81842_Auto_Edited.docx

WORD COUNT 5572
TIME SUBMITTED 15-MAR-2023 09:59PM
PAPER ID 97649810

Name of Journal: *World Journal of Gastrointestinal Surgery*
Manuscript NO: 81842
Manuscript Type: SYSTEMATIC REVIEWS

Combined and intraoperative risk modelling for oesophagectomy: A systematic review

Preoperative and Intraoperative Oesophagectomy Risk Modelling

James Paul Grantham, Amanda Hii, Jonathan Shenfine

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 ASA score were also excluded. The review was conducted following the PRISMA 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 being promising in predicting mortality, namely the P-POSSUM and POSSUM scores. A further two studies, the IPF and eSAS 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

Grantham JP, Hii A, Shenfine J. Combined and intraoperative risk modelling for oesophagectomy: A systematic review. *World J Gastrointest Surg* 2023; In press

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.

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 and colleagues found that 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 and colleagues 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 and colleagues[13] attempted to appraise the scientific rigor with which these models were developed. There have been a plethora 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 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 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 and incorporates seven criteria[17, 18, 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 Minnie and colleagues[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 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 receiver operating characteristics (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 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)[27-31,46-51]. 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 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 and colleagues top-scored with a median score of 5. Six of the eleven models in this group were specifically developed for oesophagectomy with O-POSSUM also incorporating gastrectomy into its development set. The 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 patient mix with samples exceeding 100. Only the study from Huang and colleagues failed to outline their exclusion criteria and, along with the study from Yoshida, they were one of two studies not to report mortality rates. The study from Yoshida 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.

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.

Methodological Quality – Overall Performance

Overall, the median score of methodological quality for the 11 studies reviewed was 17 out 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 and colleagues. 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 days. 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 complication on the Clavien-Dindo scale. Owing to some studies assessing multiple models, there were thirteen instances of a model being tested for prediction of major complication. Six unique models were tested in total, with five of these having reported discrimination statistics as an area under the ROC curve. Two combined models, the eSAS and IPF nomograms developed by Xi and colleagues, 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 reporting with respect to calibration. Only the model proposed by Huang and colleagues 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 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 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 and colleagues. 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[52]. The pathway may be

altered or augmented to prophylactically address anticipated challenges, such as preoperative tube feeding in patients deemed susceptible to postoperative malnutrition[53]. Other high-risk individuals may even be deemed inappropriate for the standard enhanced recovery pathway altogether and receive a more conservative treatment strategy[54]. 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[55, 56, 57].

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[58, 59, 60]. 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[61,62]. This reflects the desire to individualise care and target resources as postoperative complications substantially increase expenditure per patient[64]. Therefore, if more intensive allocation of resources can be afforded to high-risk individuals, it may circumvent negative outcomes and prove economically beneficial[64].

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[65]. An oesophageal specific systematic review conducted in 2014 by Findlay and colleagues 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.

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[66]. 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[66]. 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 focussing 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 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

+ADw-html+AD4APA-p+AD4-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 exisitng 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.+ACY-nbsp+ADsAPA-/p+AD4APA-/html+AD4-

Research motivation

+ADw-html+AD4APA-p+AD4-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.+ACY-nbsp+ADsAPA-/p+AD4APA-/html+AD4-

Research objectives

+ADw-html+AD4APA-p+AD4-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 seconary 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.+ACY-nbsp+ADsAPA-/p+AD4APA-/html+AD4-

Research methods

+ADw-html+AD4APA-p+AD4-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 PRISMA guidelines. All of the included risk models

were appraised across five categories, namely clinical credibility, methodological quality, model performance, external validation and clinical effectiveness.+ACY-nbsp+ADsAPA-/p+AD4APA-/html+AD4-

Research results

+ADw-html+AD4APA-p+AD4-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 pubished 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 P-POSSUM and POSSUM scores. A further two studies, the IPF and eSAS 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.+ACY-nbsp+ADsAJg-nbsp+ADsAPA-/p+AD4APA-/html+AD4-

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

+ADw-html+AD4APA-p+AD4-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.+ADw-/p+AD4APA-/html+AD4-

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.

81842_Auto_Edited.docx

ORIGINALITY REPORT

0%
SIMILARITY INDEX

PRIMARY SOURCES

1	publishing.rcseng.ac.uk Internet	12 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	< 12 WORDS
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	< 12 WORDS