Prediction models for patients with esophageal or gastric cancer: A systematic review and meta-analysis

H. G. van den Boorn[1,2]*, E. G. Engelhardt[3,4], J. van Kleef[1,2], M. A. G. Sprangers[3,5], M. G. H. van Oijen[1,2], A. Abu-Hanna[6], A. H. Zwinderman[7], V. M. H. Coupé[4], H. W. M. van Laarhoven[1,2]

1 Cancer Center Amsterdam, Amsterdam, The Netherlands, 2 Department of Medical Oncology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands, 3 Amsterdam Public Health Research Institute, Amsterdam, The Netherlands, 4 Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands, 5 Department of Medical Psychology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands, 6 Department of Medical Informatics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands, 7 Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

* h.g.vandenboorn@amc.uva.nl

Abstract

Background

Clinical prediction models are increasingly used to predict outcomes such as survival in cancer patients. The aim of this study was threefold. First, to perform a systematic review to identify available clinical prediction models for patients with esophageal and/or gastric cancer. Second, to evaluate sources of bias in the included studies. Third, to investigate the predictive performance of the prediction models using meta-analysis.

Methods

MEDLINE, EMBASE, PsycINFO, CINAHL, and The Cochrane Library were searched for publications from the year 2000 onwards. Studies describing models predicting survival, adverse events and/or health-related quality of life (HRQoL) for esophageal or gastric cancer patients were included. Potential sources of bias were assessed and a meta-analysis, pooled per prediction model, was performed on the discriminative abilities (c-indices).

Results

A total of 61 studies were included (45 development and 16 validation studies), describing 47 prediction models. Most models predicted survival after a curative resection. Nearly 75% of the studies exhibited bias in at least 3 areas and model calibration was rarely reported. The meta-analysis showed that the averaged c-index of the models is fair (0.75) and ranges from 0.65 to 0.85.

Conclusion

Most available prediction models only focus on survival after a curative resection, which is only relevant to a limited patient population. Few models predicted adverse events after resection, and none focused on patient's HRQoL, despite its relevance. Generally, the quality of reporting is poor and external model validation is limited. We conclude that there is a need for prediction models that better meet patients' information needs, and provide information on both the benefits and harms of the various treatment options in terms of survival, adverse events and HRQoL.

Introduction

Worldwide, esophageal and gastric cancer account for 3.2% and 6.8% of all new cancer cases, respectively. The prognosis is dismal: 1% of patients with esophageal cancer and 5% of patients with gastric cancer survive at least 5 years after being diagnosed[1]. However, survival rates for both entities vary greatly[1–4] and metastasis is one of the decisive factors for curative or palliative treatment. In both the curative and palliative setting, patients may choose between various treatment options that differ in terms of efficacy, adverse events and impact on health-related quality of life (HRQoL).

Many patients with potentially curable esophageal or gastric cancer report loss of HRQoL [5, 6] during the first year after surgery, even though patients indicate that an improved HRQoL may be their primary outcome of treatment[7]. Likewise, one in four patients with metastatic esophageal cancer state that HRQoL is their main treatment goal[8]. Since life prolonging treatment may come at a cost as it may induce adverse events and impair HRQoL[5, 6], patients need to be informed at an early stage about the projected survival, adverse events and HRQoL.

To make well-informed treatment choices that match patients' preferences and goals, information about treatment outcomes in terms of survival, treatment-related adverse events and HRQoL is necessary[9]. Statistical prediction models that provide personalized estimates of such outcomes can help inform patients and clinicians consequently supporting shared decision-making. Such statistical models are generally derived from large historical patient cohorts. Examples of such models in oncology are Adjuvant![10] and PREDICT[11], which are broadly used in the field of breast cancer. However, a comprehensive overview of available models for esophageal and gastric cancer, and their predictive performance is currently lacking. Therefore, the aim of this systematic review was first to provide an overview of published prediction models that provide personalized estimates of survival probabilities (i.e., overall, disease-specific, progression-free or disease-free survival), the probability of developing treatment-related adverse events, and/or the impact of treatment on HRQoL. Secondly, we aimed to examine the quality of the development and validation studies conducted for the identified prediction models. Finally, we evaluated the reported performance of the prediction models in terms of discriminative ability and calibration.

Methods

Systematic literature search

A systematic literature search was performed to identify all relevant publications in the bibliographic databases MEDLINE, EMBASE, PsycINFO, CINAHL, and The Cochrane Library (no

protocol available). To increase the relevance of the findings of this review for current clinical practice, we only included papers published from January, 1st 2000 up to February 6th, 2017. Search terms for 'esophageal cancer' or 'gastric cancer' were used in combination with search terms for 'prediction model', 'survival', 'adverse events' and 'quality of life' (see S1 Table for the detailed search strategy). The reference list of relevant articles identified were also searched for additional relevant publications.

The aim of our search was to identify prediction models that provide personalized estimates of survival, the probability of experiencing an adverse event and/or the impact of disease or treatment on HRQoL for esophageal and gastric cancer patients. Models intended to support treatment decisions in both the curative or the palliative setting were eligible for inclusion. Studies validating models in patients with esophageal or gastric cancer that were not originally developed for use in these populations, were also eligible for inclusion. Also, only papers published in English were assessed. We excluded studies describing prediction models that aimed to classify patients into risk categories (such as "low risk" and "high risk"), rather than providing personalized estimates of outcome probability. Although risk categories may be useful for discriminating between outcome severity, it is difficult to quantify the calibration of such prediction models (i.e., how does the expected outcome compare to the actual observed outcome). This is an important aspect of model validation, as the absolute outcome probabilities are needed to determine model fit, and therefore, the quality of the model.

The selection process consisted of two phases. First, all titles and abstracts were screened by two reviewers (HvdB and EE) independently. Discrepancies were resolved through consensus, and when necessary by consulting a third arbiter (HvL). Studies were also selected if eligibility could not be determined on the basis of the titles and abstracts. In the second phase, two reviewers (HvdB and EE) independently screened full texts of the studies selected in phase 1 to determine eligibility conclusively.

Data extraction

Data were extracted from the full text papers according to the CHARMS[12] statement, which provides a data extraction checklist for systematic reviews of prediction models. Extracted data included information about the type of article, study design, data source, characteristics of the population, aim of the model, type of outcome, sample size, methods used and presentation of the final prediction model. Model performance was also extracted and categorized as development performance (obtained when using the development dataset), internal validation performance (obtained when using data from a population similar to that of the development set), and external validation performance (when the data used differs temporally, geographically etc. from the development set). Model performance was described using measures for discriminative ability and measures for calibration. Discriminative ability is defined as a model's ability to differentiate between patients who experience an event (such as death or an adverse event) and those who do not[13]. This can be quantified by calculating an index of predictive discrimination, the concordance index (c-index). This c-index typically has values ranging from 0.5 (no discrimination at all) to 1 (perfect discrimination), and is the generalization of the area under the curve, a well-known measure of discrimination. Typically c-indices can be interpreted by the following rule of thumb: 0.5–0.6 no discrimination, 0.6–0.7 poor, 0.7–0.8 fair, 0.8–0.9 good and 0.9–1 excellent discrimination. Model calibration, in contrast, conveys the goodness of fit, i.e., the congruence between observed and average predicted outcomes [13]. Calibration can be displayed visually in a calibration plot.

The levels of evidence of the discriminatory accuracy of the prediction model as described by Reilly and Evans[14], indicates how extensively a prediction model has been validated and

to what extent a model is ready for clinical use. Level 1 refers to model development, level 2 to narrow validation, level 3 to broader validation and level 4 and 5 to respectively narrow or broad impact analysis. Each identified study was categorized according to the Reilly-Evans levels. For the assessment of bias, there are no established checklists specifically designed for use in prediction modelling studies. We therefore created a classification system for several areas of possible bias, which were derived from the TRIPOD-statement (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis)[15]. S2 Table presents an overview of the classification system used for potential risk of bias.

Data extraction was performed by two researchers (HvdB and EE). First, a subset of 10 articles was used as a training set. The training set was coded by both researchers independently and discrepancies in coding were resolved during a consensus meeting. The percentage overall agreement between the two coders was approximately 90% across individual items. The coding scheme was revised where necessary as a result of the training set findings. Thereafter, each researcher coded half of the remaining articles. Classification of the potential for bias was done in two stages; each researcher made notes of potential sources of bias per category separately, and together they (HvdB and EE) then categorized the identified potential sources of bias. The bias was determined in six areas: population-related (such as selection bias), predictor-related (such as ill-defined predictors), outcome-related (such as an unclear outcome), sample size-related, missing data-related (such as only complete case analysis) and statistical analysis-related (such as underreporting of statistics).

Bias analyses

Descriptive analyses were used to summarize study and model characteristics. We expected that the higher the impact factor of a journal in which the study was published, the more stringent the internal screening and peer review procedures would be and, hence, the lower the risk of bias. Further, we hypothesized that the higher the impact factor of the journal a prediction model was published in, the better its performance in terms of c-index would be. Both hypotheses were assessed through the Spearman rank correlation between the journal impact factor [16] (in the year of publication, or the closest to publication year available) and the reported c-index as well as between journal impact factor and the potential sources of bias (assessed using the classification of potential sources of bias presented in S2 Table), respectively. Due to differences in esophageal carcinoma histology in different geographical populations[17], we examined whether models were constructed and validated with patient cohorts from different continents using the Fisher's exact test. Finally, we hypothesized that the reported c-indices would be larger during model development than during validation due to overfitting. This was assessed using a one-tailed Wilcoxon signed-rank test. These analyses were performed in the R-studio environment with R version 3.3.3 (R Foundation for Statistical Computing, Vienna, Austria, https://www.r-project.org).

Meta-analyses of c-indices

To gain insight in the discriminative abilities of the prediction models, we performed meta-analyses. The c-indices were pooled per prediction model using random effects modelling for models for which at least two concordance indices were available. Analyses were performed using linear restricted maximum-likelihood estimation. In most articles, the c-index confidence interval or variance was not reported. In those cases, the study weights in the meta-analysis were determined as the inverse square root of the sample size. The logistic transformation as described in Kottas et al.[18] was applied to all c-index estimates during calculations and then transformed back; this procedure ensures that all estimates are bounded by 0 and 1 after
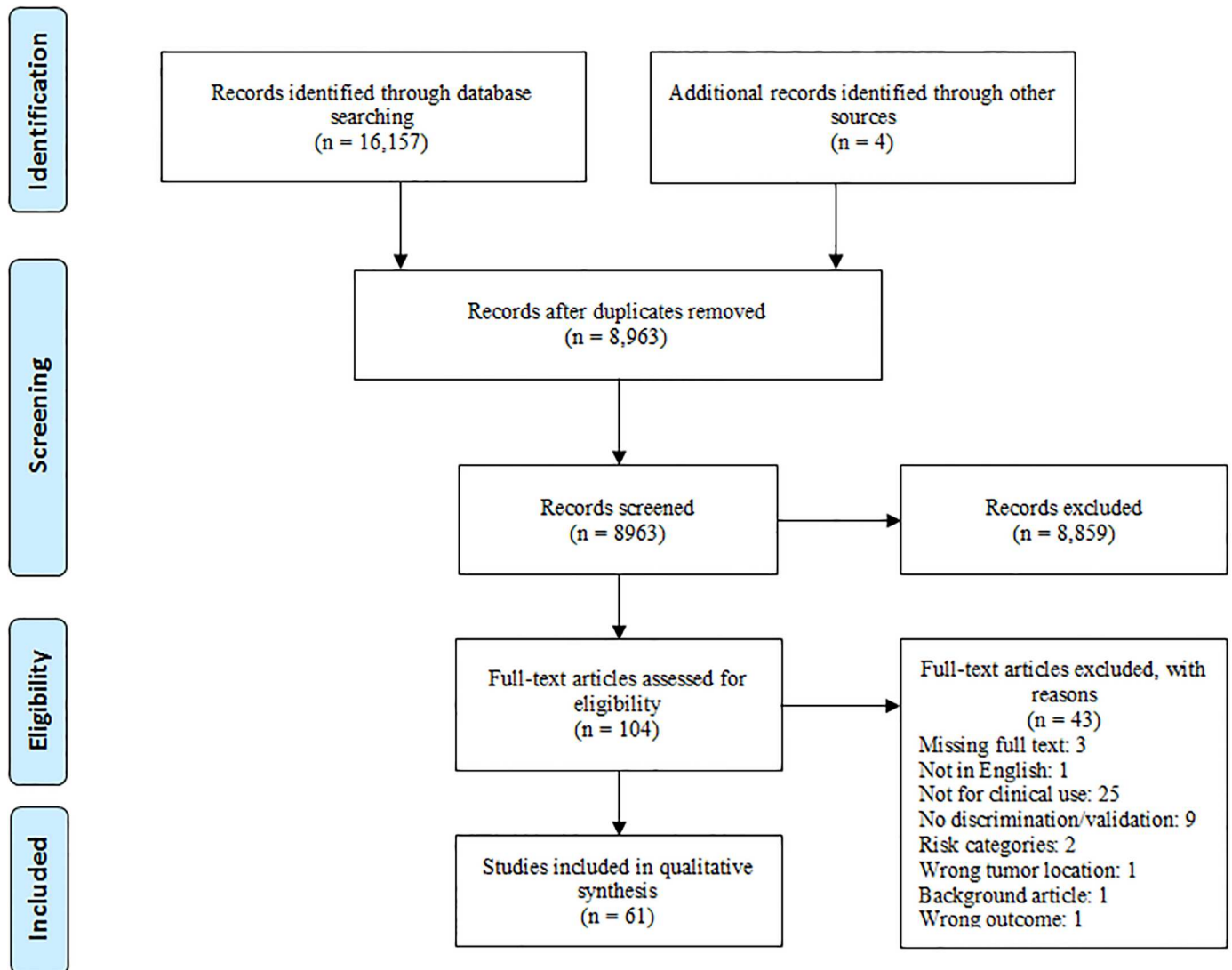
Fig 1. Overview of study selection according to the "Preferred Reporting Items for Systematic Reviews and Meta-Analyses" (PRISMA) statement[21].

https://doi.org/10.1371/journal.pone.0192310.g001

pooling, which is a property of the c-index. These analyses were performed using the Metafor package in the R-studio environment (R version 3.3.3).

Results

A total of 8,963 articles was identified, of which 61 were eligible for inclusion in this systematic review (Fig 1). These studies described a grand total of 47 prediction models for patients with esophageal or gastric cancer. Two studies describing the development of a prediction model, were not included in our systematic review due to the publication year (POSSUM[19]), and incorrect patient population (P-POSSUM[20]). The remaining 45 development studies are shown in Table 1. Further, we found 16 validation studies on a total of 10 prediction models. These studies are shown in Table 2.

Of the models described in the 45 development studies, six predict adverse events; one predicts the recurrence of malignancy; and most studies (N = 39) predict various types of survival

Table 1. Overview of selected studies which describe the creation of a novel prediction model.

Study	N	Country	Tumor location	Treatment intention	Outcome	Model c-indices	Model presentation	Reilly-Evans level
Biglarian 2011[22]	300	Iran	Stomach	Unclear	OS	OS: 0.88 (dev), 0.92 (int)	None	1
Cao 2016[23]	4281	USA, China	Esophagus	Unclear	CSS	DSS: 0.72 (dev), 0.699 (ext)	Nomogram	2
Chen, S. 2016[24]	308	China	Esophagus	Curative	DSS	DSS: 0.688 (dev)	Nomogram	1
Deans 2007[25]	220	UK	Esophagogastric	Curative/ Palliative	OS	OS: 0.84 (dev), 0.85 (dev)	Formula	1
Dhir 2012[26]	14235	USA	Stomach	Curative/ Palliative	POM	POM: 0.75 (ext)	Nomogram	2
Dikken 2013[27]	1642	USA/NL	Stomach	Curative	DSS	DSS: 0.77 (dev)	Nomogram	1
Duan 2016[28]	328	China	Esophagus	Curative	OS, DFS	OS: 0.71 (dev), 0.77 (int); DFS 0.71 (dev), 0.65 (int)	Nomogram	1
Eil 2014[29]	824	USA	Esophagus	Unclear	OS	OS: 0.72 (dev)	Online tool	1
Eom 2015[30]	1579	Korea	Stomach	Curative	OS	OS: 0.831 (ext)	Nomogram	3
Filip 2015[31]	167	Italy	Esophagus	Unclear	AE	AE: 0.8 (dev)	Formula	3
Fischer 2016[32]	4882	UK	Esophagogastric	Curative	POM, AE	POM: 0.698 (dev), 0.694 (dev); AE: 0.631 (dev)	Formula	1
Fuccio 2016[33]	267	Italy	Esophagus	Curative/ Palliative	AE	AE: 0.617 (dev), 0.617 (dev), 0.622 (dev)	Table	1
Gabriel 2017[34]	7179	USA	Esophagus	Curative	OS	OS: 0.656 (dev), 0.669 (dev), 0.63 (int), 0.682 (int)	Formula	1
Haga 2015[35]	762	Japan	Stomach	Unclear	OS	OS: 0.89 (dev)	Formula	1
Han 2012[36]	5300	Korea, Japan	Stomach	Unclear	OS	OS: 0.78 (int), 0.79 (ext)	Nomogram	2
Hirabayashi 2014[37]	3085	Japan	Stomach	Curative	OS	OS: 0.68 (ext)	Nomogram	2
Jiang 2016[38]	125	China	Stomach	Unclear	OS	OS: 0.868 (int), 0.698 (int), 0.84 (int), 0.786 (int), 0.836 (ext), 0.669 (ext), 0.832 (ext), 0.749 (ext)	Nomogram	2
Jung 2013[39]	239	Korea	Esophagus	Palliative	OS	OS: 0.69 (dev)	Nomogram	1
Kattan 2003 (MSKCC) [40]	1039	USA	Stomach	Curative	DSS	DSS: 0.8 (dev)	Nomogram, online tool	3
Kim, Y. 2015[41]	719	USA	Stomach	Curative	OS, DFS	OS: 0.711 (dev), 0.691 (ext); DFS: 0.702 (dev), 0.685 (ext)	Nomogram	1
Kunisaki 2016[42]	52770	Japan	Stomach	Unclear	AE	AE: 0.797 (int), 0.784 (int), 0.748 (int), 0.832 (int), 0.728 (int), 0.7 (int), 0.779 (int), 0.658 (int)	Formula	2
Kurita 2015[43]	33917	Japan	Stomach	Curative	POM	POM: 0.785 (dev), 0.798 (int)	None	1
Lagarde 2007b[44]	364	Unclear	Esophagus	Curative	DSS	DSS: 0.77 (dev)	Nomogram	2
Lagarde 2008a[45]	663	Netherlands	Esophagus	Curative	AE	AE: 0.65 (dev), 0.666 (int)	Nomogram	3
Lai 2009[46]	2923	Korea	Stomach	Curative	DFS	DFS: 0.79 (dev)	None	2
Liu, J. 2016a[47]	817	China	Stomach	Unclear	OS	OS: 0.79 (ext)	Nomogram	1
Liu, J. 2016b[48]	2770	USA, China	Stomach	Curative	DSS	DSS: 0.73 (int), 0.76 (ext)	Nomogram	2
Liu, J.S. 2015[49]	326	China	Esophagus	Curative	DSS	DSS: 0.72 (dev)	Nomogram	1
Marrelli 2005[50]	536	Italy	Stomach	Curative	DFS	DFS: NA (dev)	Formula	2
Mohammadzadeh 2015[51]	194	Iran	Stomach	Unclear	OS	OS: 0.8 (dev), 0.79 (int)	Decision tree	1
Muneoka 2016[52]	207	Japan	Stomach	Curative	DFS	DFS: 0.8 (dev)	Nomogram, online tool	1
Shao 2015[53]	633	China	Esophagus	Curative	OS	OS: 0.77 (dev), 0.77 (dev), 0.76 (int), 0.77 (int)	Nomogram	1
Shapiro 2016[54]	626	Netherlands	Esophagus	Curative	OS	OS: 0.63 (dev)	Nomogram	1
Shiozaki 2016[55]	64	USA	Esophagogastric	Palliative	OS	OS: 0.61 (dev)	Nomogram	1

(*Continued*)

Table 1. (Continued)

Study	N	Country	Tumor location	Treatment intention	Outcome	Model c-indices	Model presentation	Reilly-Evans level
Song 2014[56]	805	Korea	Stomach	Curative	DSS	DSS: 0.87 (dev), 0.84 (int)	Nomogram, formula	1
Steyerberg 2006[57]	1327	USA, Netherlands	Esophagus	Unclear	POM	POM: 0.66 (dev), 0.7 (ext), 0.56 (ext), 0.66 (ext)	Formula	3
Su 2015[58]	797	China	Esophagus	Unclear	OS	OS: 0.73 (dev), 0.715 (int)	Nomogram	1
Suzuki 2012[59]	196	USA	Esophagus	Unclear	OS, DFS	OS: 0.7 (dev); DFS: 0.77 (dev)	Nomogram	1
Tekkis 2004 (O-POSSUM)[60]	1042	UK	Esophagogastric	Curative/ Palliative	POM	POM: 0.8 (dev)	Formula	3
Tu 2017[61]	3632	China	Stomach	Curative	AE	AE: 0.68 (dev)	Nomogram	1
Woo 2016[62]	11851	Korea, Japan, China	Stomach	Curative/ Palliative	OS	OS: 0.824 (dev), 0.842 (ext), 0.868 (ext), 0.839 (ext), 0.798 (ext)	Formula	3
Yang 2013[63]	319	China	Esophagus	Curative	REC	Not available	Formula	1
Yu 2016[64]	1004	China	Esophagus	Curative	OS	OS: 0.7 (dev)	Nomogram	1
Zhao 2016[65]	510	China	Stomach	Curative	OS	OS: 0.834 (dev), 0.809 (int)	Nomogram	1
Zhou, Z. 2015[66]	953	USA, China	Esophagus	Curative	OS	OS: 0.69 (dev), 0.75 (ext)	Nomogram	2

N: sample size used for training. DSS: disease-specific survival, POM: post-operative mortality, OS: overall survival, AE: adverse events, DFS: disease-free survival, REC: cancer recurrence. The type of validation is indicated in brackets with the reported c-index; dev: development c-index, int: internal validation, ext: external validation.

https://doi.org/10.1371/journal.pone.0192310.t001

Table 2. Overview of studies which externally validate prediction models.

Study	Validation of	N	Country	Tumor location	Treatment intention	outcome	Model C-indices	Reilly-Evans level
Ashfaq 2015[67]	MSKCC[40]	6954	USA	Stomach	Curative	DSS	DSS: 0.68	3
Bosch 2011[68]	P-POSSUM[20], O-POSSUM [60]	278	Netherlands	Esophagus	Curative	POM	POM: 0.766, 0.756	3
Chen, D. 2013[69]	MSKCC[40]	979	China	Stomach	Curative	DSS	DSS: 0.74	3
D'Journo 2016[70]	Steyerberg 2006[57]	1039	France	Esophagus	Unclear	OS	OS: 0.63, 0.64, 0.63	3
Dikken 2014[71]	MSKCC[40]	139	USA	Stomach	Unclear	DSS	DSS: 0.64	3
Grotenhuis 2010[72]	Lagarde 2008a[45]	777	Netherlands	Esophagus	Curative	AE	AE: 0.64	3
Kim, J.H. 2012[73]	Lai 2009[46]	930	Korea	Stomach	Curative	DFS	DFS: 0.7	2
Lagarde 2007a[74]	O-POSSUM[60]	663	Netherlands	Esophagus	Curative	POM	POM: 0.6	3
Lagarde 2008b[75]	Lagarde 2007b[44]	382	Belgium	Esophagus	Curative	DSS	DSS: 0.76	2
Marrelli 2015[76]	Marrelli 2005[50]	635	Italy	Stomach	Curative	REC	REC: 0.889	2
Nagabhushan 2007 [77]	P-POSSUM[20], O-POSSUM [60]	313	UK	Esophagogastric	Curative	POM	POM: 0.68, 0.61	3
Novotny 2006[78]	MSKCC[40]	862	Germany	Stomach	Curative	DSS	DSS: 0.77	3
Peeters 2005[79]	MSKCC[40]	459	Netherlands	Stomach	Curative	DSS	DSS: 0.77	3
Reim 2015[80]	Eom 2015[30]	908	Germany	Stomach	Curative	OS	OS: 0.761	3
Zafirellis 2002[81]	POSSUM[19]	204	UK	Esophagus	Curative/Palliative	OS, AE	OS: 0.62; AE: 0.55	3
Zhou, M.L. 2016[82]	MSKCC[40]	150	China	Stomach	Curative	DSS	DSS: 0.657	3

N: sample size used for validation, DSS: disease-specific survival, POM: post-operative mortality, OS: overall survival, AE: adverse events, DFS: disease-free survival, REC: cancer recurrence.

https://doi.org/10.1371/journal.pone.0192310.t002

Overview prediction models

Fig 2. Overview of included prediction models. The shape indicates the type of study and the size of shapes indicate the pooled c-index. Larger sizes of shapes indicate higher c-indices. AE = adverse event; Reilly-Evans = levels of evidence on the discriminatory accuracy of the prediction model described by Reilly and Evans[14], which indicate how extensively a prediction model has been validated and to what extent a model is ready for clinical use.

(six disease-free survival, eight disease-specific survival, 23 overall survival and five post-operative mortality). None of the studies predict HRQoL and none predict more than one outcome, i.e., no model predicts both the harms and benefits of the treatments of interest. The majority of studies (N = 28) used a nomogram to present the prediction model, while others (N = 13) used a formula as a presentation method (see Table 1). Three prediction models were also available online. A graphical overview of the outcomes per prediction model is given in Fig 2, and includes depiction of each model's Reilly-Evans level of evidence on discriminatory accuracy.

Table 3 provides an overview of the selected studies. Most models underwent only limited validation, as the majority of development models were not validated further in later studies. This is expressed by the Reilly and Evans levels of evidence[14]. In 84% of the development

Table 3. Overview of study characteristics in development and validation studies.

	Development studies	Validation studies	P-value
N	45	16	
Reilly-Evans level (%)			
- 1	27 (60.0)	0 (0.0)	
- 2	11 (24.4)	3 (18.8)	
- 3	7 (15.6)	13 (81.2)	
Continent of patient population (%)			p = 0.003
- Asia	25 (56.8)	3 (18.8)	
- Europe	8 (18.2)	11 (68.8)	
- North-America	10 (22.7)	2 (12.5)	
- North-America and Europe	1 (2.3)	0 (0.0)	
Intended time of model use (%)			p = 0.857
- After adjuvant chemotherapy	1 (2.2)	0 (0.0)	
- After consolidation therapy	1 (2.2)	0 (0.0)	
- After definitive chemoradiation	1 (2.2)	0 (0.0)	
- After resection	32 (71.1)	14 (87.5)	
- At diagnosis	1 (2.2)	0 (0.0)	
- Before definitive chemotherapy	1 (2.2)	0 (0.0)	
- Before resection	5 (11.1)	2 (12.5)	
- Before/after resection	3 (6.7)	0 (0.0)	
Curative/palliative setting (%)			p = 0.316
- Curative	25 (55.6)	13 (81.2)	
- Curative/Palliative	5 (11.1)	1 (6.2)	
- Palliative	2 (4.4)	0 (0.0)	
- Unclear	13 (28.9)	2 (12.5)	
Calibration method (%)			p = 0.045
- Calibration plot	23 (51.1)	6 (37.5)	
- Statistical analysis	2 (4.4)	4 (25.0)	
- Calibration plot and statistical analysis	6 (13.3)	4 (25.0)	
- None	14 (31.1)	2 (12.5)	

studies the two lowest Reilly and Evans levels, namely 1 or 2, were scored indicating only narrow validation. The validation studies are limited to a select group of prediction models, which are validated more extensively. These are the prediction models developed by Eom 2015[30], Lagarde 2007[44], Lagarde 2008[45], Lai 2009[46], Marelli 2005[50], Steyerberg 2006[57], the MSKCC[83], and the Possum[19], O-Possum[60], and P-Possum[20] models. This more extensive validation resulted in a majority of these models having a Reilly and Evans level of 3.

Table 3 also indicates the study patient distribution across the continents. This differs significantly between development and validation studies (p = 0.003), indicating that different populations are used for model development and for validation. This difference is especially pronounced between Asia and Europe (p < 0.001). Models were more often developed in Asian than in European populations (56.8% vs. 18.2% respectively), however, fewer validation studies were conducted in Asian than in European populations (18.8% vs 68.8% respectively). The development and validation studies mostly concerned prediction outcomes before or after resection (89% and 100% respectively), and were mostly aimed at patients treated with curative intent (56% and 81.2% respectively).

Table 4. Overview of areas of bias in the included studies (part 1).

	Subject bias	Predictor bias	Outcome bias	Sample size bias	Missing data bias	Statistical analysis bias
Ashfaq 2015[67]	-	+	+	++	--	-
Biglarian 2011[22]	-	-	?	-	--	--
Bosch 2011[68]	+	+	-	-	-	-
Cao 2016[23]	-	+	?	+	-	+
Chen, D. 2013[69]	+	+	+	+	--	-
Chen, S. 2016[24]	--	+	+	-	-	-
D'Journo 2016[70]	--	+	+	+	-	+
Deans 2007[25]	+	-	+	-	-	-
Dhir 2012[26]	-	-	?	++	-	+
Dikken 2013[27]	-	-	+	+	--	-
Dikken 2014[71]	-	+	-	-	--	-
Duan 2016[28]	+	+	-	+	-	+
Eil 2014[29]	--	-	--	+	--	-
Eom 2015[30]	-	-	+	-	--	+
Filip 2015[31]	--	-	--	-	-	-
Fischer 2016[32]	+	-	+	+	+	-
Fuccio 2016[33]	-	+	--	-	--	-
Gabriel 2017[34]	-	-	+	++	--	-
Grotenhuis 2010[72]	+	-	-	+	+	-
Haga 2015[35]	-	+	--	+	-	-
Han 2012[36]	-	+	?	+	--	+
Hirabayashi 2014[37]	+	+	?	+	--	-
Jiang 2016[38]	--	+	--	-	-	-
Jung 2013[39]	-	+	-	-	-	+
Kattan 2003[40]	-	-	+	+	--	-
Kim, J.H. 2012[73]	+	-	-	+	--	-
Kim, Y. 2015[41]	+	+	--	+	-	-
Kunisaki 2016[42]	-	+	-	++	-	+
Kurita 2015[43]	+	-	-	+	-	-
Lagarde 2007a[74]	+	+	-	+	-	-
Lagarde 2007b[44]	--	+	+	+	?	-
Lagarde 2008a[45]	+	+	-	+	+	-
Lagarde 2008b[75]	+	-	+	-	+	-
Lai 2009[46]	+	-	+	+	--	-
Liu, J. 2016a[47]	-	+	?	+	-	+
Liu, J. 2016b[48]	+	-	?	+	--	+

A minus sign indicates possible areas of bias; a question mark indicates that bias could not be determine;. a positive sign indicates a lack of bias.

Bias analyses

We analyzed several areas of possible bias of the studies, which are shown in Tables 4 and 5. The exact definitions of the biases are presented in S2 Table. Of all selected studies, population-related bias occurred in 61%, predictor-related bias in 43%, outcome-related bias in 43%, sample size -related bias in 38%, missing data-related bias in 89% and statistical analysis-related bias in 66%. All studies have a bias in at least one area. Due to poor or inconsistent reporting, it was difficult to extract pertinent study information. For example, treatment intent was not reported in most articles. In such cases intent was deduced from other available

Table 5. Overview of areas of bias in the included studies (part 2).

	Subject bias	Predictor bias	Outcome bias	Sample size bias	Missing data bias	Statistical analysis bias
Liu, J.S. 2015[49]	--	-	-	-	-	-
Marrelli 2005[50]	+	-	?	-	-	-
Marrelli 2015[76]	+	+	+	+	--	+
Mohammadzadeh 2015[51]	--	-	--	-	--	--
Muneoka 2016[52]	-	+	-	-	-	-
Nagabhushan 2007[77]	+	+	-	-	--	-
Novotny 2006[78]	-	+	+	+	--	-
Peeters 2005[79]	-	-	+	-	--	-
Reim 2015[80]	--	+	+	+	+	+
Shao 2015[53]	+	+	+	+	-	+
Shapiro 2016[54]	+	-	-	+	-	-
Shiozaki 2016[55]	+	+	+	--	--	-
Song 2014[56]	-	-	--	+	--	+
Steyerberg 2006[57]	-	-	+	+	?	-
Su 2015[58]	--	+	+	+	--	+
Suzuki 2012[59]	--	+	?	-	-	-
Tekkis 2004[60]	+	-	+	+	-	-
Tu 2017[61]	+	+	--	+	--	-
Woo 2016[62]	-	+	?	++	-	+
Yang 2013[63]	-	+	+	-	-	-
Yu 2016[64]	+	+	+	+	-	+
Zafirellis 2002[81]	-	+	--	-	--	+
Zhao 2016[65]	-	-	--	-	-	+
Zhou, M.L. 2016[82]	-	+	+	-	--	+
Zhou, Z. 2015[66]	-	-	-	+	-	+

A minus sign indicates possible areas of bias; a question mark indicates that bias could not be determine;. a positive sign indicates a lack of bias.

information such as the presence of metastatic disease. However, in fifteen studies the treatment intent could not be established. Also, unclear descriptions of treatment and patient characteristics limited our ability to evaluate the risk of bias. The potential source of bias that was most difficult to evaluate due to poor reporting, concerns the handling of missing data. Although few studies report that their dataset was complete, most studies did not mention whether this was the case and how they handled missing data (e.g., via multiple imputation). Further, in many studies, it was unclear what outcome was being predicted. For example, authors mention 'survival' as an outcome[51], but it remained unclear whether overall survival or disease-specific survival was implied.

In most studies the model calibration was poorly reported. Although 45 out of 61 studies described some form of calibration, only 16 studies performed a formal statistical calibration analysis to support whether the predicted risk matched the observed risk. None of the studies determined the calibration slope and intercept (which represents the systematic over- or underprediction of risk).

Finally, we also investigated whether the impact factor of the journal in which the study was published influenced the amount of bias. We found no significant correlation between journal impact factor and the risk of population-related bias (rho = 0.09, p = 0.51), predictor-related bias (rho = -0.12, p = 0.37), outcome-related bias (rho = 0.17, p = 0.20), sample size-related

bias (rho = 0.13, p = 0.32), missing data-related bias (rho = 0.03, p = 0.79) or statistical analysis-related bias (rho = 0.03, p = 0.80). When we assessed whether models published in high impact journals performed better in terms of discriminative ability, again, we found no relation between the impact factor of the journal and the reported c-index (rho = 0.15, p = 0.11).

Meta analyses of c-indices

Results of the meta-analysis of available c-indices of corresponding prediction models are shown in Fig 3. Results are pooled per prediction model and are indicated by diamonds. Overall, the meta-analysis highlights that there is great uncertainty about the predictive performances of available models, given the large confidence intervals (with ranges >0.1) in most pooled estimates. Furthermore, the pooled estimates show that the models vary in discriminating ability, ranging from 0.65 (poor discrimination) to 0.85 (good discrimination), with an average pooled estimate of 0.75 (fair discrimination).

To investigate whether model overfitting occurs, that is the discriminative ability of a model is overestimated during training, we examined the difference in model c-indices. It was found that the discriminative ability of the model was indeed larger (p = 0.01) in development (average c-index: 0.76) than in validation studies (average c-index = 0.73).

Discussion

The main aim of this review was to provide an overview of prediction models aimed at predicting survival, adverse events and HRQoL in patients with esophageal or gastric cancer, and establish their predictive performance and biases.

We identified 45 articles describing the development of novel prediction models and only 16 studies validating these prediction models. We were unable to perform meta-analyses of model calibration, as studies either did not or not adequately report model calibration. The meta-analyses of model discriminative abilities indicate large heterogeneity. The pooled estimates of the discriminative abilities tended to have large confidence intervals, which can be explained by low levels of validation and small cohort sizes. The identified studies generally report a fair discriminative ability for the prediction models. Although nearly every study states that the model is potentially useful in practice, almost all studies do acknowledge the need for further external model validation. However, a mere 10 out of 47 prediction models were subsequently tested in such external validation studies. Indeed, the importance of external validation is shown by the present study as we found that the discriminative ability of models was significantly lower in the validation than in the development phase. Presenting only development results may lead to optimism bias and should be acknowledged when using the prediction models in clinical practice. Large datasets are increasingly being made (freely) available online, which may facilitate more extensive validation of prediction models in the future.

Our findings highlight that prior to using any of these prediction models in clinical practice, clinicians need to carefully consider the number and quality of available validations, the countries/populations in which the models were validated, sample sizes and study biases. In fact, the reported low Reilly and Evans levels of validation indicate that the models we have identified are not ready for widespread implementation in clinical practice. Despite the absence of clinically relevant models, the reported results are essential for future benchmarking and validation studies. Eight models have reached a Reilly and Evans level 3, with the MSKCC model being the most promising with a pooled c-index of 0.73, and extensive validation in a wide-range of populations and settings. We recommend that the MSKCC will be further investigated for its added value in clinical practice in terms of, for example, reduction of decisional conflict and increased patient participation (i.e., shared decision making). Only when the
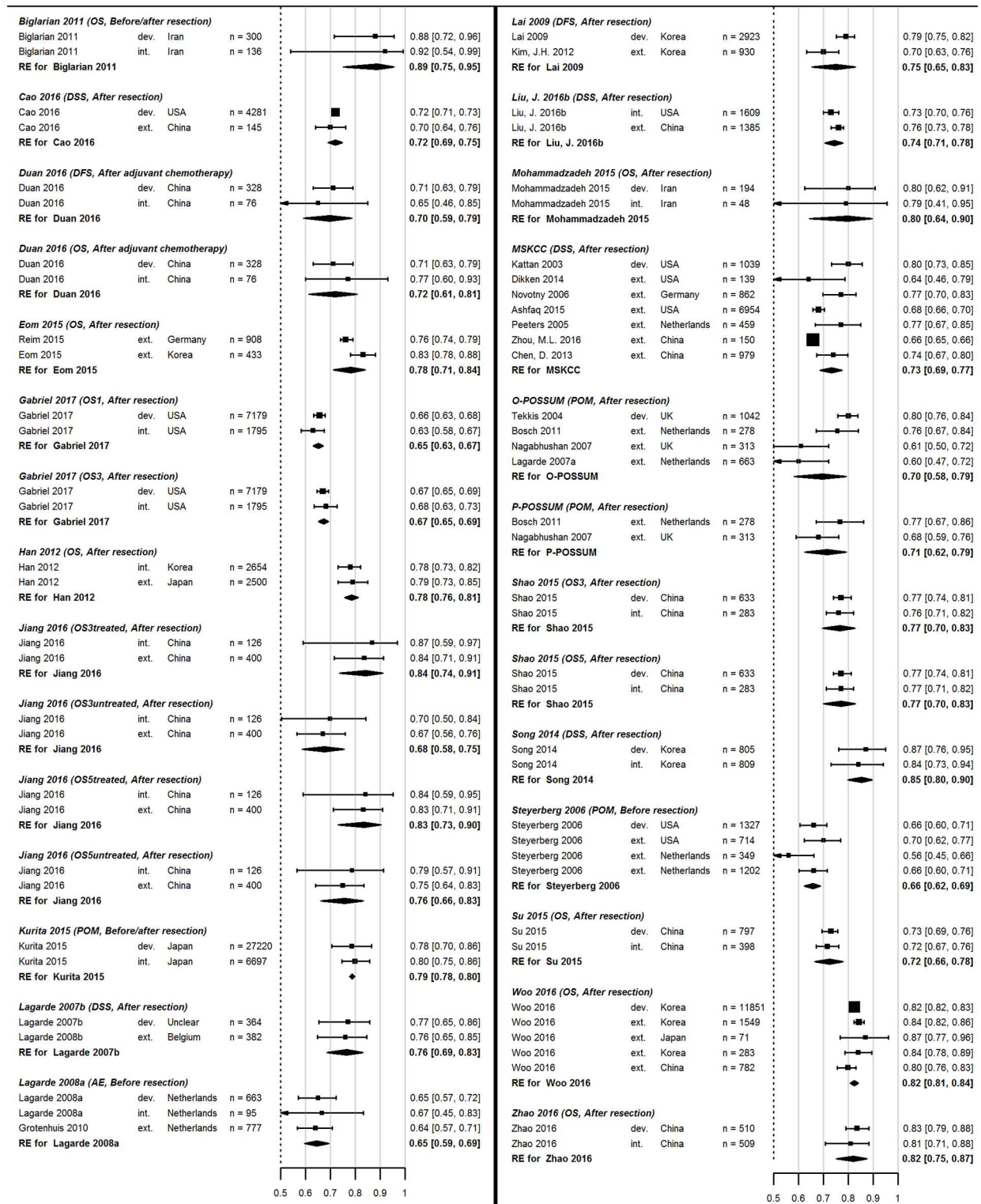
Fig 3. Random effects meta analyses of the discriminative abilities (c-indices) of the identified prediction models. DSS: disease-specific survival, POM: post-operative mortality, OS: overall survival, AE: adverse events, DFS: disease-free survival, REC: cancer recurrence, dev: development c-index, int: internal validation, ext: external validation.

quality of care is improved following implementation of the model, its widespread use in clinical practice can be recommended.

Most of the identified models focus on prediction of survival after curative resection of esophageal or gastric cancer. Although these models provide insight into prognosis of this particular group of patients, they are of limited value for treatment decisions, as treatment has largely been completed at the point of resection. Furthermore, none of the prediction models predict HRQoL, despite the established relevance of HRQoL when making treatment decisions[7], especially in the palliative setting. Finally, in order to make a well-informed treatment choice, patients need to consider both the benefits and harms of treatments to determine which option best fits their preferences and goals. However, none of the prediction models we identified provide estimates of both the benefits and harms associated with a treatment option. Thus, if clinicians opt to use the currently available models, it is imperative that they supplement the information provided by the model with evidence-based predictions concerning not only the possible increase in life-span, but also the possible adverse events and impact on HRQoL.

In order to assess the quality of the studies, we determined sources of possible bias in six different areas. Most studies had a high risk of bias, and all articles showed possible bias in at least one area. The most common bias concerned the handling of missing data. In many studies, it was unclear whether data was missing, how much was missing and how the missing data were handled. Model calibration was not mentioned in some cases and often not accompanied by statistics to provide insight into model quality. Overall, the quality of reporting was poor. Crucial information needed for the interpretation of the results was ill-reported, such as when the model should be used, if the model was to be used with patients for whom treatment has a palliative or curative intent, and what the confidence intervals of the outcomes were. We did not contact authors in cases where the reporting was incomplete, as the focus of this study was to create an overview of reported studies and not to analyze bias in prediction models per se. We strongly advocate that when reporting the development or validation of prediction models the guidance in TRIPOD-statement[15] is followed. This statement provides a checklist of necessary items to include when reporting prediction model development and validation studies, which would facilitate a consistent manner of reporting and safeguard the inclusion of important items needed for interpretation of the data.

In contrast to our expectation, we found no relation between the predictive performance of the models and the impact factor of the journal in which the study is published, nor between the impact factor and study bias. Clinicians should keep in mind that a high impact factor is not a guarantee for quality, and they should always critically assess the quality and generalizability of the prediction model for use in clinical practice. The results of the current study may aid such an evaluation.

In conclusion, we found 47 prediction models intended to predict outcomes in patients with esophageal and gastric cancer. Most models mainly aimed to predict survival after curative resection. Validation of these models is generally limited and the overall performance was fair. There is a clear need for new prediction models for patients with esophageal and gastric cancer that focus on both the potential benefits (e.g., improved survival) and harms (e.g., occurrence of adverse events and/or loss of quality of life) of treatment. Such comprehensive prediction models will likely support the decision-making process.

Supporting information

S1 Table. Search strategy per database.
(DOC)

S2 Table. Overview and categorization of potential sources of bias identified in included articles.
(DOC)

S3 Table. PRISMA[21] checklist.
(DOC)

Acknowledgments

Author Contributions

Conceptualization: H. G. van den Boorn, E. G. Engelhardt, H. W. M. van Laarhoven.

Data curation: H. G. van den Boorn, E. G. Engelhardt.

Formal analysis: H. G. van den Boorn, E. G. Engelhardt.

Funding acquisition: H. W. M. van Laarhoven.

Investigation: H. G. van den Boorn, E. G. Engelhardt.

Methodology: H. G. van den Boorn, E. G. Engelhardt, A. Abu-Hanna, A. H. Zwinderman, H. W. M. van Laarhoven.

Project administration: H. G. van den Boorn, E. G. Engelhardt, H. W. M. van Laarhoven.

Resources: H. W. M. van Laarhoven.

Software: H. G. van den Boorn.

Supervision: H. W. M. van Laarhoven.

Validation: H. G. van den Boorn, E. G. Engelhardt, J. van Kleef, M. A. G. Sprangers, M. G. H. van Oijen, A. Abu-Hanna, A. H. Zwinderman, V. M. H. Coupé, H. W. M. van Laarhoven.

Visualization: H. G. van den Boorn, E. G. Engelhardt, H. W. M. van Laarhoven.

Writing – original draft: H. G. van den Boorn, E. G. Engelhardt, H. W. M. van Laarhoven.

Writing – review & editing: H. G. van den Boorn, E. G. Engelhardt, J. van Kleef, M. A. G. Sprangers, M. G. H. van Oijen, A. Abu-Hanna, A. H. Zwinderman, V. M. H. Coupé, H. W. M. van Laarhoven.

References

1. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer, 2013.

2. Nashimoto A, Akazawa K, Isobe Y, Miyashiro I, Katai H, Kodera Y, et al. Gastric cancer treated in 2002 in Japan: 2009 annual report of the JGCA nationwide registry. Gastric Cancer. 2013; 16(1):1–27. Epub 06/26. https://doi.org/10.1007/s10120-012-0163-4 Epub 2012 Jun 23. PMID: 22729699

3. Reim D, Loos M, Vogl F, Novotny A, Schuster T, Langer R, et al. Prognostic implications of the seventh edition of the international union against cancer classification for patients with gastric cancer: the West-ern experience of patients treated in a single-center European institution. Journal of Clinical Oncology. 2013; 31(2):263–71. https://dx.doi.org/10.1200/JCO.2012.44.4315. PMID: 23213098

4. Surveillance, Epidemiology, and End Results Program: National Cancer Institute; 2013 [cited 2017 30-05-2017]. https://seer.cancer.gov/statfacts/.

5. Jacobs M, Macefield RC, Elbers RG, Sitnikova K, Korfage IJ, Smets EM, et al. Meta-analysis shows clinically relevant and long-lasting deterioration in health-related quality of life after esophageal cancer surgery. Quality of life research: an international journal of quality of life aspects of treatment, care and rehabilitation. 2014; 23(4):1097–115. Epub 2013/10/17. https://doi.org/10.1007/s11136-013-0545-z PMID: 24129668.

6. Al-Batran S-E, Ajani JA. Impact of chemotherapy on quality of life in patients with metastatic esophago-gastric cancer. Cancer. 2010; 116(11):2511–8. https://doi.org/10.1002/cncr.25064 PMID: 20301114

7. Thrumurthy SG, Morris JJA, Mughal MM, Ward JB. Discrete-choice preference comparison between patients and doctors for the surgical management of oesophagogastric cancer. British Journal of Surgery. 2011; 98(8):1124–31. https://doi.org/10.1002/bjs.7537 PMID: 21674471

8. Hitz F, Ribi K, Li Q, Klingbiel D, Cerny T, Koeberle D. Predictors of satisfaction with treatment decision, decision-making preferences, and main treatment goals in patients with advanced cancer. Supportive Care in Cancer. 2013; 21(11):3085–93. https://doi.org/10.1007/s00520-013-1886-4 PMID: 23828394

9. Hodgkinson K, Butow P, Hunt GE, Pendlebury S, Hobbs KM, Lo SK, et al. The development and evaluation of a measure to assess cancer survivors' unmet supportive care needs: the CaSUN (Cancer Survivors' Unmet Needs measure). Psycho-oncology. 2007; 16(9):796–804. Epub 2006/12/21. https://doi.org/10.1002/pon.1137 PMID: 17177268.

10. Ravdin PM, Siminoff LA, Davis GJ, Mercer MB, Hewlett J, Gerson N, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2001; 19(4):980–91. Epub 2001/02/22. https://doi.org/10.1200/jco.2001.19.4.980 PMID: 11181660.

11. Wishart GC, Bajdik CD, Azzato EM, Dicks E, Greenberg DC, Rashbass J, et al. A population-based validation of the prognostic model PREDICT for early breast cancer. European journal of surgical oncology: the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 2011; 37(5):411–7. Epub 2011/03/05. https://doi.org/10.1016/j.ejso.2011.02.001 PMID: 21371853.

12. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 2014; 11(10):e1001744. https://doi.org/10.1371/journal.pmed.1001744 PMID: 25314315.

13. Harrell FE Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in medicine. 1996; 15 (4):361–87. Epub 1996/02/28. https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4 PMID: 8668867.

14. Reilly BM, Evans AT. Translating clinical research into clinical practice: Impact of using prediction rules to make decisions. Annals of Internal Medicine. 2006; 144(3):201–9. https://doi.org/10.7326/0003-4819-144-3-200602070-00009 PMID: 16461965

15. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Medicine. 2015; 13(1):1. https://doi.org/10.1186/s12916-014-0241-z PMID: 25563062

16. 2015 Journal Citation Reports®: Clarivate Analytics; 2017 [cited 2017 March 2017]. https://jcr.incites.thomsonreuters.com.

17. Zhang H-Z, Jin G-F, Shen H-B. Epidemiologic differences in esophageal cancer between Asian and Western populations. Chinese journal of cancer. 2012; 31(6):281. https://doi.org/10.5732/cjc.011.10390 PMID: 22507220

18. Kottas M, Kuss O, Zapf A. A modified Wald interval for the area under the ROC curve (AUC) in diagnostic case-control studies. BMC Medical Research Methodology. 2014; 14(1):26. https://doi.org/10.1186/1471-2288-14-26 PMID: 24552686

19. Copeland GP, Jones D, Walters M. POSSUM: A scoring system for surgical audit. British Journal of Surgery. 1991; 78(3):355–60. https://doi.org/10.1002/bjs.1800780327 PMID: 2021856

20. Prytherch DR, Whiteley MS, Higgins B, Weaver PC, Prout WG, Powell SJ. POSSUM and Portsmouth POSSUM for predicting mortality. Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity. The British journal of surgery. 1998; 85(9):1217–20. Epub 1998/09/30. PMID: 9752863.

21. Moher D, Liberati A, Tetzlaff J, Altman DG, The PG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLOS Medicine. 2009; 6(7):e1000097. https://doi.org/10.1371/journal.pmed.1000097 PMID: 19621072

22. Biglarian A, Hajizadeh E, Kazemnejad A, Zali M. Application of artificial neural network in predicting the survival rate of gastric cancer patients. Iranian Journal of Public Health. 2011; 40(2):80–6. PMID: 23113076

23. Cao J, Yuan P, Wang L, Wang Y, Ma H, Yuan X, et al. Clinical Nomogram for Predicting Survival of Esophageal Cancer Patients after Esophagectomy. Scientific Reports. 2016; 6:26684. https://doi.org/10.1038/srep26684 PMID: 27215834

24. Chen S, Yang X, Feng JF. A novel inflammation-based prognostic score for patients with esophageal squamous cell carcinoma: The c-reactive protein/ prognostic nutritional index ratio. Oncotarget. 2016; 7 (38):62123–32. https://doi.org/10.18632/oncotarget.11389 PMID: 27557504

25. Deans DA, Wigmore SJ, de Beaux AC, Paterson-Brown S, Garden OJ, Fearon KC. Clinical prognostic scoring system to aid decision-making in gastro-oesophageal cancer. British Journal of Surgery. 2007; 94(12):1501–8. https://doi.org/10.1002/bjs.5849 PMID: 17703501

26. Dhir M, Smith LM, Ullrich F, Leiphrakpam PD, Ly QP, Sasson AR, et al. A preoperative nomogram to predict the risk of perioperative mortality following gastric resections for malignancy. Journal of Gastrointestinal Surgery. 2012; 16(11):2026–36. https://doi.org/10.1007/s11605-012-2010-7 PMID: 22948837

27. Dikken JL, Baser RE, Gonen M, Kattan MW, Shah MA, Verheij M, et al. Conditional probability of survival nomogram for 1-, 2-, and 3-year survivors after an R0 resection for gastric cancer. Annals of Surgical Oncology. 2013; 20(5):1623–30. https://doi.org/10.1245/s10434-012-2723-6 PMID: 23143591

28. Duan J, Deng T, Ying G, Huang D, Zhang H, Zhou L, et al. Prognostic nomogram for previously untreated patients with esophageal squamous cell carcinoma after esophagectomy followed by adjuvant chemotherapy. Japanese Journal of Clinical Oncology. 2016; 46(4):336–43. https://doi.org/10.1093/jjco/hyv206 PMID: 26819278

29. Eil R, Diggs BS, Wang SJ, Dolan JP, Hunter JG, Thomas CR. Nomogram for predicting the benefit of neoadjuvant chemoradiotherapy for patients with esophageal cancer: a SEER-Medicare analysis. Cancer. 2014; 120(4):492–8. https://doi.org/10.1002/cncr.28447 PMID: 24194477

30. Eom BW, Ryu KW, Nam BH, Park Y, Lee HJ, Kim MC, et al. Survival nomogram for curatively resected Korean gastric cancer patients: multicenter retrospective analysis with external validation. PLoS ONE [Electronic Resource]. 2015; 10(2):e0119671–e.

31. Filip B, Scarpa M, Cavallin F, Cagol M, Alfieri R, Saadeh L, et al. Postoperative outcome after oesophagectomy for cancer: Nutritional status is the missing ring in the current prognostic scores. European Journal of Surgical Oncology. 2015; 41(6):787–94. https://doi.org/10.1016/j.ejso.2015.02.014 PMID: 25890494

32. Fischer C, Lingsma H, Hardwick R, Cromwell DA, Steyerberg E, Groene O. Risk adjustment models for short-term outcomes after surgical resection for oesophagogastric cancer. British Journal of Surgery. 2016; 103(1):105–16. https://doi.org/10.1002/bjs.9968 PMID: 26607783

33. Fuccio L, Scagliarini M, Frazzoni L, Battaglia G. Development of a prediction model of adverse events after stent placement for esophageal cancer. Gastrointestinal Endoscopy. 2016; 83(4):746–52. https://doi.org/10.1016/j.gie.2015.08.047 PMID: 26344881

34. Gabriel E, Attwood K, Shah R, Nurkin S, Hochwald S, Kukar M. Novel Calculator to Estimate Overall Survival Benefit from Neoadjuvant Chemoradiation in Patients with Esophageal Adenocarcinoma. Journal of the American College of Surgeons. 2017; 29:29.

35. Haga Y, Ikejiri K, Wada Y, Ikenaga M, Takeuchi H. Preliminary study of surgical audit for overall survival following gastric cancer resection. Gastric Cancer. 2015; 18(1):138–46. https://doi.org/10.1007/s10120-014-0343-5 PMID: 24500678

36. Han DS, Suh YS, Kong SH, Lee HJ, Choi Y, Aikou S, et al. Nomogram predicting long-term survival after d2 gastrectomy for gastric cancer. Journal of Clinical Oncology. 2012; 30(31):3834–40. https://doi.org/10.1200/JCO.2012.41.8343 PMID: 23008291

37. Hirabayashi S, Kosugi S, Isobe Y, Nashimoto A, Oda I, Hayashi K, et al. Development and external validation of a nomogram for overall survival after curative resection in serosa-negative, locally advanced gastric cancer. Annals of Oncology. 2014; 25(6):1179–84. https://doi.org/10.1093/annonc/mdu125 PMID: 24669009

38. Jiang Y, Zhang Q, Hu Y, Li T, Yu J, Zhao L, et al. ImmunoScore Signature: A Prognostic and Predictive Tool in Gastric Cancer. Annals of Surgery. 2016;20.

39. Jung HA, Adenis A, Lee J, Park SH, Maeng CH, Park S, et al. Nomogram to predict treatment outcome of fluoropyrimidine/platinum-based chemotherapy in metastatic esophageal squamous cell carcinoma. Cancer Research and Treatment. 2013; 45(4):285–94. https://doi.org/10.4143/crt.2013.45.4.285 PMID: 24454001

40. Kattan MW, Karpeh MS, Mazumdar M, Brennan MF. Postoperative nomogram for disease-specific survival after an R0 resection for gastric carcinoma. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2003; 21(19):3647–50. Epub 2003/09/27. https://doi.org/10.1200/jco.2003.01.240 PMID: 14512396.

41. Kim Y, Spolverato G, Ejaz A, Squires MH, Poultsides G, Fields RC, et al. A nomogram to predict overall survival and disease-free survival after curative resection of gastric adenocarcinoma. Annals of Surgical Oncology. 2015; 22(6):1828–35. https://doi.org/10.1245/s10434-014-4230-4 PMID: 25388061

42. Kunisaki C, Miyata H, Konno H, Saze Z, Hirahara N, Kikuchi H, et al. Modeling preoperative risk factors for potentially lethal morbidities using a nationwide Japanese web-based database of patients undergoing distal gastrectomy for gastric cancer. Gastric Cancer. 2016; 23:23.

43. Kurita N, Miyata H, Gotoh M, Shimada M, Imura S, Kimura W, et al. Risk Model for Distal Gastrectomy When Treating Gastric Cancer on the Basis of Data From 33,917 Japanese Patients Collected Using a Nationwide Web-based Data Entry System. Annals of Surgery. 2015; 262(2):295–303. https://doi.org/10.1097/SLA.0000000000001127 PMID: 25719804

44. Lagarde SM, Reitsma JB, de Castro SM, Ten Kate FJ, Busch OR, van Lanschot JJ. Prognostic nomogram for patients undergoing oesophagectomy for adenocarcinoma of the oesophagus or gastro-oesophageal junction. British Journal of Surgery. 2007; 94(11):1361–8. https://doi.org/10.1002/bjs.5832 PMID: 17582230

45. Lagarde SM, Reitsma JB, Maris AK, van Berge Henegouwen MI, Busch OR, Obertop H, et al. Preoperative prediction of the occurrence and severity of complications after esophagectomy for cancer with use of a nomogram. Annals of Thoracic Surgery. 2008; 85(6):1938–45. https://doi.org/10.1016/j.athoracsur.2008.03.014 PMID: 18498798

46. Lai JF, Kim S, Kim K, Li C, Oh SJ, Hyung WJ, et al. Prediction of recurrence of early gastric cancer after curative resection. Annals of Surgical Oncology. 2009; 16(7):1896–902. https://doi.org/10.1245/s10434-009-0473-x PMID: 19434457

47. Liu J, Geng Q, Chen S, Liu X, Kong P, Zhou Z, et al. Nomogram based on systemic inflammatory response markers predicting the survival of patients with resectable gastric cancer after D2 gastrectomy. Oncotarget. 2016; 7(25):37556–65. https://doi.org/10.18632/oncotarget.8788 PMID: 27121054

48. Liu J, Geng Q, Liu Z, Chen S, Guo J, Kong P, et al. Development and external validation of a prognostic nomogram for gastric cancer using the national cancer registry. Oncotarget. 2016; 7(24):35853–64. https://doi.org/10.18632/oncotarget.8221 PMID: 27016409

49. Liu JS, Huang Y, Yang X, Feng JF. A nomogram to predict prognostic values of various inflammatory biomarkers in patients with esophageal squamous cell carcinoma. American Journal of Cancer Research. 2015; 5(7):2180–9. PMID: 26328248

50. Marrelli D, De Stefano A, de Manzoni G, Morgagni P, Di Leo A, Roviello F. Prediction of recurrence after radical surgery for gastric cancer: a scoring system obtained from a prospective multicenter study. Annals of Surgery. 2005; 241(2):247–55. PMID: 15650634

51. Mohammadzadeh F, Noorkojuri H, Pourhoseingholi MA, Saadat S, Baghestani AR. Predicting the probability of mortality of gastric cancer patients using decision tree. Irish Journal of Medical Science. 2015; 184(2):277–84. https://doi.org/10.1007/s11845-014-1100-9 PMID: 24626962

52. Muneoka Y, Akazawa K, Ishikawa T, Ichikawa H, Nashimoto A, Yabusaki H, et al. Nomogram for 5-year relapse-free survival of a patient with advanced gastric cancer after surgery. International Journal Of Surgery. 2016; 35:153–9. https://doi.org/10.1016/j.ijsu.2016.09.080 PMID: 27664559

53. Shao Y, Ning Z, Chen J, Geng Y, Gu W, Huang J, et al. Prognostic nomogram integrated systemic inflammation score for patients with esophageal squamous cell carcinoma undergoing radical esophagectomy. Scientific Reports. 2015; 5:18811. https://doi.org/10.1038/srep18811 PMID: 26689680

54. Shapiro J, van Klaveren D, Lagarde SM, Toxopeus EL, van der Gaast A, Hulshof MC, et al. Prediction of survival in patients with oesophageal or junctional cancer receiving neoadjuvant chemoradiotherapy and surgery. British Journal of Surgery. 2016; 103(8):1039–47. https://doi.org/10.1002/bjs.10142 PMID: 27115731

55. Shiozaki H, Slack RS, Chen HC, Elimova E, Planjery V, Charalampakis N, et al. Metastatic Gastro-esophageal Adenocarcinoma Patients Treated with Systemic Therapy Followed by Consolidative Local Therapy: A Nomogram Associated with Long-Term Survivors. Oncology. 2016; 91(1):55–60. https://doi.org/10.1159/000445998 PMID: 27120436

56. Song KY, Park YG, Jeon HM, Park CH. A nomogram for predicting individual survival of patients with gastric cancer who underwent radical surgery with extended lymph node dissection. Gastric Cancer. 2014; 17(2):287–93. https://doi.org/10.1007/s10120-013-0270-x PMID: 23712439

57. Steyerberg EW, Neville BA, Koppert LB, Lemmens VE, Tilanus HW, Coebergh JW, et al. Surgical mortality in patients with esophageal cancer: development and validation of a simple risk score. Journal of Clinical Oncology. 2006; 24(26):4277–84. https://doi.org/10.1200/JCO.2005.05.0658 PMID: 16963730

58. Su D, Zhou X, Chen Q, Jiang Y, Yang X, Zheng W, et al. Prognostic Nomogram for Thoracic Esophageal Squamous Cell Carcinoma after Radical Esophagectomy. PLoS ONE [Electronic Resource]. 2015; 10(4):e0124437–e.

59. Suzuki A, Xiao L, Hayashi Y, Blum MA, Welsh JW, Lin SH, et al. Nomograms for prognostication of outcome in patients with esophageal and gastroesophageal carcinoma undergoing definitive chemoradiotherapy. Oncology. 2012; 82(2):108–13. https://doi.org/10.1159/000335951 PMID: 22328056

60. Tekkis PP, McCulloch P, Poloniecki JD, Prytherch DR, Kessaris N, Steger AC. Risk-adjusted prediction of operative mortality in oesophagogastric surgery with O-POSSUM. Br J Surg. 2004; 91(3):288–95. https://doi.org/10.1002/bjs.4414 PMID: 14991628.

61. Tu RH, Lin JX, Zheng CH, Li P, Xie JW, Wang JB, et al. Development of a nomogram for predicting the risk of anastomotic leakage after a gastrectomy for gastric cancer. European Journal of Surgical Oncology. 2017; 43(2):485–92. https://doi.org/10.1016/j.ejso.2016.11.022 PMID: 28041649

62. Woo Y, Son T, Song K, Okumura N, Hu Y, Cho GS, et al. A Novel Prediction Model of Prognosis After Gastrectomy for Gastric Carcinoma: Development and Validation Using Asian Databases. Annals of Surgery. 2016; 264(1):114–20. PMID: 26945155

63. Yang HX, Feng W, Wei JC, Zeng TS, Li ZD, Zhang LJ, et al. Support vector machine-based nomogram predicts postoperative distant metastasis for patients with oesophageal squamous cell carcinoma. British Journal of Cancer. 2013; 109(5):1109–16. https://doi.org/10.1038/bjc.2013.379 PMID: 23942069

64. Yu S, Zhang W, Ni W, Xiao Z, Wang X, Zhou Z, et al. Nomogram and recursive partitioning analysis to predict overall survival in patients with stage IIB-III thoracic esophageal squamous cell carcinoma after esophagectomy. Oncotarget. 2016; 7(34):55211–21. https://doi.org/10.18632/oncotarget.10904 PMID: 27487146

65. Zhao LY, Chen XL, Wang YG, Xin Y, Zhang WH, Wang YS, et al. A new predictive model combined of tumor size, lymph nodes count and lymphovascular invasion for survival prognosis in patients with lymph node-negative gastric cancer. Oncotarget. 2016; 7(44):72300–10. https://doi.org/10.18632/oncotarget.11035 PMID: 27509175

66. Zhou Z, Zhang H, Xu Z, Li W, Dang C, Song Y. Nomogram predicted survival of patients with adenocarcinoma of esophagogastric junction. World Journal of Surgical Oncology. 2015; 13:197. https://doi.org/10.1186/s12957-015-0613-7 PMID: 26055624

67. Ashfaq A, Kidwell JT, McGhan LJ, Dueck AC, Pockaj BA, Gray RJ, et al. Validation of a gastric cancer nomogram using a cancer registry. Journal of Surgical Oncology. 2015; 112(4):377–80. https://doi.org/10.1002/jso.23999 PMID: 26271201

68. Bosch DJ, Pultrum BB, de Bock GH, Oosterhuis JK, Rodgers MG, Plukker JT. Comparison of different risk-adjustment models in assessing short-term surgical outcome after transthoracic esophagectomy in patients with esophageal cancer. Am J Surg. 2011; 202(3):303–9. https://doi.org/10.1016/j.amjsurg.2011.04.003 PMID: 21871985.

69. Chen D, Jiang B, Xing J, Liu M, Cui M, Liu Y, et al. Validation of the memorial Sloan-Kettering Cancer Center nomogram to predict disease-specific survival after R0 resection in a Chinese gastric cancer population. PLoS ONE [Electronic Resource]. 2013; 8(10):e76041–e.

70. D'Journo XB, Berbis J, Jougon J, Brichon PY, Mouroux J, Tiffet O, et al. External validation of a risk score in the prediction of the mortality after esophagectomy for cancer. Diseases of the Esophagus. 2016; 3:03.

71. Dikken JL, Coit DG, Baser RE, Gonen M, Goodman KA, Brennan MF, et al. Performance of a nomogram predicting disease-specific survival after an R0 resection for gastric cancer in patients receiving postoperative chemoradiation therapy. International Journal of Radiation Oncology, Biology, Physics. 2014; 88(3):624–9. https://doi.org/10.1016/j.ijrobp.2013.11.213 PMID: 24411620

72. Grotenhuis BA, Van Hagen P, Reitsma JB, Lagarde SM, Wijnhoven BPL, Van Berge Henegouwen MI, et al. Validation of a nomogram predicting complications after esophagectomy for cancer. Annals of Thoracic Surgery. 2010; 90(3):920–5. https://doi.org/10.1016/j.athoracsur.2010.06.024 PMID: 20732518

73. Kim JH, Kim HS, Seo WY, Nam CM, Kim KY, Jeung HC, et al. External validation of nomogram for the prediction of recurrence after curative resection in early gastric cancer. Annals of Oncology. 2012; 23(2):361–7. https://doi.org/10.1093/annonc/mdr118 PMID: 21566150

74. Lagarde SM, Maris AK, de Castro SM, Busch OR, Obertop H, van Lanschot JJ. Evaluation of O-POSSUM in predicting in-hospital mortality after resection for oesophageal cancer. Br J Surg. 2007; 94(12):1521–6. https://doi.org/10.1002/bjs.5850 PMID: 17929231

75. Lagarde SM, Reitsma JB, Ten Kate FJ, Busch OR, Obertop H, Zwinderman AH, et al. Predicting individual survival after potentially curative esophagectomy for adenocarcinoma of the esophagus or gastroesophageal junction. Annals of Surgery. 2008; 248(6):1006–13. https://doi.org/10.1097/SLA.0b013e318190a0a2 PMID: 19092345

76. Marrelli D, Morgagni P, de Manzoni G, Marchet A, Baiocchi GL, Giacopuzzi S, et al. External Validation of a Score Predictive of Recurrence after Radical Surgery for Non-Cardia Gastric Cancer: Results of a

Follow-Up Study. Journal of the American College of Surgeons. 2015; 221(2):280–90. https://doi.org/10.1016/j.jamcollsurg.2015.03.042 PMID: 26141465

77. Nagabhushan JS, Srinath S, Weir F, Angerson WJ, Sugden BA, Morran CG. Comparison of P-POSSUM and O-POSSUM in predicting mortality after oesophagogastric resections. Postgrad Med J. 2007; 83(979):355–8. https://doi.org/10.1136/pgmj.2006.053223 PMID: 17488869.

78. Novotny AR, Schuhmacher C, Busch R, Kattan MW, Brennan MF, Siewert JR. Predicting individual survival after gastric cancer resection: validation of a U.S.-derived nomogram at a single high-volume center in Europe. Annals of Surgery. 2006; 243(1):74–81. PMID: 16371739

79. Peeters KCMJ, Kattan MW, Hartgrink HH, Kranenbarg EK, Karpeh MS, Brennan MF, et al. Validation of a nomogram for predicting disease-specific survival after an RO resection for gastric carcinoma. Cancer. 2005; 103(4):702–7. https://doi.org/10.1002/cncr.20783 PMID: 15641033

80. Reim D, Novotny A, Eom BW, Park Y, Yoon HM, Choi IJ, et al. External Validation of an Eastern Asian Nomogram for Survival Prediction After Gastric Cancer Surgery in a European Patient Cohort. Medicine. 2015; 94(52):e2406–e. https://doi.org/10.1097/MD.0000000000002406 PMID: 26717397

81. Zafirellis KD, Fountoulakis A, Dolan K, Dexter SP, Martin IG, Sue-Ling HM. Evaluation of POSSUM in patients with oesophageal cancer undergoing resection. Br J Surg. 2002; 89(9):1150–5. https://doi.org/10.1046/j.1365-2168.2002.02179.x PMID: 12190681.

82. Zhou ML, Wang L, Wang JZ, Yang W, Hu R, Li GC, et al. Validation of the Memorial Sloan Kettering Cancer Center nomogram to predict disease-specific survival in a Chinese gastric cancer population receiving postoperative chemoradiotherapy after an R0 resection. Oncotarget. 2016; 7(40):64757–65. https://doi.org/10.18632/oncotarget.11665 PMID: 27588465

83. Kattan MW, Karpeh MS, Mazumdar M, Brennan MF. Postoperative nomogram for disease-specific survival after an R0 resection for gastric carcinoma. Journal of Clinical Oncology. 2003; 21(19):3647–50. https://doi.org/10.1200/JCO.2003.01.240 PMID: 14512396