



Combined and intraoperative risk modelling for oesophagectomy: A systematic review

James Paul Grantham, Amanda Hii, Jonathan Shenfine

Specialty type: Gastroenterology and hepatology

Provenance and peer review: Unsolicited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): 0
Grade C (Good): C
Grade D (Fair): 0
Grade E (Poor): 0

P-Reviewer: Cerwenka H, Austria

Received: November 26, 2022

Peer-review started: November 26, 2022

First decision: February 20, 2023

Revised: March 13, 2023

Accepted: May 22, 2023

Article in press: May 22, 2023

Published online: July 27, 2023



James Paul Grantham, Amanda Hii, Department of General Surgery, Modbury Hospital, Modbury 5092, South Australia, Australia

Jonathan Shenfine, Department of General Surgical Unit, Jersey General Hospital, Saint Helier JE1 3QS, Jersey, United Kingdom

Corresponding author: James Paul Grantham, MBBS, MSc, Doctor, Department of General Surgery, Modbury Hospital, Smart Road, Modbury 5092, South Australia, Australia. jamespgrantham91@gmail.com

Abstract

BACKGROUND

Oesophageal cancer is the eighth most common malignancy worldwide and is associated with a poor prognosis. Oesophagectomy remains the best prospect for a cure if diagnosed in the early disease stages. However, the procedure is associated with significant morbidity and mortality and is undertaken only after careful consideration. Appropriate patient selection, counselling and resource allocation is essential. Numerous risk models have been devised to guide surgeons in making these decisions.

AIM

To evaluate which multivariate risk models, using intraoperative information with or without preoperative information, best predict perioperative oesophagectomy outcomes.

METHODS

A systematic review of the MEDLINE, EMBASE and Cochrane databases was undertaken from 2000-2020. The search terms used were [(Oesophagectomy) AND (Model OR Predict OR Risk OR score) AND (Mortality OR morbidity OR complications OR outcomes OR anastomotic leak OR length of stay)]. Articles were included if they assessed multivariate based tools incorporating preoperative and intraoperative variables to forecast patient outcomes after oesophagectomy. Articles were excluded if they only required preoperative or any post-operative data. Studies appraising univariate risk predictors such as preoperative sarcopenia, cardiopulmonary fitness and American Society of Anesthesiologists score were also excluded. The review was conducted following the preferred reporting items for systematic reviews and meta-analyses model. All captured risk models were appraised for clinical credibility, methodological quality, performance, validation and clinical effectiveness.

RESULTS

Twenty published studies were identified which examined eleven multivariate risk models. Eight of these combined preoperative and intraoperative data and the remaining three used only intraoperative values. Only two risk models were identified as promising in predicting mortality, namely the Portsmouth physiological and operative severity score for the enumeration of mortality and morbidity (POSSUM) and POSSUM scores. A further two studies, the intraoperative factors and Esophagectomy surgical Apgar score based nomograms, adequately forecasted major morbidity. The latter two models are yet to have external validation and none have been tested for clinical effectiveness.

CONCLUSION

Despite the presence of some promising models in forecasting perioperative oesophagectomy outcomes, there is more research required to externally validate these models and demonstrate clinical benefit with the adoption of these models guiding postoperative care and allocating resources.

Key Words: Oesophagectomy; Risk model; Oesophageal cancer; Preoperative; Intraoperative; Morbidity; Mortality

©The Author(s) 2023. Published by Baishideng Publishing Group Inc. All rights reserved.

Core Tip: Performing an oesophagectomy is a technically demanding procedure for the surgeon and a physiologically demanding undertaking for the patient. Aspects relating to the operation, as well as preoperative patient characteristics both have a significant impact on perioperative outcomes. These factors have been harnessed in the construction of numerous multivariate models aimed at identifying individuals at heightened risk. Given the plethora of options available, it is important to determine which of these models is most accurate in doing this and thereby most effective in guiding resource allocation.

Citation: Grantham JP, Hii A, Shenfine J. Combined and intraoperative risk modelling for oesophagectomy: A systematic review.

World J Gastrointest Surg 2023; 15(7): 1485-1500

URL: <https://www.wjgnet.com/1948-9366/full/v15/i7/1485.htm>

DOI: <https://dx.doi.org/10.4240/wjgs.v15.i7.1485>

INTRODUCTION

Oesophageal cancer is generally associated with a poor prognosis and exacts a substantial global burden with over half a million people diagnosed annually. If detected in the early stages, curative intent may be pursued through surgical intervention in the form of oesophagectomy. Despite advances in minimally invasive techniques, hybrid surgical approaches and the use of robotics, oesophagectomy remains associated with high rates of perioperative morbidity and mortality[1]. The operation is lengthy and entails at least two cavity surgery with extended periods of single-lung ventilation, placing a significant physiological burden on patients[2]. Respiratory complications such as atelectasis, pneumonia and acute respiratory distress syndrome are common, occurring in 20-40 percent of patients[3,4]. Adverse cardiovascular outcomes such as postoperative arrhythmias, myocardial infarction or heart failure develop in around a quarter of patients[5,6]. Surgical complications are also frequently encountered including anastomotic leaks, wound infections, bleeding, chyle leaks and conduit necrosis[7,8]. As a result, the perioperative mortality rate for oesophagectomy is accepted to run between 2%-8%[9]. Surgeons recognise the importance of selecting patients who can withstand the physiological strain of the operation and there is a benefit to identify early those patients at greater risk of complications. Decision-making tools such as surgical risk prediction calculators may help to predict these[10]. The application of predictive modelling to decision making has previously been demonstrated to be superior to clinical judgement alone[11,12]. A more strategic deployment of resources through the use of these tools may lead to improved patient outcomes.

In order for a model to be adopted widely, it would need to be simple enough to use yet sufficiently accurate to discriminate post-operative outcomes. Currently, there are multiple tools of varying complexity and ability to predict outcomes. The most accurate are those which utilise perioperative data in addition to pre-operative factors. There have been two systematic reviews to identify the most promising model for predicting perioperative outcomes following oesophagectomy. In 2014, Findlay *et al*[13] found that portsmouth physiological and operative severity score for the enumeration of mortality and morbidity (P-POSSUM) demonstrated the most promise for predicting perioperative mortality but that no existing model forecast morbidity with sufficient accuracy to be of clinical use[13]. A year later, Warnell *et al*[14] were unable to find any model that could be applied to clinical practice with confidence in either regard [14]. Although neither review discriminated between those models that utilised preoperative variables only and those which also incorporated intraoperative data. In addition, only Findlay *et al*[13] attempted to appraise the scientific rigor with which these models were developed. There have been a plethora of new multivariate risk models developed and validated for oesophagectomy subsequent to these reviews.

The purpose of this systematic review, therefore, is to conduct an analysis of multivariate risk prediction models that use both preoperative and intraoperative factors or exclusively intraoperative factors to determine which model most accurately predicts post-operative outcomes following oesophagectomy for cancer. The primary objective is the predictive capacity of each model in relation to perioperative mortality. The secondary objectives are the predictive capacities of these tools in respect of major morbidity, overall morbidity, respiratory complications and anastomotic leak. By identifying which model is most accurate, the results of this systematic review may allow surgeons to more appropriately allocate resources to patients selected for surgical resection but who are deemed at higher potential risk of a complication in the perioperative period and thereby improve treatment outcomes.

MATERIALS AND METHODS

Search strategy and article selection

A systematic review of MEDLINE, EMBASE and Cochrane review databases was undertaken. The search terms used were [(Oesophagectomy) AND (Model OR Predict OR Risk OR score) AND (Mortality OR morbidity OR complications OR outcomes OR anastomotic leak OR length of stay)]. The articles captured from the search were processed with reference to the preferred reporting items for systematic reviews and meta-analyses (PRISMA) model[15]. Initial deduplication was followed by preliminary screening of titles and abstract from the relevant publications. This was performed by the primary author with reference to the inclusion criteria. Texts that were considered potentially relevant were then assessed in full for eligibility from inclusion and exclusion criteria by two of the authors.

Inclusion and exclusion criteria

Articles which assessed multivariate based tools incorporating preoperative and intraoperative variables to forecast patient outcomes after oesophagectomy, published in English after the year 2000 were included. Articles that required only preoperative or any post-operative data and publications appraising univariate risk predictors such as preoperative sarcopenia, cardiopulmonary fitness and American Society of Anesthesiologists (ASA) score were excluded. The authors also excluded articles that considered only long-term outcomes, such as overall survival or disease-free survival, and publications which incorporated cohorts undergoing procedures other than oesophagectomy. Studies which reported insufficient data for appropriate analysis were excluded as well.

Data extraction and synthesis

Every publication meeting the inclusion and exclusion criteria was collected and study characteristics were extracted. These characteristics included the study period, sample size, geographical location, number of centres involved and case mix descriptors such as type of operation, proportions of neoadjuvant therapy use and histological subtype. The model or models appraised within each article and major performance metrics including discrimination and calibration were recorded. Outcome measures such as definitions of perioperative morbidity and mortality were also extracted. Oesophagectomy approach was categorised into transthoracic, transhiatal, hybrid or totally minimally invasive. Perioperative outcomes were classified into mortality, major morbidity (Defined as Clavien-Dindo grades 3 or 4) and overall morbidity[16]. Where reported, specific morbidity such as respiratory complications and anastomotic leakage were considered separately. Each model was analysed across the following five domains: Clinical credibility, methodological quality, external validation, model performance and clinical effectiveness.

Clinical credibility

Clinical credibility is the degree to which the unique model characteristics promote utilisation by a relevant clinician. This concept was first defined by Wyatt and Altman[17] and incorporates seven criteria[17-19]. These include whether the model used oesophageal specific factors and avoided harsh thresholds categorising data. Whether the data is available prior to decision making, whether the data is objective and if the outcome can be easily generated are also considered. Finally, the degree to which the outcome is understandable to the clinician and how effectively the model stratifies risk to clinically useful outcomes. A full account of the factors is provided within the [Supplementary material](#).

Methodological quality

The methodological quality of each model was considered with reference to the quality assessment framework developed by Minne *et al*[20,21]. This framework is designed to appraise risk of bias using a twenty point checklist: Eight points allocated to study participation characteristics; four points to measurements of prognostic factors and outcomes; and the final eight to the methodological integrity of analysis[13,21]. A point was awarded when a model satisfied each component and half a point when a component was partially satisfied. No points were awarded in the instance of a criterion not being met. The details of the appraisal method are outlined within the [Supplementary material](#).

External validation

Reviewed studies either reported the development of new models or an external validation of existing models within populations separate to the model development cohort. We reviewed each model to consider whether it had been externally validated.

Model performance

The performance for each model was also compared in terms of discrimination and calibration. Discrimination is the capacity of a model to discern whether a specific outcome will occur. The accuracy with which the model predicts outcomes is recorded in terms of the area under the receiver operating characteristics (ROC) curve, otherwise known as the c-statistic. If a model demonstrates no capacity for discrimination, the c-statistic will be 0.5 whereas a c-statistic of 1 connotes perfect discrimination[22]. In instances where there was adequate reporting to allow it, weighted discrimination metrics were generated for each model in the form of area under the ROC curve. Conventionally the accepted threshold for clinical utility is a c-statistic exceeding 0.7[23]. The alignment between the actual and predicted frequency of an outcome is known as calibration. This value can be represented in observed to expected (O:E) outcome ratios or Hosmer and Lemeshow[24] goodness of fit *P* values[24,25]. An O:E of 1 suggests perfect calibration[24]. Similarly, a goodness of fit *P* value of greater than 0.05 indicates appropriate calibration[25].

Clinical effectiveness

We also appraised whether there was any evidence that the clinical application of any of these models have been proven to improve perioperative outcomes.

RESULTS

Search results

A total of 8715 articles were initially identified, but subsequently reduced to 5827 following deduplication. Following title and abstract screening: 197 potentially relevant articles were retrieved. Detailed review of the articles determined that 20 satisfied the inclusion criteria without triggering any exclusion criteria. The exclusion rationale for the 177 articles omitted is illustrated (See [Figure 1](#)). Six articles presented the development of new predictive risk models for oesophagectomy which combined preoperative and intraoperative data or exclusively intraoperative data ([Table 1](#))[26-31]. The remaining 14 articles provided external validation of existing models using new data sets ([Table 2](#))[32-45]. The 20 articles reviewed assessed 11 different multivariate risk prediction models in oesophagectomy ([Table 3](#))[26-31,46-50]. Of these 11 models, eight used a combination and preoperative and intraoperative variables and three exclusively used intra-operative data. Two models were tested for mortality prediction alone, eight assessed morbidity and a single model, the physiological and operative severity score for the enumeration of mortality and morbidity (POSSUM) score, was tested for both mortality and morbidity.

Study characteristics

The studies included were published over a twenty-year period and from three different continents: 13 from Asia, six from Europe and a single study from North America. All 11 models were developed through multivariate logistical regression and based on retrospective data collected from patient cohorts. The six articles which developed a new model had a median population size of 243 (range 168-365). The fourteen articles validating existing models had a median population size of 249 (range 43-663). The studies possessed heterogeneity in operative approach and technique. Nineteen articles studied open oesophagectomy, all including a transthoracic procedure (Ivor-Lewis, McKeown or left thoracolumbar); six included transhiatal techniques; eight also included minimally invasive oesophagectomy; and six had patients undergoing hybrid oesophagectomy. One study did not incorporate patients undergoing open techniques with this relating to patients receiving a hybrid oesophagectomy.

In total, 17 of the 20 studies reported neoadjuvant therapy use, varying significantly from 3.6% to 80.3% of study populations. The total combined samples had a proportion of 37.9% patients that received neoadjuvant therapy. Ten of the studies reported the histological subtype of oesophageal cancer. This included five of the thirteen studies originating from Asia and five of the seven from Western nations. Overall, 64.1% of the patients had squamous cell carcinoma compared to 35.6% of those with adenocarcinoma. Just 0.3% had another histological tumour type altogether. Studies from Asia reported 95.6% of patients had squamous cell carcinoma whereas in North American and European studies, 72.4% of cases were adenocarcinoma.

Clinical credibility

Across the eleven prediction models, the median score for clinical credibility was 4.5 out of a possible maximum of 7 (range 3-5) ([Table 3](#)). The tool developed by Huang *et al*[31] top-scored with a median score of 5. Six of the eleven models in this group were specifically developed for oesophagectomy with oesophagogastric-POSSUM (O-POSSUM) also incorporating gastrectomy into its development set. The estimation of physiologic ability and surgical stress (E-PASS) was the only model to avoid using thresholds but was deemed difficult to generate, owing to its complexity. This complexity also contributed to it being the only model in this group for which marks were deducted for it being hard to understand. None of these models were appraised as providing timely data. Six of the eleven combined and intraoperative risk models incorporated estimated blood loss and were thus marked partially. Ten out of eleven generated useful scoring ranges for stratifying patients.

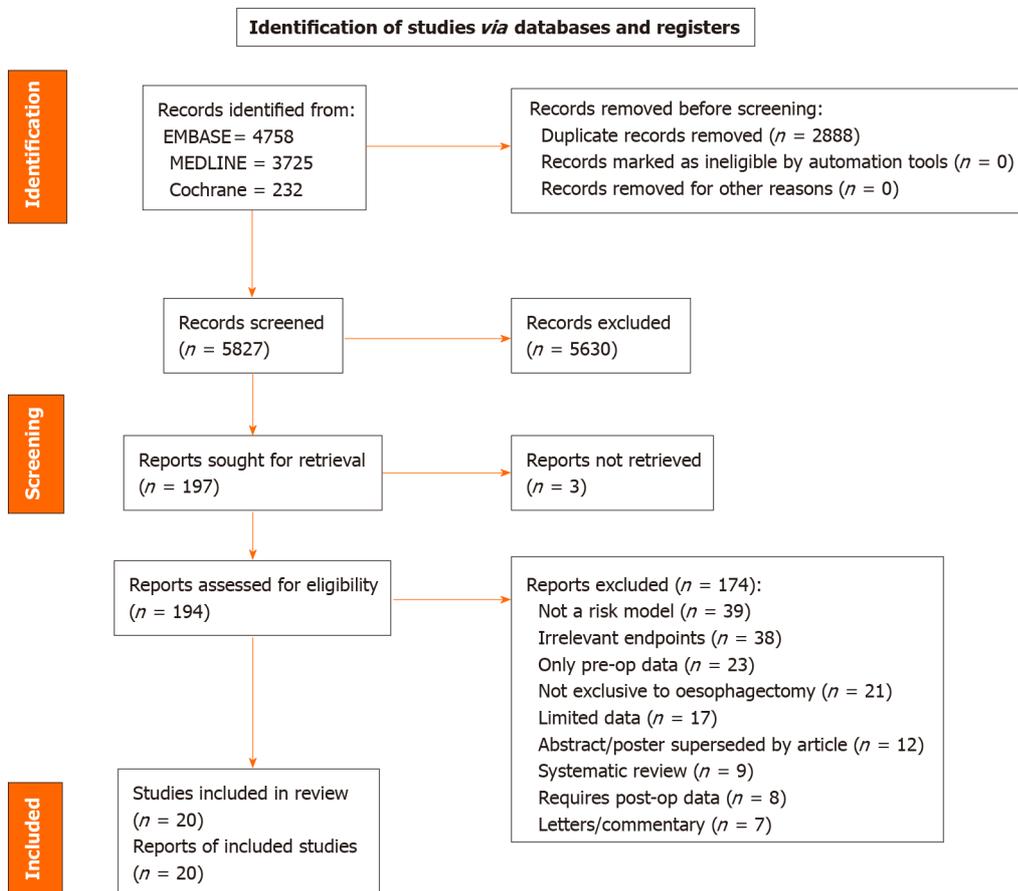
Methodological quality – study participation

Within the eleven combined and intraoperative models, the median score for study participation was 8 out of 8 (range 6-8). All of these eleven studies included the setting and period of the study. All reported the number of patients and

Table 1 Development studies of combined and intraoperative multivariate models

Ref.	Period + number	Sample region	Operation	Characteristics	Source of data	Models tested	Outcomes tested
Janowak <i>et al</i> [26], 2015	2009-2013 (168)	United States	TT/TH/Hybrid/MIE	N: 93/168; H: N/A	Single centre	eSAS	Major morbidity
Yoshida <i>et al</i> [28], 2015	2005-2013 (365)	Japan	TT	N: 113/250; H: N/A	Single centre	Yoshida score	Morbidity +; Major morbidity
Stroyer <i>et al</i> [27], 2017	2011-2014 (234)	Denmark	TT	N: 188/234; H: N/A	Single centre	Modified eSAS, SAS, eSAS	Major morbidity
Xi <i>et al</i> [30], 2019	2015-2018 (194)	China	TT/Hybrid/MIE	N: 48/194; H: N/A	Single centre	eSAS nomogram	Major morbidity
Xi <i>et al</i> [29], 2020	2015-2017 (251)	China	TT/Hybrid/MIE	N: 56/251; H: N/A	Single centre	IPF nomogram	Major morbidity
Huang <i>et al</i> [31], 2020	2016-2018 (330)	China	TT/Hybrid/MIE	N: 41/330; H: 32ISCC/9AC	Single centre	Huang nomogram	Anastomotic leak

TT: Transthoracic; TH: Transhiatal; MIE: Minimally invasive esophagectomy; N: Neoadjuvant status; H: Histopathology; N/A: Not applicable; SCC: Squamous cell carcinoma; AC: Adenocarcinoma; eSAS: Esophagectomy surgical Apgar score; SAS: Surgical Apgar score; IPF: Intraoperative factors.



DOI: 10.4240/wjgs.v15.i7.1485 Copyright ©The Author(s) 2023.

Figure 1 The preferred reporting items for systematic reviews and meta-analyses flow diagram of study selection.

patient mix with samples exceeding 100. Only the study from Huang *et al*[31] failed to outline their exclusion criteria and, along with the study from Yoshida *et al*[40], they were one of two studies not to report mortality rates. The study from Yoshida *et al*[40], was the only one of the eleven not to include a cohort representative of patients to undergo oesophagectomy. Two development studies, P-POSSUM and Yoshida did not describe patient characteristics adequately.

Table 2 Validation studies of combined and intraoperative models

Ref.	Period + number	Sample region	Operation	Characteristics	Source of data	Models tested	Outcomes tested
Zafirellis <i>et al</i> [36], 2002	1990-1999 (204)	United Kingdom	TT/TH	N: 39/204; H: 156AC/45SCC	Single centre	POSSUM	Major morbidity + Mortality
Lai <i>et al</i> [37], 2007	2001-2005 (545)	China	TT	N: N/A; H: 545SCC/0AC	Hong Kong (All)	POSSUM; O-POSSUM; P-POSSUM	Mortality
Lagarde <i>et al</i> [38], 2007	1993-2005 (663)	Netherlands	TT/TH	N: N/A; H: 476AC/187SCC	Single centre	O-POSSUM	Mortality
Baba <i>et al</i> [39], 2008	2000-2007 (142)	Japan	TT	N: 15/142; H: 142SCC/0AC	Single centre	E-PASS	Morbidity
Bosch <i>et al</i> [33], 2011	1991-2007 (278)	Netherlands	TT	N: 10/278; H: 235AC/43SCC	Single centre	ACCI; CCI; O-POSSUM; P-POSSUM	Mortality
Yoshida <i>et al</i> [40], 2013	2005-2012 (308)	Japan	TT	N: N/A; H: N/A	Single centre	E-PASS	Morbidity
Filip <i>et al</i> [34], 2014	2004-2013 (43)	Romania	TT/TH	N: 22/43; H: 33SCC/9AC	Single centre	ACCI; CCI; POSSUM; O-POSSUM; P-POSSUM	Mortality
Filip <i>et al</i> [32], 2015	2008-2012 (167)	Italy	TT/TH/MIE	N: 131/167; H: 105 AC/62SCC	Single centre	PNI-multivariate PNI; CCI; ACCI; O-POSSUM; P-POSSUM; Amsterdam score	Morbidity
Yamana <i>et al</i> [35], 2015	2005-2013 (251)	Japan	TT/MIE	N: 150/251; H: N/A	Single centre	GNRI; PNI; E-PASS; POSSUM	Respiratory complications
Xing <i>et al</i> [41], 2016	2008-2010 (189)	China	TT/TH	N: 7/189; H: 141SCC/45AC	Single centre	E-SAS	Major morbidity
Eto <i>et al</i> [42], 2016	2007-2015 (399)	Japan	Hybrid	N: 176/399; H: N/A	Single centre	SAS	Major morbidity
Nakagawa <i>et al</i> [43], 2017	2007-2017 (379)	Japan	TT/Hybrid/MIE	N: 129/379; H: 367SCC/12AC	Single centre	SAS	Major morbidity
Aoki <i>et al</i> [44], 2019	2011-2015 (246)	Japan	TT/Hybrid/MIE	N: 152/246; H: N/A	Single centre	E-SAS; Modified E-SAS	Major morbidity
Hayashi <i>et al</i> [45], 2019	2012-2016 (190)	Japan	TT	N: 115/190; H: N/A	Single centre	SAS	Major morbidity

TT: Transthoracic; TH: Transhiatal; MIE: Minimally invasive esophagectomy; N: Neoadjuvant status; H: Histopathology; N/A: Not applicable; SCC: Squamous cell carcinoma; AC: Adenocarcinoma; eSAS: Esophagectomy surgical Apgar score; SAS: Surgical Apgar score; IPF: Intraoperative factors; POSSUM: Physiological and operative severity score for the enumeration of mortality and morbidity; P-POSSUM: Portsmouth physiological and operative severity score for the enumeration of mortality and morbidity; O-POSSUM: Oesophagogastric physiological and operative severity score for the enumeration of mortality and morbidity; E-PASS: Estimation of physiologic ability and surgical stress; ACCI: Age adjusted charlson comorbidity index; CCI: Charlson comorbidity index; PNI: Prognostic nutritional index; GNRI: Geriatric nutritional risk index.

Methodological quality – prognostic factor and outcome management

The majority of the 11 development studies performed well in defining their prognostic factors and outcome measurements. The median score was 4 out of 4 (range 3-4). All development studies defined their prognostic factors and model type, as well as their outcomes. Three of the eleven models incorporating intraoperative data failed to outline their handling of missing data.

Methodological quality – analysis

The median score for methodological quality of analysis was 6 out of 8 within the eleven assessed model development studies (range 4-8). The evaluation measures, model building strategy and testing method were outlined for all studies. Model discrimination and calibration were reported in eight and seven out of eleven studies respectively. Six of the studies used a testing set to validate the performance of the model. It was felt that five studies did not include adequate statistics to assess model performance thoroughly and there was one instance of selective reporting. Just three of the eleven combined and intraoperative models assessed the reference model against an existing tool in the same population.

Table 3 Clinical credibility of combined and intraoperative models

Model	Ref.	Oesophageal specific	No thresholds	Timely data	Reliable data	Easy to generate	Understandable	Useful range	Total
POSSUM	Copeland <i>et al</i> [46], 1991	No	No	No	Yes	Yes	Yes	Yes	4
O-POSSUM	Tekkis <i>et al</i> [48], 2004	Partly	No	No	Yes	Yes	Yes	Yes	4.5
P-POSSUM	Prytherch <i>et al</i> [47], 1998	No	No	No	Yes	Yes	Yes	Yes	4
E-PASS	Haga <i>et al</i> [49], 1999	No	Yes	No	Partly	No	Partly	Yes	3
Yoshida score	Yoshida <i>et al</i> [28], 2014	Yes	No	No	Yes	Yes	Yes	No	4
Xi SAS nomogram	Xi <i>et al</i> [29], 2020	Yes	No	No	Partly	Yes	Yes	Yes	4.5
Xi IPF nomogram	Xi <i>et al</i> [30], 2019	Yes	No	No	Partly	Yes	Yes	Yes	4.5
Huang Model	Huang <i>et al</i> [31], 2020	Yes	No	No	Yes	Yes	Yes	Yes	5
SAS	Gawande <i>et al</i> [50], 2007	No	No	No	Partly	Yes	Yes	Yes	3.5
eSAS	Janowak <i>et al</i> [26], 2015	Yes	No	No	Partly	Yes	Yes	Yes	4.5
Modified eSAS	Stroyer <i>et al</i> [27], 2017	Yes	No	No	Partly	Yes	Yes	Yes	4.5

POSSUM: Physiological and operative severity score for the enumeration of mortality and morbidity; P-POSSUM: Portsmouth physiological and operative severity score for the enumeration of mortality and morbidity; O-POSSUM: Oesophagogastric physiological and operative severity score for the enumeration of mortality and morbidity; E-PASS: Estimation of physiologic ability and surgical stress; SAS: Surgical Apgar score; eSAS: Esophagectomy surgical Apgar score.

Methodological quality – overall performance

Overall, the median score of methodological quality for the 11 studies reviewed was 17 out of 20 (range 15.5-20). The highest scoring model was the O-POSSUM. Other models which scored well for methodological quality included the surgical Apgar score and the nomogram based on the surgical Apgar score developed by Xi *et al*[29,50]. The lowest scoring model was the original POSSUM. The performance of all models is represented in Table 4.

External validation

Of the six combined and intra-operative risk models which were development studies within the included articles, two were later externally validated. The other five models were externally validated within the group of articles. All in all, seven out of 11 combined or intraoperative models have had external validation. These findings are outlined in Figure 2.

Model performance – mortality

Five of the twenty included articles appraised mortality as an outcome measure. The mortality related endpoints varied slightly, with some assessing inpatient mortality and others selecting a post-operative timeframe, typically 30 or 90 d. Several articles reviewed multiple models, leading to a total of ten instances of a model being tested for predicting mortality. Some studies tested both preoperative and combined models within the same article. Overall, three of the eleven prediction models were tested against mortality with each statistically tested in terms of discrimination, represented through area under the ROC curve. Two models that combined preoperative and intraoperative variables had a weighted average exceeding 0.70, indicating clinical utility. These were the POSSUM and the P-POSSUM models. The calibration was tested in terms of Hosmer-Lemeshow goodness of fit and observed to expected ratios. Of the five studies which tested models against mortality on ten occasions, calibration was reported in eight and adequate in four. Amongst the models which combined preoperative and intraoperative data, the P-POSSUM performed best, being appropriately calibrated in two out of three studies.

Model performance – major morbidity

Eleven of the twenty included articles reported major morbidity, defined as a grade three or four complications on the Clavien-Dindo scale. Owing to some studies assessing multiple models, there were thirteen instances of a model being tested for the prediction of major complications. Six unique models were tested in total, with five of these having reported

Table 4 Methodological quality (overall performance) for combined and intraoperative models

Model	Study participation (out of 8)	Measurements (out of 4)	Analysis (out of 8)	Total (out of 20)
POSSUM	7	3	5.5	15.5
O-POSSUM	8	4	8	20
P-POSSUM	7	3	7	17
E-PASS	7.5	3	6	16.5
Yoshida score	6	4	6	16.5
Xi SAS nomogram	8	4	7	19
Xi IPF nomogram	8	4	6	18
Huang model	7.5	4	6	16.5
SAS	8	4	7	19
eSAS	8	4	4	16
Modified eSAS	8	4	6	17.5

POSSUM: Physiological and operative severity score for the enumeration of mortality and morbidity; P-POSSUM: Portsmouth physiological and operative severity score for the enumeration of mortality and morbidity; O-POSSUM: Oesophagogastric physiological and operative severity score for the enumeration of mortality and morbidity; E-PASS: Estimation of physiologic ability and surgical stress; SAS: Surgical Apgar score; eSAS: Esophagectomy surgical Apgar score.

Combined + intraoperative models with external validation



Combined + intraoperative models without external validation



DOI: 10.4240/wjgs.v15.i7.1485 Copyright ©The Author(s) 2023.

Figure 2 External validation status of combined and intraoperative models. POSSUM: Physiological and operative severity score for the enumeration of mortality and morbidity; P-POSSUM: Portsmouth physiological and operative severity score for the enumeration of mortality and morbidity; O-POSSUM: Oesophagogastric physiological and operative severity score for the enumeration of mortality and morbidity; E-PASS: Estimation of physiologic ability and surgical stress; SAS: Surgical Apgar score; eSAS: Esophagectomy surgical Apgar score; IPF: Intraoperative factors.

discrimination statistics as an area under the ROC curve. Two combined models, the esophagectomy surgical Apgar score (eSAS) and intraoperative factors (IPF) nomograms developed by Xi *et al*[29,30], had a weighted mean exceeding 0.7. Both of these were only tested in one study and neither had been externally validated in a second cohort. In terms of calibration in predicting major morbidity, there were poor levels of reporting. Only on two occasions was calibration reported, once each for the eSAS nomogram and IPF nomogram, with calibration sufficient in each instance.

Model Performance – Overall Morbidity

Four of the twenty studies cumulatively reported four instances of three unique models tested against outcomes relating

to overall morbidity. Only the POSSUM score had discrimination represented as area under the ROC curve and this did not reach the discriminatory threshold of utility (AUC ROC 0.55). The POSSUM score was also the only one of the three models reporting calibration with this also being inadequate.

Model performance – respiratory complications/anastomotic leak/readmission/return to theatre

Two articles tested model performance in forecasting adverse respiratory events. The E-PASS and POSSUM scores were each appraised on a single occasion each with neither reaching a c-statistic suggesting clinical utility. Neither model had reported with respect to calibration. Only the model proposed by Huang *et al*[31] had its performance tested in terms of predicting anastomotic leak rates. This model fell just short in terms of reaching discriminatory utility but was well calibrated. No combined or intraoperative model was tested specifically for the prediction of readmission and return to theatre rates.

Model performance – overall comments

The performance for each outcome against which the models were tested has been summarised in Table 5. In terms of discrimination, the weighted average area under the ROC curve is presented in each of the major four outcomes for every model in which this was reported (Figure 3).

Clinical effectiveness

Within the twenty articles reviewed, there was no demonstrated effort to test whether the adoption of these models improved clinical outcomes.

Overall performance

The overall performance of each model across the five domains has been represented in Table 6.

DISCUSSION

This systematic review captured twenty articles with eleven distinct risk prediction models utilising intraoperative variables: Three models exclusively used intraoperative data and the remaining eight combined both intraoperative and preoperative data. Six of the models were designed specifically for oesophagectomy patients and seven had been externally validated. The development studies demonstrated middling clinical credibility but strong methodological quality. However, in general, the models' performance in predicting clinical outcomes was underwhelming, with few instances of the threshold for clinical utility being reached.

Within the included studies, there were two models which possessed a weighted mean of discrimination exceeding the threshold of clinical utility for forecasting mortality. These two models were the POSSUM and the P-POSSUM score. Each reached a discriminatory threshold in two out of three tested studies. The P-POSSUM was appropriately calibrated in two studies, compared to a single study for POSSUM. Both models have also been externally validated. The O-POSSUM model demonstrated clinical utility in the discrimination of mortality in two out of the four considered studies but failed to meet this threshold with its weighted mean. Two other models incorporating intraoperative variables were shown to be adequate in forecasting major morbidity, these were both nomograms devised by Xi *et al*[29,30]. They are both well calibrated and utilise a similar combination of intraoperative and preoperative factors. However, neither have been externally validated. None of the models demonstrated utility in predicting overall morbidity nor respiratory morbidity.

The identification of high-risk patients immediately following the procedure could help tailor postoperative care. First and foremost, it could inform which patients may receive more intensive reviews and potentially influence the decision to investigate for evolving complications in the event of deviation from the expected course of recovery. It is now commonplace for patients following oesophagectomy to be enrolled in a standardised enhanced recovery after surgery pathway[51]. The pathway may be altered or augmented to prophylactically address anticipated challenges, such as preoperative tube feeding in patients deemed susceptible to postoperative malnutrition[52]. Other high-risk individuals may even be deemed inappropriate for the standard enhanced recovery pathway altogether and receive a more conservative treatment strategy[53]. Significant complications within the perioperative period negatively impacts long-term functional outcomes and connotes poorer survival prognosis and therefore these benefits may not be limited to only the short-term[54-56].

By identifying and stratifying perioperative risk this could permit the judicious application of resources, yielding systemic economic gain. For example, intensive care admission is a conventional practice post-oesophagectomy in many parts of the world, but there is a growing adoption of high dependency unit postoperative care[57-59]. Tools that can accurately characterise an individual's post-operative risk profile may identify individuals that should receive admission directly to intensive care. Many of the models incorporating intraoperative data identified in this review have been found to be superior to pre-operative fitness testing in terms of post-operative outcomes. An example is the use of cardiopulmonary fitness testing. This tool is being increasingly used to direct services and guide care, despite the fact that it fails to reach the threshold of clinical utility in predicting perioperative outcomes[60,61]. This reflects the desire to individualise care and target resources as postoperative complications substantially increase expenditure per patient[62]. Therefore, if more intensive allocation of resources can be afforded to high-risk individuals, it may circumvent negative outcomes and prove economically beneficial[63].

Table 5 Summary of the performance for each of the combined and intraoperative models

Predictive model (n)	Ref.	Discrimination	Calibration/other statistics	Outcome
POSSUM (5)	Zafirellis <i>et al</i> [36], 2002	0.55, 0.62	HL $P < 0.001$, 0.002	Morbidity + Mortality
	Lai <i>et al</i> [37], 2007	AUC = 0.776	O:E 0.367 ($P = 0.001$)	Mortality
	Filip <i>et al</i> [34], 2014	AUC = 0.826	O:E 1.11	Mortality
	Filip <i>et al</i> [32], 2015	AUC = 0.59	HL P value (0.08)	Morbidity
	Yamana <i>et al</i> [35], 2015	AUC = 0.601	Not reported	Resp complications
O-POSSUM (4)	Lai <i>et al</i> [37], 2007	AUC = 0.676	O:E 0.505 ($P = 0.002$)	Mortality
	Lagarde <i>et al</i> [38], 2007	AUC = 0.60	O:E 0.29 ($P < 0.001$)	Mortality
	Bosch <i>et al</i> [33], 2011	AUC = 0.756	HL P value (0.035)	Mortality
	Filip <i>et al</i> [34], 2014	AUC = 0.740	O:E 1.70	Mortality
P-POSSUM (3)	Lai <i>et al</i> [37], 2007	AUC = 0.776	O:E ($P = 0.814$)	Mortality
	Bosch <i>et al</i> [33], 2011	AUC = 0.766	HL P value (0.529)	Mortality
	Filip <i>et al</i> [34], 2014	AUC = 0.607	O:E 4.28	Mortality
E-PASS (3)	Baba <i>et al</i> [39], 2008	CRS > 1 ($P < 0.001$)	N/A	Morbidity
	Yoshida <i>et al</i> [40], 2013	CRS > 0.9 ($P 0.029$)	N/A	Morbidity
	Yamana <i>et al</i> [35], 2015	AUC = 0.659	N/A	Resp complications
Yoshida (1)	Yoshida <i>et al</i> [28], 2015	P value = < 0.001, 0.037	Not reported	Morbidity + Major morbidity
Xi SAS nomogram (1)	Xi <i>et al</i> [30], 2019	AUC = 0.903, 0.967	“Good”	Major morbidity
Xi IPF nomogram (1)	Xi <i>et al</i> [29], 2020	AUC = 0.921	“Sufficient”	Major morbidity
Huang (1)	Huang <i>et al</i> [31], 2020	AUC = 0.69	“Good”	Anastomotic leak
SAS (4)	Eto <i>et al</i> [42], 2016	AUC = 0.68	Cut off = 6 Pulm, SSI, AL	Major morbidity
	Stroyer <i>et al</i> [27], 2017	AUC = 0.561	All comp $P = 0.148$	Major morbidity
	Nakagawa <i>et al</i> [43], 2017	Not reported	Cut off > 5, $P < 0.0001$	Major morbidity
	Hayashi <i>et al</i> [45], 2019	Not reported	Cut off ≥ 6 , $P = 0.032$	Major morbidity
eSAS (4)	Janowak <i>et al</i> [26], 2015	AUC = 0.614	Cut off = 6, Com $P = 0.01$	Major morbidity
	Xing <i>et al</i> [41], 2016	AUC = 0.62	Cut off > 7 LOS	Major morbidity
	Stroyer <i>et al</i> [27], 2017	AUC = 0.572	All comp $P = 0.460$	Major morbidity
	Aoki <i>et al</i> [44], 2019	Optimal cut-off 4 (Youden’s index = 0.069)	Major morbidity ($P = 0.29$)	Major morbidity
Modified eSAS (2)	Stroyer <i>et al</i> [27], 2017	AUC = 0.532	All comp $P = 0.096$	Major morbidity
	Aoki <i>et al</i> [44], 2019	Optimal cut-off 4 (Youden’s index = 0.180)	Major morbidity ($P = 0.004$)	Major morbidity

POSSUM: Physiological and operative severity score for the enumeration of mortality and morbidity; P-POSSUM: Portsmouth physiological and operative severity score for the enumeration of mortality and morbidity; O-POSSUM: Oesophagogastric physiological and operative severity score for the enumeration of mortality and morbidity; E-PASS: Estimation of physiologic ability and surgical stress; SAS: Surgical Apgar score; eSAS: Esophagectomy surgical Apgar score; N/A: Not applicable; IPF: Intraoperative factors; AUC: Area under the curve.

The results of this study are consistent with the findings of preceding systematic reviews. In 2010, a systematic review in a combined cohort of oesophagectomy and gastrectomy patients tested various POSSUM models as predictors of morbidity and mortality. They found that the models had limited utility in relation to morbidity and overestimated mortality but nevertheless that the P-POSSUM was the most accurate in predicting post-operative mortality[64]. An oesophageal specific systematic review conducted in 2014 by Findlay *et al*[13] again concluded that no predictive model predicted morbidity with sufficient accuracy and that P-POSSUM, followed by POSSUM were most promising in forecasting postoperative mortality[13]. A more recent review undertaken in 2015 found that none of the models could be confidently used in clinical practice to forecast any perioperative outcomes[14]. The only models that were found in our study to predict major morbidity were developed after these previous reviews.

Table 6 Summary of the combination and intraoperative models across the five categories

Model	Ref.	Clinical credibility (out of 7)	Methodological quality (out of 20)	Model performance (overall utility)	External validation	Clinical effectiveness
POSSUM	Copeland <i>et al</i> [46], 1991	4	15.5	Mortality	Yes	No
O-POSSUM	Tekkis <i>et al</i> [48], 2004	4.5	20	No	Yes	No
P-POSSUM	Prytherch <i>et al</i> [47], 1998	4	17	Mortality	Yes	No
E-PASS	Haga <i>et al</i> [49], 1999	3	16.5	No	Yes	No
SAS	Gawande <i>et al</i> [50], 2007	4	16.5	No	Yes	No
eSAS	Janowak <i>et al</i> [26], 2015	4.5	19	No	Yes	No
Modified eSAS	Stroyer <i>et al</i> [27], 2017	4.5	18	No	Yes	No
Yoshida score	Yoshida <i>et al</i> [28], 2015	5	16.5	No	No	No
Xi SAS nomogram	Xi <i>et al</i> [29], 2020	3.5	19	Major morbidity	No	No
Xi IPF nomogram	Xi <i>et al</i> [30], 2019	4.5	16	Major morbidity	No	No
Huang	Huang <i>et al</i> [31], 2020	4.5	17.5	No	No	No

POSSUM: Physiological and operative severity score for the enumeration of mortality and morbidity; P-POSSUM: Portsmouth physiological and operative severity score for the enumeration of mortality and morbidity; O-POSSUM: Oesophagogastric physiological and operative severity score for the enumeration of mortality and morbidity; E-PASS: Estimation of physiologic ability and surgical stress; SAS: Surgical Apgar score; eSAS: Esophagectomy surgical Apgar score.

There are numerous potential reasons that the models fall short of meeting clinical utility. The models included were all data generated wherein the outcome predictors were derived from the cohort against which they were subsequently tested[65]. This natural bias predisposes to overfitting the development data set as a fait accompli and thus poorer performance when tested externally against a different dataset[65]. Some of the models were also limited by the size of their development data. In addition, the performance of some models may have been confounded by the fact that they were devised years earlier or for procedures other than oesophageal resection, thereby creating a divergence in clinical practice between development and validation datasets. Even in the event of recently developed models specifically designed for oesophagectomy, differences and advances in surgical techniques and perioperative management strategies between centres may lead to inconsistent results.

There are strengths to our systematic review. It was conducted in accordance with the PRISMA method of study search and selection strategy. There have only been two previous systematic reviews appraising multivariate risk models in the forecasting of perioperative outcomes with only one of these incorporating a qualitative analysis of the risk models involved and neither made the distinction between preoperative or intraoperative variables. This review focuses only on models incorporating intraoperative data, allowing the authors to focus on the implications that these models may have on resource allocation, rather than patient selection. But perhaps the greatest strength of this review is that the most recent effort was in 2015, thereby allowing consideration of a large number of models developed within the intervening years which have not yet been independently reviewed.

Despite this, several challenges were encountered. The quality of the results in this study remains dependent on the accuracy and completeness of reporting within the original publications. Bias is associated with excluding articles not published in English or prior to the year 2000. Another significant limitation of this review is that it did not consider longer term outcomes such as overall and disease-free survival. It also did not consider models using exclusively preoperative variables, meaning that the results cannot aid clinical decision making in terms of appropriate patient selection for oesophagectomy instead focusing on resource allocation post-procedure. Finally, despite efforts to standardise the assessment process, there remains a subjective component of any qualitative appraisal which can be a source of bias.

The findings and areas of limitation identified in this review inform the direction of further research. Substantial heterogeneity was observed across the studies in terms of outcome measures and the clinical outcomes achieved, reflective of temporal and regional variance in clinical practice. This limits the credibility of these studies when extending

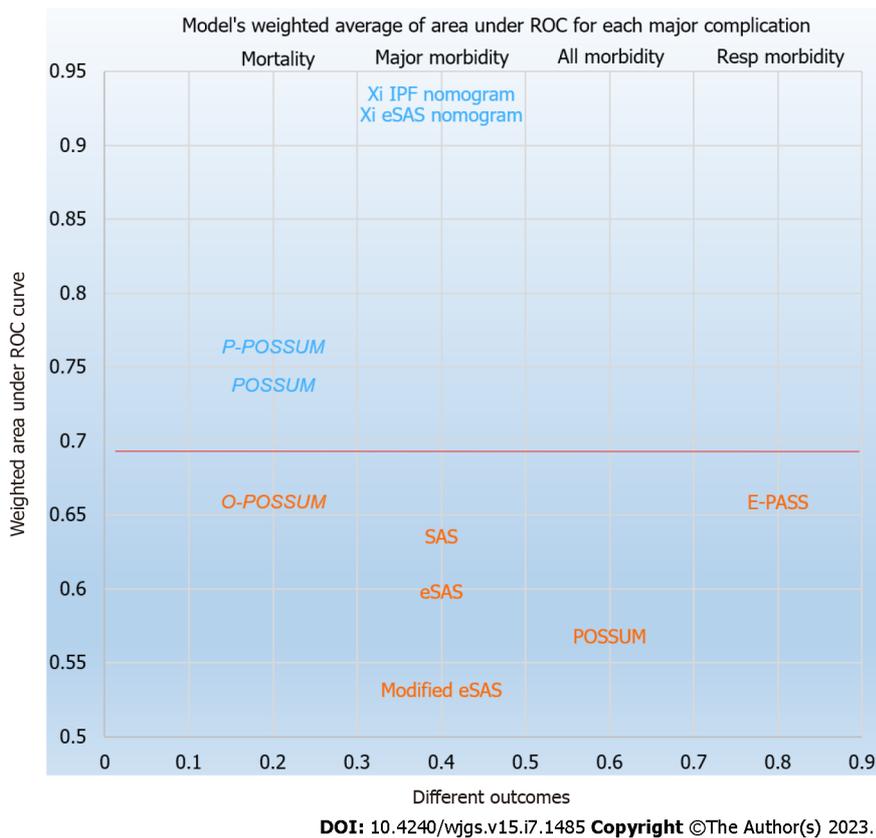


Figure 3 Weighted mean of c-statistics for each major outcome. ROC: Receiver operating characteristics; POSSUM: Physiological and operative severity score for the enumeration of mortality and morbidity; P-POSSUM: Portsmouth physiological and operative severity score for the enumeration of mortality and morbidity; O-POSSUM: Oesophagogastric physiological and operative severity score for the enumeration of mortality and morbidity; E-PASS: Estimation of physiologic ability and surgical stress; SAS: Surgical Apgar score; eSAS: Esophagectomy surgical Apgar score; IPF: Intraoperative factors.

their conclusions to external populations. The review highlighted that many of the models identified, including those shown to predict major morbidity, still require external validation and all need to be tested to prove that their application leads to improvement in clinical outcomes. If these models were to be demonstrated as effective on another population, it would greatly strengthen the case for their application to clinical practice. Due to the relative infrequency of mortality related outcomes following oesophagectomy, this would require a large-scale multi-centre prospective clinical trial conducted over many years. If a model was demonstrated to be beneficial in guiding most appropriate postoperative treatment course, it would incentivise surgeons and intensivists to perform the risk assessment routinely.

CONCLUSION

There have been numerous multivariate risk models either adapted to or specifically developed for predicting perioperative outcomes mortality following oesophagectomy. An individualised assessment can assist in identifying patients at high risk with more intensive resource allocation directed accordingly. Some models utilise only patient characteristics evident prior to the operation. However, many models incorporate preoperative data with intraoperative variables whilst others are generated exclusively from intraoperative features with these two model groups being the subject of this review. This study has demonstrated that the majority of these models are clinically credible and have been developed with sound methodological quality. Unfortunately, only two models adequately predicted mortality, namely the P-POSSUM and POSSUM scores. There were also only two studies, the IPF and eSAS based nomograms, that forecasted major morbidity and neither of these had been externally validated or tested for clinical effectiveness. Further research is therefore required before prediction models can be utilised in clinical practice with confidence to guide postoperative resource allocation.

ARTICLE HIGHLIGHTS

Research background

Oesophageal cancer is a major contributor to the worldwide cancer-related morbidity and mortality disease burden.

Undertaking an oesophagectomy can offer a realistic curative option if the disease is detected in the early stages. The most significant drawback with respect to oesophagectomy is the considerable associated risk of major complications and even mortality throughout the perioperative period. Because of this, it is imperative to appropriately select surgical candidates and allocate resources closely to those at heightened risk. A vast number of multivariate risk prediction models have been constructed to assist in this decision-making which incorporate both preoperative and intraoperative factors, with some doubt existing as to which model is most reliable. This publication is the first systematic review to focus solely on models incorporating, at least in part, intraoperative factors with its ultimate goal being to determine which model most accurately forecasts perioperative outcomes.

Research motivation

The identification of the best risk prediction model which incorporates intraoperative data in isolation and in combination with preoperative factors would allow surgeons to utilise this model in clinical practice. Such a risk model could serve to augment clinical decision-making both in terms of prudently selecting surgical candidates and allocating resources in the postoperative period. It is expected that improved patient selection and more judicious resource allocation could lead to improved perioperative outcomes for patients with oesophageal cancer.

Research objectives

The objective of this research is to perform a systematic review assessing which multivariate risk model incorporating intraoperative variables, either in isolation or in conjunction with preoperative factors, best forecasts perioperative outcomes following oesophagectomy. The primary objective pertains to assessing predictive performance for mortality outcomes. The secondary objectives are to assess the predictive capacity of these models in forecasting major morbidity, overall morbidity and other key complications such as respiratory complications and anastomotic leak.

Research methods

A systematic review incorporating the MEDLINE, Embase and Cochrane databases was performed from 2000-2020. The search terms were [(Oesophagectomy) AND (Risk OR predict OR score OR model) AND (Outcomes OR mortality OR morbidity OR complications OR anastomotic leak OR length of stay)]. Only multivariate based prediction models which utilised intraoperative factors, either in isolation or in combination with preoperative variables to predict perioperative outcomes following oesophagectomy were included. Articles were generated, collated then reported in accordance with preferred reporting items for systematic reviews and meta-analyses guidelines. All of the included risk models were appraised across five categories, namely clinical credibility, methodological quality, model performance, external validation and clinical effectiveness.

Research results

The initial search captured 8715 articles which was refined to 197 texts considered to be potentially relevant after deduplication, title and abstract screening. Following a detailed reading of these articles, 20 published studies were ultimately incorporated with these examining 11 multivariate risk prediction models. Eight of these combined preoperative and intraoperative data, with the other three models exclusively utilising intraoperative variables. The majority of these models were clinically credible and developed with sound methodological quality but many models had not been externally validated and none had been proven to be clinically effective in improving outcomes. Two models adequately predicted mortality, namely the Portsmouth physiological and operative severity score for the enumeration of mortality and morbidity (POSSUM) and POSSUM scores. A further two studies, the intraoperative factors and esophagectomy surgical Apgar score based nomograms were effective at predicting outcomes related to major morbidity. None of the included models were sufficiently accurate in predicting overall morbidity, respiratory complications or anastomotic leak rates.

Research conclusions

There are a handful of credible and well-developed multivariate risk prediction models which demonstrate the capacity to discriminate perioperative mortality and major morbidity outcomes following oesophagectomy. However, there is a need to undertake more research in terms of external validation and demonstrating improved clinical outcomes by guiding patient selection and postoperative resource allocation with the use of these models.

Research perspectives

There is an existing research gap to externally validate some of these models which are yet to be tested outside their development cohort. Further research should also take the form of a prospective randomised control trial in to compare the accuracy of clinical discretion against the results of the clinical risk prediction models in selecting appropriate surgical candidates and guiding postoperative resource allocation. Such a study could act as a catalyst to emphasise the importance of these tools which can augment decision-making and potentially lead to their widespread adoption in the care of patients undergoing oesophagectomy.

ACKNOWLEDGEMENTS

We would like to acknowledge the assistance of Nikki May, SA Health librarian in the construction and execution of the

search strategy. This work was initially undertaken as part of the University of Edinburgh, Masters of Surgical Science.

FOOTNOTES

Author contributions: Grantham JP and Shenfine J designed the research; Grantham JP and Hii A performed the research and analysed the data; Grantham JP, Hii A and Shenfine J all contributed to writing and reviewing the paper.

Conflict-of-interest statement: All the authors report no relevant conflicts of interest for this article.

PRISMA 2009 Checklist statement: The authors have read the PRISMA 2009 Checklist, and the manuscript was prepared and revised according to the 2009 PRISMA Checklist.

Open-Access: This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <https://creativecommons.org/licenses/by-nc/4.0/>

Country/Territory of origin: Australia

ORCID number: James Paul Grantham 0000-0002-5267-5646; Amanda Hii 0000-0003-3056-0087; Jonathan Shenfine 0000-0002-1862-4693.

S-Editor: Li L

L-Editor: A

P-Editor: Ma YJ

REFERENCES

- 1 **Flanagan JC**, Batz R, Saboo SS, Nordeck SM, Abbara S, Kernstine K, Vasani V. Esophagectomy and Gastric Pull-through Procedures: Surgical Techniques, Imaging Features, and Potential Complications. *Radiographics* 2016; **36**: 107-121 [PMID: 26761533 DOI: 10.1148/rg.2016150126]
- 2 **Veelo DP**, Geerts BF. Anaesthesia during oesophagectomy. *J Thorac Dis* 2017; **9**: S705-S712 [PMID: 28815066 DOI: 10.21037/jtd.2017.03.153]
- 3 **D'journo XB**, Michelet P, Avaro JP, Trousse D, Giudicelli R, Fuentes P, Doddoli C, Thomas P. [Respiratory complications after oesophagectomy for cancer]. *Rev Mal Respir* 2008; **25**: 683-694 [PMID: 18772826 DOI: 10.1016/s0761-8425(08)73798-7]
- 4 **Molena D**, Mungo B, Stem M, Lidor AO. Incidence and risk factors for respiratory complications in patients undergoing esophagectomy for malignancy: a NSQIP analysis. *Semin Thorac Cardiovasc Surg* 2014; **26**: 287-294 [PMID: 25837540 DOI: 10.1053/j.semthor.2014.12.002]
- 5 **Klevebro F**, Elliott JA, Slaman A, Vermeulen BD, Kamiya S, Rosman C, Gisbertz SS, Boshier PR, Reynolds JV, Rouvelas I, Hanna GB, van Berge Henegouwen MI, Markar SR. Cardiorespiratory Comorbidity and Postoperative Complications following Esophagectomy: a European Multicenter Cohort Study. *Ann Surg Oncol* 2019; **26**: 2864-2873 [PMID: 31183640 DOI: 10.1245/s10434-019-07478-6]
- 6 **Al-Tarshih MI**, Ghanma IM, Khamash FA, Allbrahim AE. Cardiac complications in the first week post transhiatal esophagectomy for esophageal cancer. *JRMS* 2015; **15**: 29-33 [DOI: 10.1093/dote/doab022]
- 7 **Fabbi M**, Hagens ERC, van Berge Henegouwen MI, Gisbertz SS. Anastomotic leakage after esophagectomy for esophageal cancer: definitions, diagnostics, and treatment. *Dis Esophagus* 2021; **34** [PMID: 32476017 DOI: 10.1093/dote/daaa039]
- 8 **Mboumi IW**, Reddy S, Lidor AO. Complications After Esophagectomy. *Surg Clin North Am* 2019; **99**: 501-510 [PMID: 31047038 DOI: 10.1016/j.suc.2019.02.011]
- 9 **Rouvelas I**, Lagergren J. The impact of volume on outcomes after oesophageal cancer surgery. *ANZ J Surg* 2010; **80**: 634-641 [PMID: 20840408 DOI: 10.1111/j.1445-2197.2010.05406.x]
- 10 **Collins GS**, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, Voysey M, Wharton R, Yu LM, Moons KG, Altman DG. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol* 2014; **14**: 40 [PMID: 24645774 DOI: 10.1186/1471-2288-14-40]
- 11 **Ross PL**, Gerigk C, Gonen M, Yossepowitch O, Cagiannos I, Sogani PC, Scardino PT, Kattan MW. Comparisons of nomograms and urologists' predictions in prostate cancer. *Semin Urol Oncol* 2002; **20**: 82-88 [PMID: 12012293 DOI: 10.1053/suro.2002.32490]
- 12 **Kattan MW**, Yu C, Stephenson AJ, Sartor O, Tombal B. Clinicians versus nomogram: predicting future technetium-99m bone scan positivity in patients with rising prostate-specific antigen after radical prostatectomy for prostate cancer. *Urology* 2013; **81**: 956-961 [PMID: 23375915 DOI: 10.1016/j.urology.2012.12.010]
- 13 **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**: 1532-1542 [PMID: 24760219 DOI: 10.1007/s11605-014-2524-2]
- 14 **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**: 32-43 [PMID: 25231768 DOI: 10.1093/bja/aeu294]
- 15 **Moher D**, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009; **6**: e1000097 [PMID: 19621072 DOI: 10.1371/journal.pmed.1000097]
- 16 **Clavien PA**, Barkun J, de Oliveira ML, Vauthey JN, Dindo D, Schulick RD, de Santibañes E, Pekolj J, Slankamenac K, Bassi C, Graf R, Vonlanthen R, Padbury R, Cameron JL, Makuuchi M. The Clavien-Dindo classification of surgical complications: five-year experience. *Ann Surg* 2009; **250**: 187-196 [PMID: 19638912 DOI: 10.1097/SLA.0b013e3181b13ca2]
- 17 **Wyatt JC**, Altman DG. Commentary: Prognostic models: clinically useful or quickly forgotten? *BMJ* 2011; **343**: 1539-1541 [DOI: 10.1136/bmj.d6442]

- 10.1136/bmj.311.7019.1539]
- 18 **Moons KG**, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ* 2009; **338**: b375 [PMID: 19237405 DOI: 10.1136/bmj.b375]
 - 19 **Altman DG**, Royston P. What do we mean by validating a prognostic model? *Stat Med* 2000; **19**: 453-473 [PMID: 10694730 DOI: 10.1002/(sici)1097-0258(20000229)19:4<453::aid-sim350>3.0.co;2-5]
 - 20 **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**: 1258-1268 [PMID: 21647716 DOI: 10.1007/s00134-011-2265-6]
 - 21 **Minne L**, Abu-Hanna A, de Jonge E. Evaluation of SOFA-based models for predicting mortality in the ICU: A systematic review. *Crit Care* 2008; **12**: R161 [PMID: 19091120 DOI: 10.1186/cc7160]
 - 22 **Mandrekar JN**. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol* 2010; **5**: 1315-1316 [PMID: 20736804 DOI: 10.1097/JTO.0b013e3181ec173d]
 - 23 **Schummers L**, Himes KP, Bodnar LM, Hutcheon JA. Predictor characteristics necessary for building a clinically useful risk prediction model: a simulation study. *BMC Med Res Methodol* 2016; **16**: 123 [PMID: 27655140 DOI: 10.1186/s12874-016-0223-2]
 - 24 **Crowson CS**, Atkinson EJ, Therneau TM. Assessing calibration of prognostic risk scores. *Stat Methods Med Res* 2016; **25**: 1692-1706 [PMID: 23907781 DOI: 10.1177/0962280213497434]
 - 25 **Hosmer DW**, Lemeshow S. Goodness of fit tests for the multiple logistic regression model. *Commun Stat Theory Methods* 1980; **9**: 1043-1069 [DOI: 10.1080/03610928008827941]
 - 26 **Janowak CF**, Blasberg JD, Taylor L, Maloney JD, Macke RA. The Surgical Apgar Score in esophagectomy. *J Thorac Cardiovasc Surg* 2015; **150**: 806-812 [PMID: 26234458 DOI: 10.1016/j.jtcvs.2015.07.017]
 - 27 **Strøyer S**, Mantoni T, Svendsen LB. Evaluation of the surgical apgar score in patients undergoing Ivor-Lewis esophagectomy. *J Surg Oncol* 2017; **115**: 186-191 [PMID: 28054355 DOI: 10.1002/jso.24483]
 - 28 **Yoshida N**, Baba Y, Watanabe M, Ida S, Ishimoto T, Karashima R, Iwagami S, Imamura Y, Sakamoto Y, Miyamoto Y, Baba H. Original scoring system for predicting postoperative morbidity after esophagectomy for esophageal cancer. *Surg Today* 2015; **45**: 346-354 [PMID: 24997754 DOI: 10.1007/s00595-014-0958-5]
 - 29 **Xi Y**, Shen W, Wang L, Yu C. An esophagectomy Surgical Apgar Score (eSAS)-based nomogram for predicting major morbidity in patients with esophageal carcinoma. *Transl Cancer Res* 2020; **9**: 1732-1741 [PMID: 35117520 DOI: 10.21037/tcr.2020.02.56]
 - 30 **Xi Y**, Jin C, Wang L, Shen W. Predictive value of intraoperative factors for complications after oesophagectomy. *Interact Cardiovasc Thorac Surg* 2019; **29**: 525-531 [PMID: 31553799 DOI: 10.1093/icvts/ivz150]
 - 31 **Huang C**, Yao H, Huang Q, Lu H, Xu M, Wu J. A novel nomogram to predict the risk of anastomotic leakage in patients after oesophagectomy. *BMC Surg* 2020; **20**: 64 [PMID: 32252738 DOI: 10.1186/s12893-020-00726-7]
 - 32 **Filip B**, Scarpa M, Cavallin F, Cagol M, Alfieri R, Saadeh L, Ancona E, Castoro C. Postoperative outcome after oesophagectomy for cancer: Nutritional status is the missing ring in the current prognostic scores. *Eur J Surg Oncol* 2015; **41**: 787-794 [PMID: 25890494 DOI: 10.1016/j.ejso.2015.02.014]
 - 33 **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**: 303-309 [PMID: 21871985 DOI: 10.1016/j.amjsurg.2011.04.003]
 - 34 **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**: 480-485 [PMID: 25149610]
 - 35 **Yamana I**, Takeno S, Shibata R, Shiwaku H, Maki K, Hashimoto T, Shiraishi T, Iwasaki A, Yamashita Y. Is the Geriatric Nutritional Risk Index a Significant Predictor of Postoperative Complications in Patients with Esophageal Cancer Undergoing Esophagectomy? *Eur Surg Res* 2015; **55**: 35-42 [PMID: 25790838 DOI: 10.1159/000376610]
 - 36 **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**: 1150-1155 [PMID: 12190681 DOI: 10.1046/j.1365-2168.2002.02179.x]
 - 37 **Lai F**, Kwan TL, Yuen WC, Wai A, Siu YC, Shung E. Evaluation of various POSSUM models for predicting mortality in patients undergoing elective oesophagectomy for carcinoma. *Br J Surg* 2007; **94**: 1172-1178 [PMID: 17520711 DOI: 10.1002/bjs.5793]
 - 38 **Lagarde SM**, Maris AK, de Castro SM, Busch OR, Obertop H, van Lanschoot JJ. Evaluation of O-POSSUM in predicting in-hospital mortality after resection for oesophageal cancer. *Br J Surg* 2007; **94**: 1521-1526 [PMID: 17929231 DOI: 10.1002/bjs.5850]
 - 39 **Baba Y**, Haga Y, Hiyoshi Y, Imamura Y, Nagai Y, Yoshida N, Hayashi N, Toyama E, Miyanari N, Baba H. Estimation of Physiologic Ability and Surgical Stress (E-PASS system) in patients with esophageal squamous cell carcinoma undergoing resection. *Esophagus* **5**: 81-86 [DOI: 10.1007/s10388-008-0153-0]
 - 40 **Yoshida N**, Watanabe M, Baba Y, Iwagami S, Ishimoto T, Iwatsuki M, Sakamoto Y, Miyamoto Y, Ozaki N, Baba H. Estimation of physiologic ability and surgical stress (E-PASS) can assess short-term outcome after esophagectomy for esophageal cancer. *Esophagus* **10**: 86-94 [DOI: 10.1007/s10388-013-0369-5]
 - 41 **Xing XZ**, Wang HJ, Qu SN, Huang CL, Zhang H, Wang H, Yang QH, Gao Y. The value of esophagectomy surgical apgar score (eSAS) in predicting the risk of major morbidity after open esophagectomy. *J Thorac Dis* 2016; **8**: 1780-1787 [PMID: 27499969 DOI: 10.21037/jtd.2016.06.28]
 - 42 **Eto K**, Yoshida N, Iwatsuki M, Kurashige J, Ida S, Ishimoto T, Baba Y, Sakamoto Y, Miyamoto Y, Watanabe M, Baba H. Surgical Apgar Score Predicted Postoperative Morbidity After Esophagectomy for Esophageal Cancer. *World J Surg* 2016; **40**: 1145-1151 [PMID: 26801506 DOI: 10.1007/s00268-016-3425-1]
 - 43 **Nakagawa A**, Nakamura T, Oshikiri T, Hasegawa H, Yamamoto M, Kanaji S, Matsuda Y, Yamashita K, Matsuda T, Sumi Y, Suzuki S, Kakeji Y. The Surgical Apgar Score Predicts Not Only Short-Term Complications But Also Long-Term Prognosis After Esophagectomy. *Ann Surg Oncol* 2017; **24**: 3934-3946 [PMID: 28986819 DOI: 10.1245/s10434-017-6103-0]
 - 44 **Aoki Y**, Ide K, Nakajima F, Kawasaki Y, Fujita Y, Morimoto E, Yokoyama J. Esophagectomy Surgical Apgar Score May Not Be Associated With Postoperative Morbidity. *Semin Thorac Cardiovasc Surg* 2019; **31**: 574-580 [PMID: 30529159 DOI: 10.1053/j.semtevs.2018.12.002]
 - 45 **Hayashi M**, Kawakubo H, Mayanagi S, Nakamura R, Suda K, Wada N, Kitagawa Y. A low surgical Apgar score is a predictor of anastomotic leakage after transthoracic esophagectomy, but not a prognostic factor. *Esophagus* 2019; **16**: 386-394 [PMID: 31165934 DOI: 10.1007/s10388-019-00678-9]
 - 46 **Copeland GP**, Jones D, Walters M. POSSUM: a scoring system for surgical audit. *Br J Surg* 1991; **78**: 355-360 [PMID: 2021856 DOI: 10.1002/bjs.1800780327]

- 47 **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. *Br J Surg* 1998; **85**: 1217-1220 [PMID: 9752863 DOI: 10.1046/j.1365-2168.1998.00840.x]
- 48 **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**: 288-295 [PMID: 14991628 DOI: 10.1002/bjs.4414]
- 49 **Haga Y**, Ikei S, Ogawa M. Estimation of Physiologic Ability and Surgical Stress (E-PASS) as a new prediction scoring system for postoperative morbidity and mortality following elective gastrointestinal surgery. *Surg Today* 1999; **29**: 219-225 [PMID: 10192731 DOI: 10.1007/BF02483010]
- 50 **Gawande AA**, Kwaan MR, Regenbogen SE, Lipsitz SA, Zinner MJ. An Apgar score for surgery. *J Am Coll Surg* 2007; **204**: 201-208 [PMID: 17254923 DOI: 10.1016/j.jamcollsurg.2006.11.011]
- 51 **Rubinkiewicz M**, Witowski J, Su M, Major P, Pędzwiatr M. Enhanced recovery after surgery (ERAS) programs for esophagectomy. *J Thorac Dis* 2019; **11**: S685-S691 [PMID: 31080645 DOI: 10.21037/jtd.2018.11.56]
- 52 **Low DE**, Allum W, De Manzoni G, Ferri L, Immanuel A, Kuppusamy M, Law S, Lindblad M, Maynard N, Neal J, Pramesh CS, Scott M, Mark Smithers B, Addor V, Ljungqvist O. Guidelines for Perioperative Care in Esophagectomy: Enhanced Recovery After Surgery (ERAS[®]) Society Recommendations. *World J Surg* 2019; **43**: 299-330 [PMID: 30276441 DOI: 10.1007/s00268-018-4786-4]
- 53 **Yin M**, Yan Y, Fan Z, Fang N, Wan H, Mo W, Wu X. The efficacy of Enhanced Recovery after Surgery (ERAS) for elderly patients with intertrochanteric fractures who received surgery: study protocol for a randomized, blinded, controlled trial. *J Orthop Surg Res* 2020; **15**: 91 [PMID: 32138760 DOI: 10.1186/s13018-020-01586-w]
- 54 **Quinton E**, Nightingale P, Hudsmith L, Thorne S, Marshall H, Clift P, de Bono J. Prevalence of atrial tachyarrhythmia in adults after Fontan operation. *Heart* 2015; **101**: 1672-1677 [PMID: 26289423 DOI: 10.1136/heartjnl-2015-307514]
- 55 **Fransen LFC**, Luyer MDP. Effects of improving outcomes after esophagectomy on the short- and long-term: a review of literature. *J Thorac Dis* 2019; **11**: S845-S850 [PMID: 31080668 DOI: 10.21037/jtd.2018.12.09]
- 56 **Motoori M**, Ito Y, Miyashiro I, Sugimura K, Miyata H, Omori T, Fujiwara Y, Yano M. Impact of Age on Long-Term Survival in Patients with Esophageal Cancer Who Underwent Transthoracic Esophagectomy. *Oncology* 2019; **97**: 149-154 [PMID: 31195394 DOI: 10.1159/000500604]
- 57 **Iscimen R**, Brown DR, Cassivi SD, Keegan MT. Intensive care unit utilization and outcome after esophagectomy. *J Cardiothorac Vasc Anesth* 2010; **24**: 440-446 [PMID: 18834773 DOI: 10.1053/j.jvca.2008.02.002]
- 58 **Voeten DM**, van der Werf LR, Gisbertz SS, Ruurda JP, van Berge Henegouwen MI, van Hillegersberg R; Dutch Upper Gastrointestinal Cancer Audit (DUCA) Group. 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**: 1961-1968 [PMID: 33485673 DOI: 10.1016/j.ejso.2021.01.005]
- 59 **Ghosh S**, Steyn RS, Marzouk JF, Collins FJ, Rajesh PB. The effectiveness of high dependency unit in the management of high risk thoracic surgical cases. *Eur J Cardiothorac Surg* 2004; **25**: 123-126 [PMID: 14690743 DOI: 10.1016/s1010-7940(03)00610-9]
- 60 **Patel N**, Powell AG, Wheat JR, Brown C, Appadurai IR, Davies RG, Bailey DM, Lewis WG. Cardiopulmonary fitness predicts postoperative major morbidity after esophagectomy for patients with cancer. *Physiol Rep* 2019; **7**: e14174 [PMID: 31342676 DOI: 10.14814/phy2.14174]
- 61 **Forshaw MJ**, Strauss DC, Davies AR, Wilson D, Lams B, Pearce A, Botha AJ, Mason RC. Is cardiopulmonary exercise testing a useful test before esophagectomy? *Ann Thorac Surg* 2008; **85**: 294-299 [PMID: 18154826 DOI: 10.1016/j.athoracsur.2007.05.062]
- 62 **Dimick JB**, Pronovost PJ, Cowan JA, Lipsett PA. Complications and costs after high-risk surgery: where should we focus quality improvement initiatives? *J Am Coll Surg* 2003; **196**: 671-678 [PMID: 12742194 DOI: 10.1016/S1072-7515(03)00122-4]
- 63 **Rozeboom PD**, Dyas AR, Bronsert MR, Bhagat R, Meguid RA. Improving postoperative outcomes in esophagectomy for cancer-what is the role of institutional data? *J Thorac Dis* 2020; **12**: 1750-1753 [PMID: 32642077 DOI: 10.21037/jtd-20-1681]
- 64 **Dutta S**, Horgan PG, McMillan DC. POSSUM and its related models as predictors of postoperative mortality and morbidity in patients undergoing surgery for gastro-oesophageal cancer: a systematic review. *World J Surg* 2010; **34**: 2076-2082 [PMID: 20556607 DOI: 10.1007/s00268-010-0685-z]
- 65 **Hayden JA**, Côté P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med* 2006; **144**: 427-437 [PMID: 16549855 DOI: 10.7326/0003-4819-144-6-200603210-00010]



Published by **Baishideng Publishing Group Inc**
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-3991568
E-mail: bpgoffice@wjgnet.com
Help Desk: <https://www.f6publishing.com/helpdesk>
<https://www.wjgnet.com>

