




# Prediction of Morbidity and Mortality After Esophagectomy: A Systematic Review

M. P. van Nieuw Amerongen, RN, MSc<sup>1</sup> , H. J. de Grooth, MD, PhD<sup>1</sup>, G. L. Veerman, RN<sup>1</sup>, K. A. Ziesemer, MSc<sup>2</sup>, M. I. van Berge Henegouwen, MD, PhD<sup>3,4</sup>, and P. R. Tuinman, MD, PhD<sup>1,5</sup>

<sup>1</sup>Department of Adult Intensive Care Medicine, Amsterdam UMC (VUmc), Amsterdam, The Netherlands; <sup>2</sup>Medical Library, Vrije Universiteit, Amsterdam, The Netherlands; <sup>3</sup>Department of surgery, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands; <sup>4</sup>Cancer Center Amsterdam, Cancer Treatment and Quality of Life, Amsterdam, The Netherlands; <sup>5</sup>Amsterdam Institute for Immunology and Infectious Diseases, Amsterdam, The Netherlands

## ABSTRACT

**Background.** Esophagectomy for esophageal cancer has a complication rate of up to 60%. Prediction models could be helpful to preoperatively estimate which patients are at increased risk of morbidity and mortality. The objective of this study was to determine the best prediction models for morbidity and mortality after esophagectomy and to identify commonalities among the models.

**Patients and Methods.** A systematic review was performed in accordance to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement and was prospectively registered in PROSPERO (<https://www.crd.york.ac.uk/prospero/>, study ID CRD42022350846). PubMed, Embase, and Clarivate Analytics/Web of Science Core Collection were searched for studies published between 2010 and August 2022. The Prediction model Risk of Bias Assessment Tool was used to assess the risk of bias. Extracted data were tabulated and a narrative synthesis was performed.

**Results.** Of the 15,011 articles identified, 22 studies were included using data from tens of thousands of patients. This systematic review included 33 different models, of which 18 models were newly developed. Many studies showed a high risk of bias. The prognostic accuracy of models differed between 0.51 and 0.85. For most models, variables are readily available. Two models for mortality and one model for

pulmonary complications have the potential to be developed further.

**Conclusions.** The availability of rigorous prediction models is limited. Several models are promising but need to be further developed. Some models provide information about risk factors for the development of complications. Performance status is a potential modifiable risk factor. None are ready for clinical implementation.

**Keywords** Esophageal cancer · Esophagectomy · Prediction models · Mortality · Morbidity

Esophageal cancer is the sixth leading cause of cancer death worldwide.<sup>1</sup> Treatment for patients with esophageal malignancies generally consists of neoadjuvant chemoradiation followed by esophagectomy, a high-risk procedure with a complication rate of up to 60%.<sup>2–5</sup> These postoperative complications are associated with significant morbidity, mortality, and health economic effects. In addition to worse patient outcomes, healthcare costs for a complicated course can be 2.5 times higher than that for an uncomplicated course and 5% of patients are responsible for about 20% of total hospital costs.<sup>6–11</sup>

Early identification of patients at high risk of severe complications has three potentially important healthcare benefits. First, patients at high risk of complications or death can be better informed about potential adverse consequences of surgery, which may lead to alternative treatment strategies. Second, potential preventative measures can be tailored to the risk profile by influencing potentially modifiable risk factors. Third, patients at the highest risk levels can be monitored more closely (for example, using remote monitoring

© The Author(s) 2024

First Received: 11 November 2023

Accepted: 18 January 2024

Published online: 21 February 2024

M. P. van Nieuw Amerongen, RN, MSc  
e-mail: m.vannieuwamerongen2@amsterdamumc.nl

or high-care ward admission) for early detection and treatment of complications—interventions that might not be cost-effective for the whole population.

A large number of preoperative prediction models have been developed in recent years on morbidity and mortality after esophagectomy. These models could potentially be helpful in identifying high-risk patients, but their usefulness has not yet been assessed systematically. In many research areas, the number of developed prediction models far outpaces their implementation rate, and novelty often takes precedence over validity, robustness, and usefulness.<sup>12–16</sup>

The primary aim of this study was to evaluate which of the existing prediction models are most suitable for potential widespread implementation. To evaluate the potential usefulness and readiness for clinical practice, we integrated results of models' predictive performance with methodological quality assessment and availability of the input variables.

The secondary aim was identification of commonalities among the best-performing models, in which the focus was on models predicting mortality and pulmonary complications.

## PATIENTS AND METHODS

### *Study Design*

This is a systematic review. The conduct and reporting of this review adhere to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement ([www.prisma-statement.org](http://www.prisma-statement.org)) and was prospectively registered in PROSPERO (<https://www.crd.york.ac.uk/prosp/ero/>, study ID CRD42022350846).<sup>17</sup>

### *Literature Search Strategy*

Three bibliographic databases, PubMed, Embase.com, and Clarivate Analytics/Web of Science Core, were searched for relevant literature from inception to 25 August 2022. Searches were devised in collaboration with a medical information specialist (KAZ). Search terms, including synonyms, closely related words, and keywords, were used as index terms or free-text words: 'esophagectomy' and 'prediction'. No methodological search filters, date, or language restrictions were applied that would limit results.

ASReview (version 1.0) was used to rank potentially relevant titles and abstracts. Screening in ASReview was carried out independently by two reviewers (MPvNA and GLV). All references marked as relevant were manually screened for eligibility by both reviewers. If necessary, the full-text article was checked for the eligibility criteria. Differences in judgement were resolved through a consensus procedure. If no consensus was reached, a third reviewer was consulted (PRT).

The full search strategy is detailed in Supplementary Material 1.

### *Eligibility Criteria*

Studies were included in which prognostic models/scales/indexes were developed and/or validated with respect to the preoperative prediction of morbidity (Clavien–Dindo score of at least 3) and/or mortality within 90 days after esophagectomy owing to esophageal cancer (regardless of histology or surgery type).<sup>18</sup> All types of prediction modeling studies were included. We excluded articles written before 2010. To assess models that can be used today, it is desirable that the study population match the current patient population as much as possible. Models consisting of one type of variable (such as blood markers, nutritional status, or cardiopulmonary exercise testing) were excluded. Articles that only examined association and/or correlation between the score of a model and morbidity/mortality were excluded.

To compare the accuracy of models, only models that examined accuracy and reported an outcome measure, such as area under the receiver operating characteristic curve (AUC) or observed/expected ratio (O/E ratio), were included. For more details about the inclusion and exclusion criteria, see Supplementary Material 2. Outcome definitions were described in Supplementary Material 3.

### *Assessment of Methodological Quality*

Two reviewers (MPvNA and GLV) independently assessed methodological quality of full-text papers using the Prediction model Risk Of Bias ASsessment Tool (PROBAST).<sup>19</sup> This tool, especially designed for systematic reviews of prediction models, assesses the risk of bias in four domains (participants, predictors, outcome, and statistical analysis) and addresses the concerns of applicability in three domains (participants, predictors, and outcome). A domain was assessed as low risk when all signaling questions were answered yes or probably yes. A domain was assessed as high risk when at least one signaling question in that domain was answered no or probably no. Overall risk of bias was assessed as low when all domains were considered low risk. Overall risk of bias was assessed as high when at least one domain was considered high risk. For domain one, participants, applicability is scored as unclear if it is unclear how many patients received neoadjuvant chemoradiation or if less than half the patients received neoadjuvant therapy.

When multiple models were developed and/or validated in a single study, a separate PROBAST form for each model or for both development and validation was needed. However, if results were completely similar, then this is reflected as one result in the Supplementary Material.

### Data Extraction

Data extraction of the identified studies was performed using the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) checklist (MPvNA).<sup>20</sup> Extracted data consisted of study characteristics (first author, country, study type, pretreatment, surgery type, and cohort years), study outcomes (outcome, number of events/sample size, outcome measures used regarding discrimination, and calibration), and the variables used in the different models.

### Data Synthesis and Statistical Analysis

The results were tabulated and a narrative synthesis was performed.

For prognostic accuracy (e.g., discrimination) the AUC value was often used. An AUC value under 0.60 was rated as poor, a value between 0.60 and 0.75 as possibly helpful discrimination, and more than 0.75 as useful discrimination.<sup>21</sup> For calibration, the Hosmer–Lemeshow test was often used. Small *p*-values mean that the model has poor calibration.

To assess potential usefulness and readiness for clinical practice, we integrated the results of the methodological quality assessment with the predictive performance of the models (lower limit confidence interval AUC), presence of external validation, sample size, and availability of the input variables. We rated the models ranked high in the tables as the better models.

Commonalities of the models (predictor variables) were presented in a figure and quality assessments were transformed into figures using Rstudio, version 4.2.1.

## RESULTS

### Systematic Search

Details regarding the literature search are shown in Fig. 1. Of the original 15,011 references identified by the search, 22 articles met the inclusion criteria and were included in this systematic review, using data from 108,208 patients.

### Study Characteristics

Study characteristics are presented in Table 1. Seven studies were conducted in Asia, seven in Europe, six studies in the USA, and two studies were intercontinental studies. In general, data were retrospectively collected from an existing database. Of the 22 studies, 12 collected their data entirely from 2010 onward, and the remaining studies partly after 2010. Eleven studies included a population in which at least

half of the patients were pretreated with chemotherapy, radiotherapy, chemoradiotherapy, or immunochemotherapy.

Most articles described the development of one or more new models or described the validation of one or more existing models. We assessed 39 models, including 8 models in the study by Ohkura et al.<sup>22</sup> From this study, only the two most relevant models were used in our analysis (anastomotic leakage and pneumonia). Finally, 33 models were included in this systematic review, of which 18 models were newly developed.

### Quality Assessment

The overall risk of bias (ROB) and overall concerns regarding applicability are presented in Table 2. For more details about ROB and concerns regarding applicability, see Supplementary Material 4.

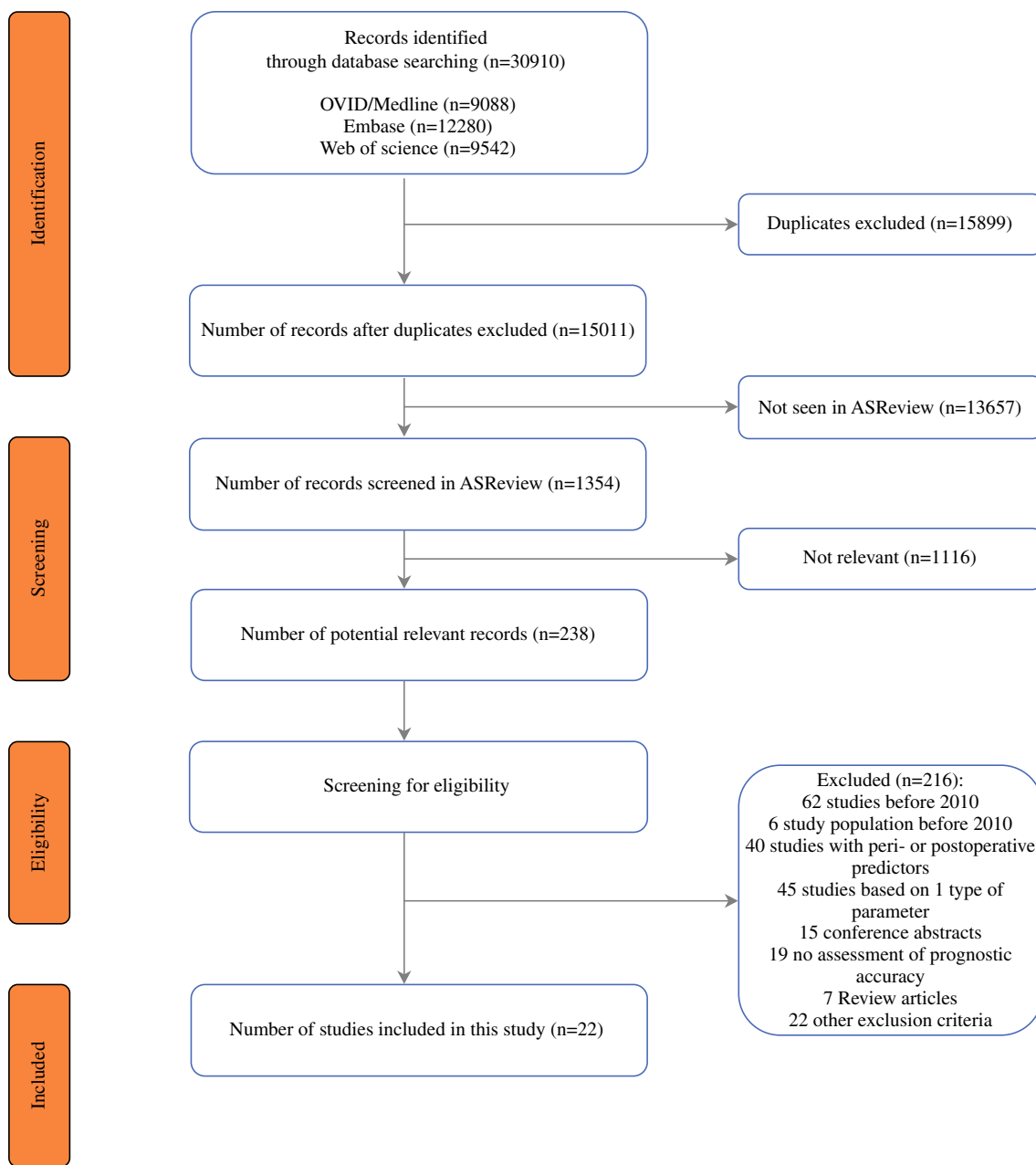
#### Risk of bias:

- **Participants:** For the development and validation of the models, retrospective cohort and/or existing (national) databases were generally used. Most studies showed a low risk of bias at this point.
- **Predictors:** There does not appear to be a risk of bias in any of the studies.
- **Outcome:** A few studies had unclear risk of bias owing to unclear definitions of outcome.
- **Analysis:** Most studies had a high risk of bias, in developmental studies mainly owing to an insufficient event-to-variable rate (candidate variable), and in validation studies mainly owing to an insufficient event rate (less than 100 events). Development studies on mortality have been carried out in large populations. This is in contrast to some development studies related to pulmonary complications. With the exception of the validation study by D'Journo et al., the studies validating only preexisting models were carried out in small populations.<sup>23</sup>

Other common causes of risk of bias in development studies were lack of relevant model performance measures or lack of correction for overfitting or optimism.

A model may be overfit when it makes good predictions on the study sample (owing to certain typical factors in the study population) but poor predictions outside of the study sample. This can be corrected through techniques such as bootstrapping or cross-validation.

**Concerns regarding applicability:** Concerns regarding applicability were in general related to no information whether neoadjuvant chemoradiation was given or because there were relatively few people in the population that underwent neoadjuvant treatment with chemoradiation.



**FIG. 1** Flowchart of the study search and selection procedure

### *Discrimination, Calibration, and Validation*

Study outcomes are presented in Table 2. More detailed information can be found in Supplementary Material 5.

#### *Discrimination*

Discrimination of models for mortality: Only Takeuchi's models had a prognostic accuracy of about or greater than 0.75; Sasaki's models were just below that.<sup>24,25</sup> The remaining models had accuracies between 0.60 and 0.75.

Discrimination of models for morbidity and/or mortality: Only Saito's model, the Padua model, the aCCI validated by Filip, and the O-POSSUM model validated by Fodor found prognostic accuracies between 0.74 and 0.80.<sup>26–29</sup> Other models had poor performance, including the aCCI validated in studies other than Filip's.<sup>27,30,31</sup>

Discrimination of models for pulmonary complications: The development of Wang's model, the validation of Thomas' model, and the Ferguson model found accuracies above 0.75. All were done using small sample sizes.<sup>32–34</sup> The remaining models found accuracies between 0.60 and 0.75.

**TABLE 1** Study characteristics

Study ID	Country/cohort years	Pretreatment (%)	Surgery type	Model: development and/or validation; sample size ( <i>n</i> )	Validation of
<i>Mortality</i>					
D'Journo et al., 2021 <sup>36</sup>	19 countries world-wide 2015–2019	CRTx: 47, CTx: 29, RTx: 0.1	MIE/open, TT/TH	Dev: 4172, Ext val: 4231	Dev model
Takeuchi et al., 2014 <sup>24</sup>	Japan 2011	CTx: 19, RTx: 5	total/hybrid MIE/OE, TT/TH, partial unknown	Dev: 4261, Ext val: 1093	Dev models
Sasaki et al., 2023 <sup>25</sup>	Japan 2012–2017	RTx: 7, CTx: Unclear	TT	Dev: 29501, Ext val: 3278	Dev models
Fischer et al., 2016 <sup>51</sup>	UK 2011–2013	?	?	Dev: 4882, Int val: 800 bootstrap-samples	Dev models
Fuchs et al., 2017 <sup>52</sup>	USA 1998–2011	?	MIE/open, TT/TH	Dev: 23751	NA
D'Journo et al., 2017 <sup>23</sup>	France 2004–2013	CTx: 17, RTx: 0.4, CRTx: 23	TH/TT/TA	Val: 1039	Steyerberg
Raymond et al., 2016 <sup>53</sup>	USA 2012–2014	CTx or RTx: 68	MIE/open, TT/TH/TA	Rev STS GTSD model: 3942	NA
Wan et al., 2022 <sup>44</sup>	USA 2006–2017	?	?	Val: 10602	RAI-rev, RAI-A, mFI-5, Rev of RAI-rev (cancer corrected)
Fodor et al., 2015 <sup>29</sup>	Romania 2011–2014	?	?	Val: 55	O-POSSUM, ASA
<i>Morbidity or both morbidity and mortality</i>					
Filip et al., 2015 <sup>27</sup>	Italy 2008–2012	CRTx: 78	MIE, TT/TH	Dev new model and val existing models: 167/500 bootstrap-samples	aCCI, CCI, O-POSSUM, ASA, Lagarde, new developed Padua model
Raymond et al., 2016 <sup>53</sup>	USA 2012–2014	CTx or RTx: 68	MIE/open, TT/TH/TA	Rev STS GTSD model: 3942	NA
Wan et al., 2022 <sup>44</sup>	USA 2006–2017	?	?	Val: 10602	RAI-rev, RAI-A, mFI-5, Rev of RAI-rev (cancer corrected)
Saito et al., 2019 <sup>26</sup>	Japan 2007–2015	CTx: 32	MIE, TT	Dev: 90	NA
Scarpa et al., 2016 <sup>30</sup>	Italy 2008–2012	CRTx: 79	MIE, TT	Val: 181	aCCI, CCI, ASA
Filip et al., 2014 <sup>28</sup>	Romania 2004–2013	CRTx: 51	TH/TT	Val: 43	aCCI, CCI, O-POSSUM
Mora et al., 2021 <sup>31</sup>	Japan 2010–2016	CTx: 14, CRTx: 5	MIE/open, TT (3FLD, McKeown)	Val: 230	aCCI, CCI, O-POSSUM, Steyerberg
Gray et al., 2023 <sup>54</sup>	USA 2016–2018	?	MIE, TT	Val: 240	ACS NSQIP calculator, mFI-5
Ravindran et al., 2020 <sup>55</sup>	USA 2013–2017	Neoadjuvant: 90	TT	Val: 100	ACS NSQIP calculator
<i>Anastomotic leakage</i>					
Fischer et al., 2016 <sup>51</sup>	UK 2011–2013	?	?	Dev: 4882, Int val: 800 bootstrap-samples	Dev models
Van Kooten et al., 2022 <sup>35</sup>	The Netherlands 2011–2017	CTx: 7%, CRTx: 86%	TT/TH	Dev: 3171, Ext val: 1057	Dev models
Ohkura et al., 2019 <sup>22</sup>	Japan 2011–2012	?	?	Dev: 8715, Ext val: 2147	Dev models

**Table 1** (continued)

Study ID	Country/cohort years	Pretreatment (%)	Surgery type	Model: development and/or validation; sample size ( <i>n</i> )	Validation of
<i>Pulmonary complications</i>					
Thomas et al., 2019 <sup>33</sup>	Belgium & USA 2002–2017	CRTx: 100	MIE/open, TT	Dev: 601, Ext val: 90	Dev model
Van Kooten et al., 2022 <sup>35</sup>	The Netherlands 2011–2017	CTx: 7%, CRTx: 86%	TT/TH	Dev: 3171, Ext val: 1057	Dev models
Ohkura et al., 2019 <sup>22</sup>	Japan 2011–2012	?	?	Dev: 8715, Ext val: 2147	Dev models
Kanda et al., 2019 <sup>56</sup>	Japan 2005–2017	CTx: 52	MIE/open	Dev: 355	NA
Wang et al., 2022 <sup>32</sup>	China 2019–2021	ICTx: 100	MIE/open, TT	Dev: 78	NA
Reinersman et al., 2016 <sup>34</sup>	USA 2009–2012	CRTx: 80	Total/hybrid MIE, TH/TT	Val: 136	Ferguson

*Dev* development, *Val* validation, *Int* internal, *Ext* external, *NA* not applicable, *CTx* chemotherapy, *CRTx* chemoradiotherapy, *ICTx* immunochemotherapy, *MIE* minimal invasive esophagectomy, *OE* open esophagectomy, *TT* transthoracic, *TH* transhiatal, *TA* thoracoabdominal, *RAI* risk analysis index, *Rev* revised, *RAI-A* administrative risk analysis index, *mFI-5* 5-factor modified frailty index, *STS* *GTSD* Society of Thoracic Surgeons General Thoracic Surgeons Database, *aCCI* age adjusted Charlson comorbidity index, *CCI* Charlson Comorbidity Index, *ASA* American Society of Anesthesiologists, *O-POSSUM* physiological and operative severity score for the enumeration of mortality and morbidity adjusted for oesophagogastric surgery, *ACS NSQIP* American College of Surgeons National Surgical Quality Improvement Program. ? unknown

Discrimination of models for anastomotic leakage: All three newly developed models found accuracies between 0.53 and 0.63.

#### Calibration

Calibration was reported most as a non-significant Hosmer–Lemeshow test<sup>6</sup> or a figure such as a scatterplot or calibration plot.<sup>5</sup> Eight studies did not report on calibration and one study indicated a favorable correlation between predicted and observed events, but the data were not shown.<sup>24</sup> For more details about calibration, see Supplementary Material 5.

#### Validation

None of the 18 newly developed models were validated by another research group, 14 models were validated by the author's own research group (in a new population or by bootstrapping), and 4 models were developed but not validated. Additionally, 11 existing models were validated one or more times. For more details, see Supplementary Material 6.

#### Predictor Variables

An overview of the different predictor variables is presented in Fig. 2A,B.

For the prediction models on mortality, 52 different variables were used with a median of 10 variables per model (range 4–20). Most studies used eight or more predictor variables. The easiest model to use is Steyerberg's model with just four easily available predictor variables.<sup>23</sup> The

predictor variables could be classified as patient characteristics, medical history, tumor and treatment, test results, or other. Age, sex, and performance status were the predictor variables most used regarding patient characteristics. ASA/comorbidity in general, (congestive) heart failure, and preoperative dialysis/renal dysfunction were the predictor variables most used regarding medical history. Histology, N-classification, and cancer metastasis/relapse were most used regarding tumor and treatment. PT-INR, white blood cell count, and sodium were most used regarding test results. Hospital volume was included in three models.

For the pulmonary complication prediction models, 43 different variables were used with a median of 5 predictors per model (range 2–28). Most models used less than five predictors. Exceptions were Van Kooten's model regarding pulmonary complications which had 28 predictor variables, and Ohkura's model, which had 17 predictor variables.<sup>22,35</sup> Again, variables could be classified as patient characteristics, medical history, tumor and treatment, or test results. Only age and histology were used in more than two models.

For most models, variables are easily available, making the models relatively easy to use (Table 2). Only the models of Ferguson (studied by Reinersman et al.) and Wang require a pulmonary function test, which is not performed routinely in all patients.<sup>32,34</sup>

#### Assessment of the Best Models

On the basis of these results, we could conclude that there are a number of models that have the potential to be

**TABLE 2** Summary of evaluation of prediction models for mortality and pulmonary complications

Study ID	Outcome	Events/sample size	AUC development	AUC validation	External validation	Calibration*	Availability variables	Overall ROB	Concerns applicability
Mortality									
D'Journo et al., 2021 <sup>36</sup>	90-day	4.2%: 353/8403	0.68 (0.64–0.72)	0.64 (0.60–0.69)	● (own study)	●	●	●	●
Takeuchi et al., 2014 <sup>24</sup>	30-day operative mortality	1.2%: 63/5354 3.4%: 181/5354	0.79 (0.73–0.86) 0.78 (0.74–0.81)	0.77 (0.65–0.88) 0.74 (0.67–0.82)	● (own study)	●	●	●●	●●
Sasaki et al., 2023 <sup>25</sup>	30-day operative	1.0%: 314/32779 2.3%: 770/32779	0.74 (0.71–0.77) 0.74 (0.73–0.76)	0.69 (0.60–0.78) 0.71 (0.65–0.77)	● (own study)	●	●	●●	●●
Fischer et al., 2016 <sup>51</sup>	30-day 90-day	2.3%: 112/4882 4.4%: 216/4882	0.698 0.694	0.646 0.664	●	●	●	●●	●●
Fuchs et al., 2017 <sup>52</sup>	In-hospital	7.7%: 1829/23751	Figure	NA	●	●	●	●	●
D'Journo et al., 2017 <sup>23</sup>	30-day 90-day In-hospital	5.7%: 59/1039 9.2%: 96/1039 9.6%: 100/1039	NA	0.64 (0.57–0.71) 0.63 (0.57–0.68) 0.63 (0.58–0.68)	●	●	●	●●●	●●●
Raymond et al., 2016 <sup>53</sup>	30 day or in-hospital	3.4%: 135/3942	0.71	NA	●	●	●	●	●
Wan et al., 2022 <sup>44</sup>	30-day	3.7%: 391/10602	NA	RAI-rev (cancer corr) 0.60 (0.57–0.63) RAI-rev 0.62 (0.59–0.64) RAIA 0.58 (0.55–0.61) mFI-5 0.58 (0.55–0.61)	●	●	●	●●●	●●●
Fodor et al., 2015 <sup>29</sup>	In-hospital	36.3%: 20/55	NA	O-POSSUM 0.74 ASA 0.64	●	●	●	●	●
Pulmonary complications									

developed further. For models that predict mortality, the most promising models are the models by D'Journo et al. and Takeuchi et al.<sup>24,36</sup> On the basis of quality assessment,

there is a risk of bias, but by weighing this against the other points of assessment (validation of a model in a sample separate from the development cohort, height of lower limit 95%

**Table 2** (continued)

Thomas et al., 2019 <sup>33</sup>	Pulmonary Complications	22.7%: 157/691	0.71 (0.67–0.73)	0.79 (0.64–0.88)	●	●	●	●	●
Van Kooten et al., 2022 <sup>35</sup>	Pulmonary Complications	32.6%: 1380/4228	0.64 (0.61–0.68)	No data	●	●	●	●	●
Ohkura et al., 2019 <sup>22</sup>	Pneumonia	14.6%: 1588/10862	No data	0.63 (0.60–0.67)	●	●	●	●	●
Kanda et al., 2019 <sup>56</sup>	Pneumonia	11.5%: 41/355	0.70	No data	●	●	●	●	●
Wang et al., 2022 <sup>32</sup>	Pneumonia	33.3%: 26/78	0.85 (0.75–0.95)	No data	●	●	●	●	●
Reinersman et al., 2016 <sup>34</sup>	Pulmonary complications	34.6%: 47/136	NA	0.76 (0.72–0.81)	●	●	●	●	●

\*The studies by Takeuchi et al. and Raymond et al. described that calibration had been done, but no data were shown

NA not available

confidence interval AUC, generalizability of the study, and sample size), these seem to be the better models. On the basis of the results, for models predicting pulmonary complications, the model by Thomas et al. is the best performing model.<sup>33</sup> For anastomotic leakage, a model with potential has yet to be developed (partly owing to the fact that all three models had an AUC lower than 0.64).

## DISCUSSION

The major findings of this systematic review assessing prediction models for complications after esophagectomy are that there are several models that are either promising to be further developed or provide us with the information about risk factors for the development of complications. Models with the most potential regarding prediction of mortality are the models by D'Journo and Takeuchi, while Thomas's model has the most potential regarding pulmonary complications. However, none of these three models have been validated by independent investigators yet.

Although it may be too early to implement complication prediction models in clinical practice, given the often relatively low AUC, the risk of bias, etc., the mortality models do at least provide us with relevant information about variables that influence the mortality risk. Common predictor

variables in mortality models include age, sex, performance status, ASA score/comorbidity in general, and cancer metastasis/relapse. Of these factors, possibly only performance status could be influenced prior to surgery. This could mean two things: either a poor performance status could be examined preoperatively to see whether it could be improved before the esophagectomy, or performance status, if it is not yet, could be given a role, as with the non-influenceable factors age, sex, comorbidity, and cancer metastasis/relapse, in the preoperative assessment as to whether a surgery is the best option for the patient. In conversation with the patient, however, we should remain cautious regarding statements about the severity of the risk of complications. While we do know that a number of single factors can affect risk, we do not know what the exact level of risk is when multiple risk factors are present.

It is notable that in models for pulmonary complications, there is no uniformity in which variables should be included in models. More research is needed for this.

When examining model variables, it is noteworthy that not all models incorporate factors known to be linked with mortality and morbidity, such as surgical technique and hospital volume.<sup>2,37–39</sup> A possible explanation is that surgical technique is not regarded as a preoperative variable, and hospital volume is not considered a patient-specific





### Strengths and Limitations

This study has several strengths. To our knowledge, this systematic review is the most recent and thorough systematic review on preoperative morbidity and mortality prediction models to date. Moreover, we registered our study at PROSPERO in advance. Previous systematic reviews were either written several years ago, cover prognosis of esophageal cancer in general, describe only mortality as outcome, are about long-term survival, or include models with perioperative rather than preoperative variables.<sup>46–50</sup>

One of the limitations of this study is that the pretreatment was not clearly stated, or the pretreatment turned out to be radiation, chemotherapy or immunochemotherapy, or chemoradiation was given in just a small proportion of patients.

In conclusion, the availability of rigorous prediction models is limited and none are ready for clinical implementation. Several models are promising but need to be further developed. In addition, some models provide us with the information regarding risk factors for the development of complications. Performance status is a potential modifiable risk factor when it comes to reducing risk of morbidity and mortality.

**SUPPLEMENTARY INFORMATION** The online version contains supplementary material available at <https://doi.org/10.1245/s10434-024-14997-4>.

**AUTHOR CONTRIBUTIONS** The authors have made substantial contributions to the design of the study (MPvNA, HJdG, MivBH, and PRT), data collection (MPvNA, KAZ, and GLV), analysis of data (MPvNA, HJdG, MivBH, and PRT) and drafting the article (MPvNA, HJdG, MivBH, and PRT).

**FUNDING** MivBH received an unrestricted research Grant from Stryker and consulted for Medtronic, Johnson & Johnson, BBraun, Mylan, and Alesi Surgical. All fees were paid to the institution.

**OPEN ACCESS** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

### REFERENCES

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209–49.
- Bras Harriott C, Angeramo CA, Casas MA, Schlottmann F. Open versus hybrid versus totally minimally invasive Ivor Lewis esophagectomy: systematic review and meta-analysis. *J Thorac Cardiovasc Surg.* 2022;164(6):e233–54.
- Zheng XD, Li SC, Lu C, Zhang WM, Hou JB, Shi KF, Zhang P. Safety and efficacy of minimally invasive McKeown esophagectomy in 1023 consecutive esophageal cancer patients: a single-center experience. *J Cardiothorac Surg.* 2022;17(1):36.
- DiSiena M, Perelman A, Birk J, Rezaizadeh H. Esophageal cancer: an updated review. *South Med J.* 2021;114(3):161–8.
- Voeten DM, den Bakker CM, Heineman DJ, Ket JCF, Daams F, van der Peet DL. Definitive chemoradiotherapy versus trimodality therapy for resectable oesophageal carcinoma: meta-analyses and systematic review of literature. *World J Surg.* 2019;43(5):1271–85.
- Goense L, van Dijk WA, Govaert JA, van Rossum PS, Ruurda JP, van Hillegersberg R. Hospital costs of complications after esophagectomy for cancer. *Eur J Surg Oncol.* 2017;43(4):696–702.
- Yanasoot A, Yolsuriyanwong K, Ruangsri S, Laohawiriyakamol S, Sunpaweravong S. Costs and benefits of different methods of esophagectomy for esophageal cancer. *Asian Cardiovasc Thorac Ann.* 2017;25(7–8):513–7.
- Hagens ERC, Reijntjes MA, Anderegg MCJ, Eshuis WJ, van Berge Henegouwen MI, Gisbertz SS. Risk factors and consequences of anastomotic leakage after esophagectomy for cancer. *Ann Thorac Surg.* 2021;112(1):255–63.
- Voeten DM, van der Werf LR, Gisbertz SS, Ruurda JP, van Berge Henegouwen MI, van Hillegersberg R, Dutch Upper Gastrointestinal Cancer Audit G. Postoperative intensive care unit stay after minimally invasive esophagectomy shows large hospital variation. Results from the Dutch Upper Gastrointestinal Cancer Audit. *Eur J Surg Oncol.* 2021;47(8):1961–8.
- Gooszen JAH, Eshuis WJ, Blom R, van Dieren S, Gisbertz SS, van Berge Henegouwen MI, Dutch Upper Gastrointestinal Cancer Audit G. The effect of preoperative body mass index on short-term outcome after esophagectomy for cancer: a nationwide propensity score-matched analysis. *Surgery.* 2022;172(1):137–44.
- van der Werf LR, Busweiler LAD, van Sandick JW, van Berge Henegouwen MI, Wijnhoven BPL, Dutch Upper GICAg. Reporting national outcomes after esophagectomy and gastrectomy according to the Esophageal Complications Consensus Group (ECCG). *Ann Surg.* 2020;271(6):1095–101.
- Van Calster B, Steyerberg EW, Wynants L, van Smeden M. There is no such thing as a validated prediction model. *BMC Med.* 2023;21(1):70.
- Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal–external, and external validation. *J Clin Epidemiol.* 2016;69:245–7.
- Guerra B, Gaveikaite V, Bianchi C, Puhan MA. Prediction models for exacerbations in patients with COPD. *Eur Respir Rev.* 2017;26(143):160061
- Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ.* 2020;369:1328.
- Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Topic Group “Evaluating diagnostic t, prediction models” of the Si. Calibration: the Achilles heel of predictive analytics. *BMC Med.* 2019;17(1):230.
- Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6(7):e1000097.

18. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg.* 2004;240(2):205–13.
19. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med.* 2019;170(1):W1–33.
20. 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.
21. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, et al. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. *JAMA.* 2017;318(14):1377–84.
22. Ohkura Y, Miyata H, Konno H, Udagawa H, Ueno M, Shindoh J, et al. Development of a model predicting the risk of eight major postoperative complications after esophagectomy based on 10826 cases in the Japan National Clinical Database. *J Surg Oncol.* 2019;121(2):313–21.
23. 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. *Dis Esophagus.* 2017;30(1):1–8.
24. Takeuchi H, Miyata H, Gotoh M, Kitagawa Y, Baba H, Kimura W, et al. A risk model for esophagectomy using data of 5354 patients included in a Japanese nationwide web-based database. *Ann Surg.* 2014;260(2):259–66.
25. Sasaki A, Tachimori H, Akiyama Y, Oshikiri T, Miyata H, Kakeji Y, Kitagawa Y. Risk model for mortality associated with esophagectomy via a thoracic approach based on data from the Japanese National Clinical Database on malignant esophageal tumors. *Surg Today.* 2023;53(1):73–81.
26. Saito T, Tanaka K, Ebihara Y, Kurashima Y, Murakami S, Shichinohe T, Hirano S. Novel prognostic score of postoperative complications after transthoracic minimally invasive esophagectomy for esophageal cancer: a retrospective cohort study of 90 consecutive patients. *Esophagus.* 2019;16(2):155–61.
27. 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 link in the current prognostic scores. *Eur J Surg Oncol.* 2015;41(6):787–94.
28. Filip B, Hutanu I, Radu I, Anitei MG, Scripcariu V. Assessment of different prognostic scores for early postoperative outcomes after esophagectomy. *Chirurgia (Bucur).* 2014;109(4):480–5.
29. Fodor R, Cioc A, Grigorescu B, Lazescu B, Copotioiu SM. Evaluation of O-POSSUM vs ASA and APACHE II scores in patients undergoing oesophageal surgery. *Rom J Anaesth Intensive Care.* 2015;22(1):7–12.
30. Scarpa M, Filip B, Cavallin F, Alfieri R, Saadeh L, Cagol M, Castoro C. Esophagectomy in elderly patients: which is the best prognostic score? *Dis Esophagus.* 2016;29(6):589–97.
31. Mora A, Nakajima Y, Okada T, Hoshino A, Tokairin Y, Kawada K, Kawano T. Morbidity after esophagectomy with three-field lymph node dissection in patients with esophageal cancer: looking for the best predictive model. *Int Surg.* 2021;105(1–3):402–10.
32. Wang W, Yu YK, Sun HB, Wang ZF, Zheng Y, Liang GH, et al. Predictive model of postoperative pneumonia after neoadjuvant immunochemotherapy for esophageal cancer. *J Gastrointest Oncol.* 2022;13(2):488–98.
33. Thomas M, Defraene G, Lambrecht M, Deng W, Moons J, Nafteux P, et al. NTCP model for postoperative complications and one-year mortality after trimodality treatment in esophageal cancer. *Radiother Oncol.* 2019;141:33–40.
34. Reinersman JM, Allen MS, Deschamps C, Ferguson MK, Nichols FC, Shen KR, et al. External validation of the Ferguson pulmonary risk score for predicting major pulmonary complications after oesophagectomy. *Eur J Cardiothorac Surg.* 2016;49(1):333–8.
35. van Kooten RT, Bahadoer RR, de Vries BT, Wouters MWJM, Tollenaar RAEM, Hartgrink HH, et al. Conventional regression analysis and machine learning in prediction of anastomotic leakage and pulmonary complications after esophagogastric cancer surgery. *J Surg Oncol.* 2022;126(3):490–501.
36. D'Journo XB, Boulate D, Fourdrain A. Risk prediction model of 90-day mortality after esophagectomy for cancer. *JAMA Surg.* 2021;156(9):894.
37. Baranov NS, Slootmans C, van Workum F, Klarenbeek BR, Schoon Y, Rosman C. Outcomes of curative esophageal cancer surgery in elderly: a meta-analysis. *World J Gastrointest Oncol.* 2021;13(2):131–46.
38. Rahouma M, Baudo M, Mynard N, Kamel M, Khan FM, Shmushkevich S, et al. Volume outcome relationship in post-esophagectomy leak: a systematic review and meta-analysis. *Int J Surg.* 2023. <https://doi.org/10.1097/JS9.0000000000000420>
39. Meng R, Bright T, Woodman RJ, Watson DI. Hospital volume versus outcome following oesophagectomy for cancer in Australia and New Zealand. *ANZ J Surg.* 2019;89(6):683–8.
40. Grantham JP, Hii A, Shenfine J. Preoperative risk modelling for oesophagectomy: a systematic review. *World J Gastrointest Surg.* 2023;15(3):450–70.
41. Minne L, Ludikhuizen J, de Jonge E, de Rooij S, Abu-Hanna A. Prognostic models for predicting mortality in elderly ICU patients: a systematic review. *Intensive Care Med.* 2011;37(8):1258–68.
42. Pavlou M, Qu C, Omar RZ, Seaman SR, Steyerberg EW, White IR, Ambler G. Estimation of required sample size for external validation of risk models for binary outcomes. *Stat Methods Med Res.* 2021;30(10):2187–206.
43. Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol.* 2005;58(5):475–83.
44. Wan MA, Clark JM, Nuno M, Cooke DT, Brown LM. Can the risk analysis index for frailty predict morbidity and mortality in patients undergoing high-risk surgery? *Ann Surg.* 2022;276(6):E721–7.
45. Siontis GCM, Tzoulaki J, Castaldi PJ, Ioannidis JPA. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol.* 2015;68(1):25–34.
46. Findlay JM, Gillies RS, Sgromo B, Marshall RE, Middleton MR, Maynard ND. Individual risk modelling for esophagectomy: a systematic review. *J Gastrointest Surg.* 2014;18(8):1532–42.
47. 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(2):e0192310.
48. Warnell I, Chincholkar M, Eccles M. Predicting perioperative mortality after oesophagectomy: a systematic review of performance and methods of multivariate models. *Br J Anaesth.* 2015;114(1):32–43.
49. Zheng C, Luo C, Xie K, Li JS, Zhou H, Hu LW, et al. Surgical Apgar score could predict complications after esophagectomy: a systematic review and meta-analysis. *Interact Cardiovasc Thorac Surg.* 2022;35(1):ivac045.
50. Boshier PR, Swaray A, Vadhwan B, O'Sullivan A, Low DE, Hanna GB, et al. Systematic review and validation of clinical models predicting survival after oesophagectomy for adenocarcinoma. *Br J Surg.* 2022;109(5):418–25.

51. 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. *Brit J Surg.* 2016;103(1):105–16.
52. Fuchs HF, Harnsberger CR, Broderick RC, Chang DC, Sandler BJ, Jacobsen GR, et al. Simple preoperative risk scale accurately predicts perioperative mortality following esophagectomy for malignancy. *Dis Esophagus.* 2017;30(1):1–6.
53. Raymond DP, Seder CW, Wright CD, Magee MJ, Kosinski AS, Cassivi SD, et al. Predictors of major morbidity or mortality after resection for esophageal cancer: a society of thoracic surgeons general thoracic surgery database risk adjustment model. *Ann Thorac Surg.* 2016;102(1):207–14.
54. Gray KD, Nobel TB, Hsu ME, Tan KS, Chudgar N, Yan S, et al. Improved preoperative risk assessment tools are needed to guide informed decision making before esophagectomy. *Ann Surg.* 2023;277(1):116–20.
55. Ravindran K, Escobar D, Gautam S, Puri R, Awad Z. Assessment of the American College of Surgeons National Surgical Quality Improvement Program Calculator in predicting outcomes and length of stay after Ivor Lewis esophagectomy: a single-center experience. *J Surg Res.* 2020;255:355–60.
56. Kanda M, Koike M, Tanaka C, Kobayashi D, Hayashi M, Yamada S, et al. Risk prediction of postoperative pneumonia after subtotal esophagectomy based on preoperative serum cholinesterase concentrations. *Ann Surg Oncol.* 2019;26(11):3718–26.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.