore, it is imperative that the methodological toolkit of epidemiologist and clinical researchers keeps expanding. Fortunately, the new techniques keep emerging and one of the more interesting developments in recent years has been the introduction of the E-value (Evidence value) for sensitivity analyses in observational research.[12] The E-value, reflects the minimum strength of an association that an unmeasured confounder would need to have with both the outcome and the selection mechanism to explain away the found association. It can serve as a sensitivity analyses to test the robustness of an estimated association to unobserved confounding, as we demonstrated in Chapter 6 and 7. Higher E-values imply a higher robustness to unobserved confounding. Reporting of the E-value can therefore improve the trustworthiness of estimated associations in real-world data studies and could be essential for promoting the use of real-world evidence among regulatory bodies.

The Need for Parametric Survival Modeling

Survival modeling has traditionally relied on Cox proportional hazards models due to their simplicity, interpretability, and widespread availability in statistical software. These models are highly effective for identifying associations between predictors and outcomes, such as hazard ratios, which are crucial for understanding survival trends. However, their utility for predicting survival probabilities is limited. The Cox model is semi-parametric, meaning it does not estimate the baseline hazard function, a crucial component for making accurate survival predictions. While methods like the Breslow estimator can approximate the baseline hazard, this process can be cumbersome. This limitation became clear in Chapter 8 when we applied Steyerberg's prediction model to data from the POLDER trial.[13,14] Despite having hazard ratios, the absence of baseline hazard estimates made it challenging to adapt the model for new data.

In contrast, parametric survival models offer a more robust alternative for prediction. Unlike the Cox model, parametric models estimate the baseline hazard directly, providing a comprehensive framework for survival prediction. Chapter 11 demonstrated this advantage through the development of a fully parametric survival prediction model. This approach allowed for straightforward external validation, as other researchers could easily use the reported baseline hazard estimates and model coefficients. By enabling better model transferability and validation, parametric models are particularly well-suited for survival prediction tasks.

Beyond prediction, parametric models have significant value in epidemiological studies. For instance, Chapter 1 highlighted the use of parametric models in comparing survival rates between patients from the Netherlands and Belgium. By employing excess hazard models, which are fully parametric, the analysis accounted for both population mortality and case-mix differences. This flexibility allows for nuanced insights into survival data while controlling for numerous population-level factors, offering substantial clinical relevance.

Given these strengths, future research should emphasize parametric survival modeling over traditional Cox approaches. Parametric models provide greater flexibility, facilitate external validation, and enhance the interpretability of survival data. They are valuable for prediction and epidemiological research, offering a versatile tool for addressing complex survival questions. By adopting parametric methods, researchers can overcome the limitations of Cox models.

Synthesis of Causality and Personalized Prediction: exploring the "what-if" question

One of the most promising advancements in prediction modeling is the integration of causal inference and personalized prediction. This approach addresses critical "what-if" questions, such as, *"For this specific patient, does treatment X or Y lead to better outcomes?"* In Chapter 11, we presented a proof-of-concept study that developed and validated such a model, demonstrating its ability to predict survival outcomes under different treatments. This achievement is both methodologically and clinically significant, as it showcases the potential to use real-world observational data to construct robust, patient-specific predictive models.

While our model focused on predicting survival under two treatments (neoadjuvant chemoradiotherapy and definitive chemoradiotherapy), its applicability extends much further. Expanding the model to include a broader range of treatment options could offer clinicians a powerful tool for daily practice, enabling direct comparisons of interventions. Such advancements have the potential to transform decision-making, offering a framework for selecting the most effective and personalized treatment strategies. However, the method that we have demonstrated is not easily generalizable to multiple treatments as it utilizes the propensity score. For this to be possible, other, more novel methods will need to be applied to the data.

One of those methods that could potentially transform personalized causal prediction are digital twins. A digital twin is essentially a dynamic, virtual replica of a real-world physical object, process or product that connects the real-world and the digital world.[15] Digital twins have originally been developed in the engineering sciences for simulation, stress testing and monitoring.[16] Consider, for example, a particular mechanical engine for which a digital twin has been created. Sensors on the physical machine can send real-time data about its performance or condition to its digital twin. The digital twin uses this data to simulate,

predict, or optimize the machine's operation. If the virtual model identifies a potential issue or an opportunity for improvement, it can send instructions back to the physical machine to adjust its behavior. The digital twin uses this data to simulate, predict, or optimize the machine's operation. If the virtual model identifies a potential issue or an opportunity for improvement, it can send instructions back to the physical machine to adjust its behavior.

Although its origins lie in engineering, digital twins have found their way into many different areas including the medical sciences.[17] Having a digital copy of an actual patient in the same way you can have a digital copy of a mechanical engine, opens a great number of opportunities in the context of causal individualized predictions. It enables physicians to perform different treatments on the virtual patient and observe the effects. For example, in radiation oncology there are already developments in the application of digital twins for treatment planning.[18,19]

By integrating diverse data sources such as electronic health records, imaging, genomics, and real-time monitoring devices like wearables it can be imagined that digital twins hold a very large promise for future personalized treatment in (esophagogastric) oncology. Digital twins simulate a patient's unique physiology and health status, allowing for precise prediction of how an individual might respond to various treatments. Digital twins offer the capacity to predict not only survival outcomes but also quality of life, treatment risks, and long-term impacts, providing a comprehensive view of potential clinical scenarios.[17]

While this would be very attractive, to realize the full potential of digital twins in medicine the availability and quality of data are essential. It can be imagined that creating a meaningful virtual copy of a human being is incredibly complex and requires vast array of different data types such as imaging, multi-omics data, biometric and physiological data (such as data from wearables) and historic and real-time clinical data.[20–22] In addition to the data, which would be a prerequisite for a digital twin, advanced computational techniques and well-validated machine learning algorithms are required to manage the complexity of human physiology and ensure reliable predictions. This is by no means an easy feat, but given the increasing computation power this paints a hopeful picture for the future of computer driven personalized treatment.

To fully realize the potential of counterfactual models, they must be integrated into user-friendly tools, such as the SOURCE platform. Studies have shown that SOURCE enhances the precision of treatment information during simulated consultations. Incorporating future counterfactual models into tools like SOURCE would further advance personalized treatment planning, making these innovative approaches more accessible and impactful in clinical settings. By combining survival, quality-of-life predictions, and practical software integration, these advancements can transform decision-making, empowering clinicians to provide treatments aligned with both evidence and individual patient goals.

Concluding Remarks

This thesis highlights the potential of advanced statistical and machine learning approaches in understanding and predicting treatment outcomes for esophageal and gastric cancer patients. The prediction tools we developed enable physicians to make personalized treatment decisions, enhancing patient outcomes and quality of life. Looking ahead, integrating personalized tumor-related information, utilizing EHRs, and refining methodological approaches will further improve the accuracy and applicability of prediction models. This progress predicts a promising future for the personalized care of esophageal and gastric cancer.

References

1. Fletcher C, Wilson C, Hutchinson AD, Grunfeld EA. The relationship between anticipated response and subsequent experience of cancer treatment-related side effects: A meta-analysis comparing effects before and after treatment exposure. *Cancer Treat Rev*. 2018;68:86-93. doi:10.1016/j.ctrv.2018.06.009
2. Wan GJ, Counte MA, Cella DF. The Influence of Personal Expectations on Cancer Patients' Reports of Health-Related Quality of Life. *Psychooncology*. 1997;6(1):1-11. doi:https://doi.org/10.1002/(SICI)1099-1611(199703)6:1<1::AID-PON230>3.0.CO;2-C
3. Cockle S, Ogden J. Patients' expectations of cancer treatment and their perceived link to subsequent experiences: A qualitative study. *Br J Health Psychol*. 2022;27(2):267-282. doi:https://doi.org/10.1111/bjhp.12544
4. Bottomley A, Flechtner H, Efficace F, et al. Health related quality of life outcomes in cancer clinical trials. *Eur J Cancer*. 2005;41(12):1697-1709. doi:10.1016/j.ejca.2005.05.007
5. Hsu CC, Karnwal S, Mullainathan S, Obermeyer Z, Tan C. *Findings of the Association for Computational Linguistics Characterizing the Value of Information in Medical Notes*. https://github.com/BoulderDS/
6. Zhou W, Bitterman D, Afshar M, Miller TA. *Considerations for Health Care Institutions Training Large Language Models on Electronic Health Records*.
7. Yang X, Chen A, PourNejatian N, et al. A large language model for electronic health records. *NPJ Digit Med*. 2022;5(1). doi:10.1038/s41746-022-00742-2
8. Nashwan AJ, AbuJaber AA. Harnessing the Power of Large Language Models (LLMs) for Electronic Health Records (EHRs) Optimization. *Cureus*. 2023;15(7):e42634. doi:10.7759/cureus.42634
9. Purpura CA, Garry EM, Honig N, Case A, Rassen JA. The Role of Real-World Evidence in FDA-Approved New Drug and Biologics License Applications. *Clin Pharmacol Ther*. 2022;111(1):135-144. doi:10.1002/cpt.2474
10. *Framework for FDA's Real-World Evidence Program*.; 2018. www.fda.gov
11. Medicines Agency E. *Real-World Evidence Framework to Support EU Regulatory Decision-Making*.
12. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Ann Intern Med*. 2017;167(4):268-274. doi:10.7326/M16-2607
13. Steyerberg EW, Homs MYV, Stokvis A, Essink-Bot ML, Siersema PD. Stent placement or brachytherapy for palliation of dysphagia from esophageal cancer: A prognostic model to guide treatment selection. *Gastrointest Endosc*. 2005;62(3):333-340. doi:10.1016/S0016-5107(05)01587-7
14. Jeene PM, Vermeulen BD, Rozema T, et al. Short-Course External Beam Radiotherapy Versus Brachytherapy for Palliation of Dysphagia in Esophageal Cancer: A Matched Comparison of Two Prospective Trials. *Journal of Thoracic Oncology*. 2020;15(8):1361-1368. doi:10.1016/j.jtho.2020.04.032
15. Jones D, Snider C, Nassehi A, Yon J, Hicks B. Characterising the Digital Twin: A systematic literature review. *CIRP J Manuf Sci Technol*. 2020;29:36-52. doi:10.1016/j.cirpj.2020.02.002
16. Epp Editor DS. *Conference Proceedings of the Society for Experimental Mechanics Series*.; 2021. http://www.springer.com/series/8922
17. Kamel Boulos MN, Zhang P. Digital twins: From personalised medicine to precision public health. *J Pers Med*. 2021;11(8). doi:10.3390/jpm11080745
18. Rahmim A, Brosch-Lenz J, Fele-Paranj A, et al. Theranostic digital twins for personalized radiopharmaceutical therapies: Reimagining theranostics via computational nuclear oncology. *Front Oncol*. 2022;12. doi:10.3389/fonc.2022.1062592
19. Brosch-Lenz J, Uribe C, Rahmim A, Saboury B. Theranostic Digital Twins: An Indispensable Prerequisite for Personalized Cancer Care. *Journal of Nuclear Medicine*. 2023;64(3):501. doi:10.2967/jnumed.122.264929

20. Meijer C, Uh HW, el Bouhaddani S. Digital Twins in Healthcare: Methodological Challenges and Opportunities. *J Pers Med*. 2023;13(10). doi:10.3390/jpm13101522
21. Łukaniszyn M, Majka Ł, Grochowicz B, Mikołajewski D, Kawala-Sterniuk A. Digital Twins Generated by Artificial Intelligence in Personalized Healthcare. *Applied Sciences (Switzerland)*. 2024;14(20). doi:10.3390/app14209404
22. Xie S, Zhu S, Dai J. Feasibility study of intelligent healthcare based on digital twin and data mining. In: *2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI)*. ; 2021:906-911. doi:10.1109/CISAI54367.2021.00182

Appendices

Summary

In this thesis, we explored the application of statistical and machine learning methods to real-world data in esophagogastric oncology, with the overarching aim of advancing clinical understanding of the disease and improve understanding on the application of statistical and machine learning methods. By using diverse analytical techniques across multiple research areas, this work contributes to bridging gaps in the use of real-world data and modern statistical and machine learning methods to provide useful clinical insights.

Part I: Epidemiology of esophageal and gastric cancer

It is well-known that esophageal and gastric cancer are highly deadly diseases with heterogenous survival chances. This heterogeneity in survival is not only observed within populations of patients, but also between populations of patients. To illustrate this, a study performed in 2015, the EUROCARE-5 study, showed that there was a considerable difference in survival between patients from the Netherlands and Belgium without there being a clear explanation to account for this difference. In Chapter 1 of this thesis, we investigated what this difference can be ascribed to by comparing survival from patients with esophageal and gastric cancer in the Netherlands with Belgium in more detail. We employed parametric survival models (relative survival and excess hazard models), where we controlled for population mortality and investigated how treatment differences impact survival. Five-year relative survival rates for gastric (GC) and esophageal cancer (EC) were lower in the Netherlands than in Belgium (GC: 20% vs. 27%, EC: 21% vs. 24%). These differences were present across most tumor stages. Adjusting for treatment differences between both countries explained survival differences in esophageal cancer and stage IV gastric cancer but not in stage I–III gastric cancer. The difference in survival in stage IV cases and esophageal cancer was thus likely due to variations in treatment between both countries (in general a higher percentage of 'tumor directed' treatment in Belgium). However, the reason for differences in stage I-III gastric cancer remains unclear and will need further investigation.

In **Chapter 2**, we took a more detailed look at survival of patients with non-metastatic esophageal and gastric cancer in the Netherlands. While traditional epidemiological studies often only report the median overall survival (the 50th percentile on a survival curve), we went beyond the median and investigated survival across different points of the survival curve: survival scenarios. Using data from the Netherlands Cancer Registry, patients with non-metastatic esophageal or gastric cancer diagnosed between 2006-2020 were included. Survival scenarios were defined as best-case (20th percentile), upper-typical (40th), median, lower-typical (60th), and worst-case (80th percentile). For esophageal cancer, the best-case scenario showed the largest improvement, with survival increasing by 12 months annually until 2011, followed by smaller increases of up to 1.5 months per year in other scenarios. For gastric cancer, best-case survival did not change, while typical and worst-case scenarios improved modestly, with annual improvements of up to 1.0 month (until 2018). In addition, we observed survival improvements across patient subgroups, treatment types, and cancer subtypes. These results highlight the benefits of treatment advancements, with the most significant survival improvements seen in the best-case scenario for esophageal cancer. This indicated that, while the large majority of patients generally improved with respect to survival outcomes, the largest improvements were observed across the top 20% percent of patients.

Chapter 3 of this thesis also utilized non-traditional survival measures to better understand survival outcomes and to better inform patients with esophageal and gastric cancer. Conditional relative survival (CRS) is a useful tool for communicating prognosis, as it estimates life expectancy based on survival for a certain period after treatment. In this study, we examined 3-year CRS for patients with non-metastatic esophageal or gastric cancer treated with curative intent between 2006 and 2020, using data from the Netherlands Cancer Registry. We found that for patients with esophageal cancer, 3-year CRS improved from 62% after the first year to 87% after the 5th year post diagnosis. This implies that patients who have survived up to 5 year after initial diagnosis have a 87% chance to also survive the next 3 years. For gastric cancer we found higher rates of 69% and 90% after 3 and 5 years post diagnosis, respectively. These findings demonstrated that despite the generally poor prognoses of these cancer types, survival chances improve markedly over time, giving patients and physicians a more realistic and hopeful outlook during follow-up discussions.

Part II: Health-related quality of life

In **Part II** of this thesis, we dealt with a different, but highly relevant dimension of the outcome of cancer treatment: health-related quality of life (HRQoL). Unlike survival data, which in the Netherlands is obtainable for all oncological patients from the nationwide Netherlands Cancer Registry, data collection on HRQoL primarily relies on prospective observational cohort studies. Participation of patients in such prospective cohort studies is voluntary and typically requires patients to fill out questionnaires about their HRQoL. This voluntary component introduces the risk of selection-bias (i.e. some patients are more likely to participate than others), through which the real-world representativeness of the cohort studies and the external validity of analyses based on that could be affected. To this end, in **Chapter 4**, we performed a study on the Prospective Observational Cohort Study of Oesophageal-gastric cancer Patients (POCOP) to evaluate the real-world representativeness of patients included cancer cohort studies compared to the total population of patients with esophagogastric cancer in the Netherlands.

Using a novel method called Representativeness indicators (R-indicators), with which the total representativeness of the sample can be expressed in one single number between 0 (not representative) and 1.0 (completely representative), we found that the complete POCOP study had an R-indicator of 0.72 95%CI (0.71-0.73), which indicated a generally fair representativity. Palliative patients had a higher level of representativeness compared to those with potentially curable disease, with R-indicators of 0.88 95%CI(0.85-0.90) and 0.70 95%CI(0.68-0.71), respectively. When stratified into clinically relevant subgroups based on treatment, the R-indicators for each subgroup increased (≥0.8 for all groups). After both stratification and calibration weighting, survival estimates in the POCOP registry aligned more closely with those from the NCR population. The study shows that real-world data from a prospective esophagogastric cancer registry can be representative of the total population of patients if differences in treatment types are accounted for. This supports the clinical use of PROMs data, especially when corrected for selection bias through stratification or statistical calibration.

Having established that the POCOP registry was fairly representative of the total patient population, we continued building our understanding of HRQoL outcomes by focusing on the clinical prediction of HRQoL outcomes. In **Chapter 5** we developed prediction models that predict post-treatment HRQoL for patients with esophagogastric cancer, based on HRQoL

data from POCOP and the Netherlands Cancer Registry for clinical variables. The EORTC QLQ-C30 functioning scales (including the Summary Score) were used as prediction outcomes. Risk-prediction models, based on logistic elastic-net regression, predicted the probability of clinically significant HRQoL deterioration at 3, 6, and 12 months post-treatment. In addition, a machine-learning based sequential score model, using XGBoost regression, predicted future HRQoL scores over time. We found that the risk-prediction models performed well, with ICI values between 0.03 and 0.08 and Brier scores ranging from 0.09 to 0.17, effectively predicting declines in Summary Score, Physical Functioning, and Fatigue. The sequential score models explained up to 40% of the variance in HRQoL scores. Results from this study showed that both models accurately predicted HRQoL changes in esophagogastric cancer patients, offering valuable tools to improve patient care and facilitate shared decision-making through HRQoL forecasting. Simultaneously, the results also highlight room for improvement in terms of predictive power for a number of the functioning scales. Future studies will need to be conducted to further refine clinical prediction models for HRQoL and the potential of information extraction from electronic health-records to supplement data from HRQoL cohorts will need to be explored.

Part III. Clinical trials and real-world data

In **Part III** of this thesis, we demonstrated how data from clinical trials and real-world sources can complement each other to explore treatment effects that have not (yet) been evaluated in clinical trials. In these chapters, we focused on two relevant studies for which results from a randomized controlled trial was not presented.

In **Chapter 6**, we used data from the single arm DESTINY-Gastric02 (DG-02) trial in which HER2-positive unresectable or metastatic gastric or gastro-esophageal junction adenocarcinoma were treated with Trastuzumab Deruxtecan (T-DXd), and we compared it to a reference group treated with Ramucirumab and Paclitaxel (Ram+Pac) obtained from the Netherlands Cancer Registry (NCR). Using propensity matching, we matched patients from the DG-02 trial who received prior trastuzumab-based first-line therapy to similar patients who received Ram+Pac from the NCR, based on relevant clinical factors associated to treatment selection and survival. The resulting matched data was balanced (58 DG-02 patients and 78 real-world patients). Median overall survival was significantly longer in the T-DXd group (11.6 months; 95% CI(9.0–20.5)) compared to the Ram+Pac group (6.2 months; 95% CI(4.5–10.0)), p-value < 0.0001. These findings suggest that T-DXd may offer a survival advantage over standard second-line treatment in this patient population and also presented a practical use case of single-armed trials being supplemented with real-world data to emulate a randomized controlled trial.

Following the CheckMate-577 trial, which demonstrated a disease-free survival benefit for adjuvant nivolumab in patients with esophageal or gastroesophageal junction (GEJ) cancer after neoadjuvant chemoradiotherapy (nCRT) and resection, there remained uncertainty regarding its effect on overall survival (OS), as final OS data from the trial were not yet reported at the time of performing this study. To address this uncertainty, in **Chapter 7**, we conducted a nationwide real-world study to evaluate OS in patients treated with or without adjuvant nivolumab in routine clinical practice. Patients diagnosed with non-metastatic esophageal or GEJ cancer between 2020 and 2023 and who had residual pathological disease after nCRT and resection were identified from the Netherlands Cancer Registry. A total of 333 patients who received adjuvant nivolumab were compared to 486 patients

treated prior to the introduction of nivolumab who underwent nCRT and resection alone. Propensity score trimming and nearest-neighbor matching on relevant clinical characteristics resulted in two well-balanced cohorts of 311 patients each. Median follow-up was 24.4 months in the nivolumab group and 31.4 months in the control group. Two-year OS was significantly higher among patients treated with nivolumab (66.8%, 95% CI(61.6–72.4%)) compared to those without (58.8%, 95% CI(53.5–64.5%)). The estimated hazard ratio was 0.75 (95% CI(0.60–0.97); p = 0.024). These findings represent the first piece of evidence from the real-world of a potential OS benefit with adjuvant nivolumab in this setting. However, due to the ongoing follow-up and limited number of events, further analyses over time are needed to confirm these results.

Part IV. Prediction of survival outcomes

In the final part of this thesis, we focused on the utility, development and assessment of clinical prediction models in the field of esophageal and gastric cancer.

Beyond supporting shared decision-making between clinicians and patients, prediction models could also serve as tools to guide patient selection in clinical trials. In Chapter 8, we evaluated the potential utility of the SOURCE survival prediction model (Stimulating evidence-based, personalized, and tailored information provision to improve decision-making after an esophagogastric cancer diagnosis) alongside a widely used model developed by Steyerberg and colleagues. This evaluation was conducted within the context of the POLDER study, which assessed the effect of external beam radiotherapy for relieving dysphagia in patients with esophageal cancer. Although POLDER trial inclusion required an expected survival of at least three months, nearly one third of patients deceased within 3 months after being included in the trial. The two existing models were retrospectively applied to the POLDER data to predict three-month survival. The SOURCE model outperformed the Steyerberg model, with higher accuracy (79% vs. 64%) and a significantly better discriminative ability (AUC 0.76 vs. 0.60, p = 0.017). At optimal thresholds, the SOURCE model would have led to less incorrect inclusions (16 vs. 34 out of 110 patients) into the POLDER trial. These findings suggest that the SOURCE model could serve as a valuable decision support tool alongside clinical judgment in the context of palliative care trials

After demonstrating the clinical utility of the existing SOURCE model, we continued the development of new clinical prediction models in **Chapter 9**. In this chapter, we developed and internally validated the SOURCE Beyond First-Line model, a clinical prediction tool developed to estimate survival for patients with metastatic esophagogastric adenocarcinoma after failure of first-line palliative systemic therapy and thereby extending the SOURCE family of prediction models. Using data from 1,067 patients in the Netherlands Cancer Registry, we developed a Cox proportional hazards model based on 11 clinical and treatment-related variables. The model demonstrated good performance, with a C-index of 0.75 (95% CI (0.73–0.78)), calibration slope of 1.01, and calibration intercept of 0.01. Internal cross-validation confirmed consistent performance on unseen data, with a C-index of 0.79 (95%CI (0.77–0.82)), slope of 0.93, and intercept of 0.02. These results indicate that the SOURCE Beyond First-Line model offers fair discrimination and good calibration, making it a promising tool to support clinical decision-making in the beyond-first-line treatment setting.

In **Chapter 10**, we explored the potential of enhancing the existing SOURCE survival prediction model for patients with resectable esophageal adenocarcinoma by integrating radiomics and circulating-free tumor DNA (cfDNA) metrics. Using a cohort of 111 patients treated with neoadjuvant chemoradiation therapy, we developed models combining clinical variables from the SOURCE model with radiomic features from PET-CT scans and cfDNA. The best-performing baseline models for OS and time to progression (TTP) showed improved C-index scores when cfDNA was added to the SOURCE model (C-index 0.55 for OS and 0.59 for TTP, compared to 0.44–0.45 with SOURCE alone). Incorporating PET-CT radiomics into the SOURCE model further improved predictions, achieving a C-index of 0.65 for OS and 0.60 for TTP. The combination models also demonstrated significant risk stratification for both OS and TTP (log-rank $P < 0.01$). For predicting pathologic complete response, the best model achieved an area under the curve of 0.61, compared to 0.47 with clinical variables alone. These findings suggest that integrating radiomics and cfDNA into the SOURCE model enhances its predictive accuracy, though external validation and further optimization of radiomic pipelines are needed in future studies.

Although the prediction models that were applied and developed in the previous chapters demonstrated good predictive performance and clinical relevance, from a methodological point of view they did not address an important factor that is highly relevant for clinical decision making in a clinical context: causality. When patients and physicians are faced with choosing a particular treatment, prediction models—such as the model applied and developed in this thesis—can be used to aid in this. However, due to the fact that prediction models are almost always developed in observational data, the predictions from the model can usually not be causally interpreted. Therefore, in **Chapter 11**, we applied a novel approach to prediction called counterfactual prediction, which combines causal inference with personalized survival prediction and allows causally interpretable personalized treatment prediction. We applied this method in a proof-of-concept study for patients with non-metastatic esophageal cancer, comparing two treatment strategies: neoadjuvant chemoradiotherapy (nCRT) and definitive chemoradiotherapy (dCRT). Using data from the Netherlands Cancer Registry (4,388 patients for training and 1,693 for validation), we developed two different counterfactual survival models (parametric survival model and random forest model) based on inverse propensity score weighting to control for confounding. The models showed good calibration in the training and validation datasets, with the parametric model performing well for nCRT and the random forest model aligning more closely with survival predictions for dCRT. Additionally, we tested the existing SOURCE model in a counterfactual context, which demonstrated relatively good calibration despite not being designed for this purpose. This study highlights the potential of counterfactual prediction models to improve personalized decision-making in clinical settings where multiple treatment options are available, offering patients and clinicians a better understanding of individualized survival probabilities.

Conclusion

In conclusion, this thesis shows how statistical and machine learning methods can be used to make better sense of real-world data in esophagogastric cancer. By looking at survival outcomes, health-related quality of life, and prediction models, we gained insights that go beyond traditional analyses. The findings in this thesis help improve clinical understanding, can aid in shared decision making, and ultimately support more personalized care for patients with esophageal and gastric cancer.

Nederlandse samenvatting

In dit proefschrift onderzochten we de toepassing van statistische en machine learning-methoden op real-world data in de slokdarm en maag oncologie, met als overkoepelend doel het verbeteren van het klinisch begrip van de ziekte en het vergroten van het inzicht in de toepassing van statistische en machine learning-methoden. Door het gebruik van diverse analysetechnieken over meerdere onderzoeksgebieden draagt dit werk bij aan het overbruggen van hiaten in het gebruik van real-world data en moderne statistische en machine learning-methoden om bruikbare klinische inzichten te bieden.

Deel I: Epidemiologie van slokdarm- en maagkanker

Het is algemeen bekend dat slokdarm- en maagkanker zeer dodelijke ziekten zijn met uiteenlopende overlevingskansen. Deze heterogeniteit in overleving wordt niet alleen binnen populaties van patiënten waargenomen, maar ook tussen populaties van patiënten. Ter illustratie: een studie uitgevoerd in 2015, de EUROCARE-5 studie, toonde aan dat er een aanzienlijk verschil in overleving was tussen patiënten uit Nederland en België, zonder dat daar een duidelijke verklaring voor werd gevonden. In **Hoofdstuk 1** van dit proefschrift onderzochten we waaraan dit verschil mogelijk kan worden toegeschreven door de overleving van patiënten met slokdarm- en maagkanker in Nederland te vergelijken met die in België. We gebruikten relatieve overleving en excess hazard-modellen, welke het mogelijk maakte, om statistisch te controleren voor populatiesterfte. Met deze modellen onderzochten hoe behandelverschillen van invloed zijn op de overleving. De relatieve overlevingspercentages voor maag- (GC) en slokdarmkanker (EC) waren lager in Nederland dan in België (GC: 20% vs. 27%, EC: 21% vs. 24%). Deze verschillen waren aanwezig over alle tumorstadia. Correctie voor behandelverschillen tussen beide landen verklaarde de overlevingsverschillen bij slokdarmkanker en stadium IV maagkanker, maar niet bij stadium I–III maagkanker. Het verschil in overleving bij stadium IV en slokdarmkanker is dus waarschijnlijk toe te schrijven aan variatie in behandeling tussen beide landen (over het algemeen een hoger percentage 'agressieve' behandeling in België). De reden voor verschillen in potentieel curabele maagkanker blijft echter onduidelijk en vereist verder onderzoek.

In **Hoofdstuk 2** bekeken we de overleving van patiënten met niet-uitgezaaide slokdarm- en maagkanker in Nederland in meer detail. Waar traditionele epidemiologische studies vaak alleen de mediane overleving rapporteren (het 50e percentiel op een overlevingscurve), gingen wij verder en onderzochten we de overleving op verschillende punten van de overlevingscurve: overlevingsscenario's. Met behulp van gegevens uit de Nederlandse Kankerregistratie werden patiënten met niet-uitgezaaide slokdarm- of maagkanker gediagnosticeerd tussen 2006-2020 geïncludeerd. Overlevingsscenario's werden gedefinieerd als best-case (20e percentiel), upper-typical (40e), mediaan, lower-typical (60e), en worst-case (80e percentiel). Voor slokdarmkanker liet het best-case scenario de grootste verbetering zien, met een jaarlijkse toename van de overleving met 12 maanden tot 2011, gevolgd door kleinere toenames tot 1,5 maand per jaar in andere scenario's. Voor maagkanker veranderde de best-case overleving niet, terwijl typische en worst-case scenario's bescheiden verbeterden met jaarlijkse toenames tot 1,0 maand (tot 2018). We observeerden ook verbeteringen in overleving in patiëntensubgroepen, behandelingen en kankersubtypes. Deze resultaten benadrukken de voordelen van verbetering in de behandelopties, waarbij de meest significante overlevingswinst werd gezien in het best-case scenario voor slokdarmkanker. Dit geeft aan dat, hoewel de grote meerderheid van patiënten in het algemeen verbeterde qua overlevingsuitkomsten, de grootste verbeteringen werden waargenomen in de top 20% van patiënten.

Hoofdstuk 3 van dit proefschrift maakte ook gebruik van niet-traditionele overlevingsmaten om overlevingsuitkomsten beter te begrijpen en patiënten met slokdarm- en maagkanker beter te informeren. Conditionele relatieve overleving (conditional relative survival, CRS) is een nuttig hulpmiddel voor het communiceren van prognoses, omdat het de levensverwachting schat op basis van overleving gedurende een bepaalde periode na behandeling. In deze studie onderzochten we met behulp van gegevens uit de Nederlandse Kankerregistratie de 3-jaar CRS voor patiënten met niet-uitgezaaide slokdarm- of maagkanker die tussen 2006 en 2020 met curatieve intentie zijn behandeld. We vonden dat voor patiënten met slokdarmkanker de 3-jaar CRS toenam van 62% na het eerste jaar tot 87% na het vijfde jaar na diagnose. Voor maagkanker vonden we hogere percentages van 69% en 90% na 3 en 5 jaar, respectievelijk. Deze bevindingen tonen aan dat ondanks de doorgaans slechte prognose van deze kankertypes, de overlevingskansen in de loop van de tijd aanzienlijk verbeteren, wat patiënten en artsen een realistisch maar hoopvol perspectief biedt tijdens follow-up gesprekken.

Deel II: Gezondheidsgerelateerde kwaliteit van leven

In deel II van dit proefschrift behandelden we een andere, maar zeer relevante dimensie van de uitkomsten van kankerbehandeling: gezondheidsgerelateerde kwaliteit van leven (health related-quality of life, HRQoL). In tegenstelling tot overlevingsdata, die in Nederland beschikbaar zijn voor alle oncologische patiënten via de landelijke Nederlandse Kankerregistratie, is gegevensverzameling over HRQoL voornamelijk afhankelijk van prospectieve observationele cohortstudies. Dit zijn studies waarin patiënten actief om deelname gevraagd worden. Deelname van patiënten aan dergelijke studies is vrijwillig en vereist doorgaans dat patiënten vragenlijsten invullen over hun HRQoL. Deze vrijwillige component introduceert het risico van selectiebias (d.w.z. sommige patiënten zijn eerder geneigd deel te nemen dan anderen), waardoor de representativiteit van deze cohorten en de externe validiteit van de analyses beïnvloed kunnen worden. Daarom voerden we in **Hoofdstuk 4** een studie uit binnen het Prospectief Observationeel Cohortonderzoek van Slokdarm-maagkankerpatiënten (POCOP) om de representativiteit van patiënten in cohortstudies te evalueren ten opzichte van de totale populatie patiënten met slokdarm en maagkanker in Nederland.

Met behulp van een nieuwe methode genaamd Representativiteitsindicatoren (R-indicatoren), waarmee de totale representativiteit van een steekproef kan worden uitgedrukt in één enkel getal tussen 0 (niet representatief) en 1,0 (volledig representatief), vonden we dat de volledige POCOP-studie een R-indicator had van 0,72 (95%CI (0,71- 0,73)), wat wijst op een redelijke representativiteit. Patiënten in de palliatieve setting hadden een hogere representativiteit dan patiënten met potentieel curatieve ziekte, met R-indicatoren van respectievelijk 0,88 (95%CI (0,85-0,90)) en 0,70 (95%CI (0,68-0,71)). Wanneer gegroepeerd werd op klinisch relevante subgroepen op basis van behandeling, stegen de R-indicatoren voor elke subgroep (≥0,8 voor alle groepen). Na zowel stratificatie als kalibratieweging kwamen de overlevingsschattingen van de patiënten uit het POCOP-cohort dichter in de buurt van die uit de Nederlandse-populatie. Deze studie toont aan dat real-world data uit een prospectieve slokdarm en maagkankerregistratie representatief kunnen zijn voor de totale populatie patiënten, mits verschillen in behandelingen worden meegenomen. Dit ondersteunt het klinisch gebruik van patiënt gerapporteerde data, vooral wanneer deze gecorrigeerd zijn voor selectiebias door middel van stratificatie of statistische kalibratie.

Nadat we hadden vastgesteld dat het POCOP-register redelijk representatief was voor de totale populatie, zijn we verder gegaan met het verdiepen van ons begrip van HR-QoL-uitkomsten door ons te richten op de klinische voorspelling van HRQoL-uitkomsten. In **Hoofdstuk 5** ontwikkelden we predictiemodellen die de HRQoL na het ondergaan van behandeling voorspellen voor patiënten met slokdarm en maagkanker kanker, gebaseerd op HRQoL-gegevens uit POCOP en klinische variabelen uit de Nederlandse Kankerregistratie. De EORTC QLQ-C30 functioneringsschalen (inclusief de gemiddelde score) werden gebruikt als voorspellingsuitkomsten. Risicovoorspellingsmodellen, gebaseerd op logistische elastic-net regressie, voorspelden de kans op klinisch significante HRQoL-verslechtering op 3, 6 en 12 maanden na de behandeling. Daarnaast werd een machine learning-gebaseerd sequentieel scoremodel, met behulp van XGBoost regressie, ontwikkeld om toekomstige HRQoL-scores in de tijd te voorspellen. We vonden dat de risicovoorspellingsmodellen goed presteerden, met ICI-waarden tussen 0,03 en 0,08 en Brier-scores tussen 0,09 en 0,17, waarmee effectief achteruitgang in de gemiddelde score, Fysiek functioneren en Vermoeidheid werd voorspeld. De sequentiële scoremodellen verklaarden tot 40% van de variantie in HRQoL-scores. De resultaten van deze studie toonden aan dat beide modellen de HR-QoL-veranderingen voorspelden bij patiënten met slokdarm en maagkanker en waardevolle instrumenten bieden om de patiëntenzorg te verbeteren en gezamenlijke besluitvorming te ondersteunen via HRQoL-voorspellingen. Tegelijkertijd benadrukken de resultaten dat er nog ruimte is voor verbetering wat betreft de voorspellende kracht voor een aantal functioneringsschalen. Toekomstige studies zullen nodig zijn om de klinische voorspellingsmodellen voor HRQoL verder te verfijnen en het potentieel van informatie-extractie uit elektronische patiëntendossiers om gegevens uit HRQoL-cohorten aan te vullen zal moeten worden onderzocht.

Deel III: Klinische trials en real-world data

In Deel III van dit proefschrift lieten we zien hoe gegevens uit klinische trials en real-world bronnen elkaar kunnen aanvullen om behandelings-effecten te onderzoeken die (nog) niet in klinische trials zijn geëvalueerd. In deze hoofdstukken richtten we ons op twee relevante studies waarvoor geen volledige gerandomiseerde gecontroleerde trial werd uitgevoerd.

In **Hoofdstuk 6** gebruikten we gegevens uit de enkelarmige DESTINY-Gastric02 (DG-02) trial waarin HER2-positieve, niet-resectabele of uitgezaaide maag- of gastro-oesofageale junctiekanker werd behandeld met Trastuzumab Deruxtecan (T-DXd), en vergeleken we dit met een referentiegroep behandeld met Ramucirumab en Paclitaxel (Ram+Pac), verkregen uit de Nederlandse Kankerregistratie (NCR). Met behulp van propensity matching werden patiënten uit de DG-02 trial die eerder trastuzumab-gebaseerde eerstelijns therapie hadden gekregen, gematcht aan patiënten die Ram+Pac ontvingen uit de NCR, gebaseerd op relevante klinische factoren die verband houden met behandelkeuze en overleving. De resulterende gematchte gegevens waren in balans (58 DG-02-patiënten en 78 real-world patiënten). De mediane overleving was significant langer in de T-DXd-groep (11,6 maanden; 95% CI(9,0–20,5)) vergeleken met de Ram+Pac-groep (6,2 maanden; 95CI (4,5–10,0)), p-waarde < 0,0001. Deze bevindingen suggereren dat T-DXd een overlevingsvoordeel kan bieden ten opzichte van standaard tweedelijnsbehandeling in deze patiëntengroep en illustreren tevens het praktische gebruik van enkelarmige trials die worden aangevuld met real-world data om een gerandomiseerde trial na te bootsen.

Na de CheckMate-577 trial, die een voordeel in ziektevrije overleving aantoonde voor adjuvante nivolumab bij patiënten met slokdarm- of gastro-oesofageale junctiekanker (GEJ) na neoadjuvante chemoradiotherapie (nCRT) en resectie, bleef er onzekerheid bestaan over het effect op algehele overleving (OS), aangezien de definitieve OS-gegevens uit de trial op het moment van deze studie in dit hoofdstuk nog niet beschikbaar waren. Om deze onzekerheid aan te pakken, voerden we in **Hoofdstuk 7** een landelijke real-world studie uit om OS te evalueren bij patiënten die in de klinische praktijk al dan niet werden behandeld met adjuvante nivolumab. Patiënten met niet-gemetastaseerde slokdarm- of GEJ-kanker die tussen 2020 en 2023 werden gediagnosticeerd en geen complete pathologische response hadden na nCRT en resectie, werden geïdentificeerd uit de Nederlandse Kankerregistratie. In totaal werden 333 patiënten die adjuvante nivolumab kregen vergeleken met 486 patiënten die vóór de introductie van nivolumab werden behandeld met alleen nCRT en resectie. Propensity score trimming en nearest-neighbor matching op relevante klinische kenmerken resulteerden in twee goed gebalanceerde cohorten van elk 311 patiënten. De mediane follow-up bedroeg 24,4 maanden in de nivolumabgroep en 31,4 maanden in de controlegroep. De tweejaarsoverleving was significant hoger bij patiënten behandeld met nivolumab (66,8%, 95% CI (61,6–72,4%)) vergeleken met degenen zonder (58,8%, 95% CI (53,5–64,5%)). De geschatte hazard ratio was 0,75 (95% CI (0,60–0,97); p = 0,024). Deze bevindingen vormen het eerste bewijs uit de dagelijkse praktijk van een mogelijk OS-voordeel met adjuvante nivolumab in deze setting. Vanwege de lopende follow-up en het beperkte aantal events zijn echter vervolganalyses over langere termijn nodig om deze resultaten te bevestigen.

Deel IV. Voorspelling van overlevingsuitkomsten

In dit laatste deel van dit proefschrift richtten we ons op de implementatie, ontwikkeling en beoordeling van klinische predictiemodellen binnen het veld van slokdarm- en maagkanker.

Naast het ondersteunen van gezamenlijke besluitvorming tussen artsen en patiënten, kunnen predictiemodellen ook dienen als hulpmiddel bij patiëntselectie voor klinische trials. In **Hoofdstuk 8** evalueerden we de potentiële bruikbaarheid van het SOURCE-model (Stimulating evidence based, personalized and tailored information provision to improve decision making after Oesophagogastric Cancer diagnosis) en het veelgebruikte predictiemodel van Steyerberg en collega's, binnen de context van de POLDER-studie. Deze studie onderzocht het effect van uitwendige radiotherapie (external beam radiotherapy, EBRT) ter verlichting van dysfagie (moeite met slikken) bij patiënten met slokdarmkanker. Hoewel inclusie voor deze trial gebaseerd was op een verwachte overleving van ten minste drie maanden, bereikte bijna een derde van de patiënten deze drempel niet. Wij onderzochten of het SOURCE-model en het model van Steyerberg artsen beter hadden kunnen ondersteunen bij het selecteren van geschikte patiënten. Beide bestaande modellen werden retrospectief toegepast op de POLDER-data om de drie-maands overleving te voorspellen. Het SOURCE-model presteerde beter dan het model van Steyerberg, met een hogere accuraatheid (79% vs. 64%) en een significant betere discriminatie (AUC 0.76 vs. 0.60, p = 0.017). Bij optimale drempelwaarden zou het SOURCE-model geleid hebben tot minder foutieve inclusies (16 vs. 34 van de 110 patiënten) in de POLDER-studie. Deze bevindingen suggereren dat het SOURCE-model waardevol kan zijn als beslisondersteuning naast klinisch oordeel in palliatieve studies.

Na het aantonen van de klinische waarde van het bestaande SOURCE-model, zetten we de ontwikkeling van nieuwe klinische predictiemodellen voort in **Hoofdstuk 9**. Hier ontwikkelden en valideerden we intern het SOURCE Beyond First-Line model, een predictietool die de overleving inschat voor patiënten met uitgezaaid slokdarm adenocarcinoom na falen van eerstelijns palliatieve systemische therapie, en daarmee een uitbreiding vormt op de SOURCE-familie van predictiemodellen. Met gegevens van 1.067 patiënten uit de Nederlandse Kankerregistratie ontwikkelden we een Cox proportionele hazardmodel gebaseerd op 11 klinische en behandelingsgerelateerde variabelen. Het model liet goede prestaties zien, met een C-index van 0.75 (95% CI (0.73–0.78)), een kalibratiehelling van 1.01 en een intercept van 0.01. Interne kruisvalidering bevestigde consistente prestaties op ongeziene data, met een C-index van 0.79 (95% CI (0.77–0.82)), een helling van 0.93 en een intercept van 0.02. Deze resultaten wijzen erop dat het SOURCE Beyond First-Line model een betrouwbare voorspeller is met goede kalibratie, en daarmee een bruikbare tool vormt ter ondersteuning van klinische besluitvorming in de beyond-first-line behandelsetting.

In **Hoofdstuk 10** onderzochten we de mogelijkheid om het bestaande SOURCE-model voor patiënten met resectabel slokdarm adenocarcinoom te verbeteren door het integreren van radiomics en metingen van circulerend tumor-DNA (cfDNA). In een cohort van 111 patiënten die behandeld werden met neoadjuvante chemoradiotherapie, ontwikkelden we modellen die klinische variabelen uit het SOURCE-model combineerden met radiomics kenmerken van PET-CT-scans en cfDNA. De best presterende modellen voor algehele overleving (OS) en tijd tot progressie (TTP) toonden verbeterde C-indexscores wanneer cfDNA werd toegevoegd aan het SOURCE-model (C-index 0.55 voor OS en 0.59 voor TTP, vergeleken met 0.44–0.45 met alleen het SOURCE-model). Het toevoegen van PET-CT-radiomics aan het SOURCE-model verbeterde de voorspellingen verder, met een C-index van 0.65 voor OS en 0.60 voor TTP. De gecombineerde modellen toonden ook een significante risicostratificatie voor zowel OS als TTP (log-rank $P < 0.01$). Voor de voorspelling van pathologische complete respons behaalde het beste model een AUC van 0.61, vergeleken met 0.47 bij gebruik van alleen klinische variabelen. Deze bevindingen suggereren dat het integreren van radiomics en cfDNA de voorspellende waarde van het SOURCE-model vergroot, hoewel externe validatie en verdere optimalisatie van radiomic-pijplijnen nodig zijn in toekomstig onderzoek.

Hoewel de predictiemodellen die in de voorgaande hoofdstukken zijn toegepast en ontwikkeld goede prestaties en klinische relevantie toonden, werd er vanuit methodologisch oogpunt een belangrijk aspect nog niet meegenomen: causaliteit. Wanneer patiënten en artsen voor de keuze van een behandeling staan, kunnen predictiemodellen – zoals die in dit proefschrift – hierbij ondersteunen. Echter, omdat deze modellen meestal zijn ontwikkeld op observationele data, kunnen de voorspellingen niet causaal worden geïnterpreteerd. Daarom pasten we in **Hoofdstuk 11** een nieuwe benadering toe, namelijk counterfactual prediction. Deze methode combineert causale inferentie met gepersonaliseerde overlevingsvoorspellingen, en maakt het mogelijk om causale, persoonsgebonden behandelvoorspellingen te doen. We pasten deze methode toe in een proof-of-concept studie bij patiënten met niet-uitgezaaide slokdarmkanker, waarin twee behandelstrategieën werden vergeleken: neoadjuvante chemoradiotherapie (nCRT) en definitieve chemoradiotherapie (dCRT). Met data uit de Nederlandse Kankerregistratie (4.388 patiënten voor training en 1.693 voor validatie) ontwikkelden we twee verschillende counterfactual overlevingsmodellen (een parametrisch overlevingsmodel en een random forest-model) gebaseerd op inverse propensity

score-weighting om confounding te corrigeren. Beide modellen lieten goede kalibratie zien in de trainings- en validatiesets, waarbij het parametrisch model goed presteerde voor nCRT en het random forest-model beter aansloot bij de overlevingsvoorspellingen voor dCRT. Daarnaast testten we het bestaande SOURCE-model in een counterfactual context, waarbij het ondanks dat het daar niet voor was ontwikkeld, redelijk goed gekalibreerd bleek. Deze studie laat zien dat counterfactual prediction-modellen potentie hebben om gepersonaliseerde besluitvorming te verbeteren in klinische situaties met meerdere behandelopties. Dit ondersteund artsen en patiënten in het inschatten van overlevingskansen.

Conclusie

Samenvattend laat dit proefschrift zien hoe statistische en machine learning-methoden kunnen worden ingezet om beter gebruik te maken van real-world data bij slokdarm- en maagkanker. Door te kijken naar overlevingsuitkomsten, gezondheidsgerelateerde kwaliteit van leven en predictiemodellen, hebben we inzichten verkregen die verder gaan dan traditionele analyses. De bevindingen in dit proefschrift dragen bij aan een beter klinisch inzicht, kunnen gezamenlijke besluitvorming ondersteunen en leiden hopelijk tot meer gepersonaliseerde zorg voor patiënten met slokdarm- en maagkanker.

Dankwoord

De totstandkoming van mijn proefschrift heb ik te danken aan veel mensen in mijn directe en indirecte omgeving. In dit dankwoord wil ik, voor zover dat mogelijk is, iedereen persoonlijk bedanken die mij in de afgelopen jaren en in de jaren heeft bijgestaan.

Allereerst wil ik mijn dank uitspreken aan mijn promotores Hanneke van Laarhoven en Rob Verhoeven. Hanneke, ik heb diepe bewondering voor jou als persoon en als wetenschapper. Je hebt mij op persoonlijk vlak in een aantal cruciale momenten gedurende mijn promotietraject bijgestaan en geholpen en hebt mij ongelooflijk veel geleerd over wat het betekend om een goede wetenschapper kunt zijn met het juiste moreel- en wetenschappelijk-ethisch kompas. Dit zijn lessen die ik mijn hele carrière bij zullen blijven. Dank, dank, dank! Rob, bij jou kon ik altijd voor alles en met alles terecht. Of het nu ging om wel of niet te imputeren of hoe het ging met de hond, je was altijd bereikbaar en in voor een praatje. Wat ik ontzettend waardevol vond aan samenwerken met jou, was dat ik met jou heel informeel kon sparren bij het koffieautomaat en dat je heel makkelijk benaderbaar was. Ook jij hebt mij laten zien wat het betekend en wat er voor nodig is om een goed wetenschapper te zijn. Bijzonder veel dank daarvoor.

Daarnaast wil ik al mijn lieve collega's bedanken voor de samenwerkingsprojecten en de lunch- en koffiemomentjes. Allereest mijn lieve Chicago-vrienden Benthe Doeve en Joris Bos (Mr. J, zoals ik hem sinds dien noem). Benthe, ik vind het nog steeds niet te geloven dat je geen COVID van mij hebt gekregen na het delen van de drinkwaterflesje, maar ben heel blij dat het niet gebeurd is. Mr J., sorry dat ik je heb aangestoken met COVID, maar we hebben flink plezier gemaakt daar in de US of A. Het was een feest om met jullie op pad te gaan naar Chicago, ondanks COVID en het superspreader event wat ik daar heb veroorzaakt. Hoewel mijn R-getal niet gemeten is, was deze erg hoog vermoed ik. Om verschillende redenen was het een reis om nooit meer te vergeten.

Dan mijn San Fransisco buddies Marieke Pape, Tom van den Ende, Linde Veen en Merel van Velzen. Marieke, ook met jou was de samenwerking ook dol-fijn. In zo'n beetje alles waren wij elkaars tegenpolen, maar ik denk dat wij juist daarom zo goed samenwerkte en het zo goed met elkaar konden vinden. Of het nou ging over een artikel of over het al dan niet nemen van een hond, een stevige discussie zat er altijd wel in. Dank hiervoor! Tom, jij was (en bent) voor mij een voorbeeld. Ik denk dat ik weinig mensen ken die al zo vroeg in hun carrière zo'n goede wetenschapper zijn. Ik ben ervan overtuigd dat het niet een kwestie is van "of", maar meer "hoe lang" voordat jij hoogleraar wordt (onthoud deze voorspelling!). Ook twijfel ik er niet aan dat je een fantastisch arts zal worden. Dit laatste geld ook voor Merel. Merel, ik heb ontzettend bewondering voor de manier waarop jij je promotie hebt afgemaakt en je begonnen bent aan de opleiding tot specialist. Ik weet zeker dat je een geweldige arts zult zijn. Linde, jij was de creatieve geest van onze groep. Ik kijk er nog met plezier op terug hoe ik met een Marieke-masker een filmpje voor haar promotie aan het opnemen was. Dank voor alle koffiemomentjes en dank voor het lachen.

Ook dank aan Irene Cara (hope you don't mind me thanking you in Dutch). Het was super fijn om iemand op de afdeling te hebben met wie ik kon sparren over technische machine learning en statistische modellen. Ook wij konden het super goed vinden met elkaar en ik wens je het allerbeste toe! Sebastiaan! Ondanks wij maar relatief kort met elkaar hebben gewerkt, hebben wij wel intensief samengewerkt aan de herintroductie van oxaliplatin-studie. Ik vind je een ontzettend fijne gast en ik wens je het allerbeste toe.

Ook niet te vergeten, mijn vaste koffie-maatje Monique! Monique, jij was altijd wel te porren voor een kop koffie en een praatje, ongeacht hoe druk je was. Ik heb jou denk ik nog nooit gestrest gezien. Dank voor het lachen en de leuke tijd op kantoor!

Daarnaast wil ik ook al mijn andere onco-collega's bedanken: Marjolein, Dionne, Jelijn, Myrthe, Hylke, Sophie en Geerke. Passen jullie er wel op dat jullie nog genoeg koffie drinken? En nog een aantal oud collega's die eerder klaar waren dan ik: Charlotte, Loïs en Zoë. Dank iedereen voor de leuke tijd een mooie herinneringen.

Een bijzonder dank ook aan al mijn co-auteurs met wie ik het voorrecht had zoveel mooie artikelen te schrijven. Zonder de hulp en medewerking van jullie allemaal, had ik het niet kunnen doen.

Dan mijn paranimf to-be: Chevy. Chevy, naast een goede vriend ben jij een van de slimste mensen die ik ken (dat zegt veel over jou, of mij, jij mag kiezen). Jij wist gedurende mijn promotie altijd alles te relativeren, soms op het irritante af, maar ik heb dat altijd gewaardeerd. Dank voor je vriendschap en opdat wij nog maar naar veel concerten mogen gaan!

Mijn andere paranimf to-be: Frenk. Frenk, jij bent een van mijn oudste en beste vrienden. Wij leerden elkaar kennen bij het debatclubje van onze middelbare school (ik weet dat je zit te grinniken als je dit leest) en tot soms ergernis van ons beiden zijn we altijd bevriend gebleven (grapje natuurlijk). Het was altijd heerlijk om de frustraties van ons werk naar elkaar te spuien. Ik waardeer je eerlijk altijd en hoop dat wij altijd bevriend blijven (en meer gaan golfen).

Daarnaast heb ik door mijn hele studietijd van de middelbare school tot aan nu een aantal bijzondere mensen in mijn leven gehad die mij onder hun vleugel hebben genomen die ik ontzettend wil bedanken. De eerste die dat heeft gedaan, is René Bok. René, ik was denk ik 14 toen wij met bijlessen wiskunde begonnen aan de keukentafel. Urenlang hebben wij wiskunde opdrachten en oefentoetsen gemaakt om ervoor te zorgen dat ik eerst de mavo en later de havo zou halen. Later in ons leven zijn wij altijd bevriend gebleven en delen wij nog steeds dezelfde liefde voor muziek en alle bourgondische dingen in het leven. Dank voor alles! Zonder de kickstart aan het begin van mijn schooltijd had ik nooit geweest waar ik nu ben.

Een ander sleutelfiguur voor mijn interesse in te wetenschap, is mijn oud natuurkunde leraar Paul Logman. Paul, ik kan mij nog zo herinneren dat ik op jouw promotieceremonie te gast was en dat jij tegen mij zei "als jij doorgaat op de manier waarop je nu bezig bent, weet ik zeker dat jij ook ooit hier staat". Inmiddels zijn we heel wat jaren verder, ben ik een andere studie gaan vervolgen, maar komt je voorspelling uit: ik promoveer (als alles op de dag zelf goed gaat) bij dezelfde universiteit. Dank voor alles!

Toen ik eenmaal wist dat ik een carrière als onderzoeker ambieerde, liep ik als stagestudent tijdens mijn bachelor binnen bij het Kohnstamm instituut, waar ik Joost Meijer leerde kennen. Joost was een methodoloog pur sang en heeft mij ontzettend veel geleerd. Onder anderen, dat af en toe een beetje tegendraads en recalcitrant geen slechte, maar een goed eigenschap is in de onderzoekswereld. Joost, ontzettend bedank voor alles. Ik denk nog met veel plezier terug aan alle uren programmeren (en vloeken) met Mplus.

In het staartje van mij studententijd heb ik ontzettend fijn samengewerkt met Peter Hoffenaar, Hennie Bos, Geertjan Overbeek en Kees Jan Kan. Allemaal ontzettend bedankt voor de fijne samenwerking en wijze lessen.

Tot slot wil ik mijn lieve familie en schoonfamilie bedanken. Allereest mijn vader en moeder, Peter en Jenny. Ik denk dat het niet in woorden te vatten is wat jullie voor mij hebben betekend (en betekenen). Ik schrijf mijn succes toe aan alles wat jullie mij hebben meegegeven in het leven. Ik ben jullie onbeschrijfelijk dankbaar voor de liefde en steun die ik van jullie in alles heb gekregen en ik kijk er ontzettend naar uit om jullie niet alleen als ouders in mijn leven te hebben, maar ook als opa en oma. Pap, mam, dank voor alles; ik houd van jullie! Dan mijn lieve broer Sam, mijn schoonzus Geena mijn bonusmoeder Ilse. Ik ben ieder van jullie dankbaar voor wat jullie hebben betekend (en betekenen) in mijn leven. Ik houd van jullie allemaal!

Dan mijn lieve schoonfamilie: Frank, Claudia, Jamie, Sep, Mick en Fabian. Inmiddels kennen wij elkaar al 13 jaar en ben ik jullie ontzettend dankbaar voor alles! Sep, ik ken jou al vanaf je 6e. Inmiddels ben je 19 (ik schrijf dit in augustus 2025) en studeer je natuurkunde. Wie weet, misschien sta ik over heel wat jaren wel in het dankwoord van jouw boekje. Wat je ook kiest, ik weet zeker dat je er succesvol in zal zijn.

Als allerlaatste wil ik mijn lieve Suus bedanken. Wij kennen elkaar al sinds de middelbare school en nu, 13 jaar later, staan we op het punt om te trouwen en zijn wij de trotse ouders van onze prachtige dochter Isabella. Ik ben je onwijs dankbaar voor alles en kan niet wachten op de rest van ons leven. Ik houd van je!

About the author

Steven Kuijper was born on December 11, 1993, in Haarlem, the Netherlands. After completing his secondary education at the Vellesan College in IJmuiden, he initially set out to become a physics teacher, earning his propaedeutic diploma at the Hogeschool van Amsterdam. His curiosity for how people learn and develop led him to the University of Amsterdam, where he obtained a bachelor's degree in Educational Sciences and later a Research Master's degree in Child Development and Education.

During his studies, Steven discovered an interest for statistical modelling. In 2020, he began his PhD at the Department of Medical Oncology, focusing on prediction models and real-world data in esophagogastric oncology. This work combined his interest in statistics, data science and methodology with a drive to improve healthcare outcomes.

After completing his PhD, Steven continued this path at Danone Research & Innovation, where he works as a Clinical Data Scientist, applying data-driven approaches to advance nutritional science and health.

PhD portfolio

Name PhD student: Steven C. Kuijper **PhD period: August 2020 – November 2024** **Names of PhD supervisor(s) & co-supervisor(s):**	**Hanneke W.M. van Laarhoven; Rob H.A. Verhoeven**	
	Year	ECTS
Courses		
Advanced topics in biostatistics	2021	1.5
Observational epidemiology – Effects and Effectiveness	2022	0.6
Clinical epidemiology: Randomized controlled trials	2022	0.6
Basic oncology	2021	2.0
Oral presentations		
Scientific meetings Dutch Upper-GI Cancer Group (DUCG)		
"Overleving in de POLDER studie	2021	0.5
"Vergelijking in overleving van patiënten met slokdarm- en maagkanker tussen Nederland en België	2024	0.5
Netherlands Comprehensive Cancer Organisation (IKNL):		
"Survival in percentiles"	2022	0.5
"SOURCE Beyond first-line"	2023	0.5
"Real-world representativeness of POCOP"	2023	0.5
"SOURCE, Radiomics & cfDNA"	2024	0.5
"Quality of life prediction"	2023	0.5
Symposium: "Causal inference, causal prediction and impactions for health-related quality of life"	2025	1.0

(Continues on next page)

(Inter)national conferences		
ASCO GI (Poster, online)	2022	1.4
S.C. Kuijper, M. Pape, P.A.J. Vissers, P.M. Jeene, E.A. Kouwenhoven, N. Haj Mohammad, M. Sosef, R.H.A. Verhoeven, H.W.M. van Laarhoven, Survival expressed in best-case, typical and worst-case scenarios for patients with non-metastatic esophagogastric cancer: a population-based study, ASCO GI, January 20-22, 2022 (poster)		
P.M. Jeene, S.C. Kuijper, H.G. van den Boorn, MSc, S.Y. El Sharouni, P.M Braam, V. Oppedijk, R.H.A.Verhoeven, M.C.C. Hulshof, H.W.M. van Laarhoven, Improving survival prediction in patients treated with external beam radiotherapy for dysphagia in esophageal cancer using prediction models, ASCO GI, January 20-22, 2022 (poster)		
ASCO (Poster, Chicago)	2022	0.7
S.C. Kuijper, M. Pape, N. Haj-Mohammad, T. van Voorthuizen, R.H.A. Verhoeven, H.W.M. van Laarhoven. SOURCE Beyond First-line: A survival prediction model for patients with metastatic esophagogastric adenocarcinoma after failure of first-line therapy, ASCO , 2022 (Poster)		
Internation Association of Cancer Registries (IACR) (Oral presentation, online)	2022	0.9
S.C. Kuijper, M. Pape, N. Haj-Mohammad, T. van Voorthuizen, R.H.A. Verhoeven, H.W.M. van Laarhoven. A survival prediction model for patients with metastatic esophagogastric adenocarcinoma after failure of first-line therapy, IARC, 2022 (oral presentation)		
ASCO GI (Poster, San Francisco)	2023	1.0
S.C. Kuijper, J. Besseling, T. Klausch, M. Slingerland, C.J. van der Zijden, E.A. Kouwenhoven, L.V. Beerepoot, N. Haj Mohammad, B.R Klarenbeek, R.H.A. Verhoeven, H.W.M. van Laarhoven. Real-world representativeness of patient reported outcome measures of patients with esophagogastric cancer. ASCO GI, January 19-21, 2023 (poster)		

(Continues on next page)

Teaching		
Supervision internship and bachelor thesis	2022-2023	3.0
Statistiekclub workshop: "Predictiemodellen"	2023	0.25
Masterclass: "The methodology of clinical prediction models"	2023	0.25
Other acitivities		
OOA retreat (Poster)	2022	0.5
CCA retreat (Poster)	2023	0.5
APH: member of the junior board	2020-2024	7.0

List of publications

Publications included in this thesis

Ende TVD, **Kuijper SC**, Widaatalla Y, Noortman WA, van Velden FHP, Woodruff HC, van der Pol Y, Moldovan N, Pegtel DM, Derks S, Bijlsma MF, Mouliere F, de Geus-Oei LF, Lambin P, van Laarhoven HWM. Integrating Clinical Variables, Radiomics, and Tumor-derived Cell-Free DNA for Enhanced Prediction of Resectable Esophageal Adenocarcinoma Outcomes. Int J Radiat Oncol Biol Phys. 2025 Mar 15;121(4):963-974. doi: 10.1016/j.ijrobp.2024.10.010. Epub 2024 Oct 16. PMID: 39424077.

Kuijper SC, Besseling J, Klausch T, Slingerland M, van der Zijden CJ, Kouwenhoven EA, Beerepoot LV, Mohammad NH, Klarenbeek BR, Verhoeven RHA, van Laarhoven HWM. Assessing real-world representativeness of prospective registry cohorts in oncology: insights from patients with esophagogastric cancer. J Clin Epidemiol. 2023 Dec;164:65-75. doi: 10.1016/j.jclinepi.2023.10.009. Epub 2023 Oct 21. PMID: 37871837.

Kuijper SC, Pape M, Vissers PAJ, Jeene PM, Kouwenhoven EA, Haj Mohammad N, Ruurda JP, Sosef MN, Verhoeven RHA, van Laarhoven HWM. Trends in best-case, typical and worst-case survival scenarios of patients with non-metastatic esophagogastric cancer between 2006 and 2020: A population-based study. Int J Cancer. 2023 Jul 1;153(1):33-43. doi: 10.1002/ijc.34488. Epub 2023 Mar 24. PMID: 36855965.

Pape M, **Kuijper SC**, Vissers PAJ, Ruurda JP, Neelis KJ, van Laarhoven HWM, Verhoeven RHA. Conditional relative survival in nonmetastatic esophagogastric cancer between 2006 and 2020: A population-based study. Int J Cancer. 2023 Jun 15;152(12):2503-2511. doi: 10.1002/ijc.34480. Epub 2023 Mar 5. PMID: 36840612.

Kuijper SC, Pape M, Haj Mohammad N, van Voorthuizen T, Verhoeven RHA, van Laarhoven HWM. SOURCE beyond first-line: A survival prediction model for patients with metastatic esophagogastric adenocarcinoma after failure of first-line palliative systemic therapy. Int J Cancer. 2023 Mar 15;152(6):1202-1209. doi: 10.1002/ijc.34385. Epub 2022 Dec 9. PMID: 36451334; PMCID: PMC10107625.

Jeene PM, **Kuijper SC**, van den Boorn HG, El Sharouni SY, Braam PM, Oppedijk V, Verhoeven RHA, Hulshof MCCM, van Laarhoven HWM. Improving survival prediction of oesophageal cancer patients treated with external beam radiotherapy for dysphagia. Acta Oncol. 2022 Jul;61(7):849-855. doi: 10.1080/0284186X.2022.2079385. Epub 2022 Jun 1. PMID: 35651320.

Other publications

Kuijper SC, Gehrels AM, van der Geest LG, Verhoeven RHA, Koerkamp BG, Molenaar IQ, Stommel MWJ, de Meijer VE, de Vos-Geelen J, Wumkes ML, Besselink MG, Wilmink JW, van Laarhoven HWM; Dutch Pancreatic Cancer Group. Survival scenarios of patients with localized and metastatic pancreatic adenocarcinoma: A population-based study. Int J Cancer. 2025 May 1;156(9):1726-1735. doi: 10.1002/ijc.35267. Epub 2024 Nov 29. PMID: 39614657; PMCID: PMC11887001.

van de Water LF, Bos-van den Hoek DW, **Kuijper SC**, van Laarhoven HWM, Creemers GJ, Dohmen SE, Fiebrich HB, Ottevanger PB, Sommeijer DW, de Vos FYF, Smets EMA, Henselmans I. Potential Adverse Outcomes of Shared Decision Making about Palliative Cancer Treatment: A Secondary Analysis of a Randomized Trial. Med Decis Making. 2024 Jan;44(1):89-101. doi: 10.1177/0272989X231208448. Epub 2023 Nov 12. PMID: 37953598; PMCID: PMC10712204.

van de Water LF, **Kuijper SC**, Henselmans I, van Alphen EN, Kooij ES, Calff MM, Beerepoot LV, Buijsen J, Eshuis WJ, Geijsen ED, Havenith SHC, Heesakkers FFBM, Mook S, Muller K, Post HC, Rütten H, Slingerland M, van Voorthuizen T, van Laarhoven HWM, Smets EMA. Effect of a prediction tool and communication skills training on communication of treatment outcomes: a multicenter stepped wedge clinical trial (the SOURCE trial). EClinicalMedicine. 2023 Sep 25;64:102244. doi: 10.1016/j.eclinm.2023.102244. PMID: 37781156; PMCID: PMC10539636.

Pape M, Veen LM, Smit TM, **Kuijper SC**, Vissers PAJ, Geijsen ED, van Rossum PSN, Sprangers MAG, Derks S, Verhoeven RHA, van Laarhoven HWM. Late Toxicity and Health-Related Quality of Life Following Definitive Chemoradiotherapy for Esophageal Cancer: A Systematic Review and Meta-analysis. Int J Radiat Oncol Biol Phys. 2023 Sep 1;117(1):31-44. doi: 10.1016/j.ijrobp.2023.05.025. Epub 2023 May 22. PMID: 37224927.

Pape M, **Kuijper SC**, Vissers PAJ, Beerepoot LV, Creemers GJ, van Laarhoven HWM, Verhoeven RHA. Beyond Median Overall Survival: Estimating Trends for Multiple Survival Scenarios in Patients With Metastatic Esophagogastric Cancer. J Natl Compr Canc Netw. 2022 Dec;20(12):1321-1329.e4. doi: 10.6004/jnccn.2022.7066. PMID: 36509070.

van der Velden NCA, van Laarhoven HWM, Nieuwkerk PT, **Kuijper SC**, Sommeijer DW, Ottevanger PB, Fiebrich HB, Dohmen SE, Creemers GJ, de Vos FYFL, Smets EMA, Henselmans I. Attitudes Toward Striving for Quality and Length of Life Among Patients With Advanced Cancer and a Poor Prognosis. JCO Oncol Pract. 2022 Nov;18(11):e1818-e1830. doi: 10.1200/OP.22.00185. Epub 2022 Oct 6. PMID: 36201709.

Hensums M, de Mooij B, **Kuijper SC;** BIRC: the anti-Bullying Interventions Research Consortium; Fekkes M, Overbeek G. What Works for Whom in School-Based Anti-bullying Interventions? An Individual Participant Data Meta-analysis. Prev Sci. 2023 Nov;24(8):1435-1446. doi: 10.1007/s11121-022-01387-z. Epub 2022 Jul 7. PMID: 35796879; PMCID: PMC10678813.

Manuscripts in preperation

SC Kuijper, BH Doeve, G Silversmit, L van Walle, P Nafteux, C Rosman, PAJ Vissers, P Jeene, L Beerepoot, S Derks, A Karimi, MF Bijlsma, HWM van Laarhoven, RHA Verhoeven. Treatment and survival of patients with gastric and esophageal cancer in the Netherlands and Belgium: a population-based comparison. *Under review at International Journal of Cancer.*

SC Kuijper, I Cara, G Geleijnse, M Slingerland, GAP Nieuwenhuijzen, S Lagarde, BR Klarenbeek, E Kouwenhoven, R van Hillegersberg, RHA Verhoeven, HWM van Laarhoven. Predicting health-related quality of life for patients with gastroesophageal cancer. *Under review at Quality of Life Research*.

RHA Verhoeven, **SC Kuijper**, M Slingerland, B Wijnhoven, MI van Berge Henegouwen, PSN van Rossum, S Derks, B Mostert, N Haj Mohammad, HWM van Laarhoven. Adjuvant nivolumab after chemoradiotherapy and resection for patients with esophageal cancer: a real-world matched comparison of overall survival. *Manuscript submitted to the International Journal of Cancer.*

RHA Verhoeven, **SC Kuijper**, F Lordick, M Slingerland, A Qin, HWM van Laarhoven. Trastuzumab Deruxtecan vs Ramucirumab-Paclitaxel as second-line therapy for patients with HER2-positive gastric or GEJ adenocarcinoma. *Manuscript accepted for publication in the Journal of the National Comprehensive Cancer Network (JNCCN).*

SC Kuijper, T Klausch, BR Klarenbeek, PSN van Rossum, RHA Verhoeven, HWM van Laarhoven. Neoadjuvant chemoradiotherapy versus definitive chemoradiotherapy for patients with esophageal cancer: Development and validation of a counterfactual prediction model on real-world data. *Manuscript in preparation.*