

AUGIS Surv-G: Prediction of long-term survival after gastrectomy using Random Survival Forests

First author: SA Rahman^{1,5},

Co-authors: N Maynard², N Trudgill³, T Crosby⁴, Park M⁵, Wahedally H⁵, TJ Underwood¹, D A Cromwell⁵ on behalf of the NOGCA project team and AUGIS

1. School of Cancer Sciences, Faculty of Medicine, University of Southampton
2. Oxford University Hospitals NHS Trust
3. Sandwell and West Birmingham NHS Trust
4. Velindre Cancer Centre, Cardiff
5. Clinical Effectiveness Unit, Royal College of Surgeons of England

Corresponding Author: Mr S A Rahman

Email: srahman@soton.ac.uk

Tel: +44 (0)2381206923

Running Head

Augis-SURV-G: Prediction of survival after gastrectomy

Key Words

Gastric Cancer, Prognostic Model, Machine Learning

Disclosures

The authors present no conflicts of interest.

Article Type

Original article

Patient and public involvement

Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research. Patient consent for publication was not required.

Ethics approval

The study is exempt from UK National Research Ethics Committee approval as it involved secondary analysis of an existing dataset of anonymised data. The National Oesophago-Gastric (OG) Cancer Audit has approval for processing health care information under Section 251 (reference number: ECC 1-06 (c)/2011) for all National Health Service (NHS) patients diagnosed with OG cancer in England and Wales. Data for this study are based on patient-level information collected by the NHS, as part of the care and support of patients with cancer.

Acknowledgements and Funding

This study was undertaken as part of the work by the National Oesophago-Gastric (OG) Cancer Audit. The Audit is commissioned by the Healthcare Quality Improvement Partnership (HQIP) as part of the National Clinical Audit and Patient Outcomes Programme and funded by NHS England and the Welsh Government (www.hqip.org.uk/national-programmes). The authors had full independence from the Healthcare Quality Improvement Partnership. The aim of the National Oesophago-Gastric Cancer Audit is to evaluate the care of patients with OG cancer in England and Wales, and support NHS providers to improve the quality of hospital care for these patients. More information can be found at: www.nogca.org.uk

SAR is supported by a Royal College of Surgeons of England Research Fellowship and a British Association of Surgical Oncologists Research Project Grant.

TJU is supported by a Cancer Research UK and Royal College of Surgeons of England Advanced Clinician Scientist Fellowship, ID:A23924.

This project has been supported by the Association of Upper Gastrointestinal Surgery (AUGIS), Heartburn Cancer UK, and The Royal College of Surgeons of England Surgical Specialty Lead Programme.

Disclaimer

Neither HQIP nor the funders had any involvement in the study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Abstract

Background

No well validated and contemporaneous tools for personalised prognostication of gastric adenocarcinoma exist. This study aimed to derive and validate a prognostic model for overall survival after surgery for gastric adenocarcinoma using a large national dataset and a non-linear Random Survival Forest (RSF) methodology.

Patients and methods

National audit data from England and Wales were used to identify patients who underwent a potentially curative gastrectomy for adenocarcinoma of the stomach. A total of 2931 patients were included and 29 clinical and pathological variables considered for their impact on survival. A RSF was then trained and validated internally using bootstrapping with calibration and discrimination (time dependent AUC) assessed.

Results

The median survival of the cohort was 69 months, with a 5-year survival of 53.2%. Ten variables were found to significantly influence survival and included in the final model, with the most important being lymph node positivity, pT stage and achieving an R0 resection. Patient characteristics including ASA grade and age were also influential. On validation the model achieved excellent performance with a five-year tAUC of 0.80 (95%CI 0.78-0.82) and good agreement between observed and predicted survival probabilities. A wide spread of predictions for three- (14.8-98.3%, IQR 43.2-84.4%) and five-year (9.4-96.1%, IQR 31.7-73.8%) survival were seen.

Conclusions

A prognostic model for survival after a potentially curative resection for gastric adenocarcinoma was derived and exhibited excellent discrimination and calibration of predictions. After appropriate external validation, it could provide utility in both prognostication for patients and for benchmarking of treatment responses.

Highlights

- No well validated contemporaneous prognostic model for Gastric Adenocarcinoma is in widespread clinical use
- This study describes the derivation of a random survival forest model using routine data from a large population dataset
- The model performed well on internal validation with a tAUC of 0.80 and excellent calibration
- A wide range of predictions were yielded for each TNM stage
- After appropriate external validation, it could provide utility in both prognostication for patients and for benchmarking of treatment responses.

Introduction

Gastric cancer is among the most common causes of cancer and cancer mortality worldwide, with an estimated 1,000,000 cases and 783,000 deaths in 2018.(1) Similar to oesophageal cancer, gastric cancer is more common among men than women, and the majority of cases occur in East Asia, where an incidence of up to 32 per 100,000 is seen overall. In comparison, in Northern Europe, the incidence is to 6.2 per 100,000 in men and 3.1 per 100,000 in women. In England and Wales there is a significant burden of disease, with 5,972 cases of gastric adenocarcinoma diagnosed between April 2017 and March 2019,(2) and among those only around one third suitable for curative treatment at presentation.

Among western populations, stratification of patient outcomes is limited to TNM stage, with tools for personalised prognostication which incorporate other variables known to influence survival lacking. In a recent systematic review of prognostic tools in oesophageal and gastric cancer(3) only one model suitable for gastric cancer was considered to be methodologically sound,(4) however this study was conducted in 2003 before the widespread use of neoadjuvant treatment and was limited to patients undergoing R0 resection. A further review(5) reached similar conclusions, identifying generally poor methodology and poor validation strategies among studies. Accurate postoperative prognostication is important as it allows personalised planning of both follow up and potential adjuvant treatment in addition to accurate comparison of different treatment regimens between groups of patients. No such tool to achieve this exists to date.

It is likely that in the future in-depth analysis of patients' cancers will allow for a high level of accuracy of prognostication both in the pre- and post- treatment settings, however these methodologies are not yet widely available, are time consuming and expensive. Optimal use of clinical data is therefore key. Machine learning techniques which incorporate non-linear effects, interactions between variables and time-varying effects have the potential to capture additional information from routine clinical data that may be missed by traditional prognostic models such as the Cox proportional hazards.

Recently, data from the England and Wales National Oesophago-gastric Cancer Audit (NOGCA) has been used to derive a Random Survival Forest(6) model for prognosis after oesophagectomy with considerable accuracy in excess of a Cox Proportional Hazards model(7). This study aims to apply a similar methodology to patients diagnosed with gastric adenocarcinoma in England and Wales between 2012-2018 with the goal of deriving an accurate prediction tool for overall survival after surgery.

Methods

Patient population

This study used a dataset of cases identified from the National Oesophago-gastric Cancer audit (NOGCA) as has been described previously(7). Data entry into the NOGCA has been compulsory for all patients diagnosed with epithelial cancer of the stomach or oesophagus since 2012, with named clinicians responsible for its collection as part of the multidisciplinary team. Each year, centres and surgeons are sent their results prior to publication and are asked to update incomplete or inaccurate data. Case ascertainment is evaluated using the national administrative hospital databases (Hospital Episode Statistics/HES in England and its Welsh equivalent), and is estimated to exceed 99% for patients who undergo curative surgery. The dataset used for this study included patients diagnosed between April 2012 and March 2018 (8). Details of neoadjuvant and adjuvant treatment were cross-referenced with the Systemic Anti-Cancer Therapy dataset (SACT). A total of 4,238 cases who underwent a gastrectomy for adenocarcinoma of the stomach or gastro-oesophageal junction (Siewert III) were identified. Exclusion criteria included overt metastatic disease at resection (pM1), death prior to discharge from hospital or if fewer than 15 lymph nodes were examined from the resection specimen (suggesting the patient may have been incompletely staged).(9) A comparison of these patients to the main study cohort is provided in Table S.1 and Figure S.1. A complete list and details of exclusions to reach the final sample size of 2931 cases is given in Figure S.2. The primary outcome was defined as overall survival from time of hospital discharge, with survival confirmed using the Office for National Statistics (ONS) death register.

Variables collected in the audit were considered for inclusion if there was a plausible relationship with survival, completeness in excess of 50% and a frequency of at least 1% in the cohort. For this study a total of 29 variables were identified as potential predictors (Table S.2), including patient characteristics, preoperative tumour staging, complications of surgery, postoperative pathology and neoadjuvant/ adjuvant treatment. Type of operation (e.g., Distal Gastrectomy, Total Gastrectomy, Extended total gastrectomy) was considered but omitted as it was almost exclusively correlated with site of tumour. Anastomotic leak was defined as severe disruption to the anastomosis (detected clinically or radiologically) including those patients managed actively and conservatively. An R0 resection was defined as complete macro/microscopic resection of tumour with negative longitudinal and circumferential resection margins. We considered unit volume as combination of major upper gastrointestinal resections per year (major gastrectomy and oesophagectomy) as per published research(10) and in line with NHS commissioning guidelines(11) and also separately for gastrectomy alone.

TNM staging was conducted using the 8th edition staging manual. There was at least one data point missing in 671 cases (22.9%). The most frequently missing characteristics were return to theatre (15.1%), cT stage (12.4%) and differentiation grade (6.5%). All other variables had <5% missing data. Missing data was assumed to be missing at random and handled using multiple imputation by chained equations(12) with 10 imputed datasets.

In order to produce a more concise model with increased generalisability, a variable selection step was conducted using the Boruta method.(13) Boruta identifies core variables by comparing the importance of candidate variables in a Random Forest to a corresponding set of 'shadow' variables, which are versions of each variable with their data

randomised. Variables with importance significantly greater than all of the shadow variables are selected as important and retained, and variables with importance significantly less than the highest shadow variable are selected as unimportant and removed. This process is repeated with the decreasing number of uncertain variables until all are sorted into important or unimportant. It has been found to be more accurate than other approaches in variable selection in high dimensional data,(14) particularly in large datasets,(15) and has been used in a variety of settings.(16–18) In this study variables were selected from complete cases only (n=2304).

Identified important variables were then used to train a Random Survival Forest (RSF) using the Ranger(19) package in R. A random forest here is comprised of several hundred survival trees, each derived from different subpopulations of the cohort. Within each tree the binary split (e.g., zero positive lymph nodes vs one or more positive lymph nodes) that gives the biggest difference in survival (as measured by the Log-Rank test) is identified. The tree undergoes progressively more splits until a predetermined end point is reached. The random forest is then the mean output of all the decision trees. Parameters of the RSF that influence how it generates predictions i.e., number of trees, number of variables per tree and minimum node size were selected to minimise out-of-sample error within the random forest. As multiple imputation was employed to address missing data, a means of pooling the outcomes from the imputed datasets is required. Here, as previously(7), models were generated on each imputed dataset and predictions from each were combined after a log-log transformation(20,21).

As the model incorporates both variable interaction and non-linear time effects, expressing the effect of individual variables is difficult. The hazard ratio is less appropriate as it assumes an exponential survival distribution and proportional hazards (i.e. consistency of effect of variables over time)(22). Use of the restricted mean survival time (RMST) has been proposed to address scenarios where the proportional hazards assumption does not hold true(23), allowing for comparisons by absolute difference or ratio(24) and is increasingly thought to be a more appropriate means of comparing survival outcomes and treatment effects(25–28). Survival curves are first generated for each variable as the average predictions yielded for that variable. The RMST is then the area under each survival curve, the absolute difference in RMST between two factors (e.g., R0 vs R1) is termed the life expectancy difference (LED) and the ratio between them the life expectancy ratio (LER). The LED and LER readily provide the absolute or relative gain/loss of life for each variable for the period of follow up.

The internal validity of the model was quantified using 1000 replications of the bootstrap with replacement and the 0.632 estimator.(29) Discrimination was assessed using the time dependent area under the receiver operator curve (tAUC),(30) which corresponds to the proportion of random pairs of cases where one patient is alive and one dead at a specified time point where the model has correctly ordered their probability of survival having weighted for censoring. Calibration was assessed quantitatively using the integrated Brier score,(31,32) as a measure of overall error of predictions with a value closer to zero being better. Visual assessment of calibration was conducted by comparing predicted survival to observed (Kaplan-Meier) survival at specified time points. All analyses were conducted in R,(33) and the study was conducted to comply with the TRIPOD criteria(34). Complete code to reproduce the analysis is available on request, and instructions for external validation provided in the supplementary materials.

Results

The study population included 2931 patients who underwent a gastrectomy with a histologically proven diagnosis of adenocarcinoma. Patients were followed up for a median of 44 months, there were 1071 recorded deaths and the median survival was 69 months. At 3- and 5-years, survival was 63.5% and 53.2% respectively (Figure 1).

A median of 27 lymph nodes were examined (range 15-109) and at least one node contained tumour in the majority of cases (1635/2931, 55.8%). Extent of nodal dissection was recorded as D2 in 2425 cases (82.7%). Neoadjuvant chemotherapy was used in 48.0%. Demographics of the population were as expected with 65.3% males and a median age at diagnosis of 71. The vast majority of cases were undertaken in high volume centres, with 91.6% occurring in centres performing >30 major upper gastrointestinal resections per year. These characteristics are summarised in Table 1.

A total of ten variables were identified as important and included in the final model. These were Age, cT stage, cN stage, WHO Performance Status, ASA grade, pT/ypT, Total number of positive lymph nodes, grade of differentiation (good, moderate, poor/anaplastic), completeness of resection (R0/R1) and neoadjuvant treatment received (Figure S.3).

The model demonstrated excellent discrimination on internal validation, with a tAUC of 0.80 (95% CI 0.78-0.82) at 5 years and a C-index of 0.76 (95% CI 0.75-0.77). The tAUC using pTNM stage alone was 0.75. Agreement between predicted and observed survival was also excellent, with a wide spread of predictions observed for both three-year (14.8-98.3%, IQR 43.2-84.4%) and five-year (9.4-96.1%, IQR 31.7-73.8%) survival (Figure 2, Figure S.4). The integrated brier score was 0.137 (95% CI 0.133-0.140). Importantly, the discrimination of the model exceeds that achieved using TNM stage (tAUC 0.81 vs 0.76 $p < 0.001$). A wide range of survival estimates are also seen for each TNM stage group (Figure S.5).

Table 1 Clinical and Pathological Characteristics of study cohort

Characteristic	Count (%)	Survival at 5 years (%)	Characteristic	Count (%)	Survival at 5 years (%)		
Age	18-50	263 (9.0)	Annual volume of major upper gastrointestinal resections*	1 to 30	245 (8.4)		
	51-60	417 (14.2)		31 to 60	1563 (53.3)		
	61-70	741 (25.3)		60+	1123 (38.3)		
	71-80	1169 (39.9)		1 to 15	834 (28.1)		
	80+	341 (11.6)		16 to 30	1654 (55.7)		
Site of Tumour	Female	1017 (34.7)	Surgical Approach	Laparoscopic	439 (15.0)		
	Male	1914 (65.3)		Open	2492 (85.0)		
	Siewert III	416 (14.2)		No	2240 (76.4)		
	Fundus	195 (6.7)		Yes	679 (23.2)		
	Body	1250 (42.6)		Missing	12 (0.4)		
cT	Antrum	689 (23.5)	Surgical Complication	No	2826 (96.4)		
	Pylorus	381 (13.0)		Yes	93 (3.2)		
	T0/is/1	295 (10.1)		Missing	12 (0.4)		
	T2	579 (19.8)		Anastomotic Leak	T0	116 (4.0)	
	T3	1218 (41.6)			T1	591 (20.2)	
T4	476 (16.2)	T2	454 (15.5)				
Missing	363 (12.4)	T3	1004 (34.3)				
cN	N0	1460 (49.8)	pT/ypT Stage		T4	766 (26.1)	
	N1	882 (30.1)		N0	1296 (44.2)		
	N2	369 (12.6)		N1	495 (16.9)		
	N3	108 (3.7)		N2	513 (17.5)		
	Missing	112 (3.8)		N3	627 (21.4)		
WHO Performance Status	0	1409 (48.1)	pN/ypN Stage	Yes	2663 (90.1)		
	1	1201 (41.0)		No	268 (9.1)		
	2	288 (9.8)		Well (G1)	69 (2.4)		
	3	31 (1.1)		Moderate (G2)	730 (24.9)		
	ASA Grade	1		359 (12.2)	R0 resection	Poor/Anaplastic (G3/G4)	1674 (57.1)
2		1604 (54.7)	Unable to determine (GX)	268 (9.1)			
3		935 (31.9)	Missing	190 (6.5)			
4		33 (1.1)	Grade of differentiation (worst)	Adjuvant Treatment		No	2280 (77.8)
Neoadjuvant Treatment		None				1525 (52.0)	Yes
	Chemotherapy	1406 (48.0)			50.5		

Data given as absolute number (%), Anastomotic Leak defined as severe disruption to anastomosis, regardless of method of detection or intervention. Major gastrointestinal resections including oesophagectomy and gastrectomy

The most important variables were number of positive lymph nodes, pT stage and completeness of resection, as visualised in survival curves shown in Figure 3. The mean predicted survival (across combinations of other variables) to five years (the restricted means survival time/RMST) varied significantly for different characteristics, for example for pN0 the RMST was 46.4 months compared to 29.3 months for N3b patients. This corresponded to a life expectancy reduction (LED) of 17.1 months and a life expectancy ratio (LER) of 0.63. Table S.3 illustrates the RMST, LED and LER for all variables. Although the magnitude of effect overall is small for several of the variables, in individual cases this may not be the case due to the nature of variable interactions. Advanced pN/pT (pN2/3a/3b &

pT3/4) and a R1 resection exhibited an LER that clearly increases throughout the period of follow up, indicating diverging survival trajectories and a persistent effect on prognosis for at least four years for pN/pT and the entirety of follow up for R1 resection (Figure S.6). Notably, a Cox model trained using the same variables clearly violates the proportional hazards assumption ($p=0.013$). One limitation of traditional estimates of importance is that variable interactions that are modelled in the RSF are ignored. To address this, Figure 4 gives an overview of the average predicted five-year survival for combinations of the most important variables in addition to patient age. The importance of age can be seen to diminish with increasing tumour burden.

Example Cases

To illustrate the utility of the model, four example cases are described below.

Case 1

A 50-year-old female patient, ASA 1, with a cT3N1 tumour, undergoes and completes neoadjuvant chemotherapy followed by a gastrectomy. Post-operative pathology reveals a pT4 well differentiated tumour with one positive lymph node and a complete resection margin (R0).

Case 2

A 60-year-old male patient, ASA 3, with a cT3N1 tumour, undergoes but does not complete neoadjuvant chemotherapy followed by a gastrectomy. Post-operative pathology reveals a pT4 poorly differentiated tumour with three positive lymph nodes and an involved resection margin (R1).

Case 3

A 50-year-old female patient, ASA 1, with a cT3N1 tumour, undergoes and completes neoadjuvant chemotherapy followed by a gastrectomy. Post-operative pathology reveals a pT1 well differentiated tumour with no positive lymph nodes and a complete resection margin (R0).

Case 4

A 60-year-old male patient, ASA 3, with a cT3N1 tumour, undergoes but does not complete neoadjuvant chemotherapy followed by a gastrectomy. Post-operative pathology reveals a pT1 poorly differentiated tumour with no positive lymph nodes and a complete resection margin (R0).

Both cases 1 and 2 fall into the same pTNM stage group (3a), however their predicted survival trajectories show considerable differences (Figure 5) with 45.7% five-year survival for case 1 and 17.0% for case 2 (compared to a stage average survival of 34.4% at five-years). Similarly, Cases 3 and 4 are both stage 1a, but exhibit substantial variation in five-year survival at 88.7% and 66.5% respectively.

Discussion

This study describes the derivation and validation of a robust machine learning model for prediction of overall survival for surgically treated non-metastatic gastric adenocarcinoma. The model utilises routine clinicopathological data which should be available for every case without additional investigations, to deliver predictions of survival to five years. The model provides accuracy in excess of traditional TNM staging to enable the delivery of personalised survival predictions, with a large spread of predictions within each TNM staging group that allows discrimination in excess of TNM staging.

Strengths of this study include the large population-based dataset used to derive the model, which is larger than those used in many previously published prognostication tools. The data are reflective of modern practice, including only patients diagnosed since 2012, with a high rate of neoadjuvant treatment (48%) and D2 nodal dissection (83%), with surgery performed in high-volume specialist centres. Observed overall survival exceeded recent trials, with more than 1 in 2 patients surviving to 5 years(35,36). A machine learning non-linear approach (RSF) allowed more accuracy than otherwise could be achieved, is technically novel and has generated insight into how the importance of variables varies over time. The tripod criteria for predictive modelling were also adhered to. Limitations include the retrospective nature of the study and lack of external validation cohort. An internal validation process was conducted using a bootstrap technique(37) to assess the degree of optimism in the model's discrimination and calibration, and its performance was maintained. External validation is still required to demonstrate its generalisability, but the importance of the T stage and Nodal positive variables suggest the model is likely to be transportable to another population. There was a moderate amount of missing data within the dataset which may introduce bias into the analysis, however this effect was minimised using multiple imputation. Lauren histological classification(38), a well-recognised and prognostic variable in gastric cancer(39), was not available for this study and may provide additional information above differentiation grade if added in the future, although as the diffuse type are poorly differentiated by definition, there will be extensive overlap with the classification employed here.

This model provides a broad range of survival estimates, with substantially more variability than TNM stage both overall and within each staging group, as is clearly illustrated in the example cases. The precision facilitates use of the model in several clinical settings. Firstly, more reliable information on long-term prognosis can be given to patients. Research to understand how to best to relay data to patients is ongoing and this is undoubtedly an ethically complex area, particularly when the prognosis is poor. However, withholding accurate information from patients is unlikely to be prudent. Secondly, targeting follow up and/or additional treatment to those who most require it is vital to improving outcomes and accurate prognostication with low burden of data collection (as is the case with clinicopathological models) vital to achieve that. This is particularly important when introducing novel agents or when effect sizes appear small, as they are with current agents.

The most important variables identified (lymph node status, pT stage, resection margin) are well recognised as highly prognostic(40–42). The demonstration of effects for these variables that persist throughout follow-up is however novel and informative in the context of a modelling strategy that allows for time-varying effects. In this study, only a

small overall magnitude of effect of neoadjuvant treatment was identified, with no benefit seen for cases where chemotherapy was not completed. This is in contrast to the Medical Research Council Adjuvant Gastric Infusional Chemotherapy (MAGIC)(43) and Actions Concertées dans les Cancer Colorectaux et Digestifs (ACCORD) trials(44), which demonstrated a substantial survival benefit of neoadjuvant treatment, establishing the rationale for the widespread use of neoadjuvant chemotherapy for gastric adenocarcinoma in Western countries. In reality, the effect of chemotherapy varies at an individual level, with some patients gaining a substantial benefit from the treatment (i.e. those who respond) and the majority gaining no benefit at all. The non-linear nature of the RSF which includes interactions with other variables allows response (as reflected in e.g. pT stage or the resection margin) to be accurately incorporated into prognostication, which would be challenging in a linear model and not assessed by TNM stage alone. This may also be an explanation for the counterintuitive finding of pT0 tumour having both a slightly worse observed prognosis than pT1 tumours (five-year survival 79.5% pT0 vs 81.9% pT1, Table 1) and life expectancy difference of -1.65 months (Table S.2) although the number of patients with pT0 tumours was small (n=116).

We excluded patients in whom less than 15 lymph nodes were examined at resection. The purpose of this study was to derive an accurate predictive model for the most common stages of disease witnessed in clinical practice, rather than assess the prognostic importance of lymph node harvest which is of sufficient interest and complexity to warrant a separate study. After discussion we elected to exclude patients with an 'inadequate' lymph node resection because including these patients would introduce unreliable data into the model derivation process (due to possible under-staging) and make predictions on patients who were treated as per national/international standards less reliable. We were also mindful that some patients with an apparent 'inadequate' lymph node harvest would have received a planned D1 dissection for early-stage disease. Our model is not designed to be used in this patient group, rather for the classical presentation of locally advanced gastric cancer and limited to those patients with an adequate lymphadenectomy.

The variable selection method excluded variables that may have been expected to significantly influence survival, notably site of tumour and receipt of adjuvant therapy. It is reasonable to extrapolate that the difference in survival observed for different tumour sites (e.g. five-year survival 40.7% for pyloric tumours compared to 56.7% for Siewert III GOJ tumours, Table 1) is largely due to differences in tumour stage at these sites, however it is surprising to see no improvement of survival with the administration of adjuvant treatment. There are several possible explanations for this, including insensitivity of the modelling approach (particularly as survival is worse after adjuvant treatment on univariate analysis; as it is more often given only to cases with more advanced disease), however the margin of effect of adjuvant treatment seen in randomised trials does appear modest(45), particularly in the context of a cohort in which the majority of patients underwent neoadjuvant treatment(46).

Conclusions

A robust tool for prediction of overall survival after gastrectomy for adenocarcinoma has been derived using a random survival forest methodology. It provides accurate predictions of outcome in excess of TNM staging using routinely collected clinicopathological data. It is available at: uoscancer.shinyapps.io/AugisSurvG. Future work to validate our findings in external cohorts would be beneficial, and prospective validation before use to stratify treatment and follow-up important.

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;68:394–424.
2. Chadwick G, Varaganam M, Brand C, Cromwell D, Riley S, Crosby T, et al. National Oesophago-Gastric Cancer Audit: 2017 Annual Report. The Royal College of Surgeons England. London; 2017.
3. van den Boorn HG, Engelhardt EG, van Kleef J, Sprangers MAG, van Oijen MGH, Abu-Hanna A, et al. Prediction models for patients with esophageal or gastric cancer: A systematic review and meta-analysis. *PLoS One.* 2018;13:e0192310.
4. Kattan MW, Karpeh MS, Mazumdar M, Brennan MF. Postoperative nomogram for disease-specific survival after an R0 resection for gastric carcinoma. *J Clin Oncol.* 2003;21:3647–3650.
5. Feng Q, May MT, Ingle S, Lu M, Yang Z, Tang J. Prognostic Models for Predicting Overall Survival in Patients with Primary Gastric Cancer: A Systematic Review. *Biomed Res Int.* 2019;5634598.
6. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat.* 2008;2:841–860.
7. Rahman SA, Walker RC, Maynard N, Trudgill N, Crosby T, Cromwell DA, et al. The AUGIS Survival Predictor. *Ann Surg.* 2021;Publish Ah.
8. Cromwell D, Wahedally H, Park MH, Maynard N, Crosby T, Trudgill N, et al. National Oesophago-Gastric Cancer Audit: 2019 Annual Rep. London; 2019.
9. Washington K. 7th edition of the AJCC cancer staging manual: Stomach. *Ann Surg Oncol.* 2010;17:3077–3079.
10. Varaganam M, Hardwick R, Riley S, Chadwick G, Cromwell DA, Groene O. Changes in volume, clinical practice and outcome after reorganisation of oesophago-gastric cancer care in England: A longitudinal observational study. *Eur J Surg Oncol.* 2018;44:524–531.
11. NHS England. 2013/14 NHS Oesophageal and Gastric Cancer Commissioning Guidelines. 2013. Available from <https://www.england.nhs.uk/commissioning/wp-content/uploads/sites/12/2014/03/b11-cancer-oesop-gast.pdf>
12. van Buuren S, Groothuis-Oudshoorn K. MICE: Multivariate Imputation by Chained Equations in R. *J Stat Softw.* 2011;45:1–67.
13. Kursa MB, Rudnicki WR. Feature Selection with the Boruta Package . *J Stat Softw.* 2010;36.
14. Degenhardt F, Seifert S, Szymczak S. Evaluation of variable selection methods for random forests and omics data sets. *Brief Bioinform.* 2017;20:1–12.
15. Sanchez-Pinto LN, Venable LR, Fahrenbach J, Churpek MM. Comparison of variable selection methods for

- clinical predictive modeling. *Int J Med Inform.* 2018;116:10–17.
16. Huang C, Murugiah K, Mahajan S, Li SX, Dhruva SS, Haimovich JS, et al. Enhancing the prediction of acute kidney injury risk after percutaneous coronary intervention using machine learning techniques: A retrospective cohort study. *PLoS Med.* 2018;15:1–20.
 17. Chen D, Afzal N, Sohn S, Habermann EB, Naessens JM, Larson DW, et al. Postoperative bleeding risk prediction for patients undergoing colorectal surgery. *Surg (United States).* 2018;164:1209–1216.
 18. Long N, Park S, Anh N, Nghi T, Yoon S, Park J, et al. High-Throughput Omics and Statistical Learning Integration for the Discovery and Validation of Novel Diagnostic Signatures in Colorectal Cancer. *Int J Mol Sci.* 2019;20:296.
 19. Wright MN, Ziegler A. ranger : A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *J Stat Softw.* 2017;77:1–17.
 20. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: Current practice and guidelines. *BMC Med Res Methodol.* 2009;9:1–8.
 21. Wood AM, Royston P, White IR. The estimation and use of predictions for the assessment of model performance using large samples with multiply imputed data. *Biometrical J.* 2015;57:614–632.
 22. Hernán MA. The hazards of hazard ratios. *Epidemiology.* 2010;21:13–15.
 23. Royston P, Parmar MKB. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Stat Med.* 2011;30:2409–2421.
 24. Dehbi HM, Royston P, Hackshaw A. Life expectancy difference and life expectancy ratio: Two measures of treatment effects in randomised trials with non-proportional hazards. *BMJ.* 2017;357:1–7.
 25. Trinquart L, Jacot J, Conner SC, Porcher R. Comparison of treatment effects measured by the hazard ratio and by the ratio of restricted mean survival times in oncology randomized controlled trials. *J Clin Oncol.* 2016;34:1813–1819.
 26. Zhao L, Claggett B, Tian L, Uno H, Pfeffer MA, Solomon SD, et al. On the restricted mean survival time curve in survival analysis. *Biometrics.* 2016;72:215–221.
 27. A'Hern RP. Restricted mean survival time: An obligatory end point for time-to-event analysis in cancer trials? *J Clin Oncol.* 2016;34:3474–3476.
 28. A'Hern RP. Cancer Biology and Survival Analysis in Cancer Trials: Restricted Mean Survival Time Analysis versus Hazard Ratios. *Clin Oncol.* 2018;30:e75–e80.
 29. Efron B. Estimating the error rate of a prediction rule: Improvement on cross-validation. *J Am Stat Assoc.*

1983;78:316–331.

30. Kamarudin AN, Cox T, Kolamunnage-Dona R. Time-dependent ROC curve analysis in medical research: Current methods and applications. *BMC Med Res Methodol*. 2017;17:1–19.
31. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Stat Med*. 1999;18:2529–2545.
32. Kronek LP, Reddy A. Logical analysis of survival data: Prognostic survival models by detecting high-degree interactions in right-censored data. *Bioinformatics*. 2008;24:248–253.
33. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2019.
34. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med*. 2015;162:55–63.
35. Al-Batran S-E, Homann N, Pauligk C, Goetze TO, Meiler J, Kasper S, et al. Perioperative chemotherapy with fluorouracil plus leucovorin, oxaliplatin, and docetaxel versus fluorouracil or capecitabine plus cisplatin and epirubicin for locally advanced, resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4): a randomised phase 2/3 trial. *Lancet*. 2019;393:1948-1957.
36. Kelly RJ, Ajani JA, Kuzdzal J, Zander T, Van Cutsem E, Piessen G, et al. Adjuvant Nivolumab in Resected Esophageal or Gastroesophageal Junction Cancer. *N Engl J Med*. 2021;384:1191–1203.
37. Steyerberg EW, Harrell FE, Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*. 2016;69:245-247.
38. Lauren P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. an attempt at a histo-clinical classification. *Acta Pathol Microbiol Scand*. 1965;64:31–49.
39. Jiménez Fonseca P, Carmona-Bayonas A, Hernández R, Custodio A, Cano JM, Lacalle A, et al. Lauren subtypes of advanced gastric cancer influence survival and response to chemotherapy: Real-world data from the AGAMENON National Cancer Registry. *Br J Cancer*. 2017;117:775–782.
40. Siewert JR, Böttcher K, Stein HJ, Roder JD. Relevant prognostic factors in gastric cancer: Ten-year results of the German Gastric Cancer Study. *Ann Surg*. 1998;228:449–461.
41. Wang HM, Huang CM, Zheng CH, Li P, Xie JW, Wang J Bin, et al. Tumor size as a prognostic factor in patients with advanced gastric cancer in the lower third of the stomach. *World J Gastroenterol*. 2012;18:5470–5475.
42. Liang Y, Ding X, Wang X, Wang B, Deng J, Zhang L, et al. Prognostic value of surgical margin status in gastric cancer patients. *ANZ J Surg*. 2015;85:678–684.
43. Cunningham D, Allum WH, Stenning SP, Thompson JN, Van de Velde CJH, Nicolson M, et al. Perioperative

- Chemotherapy versus Surgery Alone for Resectable Gastroesophageal Cancer. *N Engl J Med.* 2006;355:11–20.
44. Ychou M, Boige V, Pignon JP, Conroy T, Bouché O, Lebreton G, et al. Perioperative chemotherapy compared with surgery alone for resectable gastroesophageal adenocarcinoma: An FNCLCC and FFCD multicenter phase III trial. *J Clin Oncol.* 2011;29:1715–1721.
 45. GASTRIC Group. Benefit of adjuvant chemotherapy for resectable gastric cancer. *JAMA - J Am Med Assoc.* 2010;303:1729–1737.
 46. Cats A, Jansen EPM, van Grieken NCT, Sikorska K, Lind P, Nordsmark M, et al. Chemotherapy versus chemoradiotherapy after surgery and preoperative chemotherapy for resectable gastric cancer (CRITICS): an international, open-label, randomised phase 3 trial. *Lancet Oncol.* 2018;19:616–628.

Figure Captions

Figure 1 Kaplan-Meier survival of study cohort. Shaded area represents 95% confidence interval. The vertical dotted line represents the median survival, 69 months.

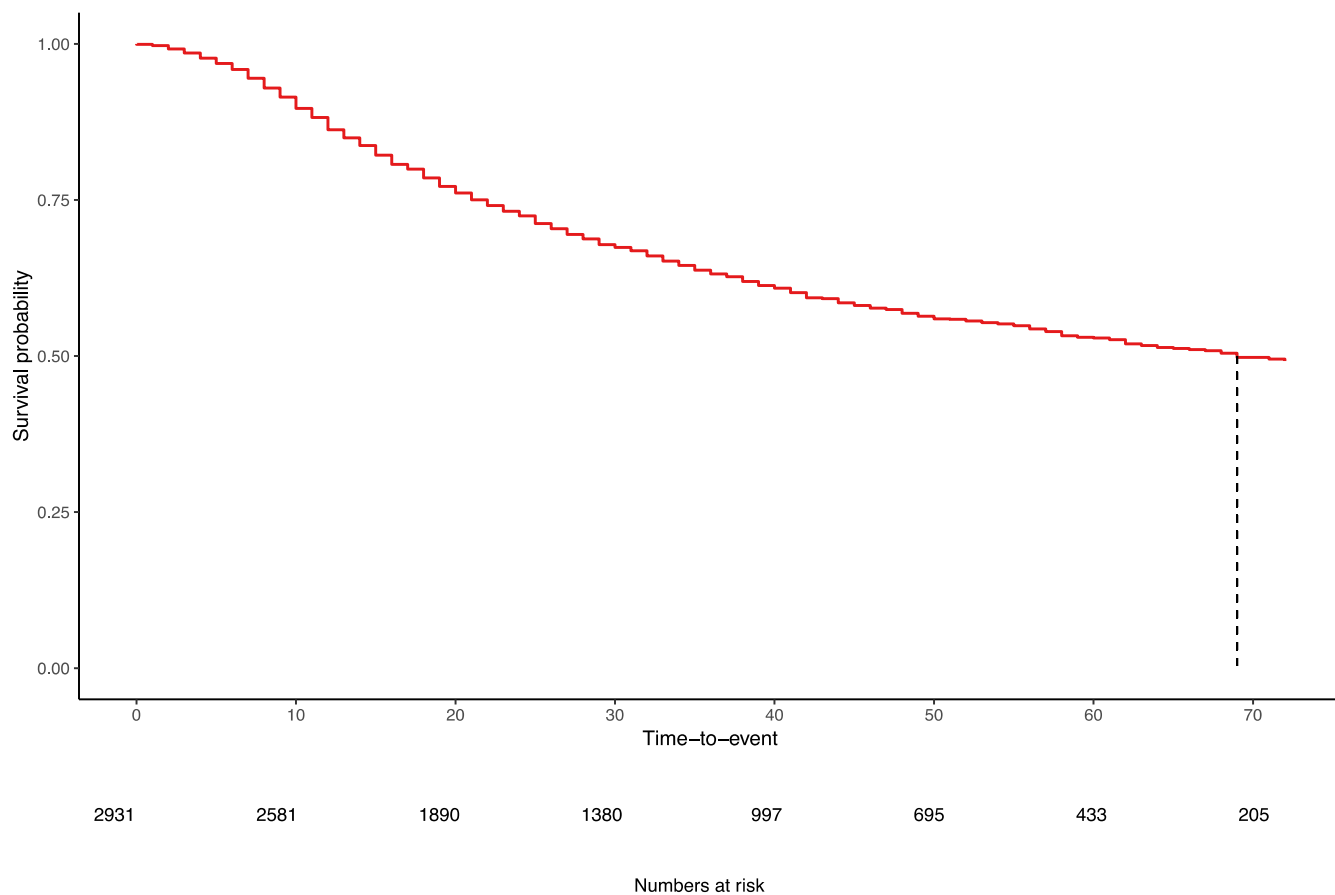


Figure 2 Calibration of predictions at (A) Three years and (B) Five years post-surgery. The dotted line represents the ideal and the solid line the model's performance on internal validation

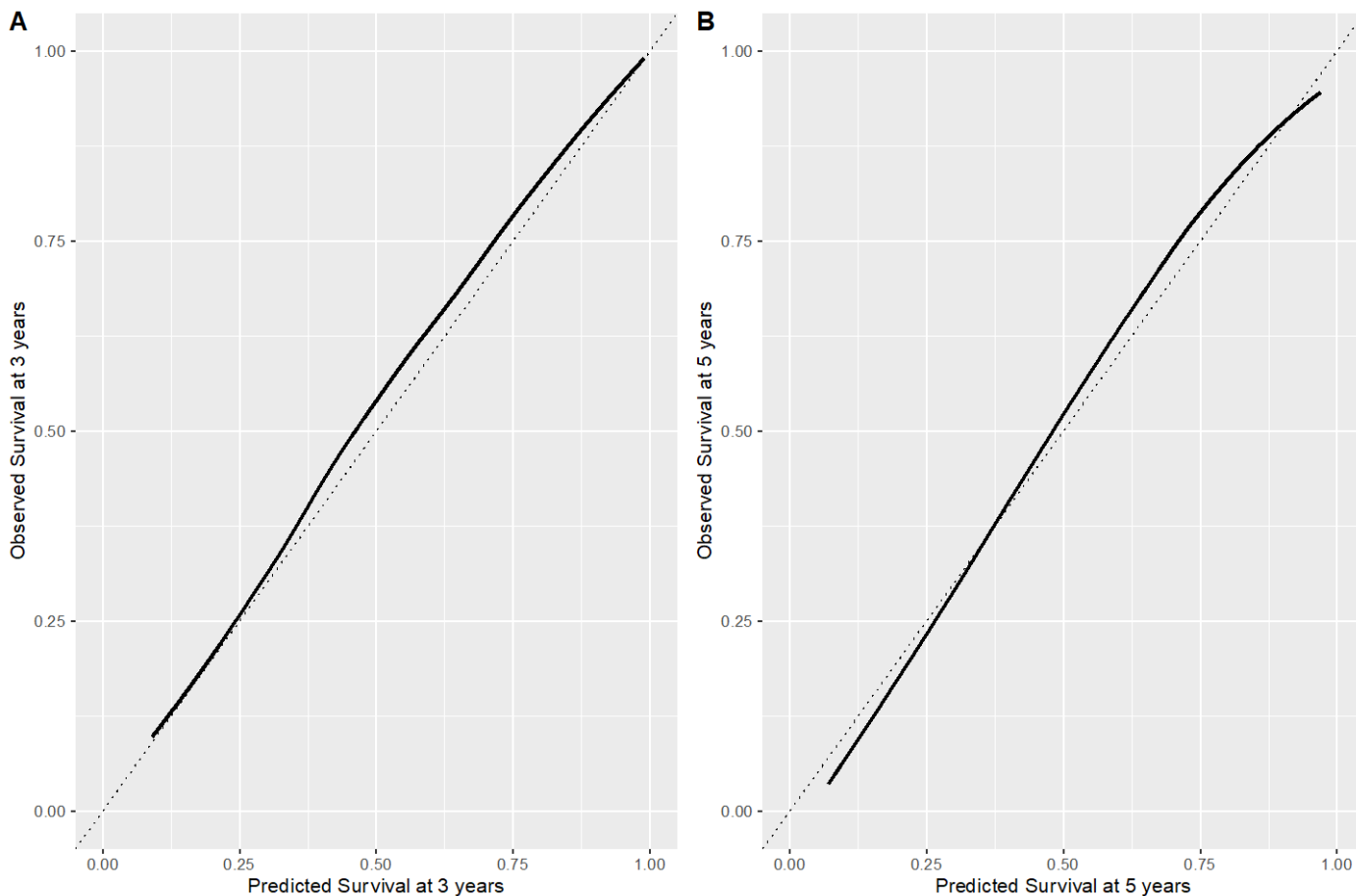


Figure 3 Variable effects on survival. Average predicted survival is shown from 0-60 months for (A) pN Stage, (B) pT Stage and (C) Resection Margin

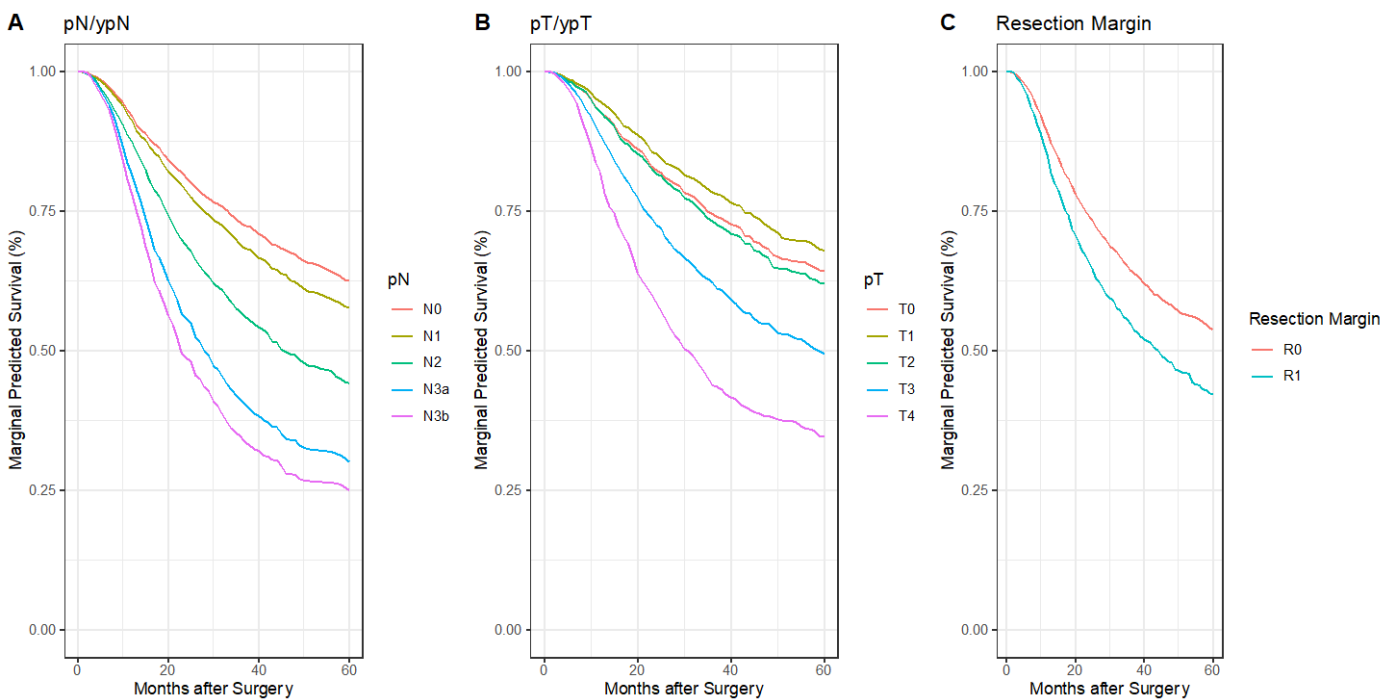
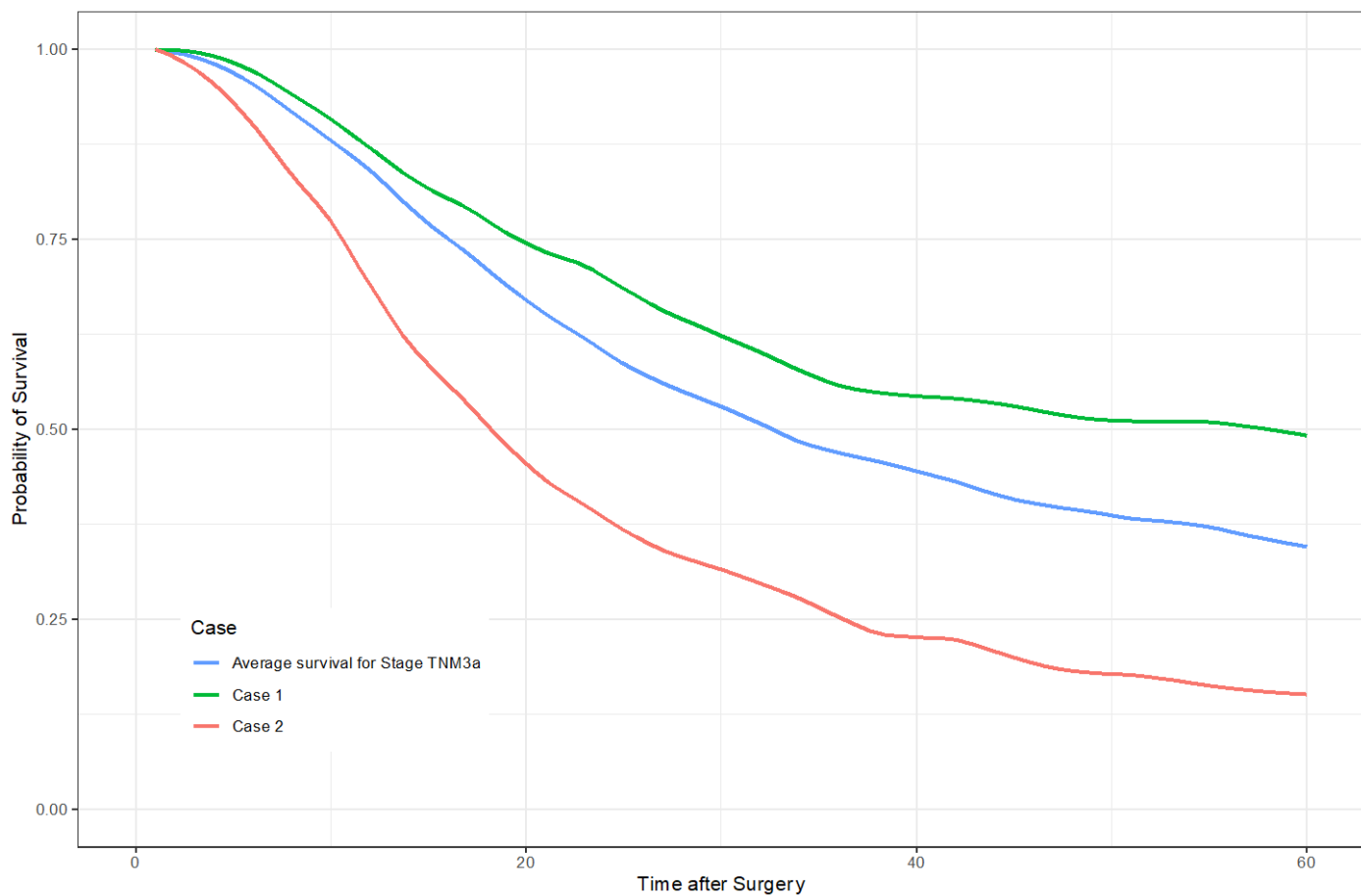


Figure 4 Predicted 5-year survival for combinations of selected variables. Colours represent differing prognosis, with green more favourable, and red less favourable.

		Lymph Node Status										
		N0		N1		N2		N3a		N3b		
		R0	R1	R0	R1	R0	R1	R0	R1	R0	R1	
		N0/-ve	N0/+ve	N1/-ve	N1/+ve	N2/-ve	N2/+ve	N3a/-ve	N3a/+ve	N3b/-ve	N3b/+ve	
T0	Age	0-50	82.1%		74.9%		60.2%		44.4%		38.2%	
		51-60	81.3%		74.3%		59.1%		43.3%		37.2%	
		61-70	81.3%		74.9%		56.3%		40.6%		34.9%	
		71-80	76.6%		70.9%		51.0%		35.5%		30.4%	
		80+	74.3%		69.0%		50.6%		35.0%		29.8%	
T1	Age	0-50	87.7%	73.4%	78.7%	66.6%	67.1%	57.3%	51.1%	42.9%	44.8%	39.1%
		51-60	87.1%	72.9%	78.2%	66.2%	66.3%	56.7%	50.2%	42.3%	43.9%	38.4%
		61-70	86.1%	71.9%	78.1%	65.9%	63.6%	54.8%	47.4%	40.2%	41.4%	36.4%
		71-80	79.8%	67.2%	72.9%	62.1%	57.9%	50.8%	42.0%	36.4%	36.6%	32.7%
		80+	77.5%	65.5%	71.3%	60.9%	57.4%	50.3%	41.3%	35.9%	35.7%	32.1%
T2	Age	0-50	78.7%	64.2%	73.6%	60.3%	58.1%	49.6%	42.5%	36.1%	36.6%	32.4%
		51-60	78.1%	63.6%	73.1%	59.8%	57.1%	48.8%	41.4%	35.4%	35.6%	31.7%
		61-70	78.3%	63.6%	73.9%	60.3%	54.6%	47.3%	39.0%	33.6%	33.5%	30.0%
		71-80	73.2%	60.2%	69.7%	57.6%	49.2%	43.7%	34.1%	30.2%	29.2%	26.8%
		80+	70.9%	58.6%	67.8%	56.4%	48.9%	43.5%	33.8%	30.0%	28.7%	26.5%
T3	Age	0-50	70.3%	39.4%	62.6%	37.4%	47.3%	32.4%	27.7%	20.7%	22.8%	19.0%
		51-60	69.8%	38.8%	61.8%	36.5%	45.6%	31.4%	26.5%	20.0%	21.8%	18.3%
		61-70	70.9%	38.3%	61.9%	35.9%	41.9%	29.5%	23.9%	18.2%	19.5%	16.5%
		71-80	71.2%	38.1%	60.8%	35.4%	38.1%	28.0%	20.2%	15.9%	16.3%	14.1%
		80+	69.7%	38.0%	59.9%	35.3%	39.6%	28.7%	20.3%	15.9%	16.2%	13.9%
T4	Age	0-50	48.5%	24.4%	44.7%	23.2%	36.2%	19.0%	26.2%	16.2%	18.1%	14.5%
		51-60	48.8%	24.4%	44.8%	23.1%	35.6%	18.6%	25.6%	15.8%	17.6%	14.1%
		61-70	49.4%	24.1%	44.8%	22.6%	34.4%	18.0%	24.2%	15.0%	16.9%	13.3%
		71-80	49.3%	24.1%	43.5%	22.4%	30.5%	17.2%	20.8%	13.5%	15.1%	12.1%
		80+	47.8%	24.1%	41.9%	22.3%	28.9%	17.0%	19.4%	13.1%	14.3%	11.7%

Figure 5 Predicted Survival for (A) stage 3A and (B) stage 1A example cases (green/red lines) and mean survival (blue line)



Supplementary

Table S.1 Characteristics of study cohort in comparison to patients with inadequate lymph node harvest

Characteristics	Lymph node Harvest			p§	Characteristics	Lymph node Harvest		
	≥15	<15				≥15	<15	
	2931	754				2931	754	
Age	71 [61, 77]	73 [64, 79]	<0.001*¶	Annual volume of major upper gastrointestinal resections*	1 to 30	245 (8.4)	91 (12.1)	0.002*
	Siewert III	416 (14.2)	68 (9.0)	<0.001*	31 to 60	1563 (53.3)	365 (48.4)	
	Fundus	195 (6.7)	41 (5.4)		>60	1123 (38.3)	298 (39.5)	
Site of Tumour	Body	1250 (42.6)	308 (40.8)		1 to 15	834 (28.1)	211 (27.6)	<0.001*
	Antrum	689 (23.5)	210 (27.9)		Annual volume of major gastrectomy	16 to 30	1654 (55.7)	479 (62.7)
	Pylorus	381 (13.0)	127 (16.8)		30+	483 (16.3)	74 (9.7)	
cT	T0/is/1	295 (11.5)	141 (22.3)	<0.001*	Laparoscopic Approach	439 (15.0)	151 (20.0)	0.001*
	T2	579 (22.5)	195 (30.9)		Any Complication	679 (23.3)	143 (19.0)	0.015*
	T3	1218 (47.4)	231 (36.6)		Anastomotic Leak	93 (3.2)	18 (2.4)	0.311
	T4	476 (18.5)	64 (10.1)		T0	116 (4.0)	41 (5.4)	<0.001*
cN	N0	1460 (51.8)	486 (68.3)	<0.001*	T1	591 (20.2)	252 (33.4)	
	N1	882 (31.3)	155 (21.8)		pT/ypT	T2	454 (15.5)	128 (17.0)
	N2	369 (13.1)	57 (8.0)		T3	1004 (34.3)	183 (24.3)	
	N3	108 (3.8)	14 (2.0)		T4	766 (26.1)	150 (19.9)	
	0	1409 (48.1)	361 (47.9)	0.126	R1	268 (9.1)	72 (9.5)	0.785
	1	1201 (41.0)	288 (38.2)		G1	69 (2.5)	30 (4.4)	0.002*
Performance Status	2	288 (9.8)	90 (11.9)		Grade of differentiation	G2	730 (26.6)	199 (29.4)
	3	31 (1.1)	14 (1.9)		G3/4	1674 (61.1)	368 (54.4)	
	4	2 (0.1)	1 (0.1)		GX	268 (9.8)	80 (11.8)	
ASA	1	359 (12.2)	118 (15.6)	0.002*	Adjuvant Treatment	872 (29.8)	165 (21.9)	<0.001*
	2	1604 (54.7)	369 (48.9)		None	234 (8.0)	98 (13.0)	<0.001*
	3	935 (31.9)	250 (33.2)		D0	21 (0.7)	12 (1.6)	
	4	33 (1.1)	17 (2.3)		Extent of Nodal Dissection	D1	240 (8.2)	169 (22.4)
Female Gender		1017 (34.7)	254 (33.7)	0.633	D2	2425 (82.7)	474 (62.9)	
Neoadjuvant Treatment		1406 (48.0)	256 (34.0)	<0.001*	D3	11 (0.4)	1 (0.1)	

Data presented as absolute number (%) and median (IQR), *<0.05, § χ^2 test, except ¶ Mann–Whitney U test **Major gastrointestinal resections including oesophagectomy or gastrectomy

Figure S.1 Kaplan Meier estimation of survival, stratified by lymph node yield

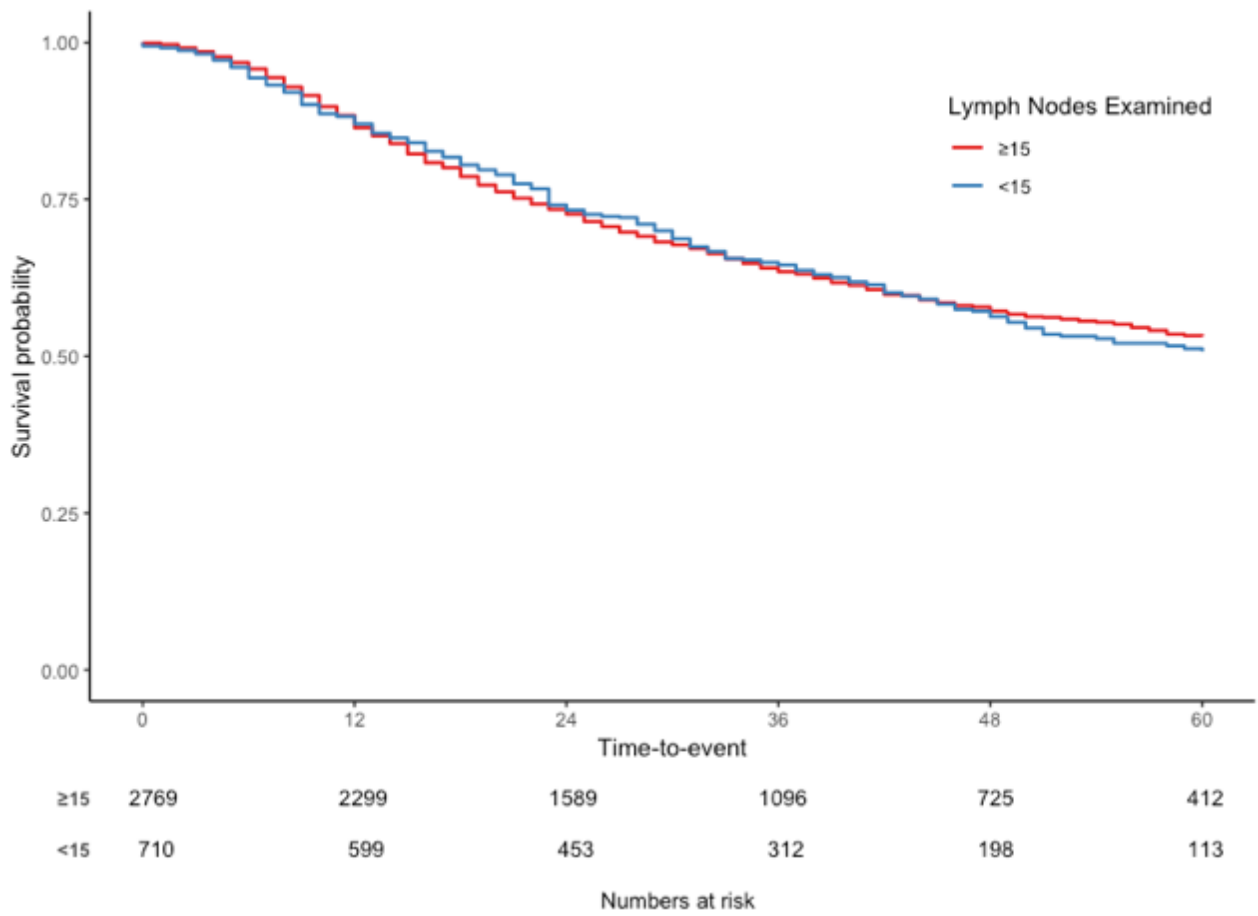


Figure S.2 Study Flow Diagram

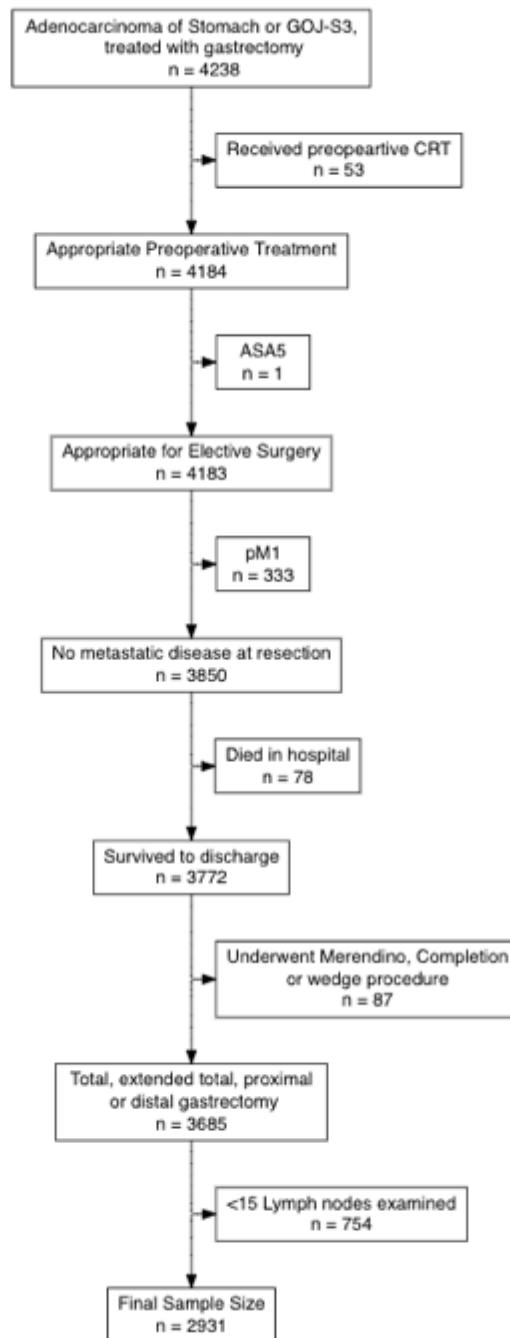


Table S.2 Candidate Predictors

Preoperative	Operative	Pathological/Post-operative
Gender	Approach	pT/ypT
Age	Number of Procedures	Total number of positive lymph nodes
Site of Tumour	Any Complications	Grade of differentiation
cT	Anastomotic Leak	Completeness of resection (R0/R1)
cN	Cardiac Complications	Adjuvant Treatment
IHD	Respiratory Complications	
COPD	Pleural Effusion	
CKD	Pneumonia	
DM	Extent of Nodal Dissection	
CVD		
PS		
ASA		
Hospital Volume (Major Upper GI Resection)		
Hospital Volume (Gastrectomy Alone)		
Neoadjuvant Treatment		

Figure S.3 Boruta variable importance. Candidate variables are compared to a set of variables that have been randomised in a Random Forest model and those with an importance to survival significantly higher than all randomised (shadow) variables are selected for inclusion.

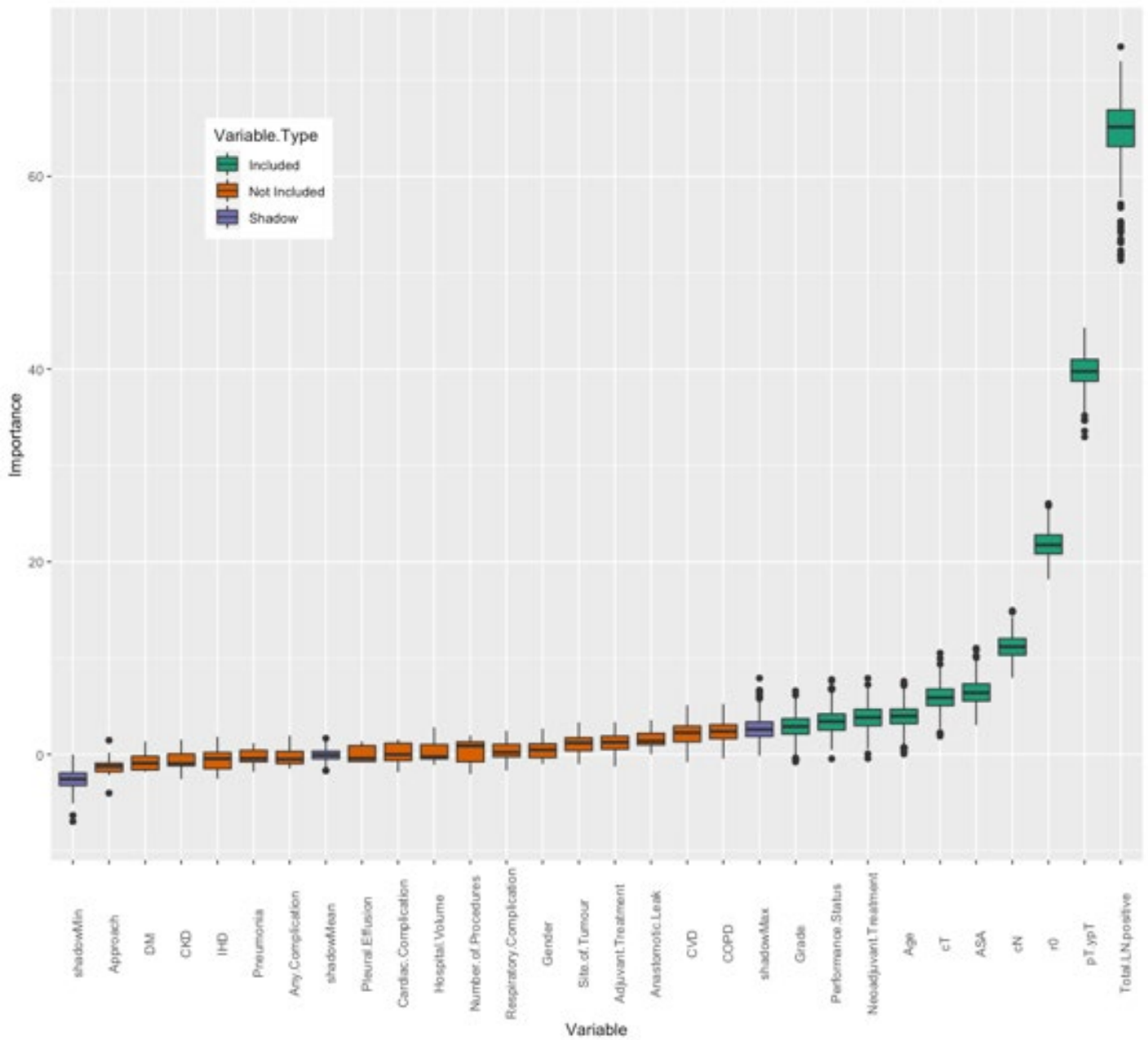


Figure S.4 Agreement between observed and predicted survival, grouped into quintiles by predicted survival at five-years post-surgery

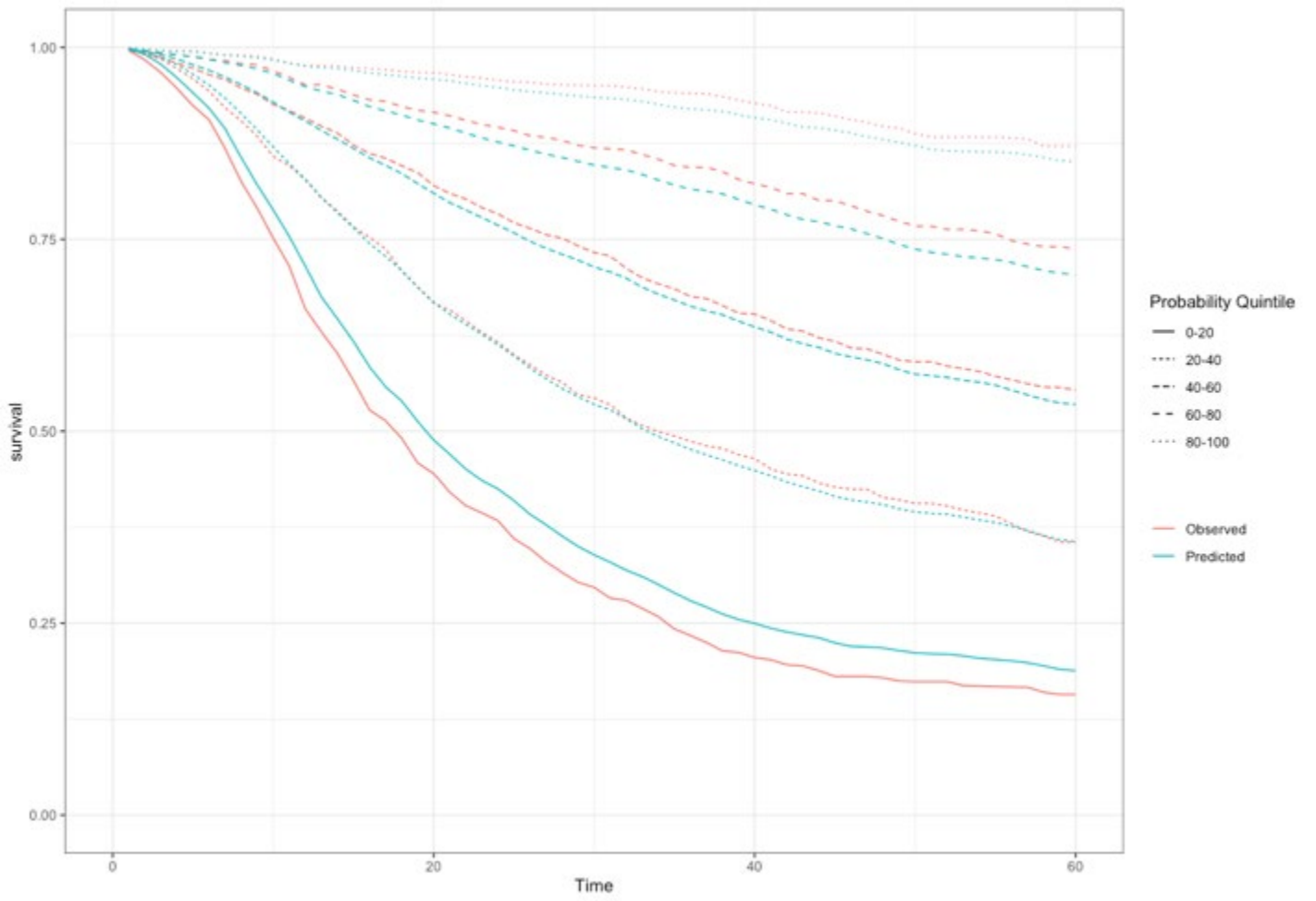


Figure S.5 Interquartile range of predictions within pTNM staging groups

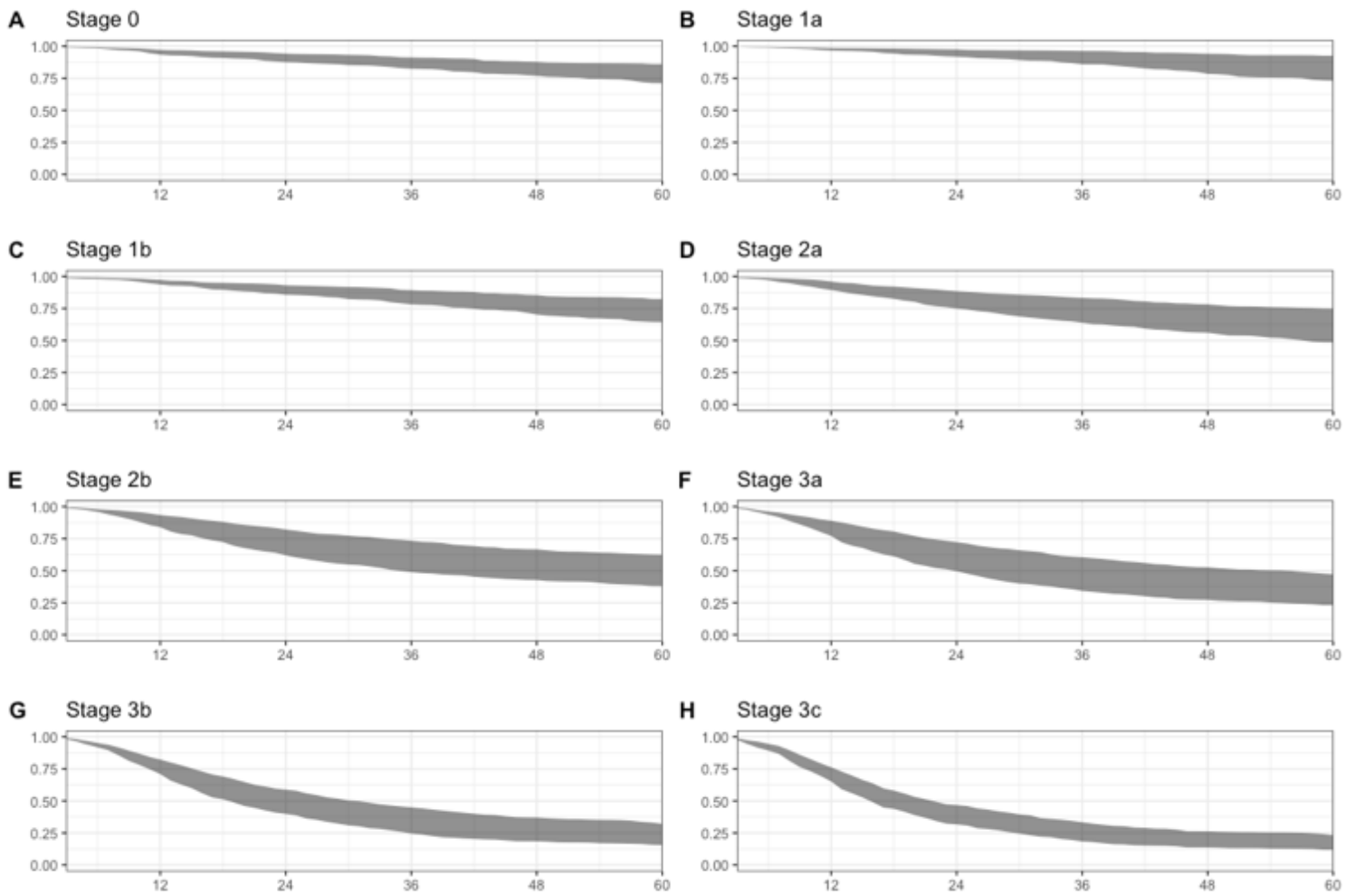
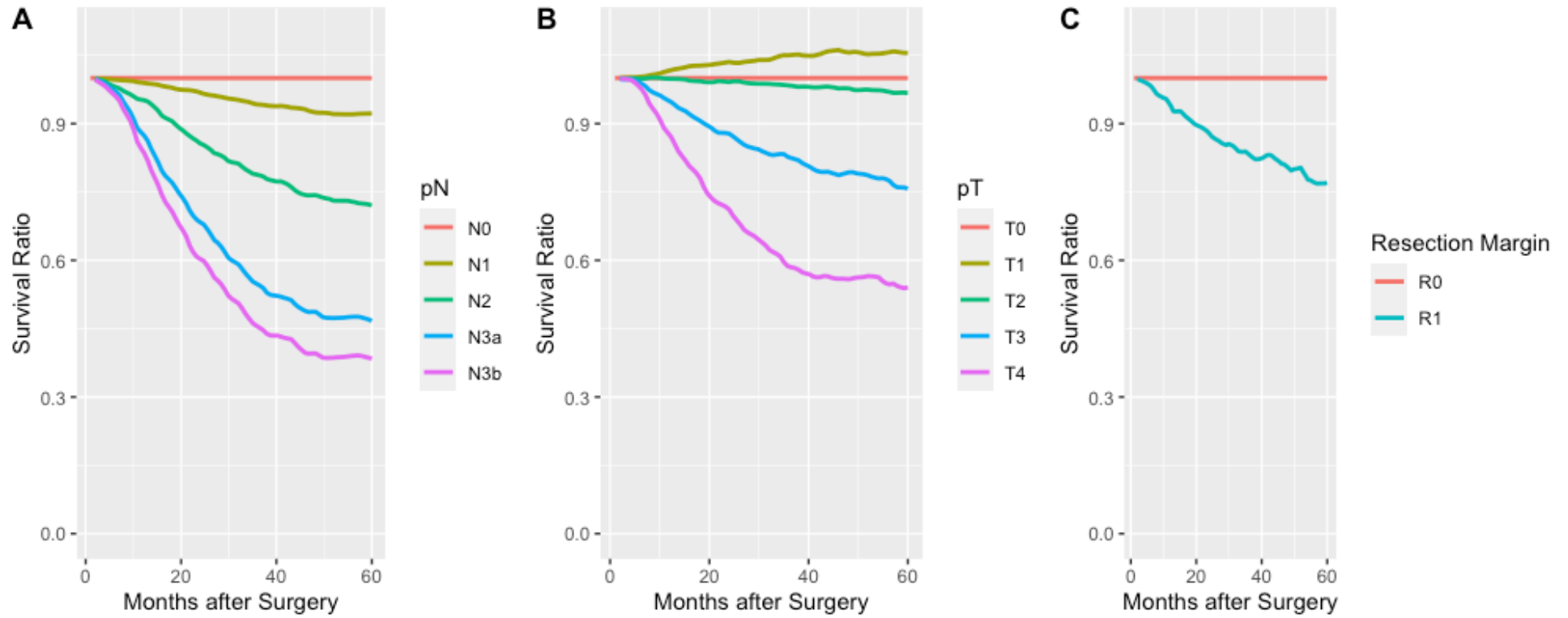


Table S.3 Predicted mean survival time according to covariate selection, restricted to five years

Characteristic		Predicted mean survival time (Months)	Life Expectancy Difference (Months 95% CI)	Life Expectancy Ratio (95% CI)
Age	18-50	43.3	Reference	1
	51-60	43.1	-0.17 (-3.99 to 3.65)	1 (0.91 to 1.08)
	61-70	42.7	-0.55 (-4.37 to 3.27)	0.99 (0.9 to 1.07)
	71-80	41.5	-1.8 (-5.62 to 2.02)	0.96 (0.87 to 1.04)
	80+	41.0	-2.25 (-6.07 to 1.57)	0.95 (0.86 to 1.03)
cT	T0/is/1	42.5	Reference	1
	T2	42.3	-0.14 (-3.21 to 2.92)	1 (0.92 to 1.07)
	T3	41.8	-0.72 (-3.78 to 2.34)	0.98 (0.91 to 1.05)
	T4	42.3	-0.14 (-3.2 to 2.93)	1 (0.92 to 1.07)
cN	N0	42.4	Reference	1
	N1	42.2	-0.27 (-3.33 to 2.79)	0.99 (0.92 to 1.07)
	N2	41.4	-0.97 (-4.02 to 2.08)	0.98 (0.91 to 1.05)
	N3	40.6	-1.81 (-4.85 to 1.23)	0.96 (0.89 to 1.03)
ASA	1	42.3	Reference	1
	2	42.6	0.25 (-2.82 to 3.32)	1.01 (0.93 to 1.08)
	3	41.3	-1.01 (-4.06 to 2.04)	0.98 (0.9 to 1.05)
Performance Status	0	42.9	Reference	1
	1	41.8	-1.1 (-4.17 to 1.97)	0.97 (0.9 to 1.05)
	2	40.3	-2.59 (-5.64 to 0.46)	0.94 (0.87 to 1.01)
	3	40.6	-2.25 (-5.31 to 0.8)	0.95 (0.88 to 1.02)
Neoadjuvant Treatment	None	42.2	Reference	1
	Chemotherapy-Completed	42.0	-0.23 (-3.28 to 2.83)	0.99 (0.92 to 1.07)
	Chemotherapy-not completed			
Completeness of Resection	R0	42.6	Reference	1
	R1	37.9	-4.71 (-7.71 to -1.7)	0.89 (0.82 to 0.96)
pT Stage	T0	47.2	Reference	1
	T1	48.8	1.69 (-1.55 to 4.93)	1.04 (0.97 to 1.11)
	T2	46.5	-0.66 (-3.85 to 2.54)	0.99 (0.92 to 1.05)
	T3	41.4	-5.8 (-8.91 to -2.68)	0.88 (0.82 to 0.94)
	T4	33.9	-13.25 (-16.29 to -10.2)	0.72 (0.66 to 0.77)
pN Stage	N0	46.4	Reference	1
	N1	44.4	-1.96 (-5.78 to 1.86)	0.96 (0.88 to 1.04)
	N2	39.1	-7.26 (-11.08 to -3.44)	0.84 (0.77 to 0.92)
	N3a	32.6	-13.8 (-17.62 to -9.98)	0.7 (0.63 to 0.77)
	N3b	29.3	-17.09 (-20.91 to -13.27)	0.63 (0.56 to 0.7)
Grade of Differentiation	G1	42.7	Reference	1
	G2	42.5	-0.19 (-3.26 to 2.87)	1 (0.92 to 1.07)
	G3/4	42.0	-0.71 (-3.77 to 2.35)	0.98 (0.91 to 1.05)
	GX	42.5	-0.17 (-3.24 to 2.89)	1 (0.92 to 1.07)

All values restricted to five years. The Life expectancy difference (LED) is the change in mean survival (months) from the reference value, and the Life expectancy ratio (LER) is the corresponding ratio of this.

Figure S.6 Life expectancy ratio (LER) over time. Persistent effects are seen for positive resection margin. pT/pN effects plateau at about 4 years



External Validation Instructions

A basic knowledge of R is required to conduct the external validation. As the model does not generate coefficients, access to the model itself is required.

First, download the file packet from the web application in the '*Model Details*' tab. This contains the models themselves and the manner in which dummy coding was conducted.

An example blank dataframe is also included showing the structure in which data must be presented to the model. Care should be taken to match the variables/names/factor-levels in this file. If the model fails to generate predictions, it is probably due to a discrepancy here.

Then access and download the R script from github:

<https://github.com/saqibrahmanUGI/AUGIS-Surv>

Running this script will firstly install and load the needed R packages, then batch generate predictions, calculate the tAUC and C-index, plot annual calibration curves and plot quintiles of prediction against observed KM estimates.