

86239_Auto_Edited-check.docx

Name of Journal: *World Journal of Gastrointestinal Surgery*

Manuscript NO: 86239

Manuscript Type: ORIGINAL ARTICLE

Retrospective Study

Establishment and application of three predictive models of anastomotic leakage after rectal cancer sphincter-preserving surgery

Li HY *et al.* Three predictive models of AL

Abstract

BACKGROUND

Anastomotic leakage (AL) occurs frequently after sphincter-preserving surgery for rectal cancer and has a significant mortality rate. There are many factors that influence the incidence of AL, and each patient's unique circumstances add to this diversity. The early identification and prediction of AL after sphincter-preserving surgery are of great significance for the application of clinically targeted preventive measures. Developing an AL predictive model coincides with the aim of personalised healthcare, enhances clinical management techniques, and advances the medical industry along a more precise and intelligent path.

AIM

To develop nomogram, decision tree, and random forest prediction models for AL following sphincter-preserving surgery for rectal cancer and to evaluate the predictive efficacy of the three models.

METHODS

The clinical information of 497 patients with rectal cancer who underwent sphincter-preserving surgery at Jincheng People's Hospital of Shanxi Province between January 2017 and September 2022 was analyzed in this study. Patients were divided into two groups: AL and no AL. Using univariate and multivariate analyses, we identified factors influencing postoperative AL. These factors were used to establish nomogram, decision tree, and random forest models. The sensitivity, specificity, recall, accuracy, and area under the receiver operating characteristic curve (AUC) were compared between the three models.

RESULTS

AL occurred in 10.26% of the 497 patients with rectal cancer. The nomogram model had an AUC of 0.922, sensitivity of 0.745, specificity of 0.966, accuracy of 0.936, recall of

0.987, and accuracy of 0.946. The above indices in the decision tree model were 0.919, 0.833, 0.862, 0.951, 0.994, and 0.955, respectively and in the random forest model were 1.000, 1.000, 1.000, 0.951, 0.994, and 0.955, respectively. The DeLong test revealed that the AUC value of the decision-tree model was lower than that of the random forest model ($P < 0.05$).

CONCLUSION

The random forest model may be used to identify patients at high risk of AL after sphincter-preserving surgery for rectal cancer owing to its strong predictive effect and stability.

Key Words: Cancer of rectum; Anastomotic leakage; Nomogram; Decision tree; Random forest

Li HY, Zhou JT, Wang YN, Zhang N, Wu SF. Establishment and application of three predictive models of anastomotic leakage after rectal cancer sphincter-preserving surgery. *World J Gastrointest Surg* 2023; In press

Core Tip: Anastomotic leakage (AL) is a very dangerous complication of rectal cancer surgery, which not only increases the recurrence rate of the tumor but also lowers the quality of life of affected patients. We examined the clinical data of 497 patients with rectal cancer to determine variables that influence AL. We established nomogram, decision tree, and random forest models to identify a prediction model tool for forecasting AL after rectal cancer surgery.

INTRODUCTION

According to the most recent data from 2020, colorectal cancer has become the third most common cancer worldwide. In China, colorectal cancer has the third highest incidence and fatality rate among all malignancies. Cases of rectal cancer account for

39% of total colorectal cancer cases, making it a serious public health issue in China^[1,2]. Radical surgery remains the first choice of treatment for rectal cancer, both for primary and secondary tumors^[3]. With advances in medical technology, the prognosis of patients undergoing rectal cancer surgery has significantly improved, and the sphincter preservation rate has also continuously improved^[4].

Nevertheless, anastomotic leakage (AL) remains the most common complication following sphincter-preserving surgery for rectal cancer. The perioperative mortality of patients with AL is as high as 18.6%, and these patients are more likely to experience other complications^[5,6]. The early identification and prediction of AL after sphincter-preserving surgery are of great significance for the application of clinically targeted preventive measures. At present, a logistic regression analysis of the factors influencing AL in clinical practice is performed^[7,8]; however, it cannot intuitively show the importance of each factor to the outcome. Therefore, this study developed nomogram, decision tree, and random forest models for predicting AL following rectal cancer sphincter-preserving surgery. The predictive power of the models were evaluated to identify a tool that would enable the identification of high-risk patients.

MATERIALS AND METHODS

Patient characteristics

For this retrospective analysis, we collected clinical data from 497 patients with rectal cancer who underwent sphincter-preserving surgery at Jincheng People's Hospital of Shanxi Province between January 2017 and September 2022. The inclusion criteria were: Rectal cancer diagnosed by colonoscopy or anal biopsy; tumor within 12 cm of the anal margin; age, ≥ 18 years old; and complete clinical data. The exclusion criteria were: Extensive tumor metastasis or the presence of other malignant tumors; a history of rectal surgery, anal stenosis or anal fistula; and conversion to laparotomy. The patients were divided into two groups, those with AL and those without (no AL). The study was approved by the Jincheng People's Hospital of Shanxi Province (JCPH.No20230407001)

and written informed consent was obtained from all study participants or their legal guardians.

Study variables

The following variables were analyzed: Sex, age, body mass index, diabetes mellitus, hypertension, smoking history, neoadjuvant treatment, hemoglobin level, albumin (Alb) level, tumor size, tumor-node-metastasis stage, American Society of Anesthesiologists score, tumor location, surgical approach, operative time, and blood loss.

Definition of AL

The diagnosis of AL was based on clinical manifestations (pain, persistent body temperature of $> 38^{\circ}\text{C}$, peritonitis, watery fecal matter, food residue in fecal matter, or pus in the drainage fluid), laboratory tests (elevated white blood cell count and neutrophil percentage), and imaging findings (computed tomography following an enema with a liquid, gas, or water-soluble contrast agent)^[9].

Statistical analysis

Statistical analyses were performed using SPSS for Windows version 26.0 (IBM Corp., Armonk, NY, United States). All continuous variable data are presented as mean \pm SD, and student's *t*-tests were used to compare differences. Data from discrete variables are presented as numbers and percentages, and the χ^2 test was used to assess differences between groups. Variables associated with AL were identified using univariate and multivariate logistic regression analyses.

The prediction models were constructed using the R software, and the data were randomly divided between a training set and a verification set in a 7:3 ratio. The nomogram was created using the 'rms' package, the decision tree with the 'rpart' package, and the random forest using the 'random Forest' package. The model with the best predictive effect was selected by comparing the sensitivity, specificity, accuracy,

recall rate, precision rate, and area under the receiver operating characteristic curve (AUC). AUCs were compared using the DeLong test. The statistical significance level was set at $P < 0.05$.

RESULTS

Patient characteristics

There were 271 men and 226 women among the 497 patients. The incidence of AL was 10.26% (51/497). Patients in the AL group had a mean age of 60.98 ± 10.83 years. The no AL group included 446 patients with a mean age of 59.27 ± 10.76 years.

Univariate analysis

Univariate analysis revealed statistically significant differences between the AL and no AL groups for the following variables: Sex, diabetes mellitus, smoking history, neoadjuvant treatment, Alb level, tumor size, and tumor location (Table 1).

Influencing factors of AL

Whether the patient had AL after surgery (not occurring = 0, occurring = 1) was used as the dependent variable, and the statistically significant factors identified by univariate analysis and shown in Table 1 (sex, diabetes mellitus, smoking history, neoadjuvant treatment, Alb level, tumor size, and tumor location) were used as independent variables. Table 2 lists the assignments of the indicators. A ² multivariate logistic regression analysis revealed sex [odds ratio (OR) = 3.656, 95% confidence interval (CI): 1.538-8.264, $P = 0.003$], diabetes mellitus (OR = 5.669, 95%CI: 2.455-13.092, $P < 0.001$), Alb level (OR = 0.898, 95%CI: 0.846-0.953, $P < 0.001$), tumor size (OR = 2.604, 95%CI: 1.840-3.684, $P < 0.001$), and tumor location (OR = 0.272, 95%CI: 0.180-0.413, $P < 0.001$) as factors that influence AL in patients with rectal cancer following sphincter-preserving surgery (Table 3).

Nomogram model

According to the results presented in Table 3, five variables (sex, diabetes mellitus, Alb level, tumor size, and tumor location) were used to construct a nomogram model for predicting AL after sphincter-preserving surgery for rectal cancer (Figure 1).

Decision tree model

A decision tree prediction model for AL after sphincter-preserving surgery for rectal cancer was also constructed, and four explanatory variables were screened: Tumor location, tumor size, Alb level, and sex. The results of the model showed that tumor location was the first-level factor influencing AL in patients with rectal cancer after sphincter-preserving surgery. The incidence of AL was 100% in patients with a distance of the tumor from the anal verge < 2.7 cm, and 78% in male patients with tumor location < 6.3 cm, Alb level < 41 g/L, and tumor size \geq 5.1 cm (Figure 2).

Random forest model

According to the change in the overall prediction accuracy of the best model, the variables affecting AL in patients with rectal cancer after sphincter-preserving surgery were tumor location, tumor size, diabetes mellitus, sex, and Alb level (Figure 3).

Evaluation of the prediction efficacy of the three models

In the training dataset, the overall performance of the random forest model in predicting AL after sphincter-preserving surgery for rectal cancer was comparable to that of the decision tree model. The AUC of the random forest model was significantly higher than that of the decision tree model ($Z = -2.836$, $P = 0.004$) (Table 4 and Figure 4). In the validation dataset, the overall effectiveness of the three models was equivalent (Table 5 and Figure 5).

DISCUSSION

Rectal carcinoma is a prevalent cancer of the digestive system. Currently, the treatment of this disease is mainly surgical^[10]. AL is one of the most common and dangerous

complications associated with rectal cancer surgery. According to studies, the incidence of AL after rectal cancer surgery ranges from 2.6% to 19.0%^[11]. AL not only affects recovery from surgery, but also leads to a variety of complications, such as intra-abdominal abscesses, diffuse peritonitis, and sepsis, and can even cause tumor recurrence. In severe cases, secondary surgery is required, which worsens patient survival rates^[12]. Therefore, it is critical to identify the factors that influence the development of AL in patients with rectal cancer after sphincter-preserving surgery and provide targeted interventions to reduce its occurrence.

The incidence of postoperative AL was analyzed in 497 patients with rectal cancer admitted to our hospital. A total of 51 patients developed postoperative AL, representing an incidence of 10.26%. This is similar to the incidence of AL reported by Degiuli *et al*^[13] in 5398 patients with rectal cancer (10.2%), but lower than that reported by Peltrini *et al*^[14] in 367 patients with rectal cancer (17.4%). These differences may be related to factors such as inclusion criteria and different populations. AL after sphincter-preserving surgery for rectal cancer results from multiple factors, including patient characteristics, tumor status, and operation-related factors. The results of this study showed that sex, diabetes mellitus, Alb level, tumor size, and tumor location influence AL.

According to this study, males were 3.656 times more likely to have postoperative AL than women, which is consistent with the results of most studies^[15,16]. As the male pelvis is narrow, the visual field is not fully exposed during the operation, making surgery more challenging; this may cause the rectal stump to retract and cause secondary injury^[17].

Penna *et al*^[18] confirmed that diabetes mellitus is an independent risk factor for AL after transanal total mesorectal excision. Additionally, this study reported that individuals with diabetes had an AL risk 5.669 times greater than that of non-diabetic patients. Diabetes mellitus can affect the anastomotic blood supply as uncontrolled hyperglycemia leads to vascular damage, reduced blood flow, and cellular

accumulation of toxic glucose-derived metabolites, resulting in a significant decrease in anastomotic healing and the ability to fight infection^[19].

According to the findings of this study, which are consistent with those of Shimura *et al*^[20], there is a direct association between preoperative Alb levels and the risk of developing postoperative AL. This may be due to a low perioperative nutritional status leading to reduced immune function and a greater risk of infection and the spread of infection^[21]. Yasui *et al*^[22] found that patients with tumor sizes ≥ 4 cm were more likely to develop AL. This study also found that the probability of postoperative AL increases with tumor size. As more tissue has to be removed during surgery, more damage occurs and the chance of developing AL subsequently increases^[23].

Both domestically and internationally, there is consensus on the impact of tumor location on AL following low anterior resection of rectal cancer; that is, the closer the tumor is to the anal margin, the higher the risk of AL^[24-26]. The same conclusion was reached in the present study. This may be because the closer the tumor is to the anus, the larger the wound during resection. In addition, intraoperative electrocoagulation damage to the tissue and blood vessels causes exudation and bleeding, which reduces the blood supply to the anastomosis and increases the risk of postoperative AL^[27].

Currently, there is no consensus regarding whether neoadjuvant therapy increases the incidence of AL. Arezzo *et al*^[12] analyzed the effects of short- and long-term radiotherapy on AL and showed that the risk of postoperative AL significantly increased in patients receiving short-term radiotherapy. However, Chang *et al*^[28] found no difference in the incidence of AL between patients with rectal cancer who underwent chemoradiotherapy before surgery, and those who did not. The present study also found that neoadjuvant therapy had no effect on AL occurrence.

With the development of computer software and artificial intelligence, machine learning has become a new direction for medical research. Studies have used machine learning models to predict the risk of anti-tuberculosis drug-induced liver injury in inpatients with tuberculosis^[29], and machine learning can also predict the risk of essential hypertension^[30]. In this study, five indicators with statistically significant

differences in the multivariate analysis were used to establish three prediction models of AL after sphincter-preserving surgery for rectal cancer using machine learning algorithms. In the training dataset, the sensitivity, specificity, accuracy, recall rate, and precision rate of the nomogram were lower than those of the random forest and decision tree models, and the prediction efficacy of the random forest model was better than that of the decision tree. But the sensitivity, specificity, and AUC of the random forest model reached 1, indicating that the random forest model may have overfitting or the generalization effect may be poor due to insufficient data in the training set. In the validation dataset, the three models exhibited similar prediction performances. Although the random forest model has better predictive performance, it also has some disadvantages. As a result, in practice, each of the three prediction models has benefits and limitations, and the most appropriate method should be chosen based on the situation.

As this was a single-center retrospective study, it has several limitations. The representativeness of single-center research is limited, and there may be some bias owing to time constraints. This study did not analyze additional risk factors for AL after sphincter-preserving surgery for rectal cancer, and the constructed model may have been overfitted. Future research will use a larger and more comprehensive sample set and multicenter studies to verify and build a more complete prediction model.

CONCLUSION

Overall, AL is a serious complication of rectal cancer surgery, with a high incidence rate. In this study, nomogram, random forest, and decision tree prediction models of AL after sphincter-preserving surgery for rectal cancer were established using machine learning algorithms. The random forest model was found to have excellent predictive effect and stability, and might serve as a reference for the clinical identification of high-risk groups for AL following sphincter-preserving surgery for rectal cancer.

ARTICLE HIGHLIGHTS

Research background

With advances in medical technology, the success rate of sphincter-preserving surgery in patients with rectal cancer is increasing. However, anastomotic leakage (AL) remains a devastating complication.

Research motivation

AL significantly lowers patients' quality of life. This study examines the elements that influence AL and establishes models to help doctors predict whether patients will develop AL, allowing the timely adoption of preventive measures.

Research objectives

This study aimed to identify the characteristics that influence AL and utilize these factors to build a prediction model for AL after sphincter-preserving surgery for rectal cancer.

Research methods

The clinical data of patients with rectal cancer who underwent sphincter-preserving surgery at our institution in the past five years were examined to analyze the factors influencing AL; nomogram, decision tree, and random forest prediction models were established; and the predictive efficacy of the three models was compared.

Research results

The factors influencing AL after sphincter-preserving surgery for rectal cancer were sex, diabetes mellitus, albumin level, tumor size, and tumor location. To predict the probability of postoperative AL, we constructed nomogram, decision tree, and random forest models.

Research conclusions

This study compared the predictive efficacy of the three prediction models. The random forest model performed the best and may be a useful alternative tool for predicting patients at a high risk of AL.

Research perspectives

Future research will include larger and more comprehensive cohorts across multiple centers, and build a more complete prediction model.

ORIGINALITY REPORT

2%

SIMILARITY INDEX

PRIMARY SOURCES

1	pure.uva.nl Internet	18 words — 1%
2	www.ncbi.nlm.nih.gov Internet	17 words — 1%
3	Yexin Su, Hongyue Zhao, Pengfei Liu, Linhan Zhang, Yuying Jiao, Peng Xu, Zhehao Lyu, Peng Fu. "A nomogram model based on MRI and radiomic features developed and validated for the evaluation of lymph node metastasis in patients with rectal cancer", Abdominal Radiology, 2022 Crossref	16 words — 1%

EXCLUDE QUOTES ON
EXCLUDE BIBLIOGRAPHY ON

EXCLUDE SOURCES < 15 WORDS
EXCLUDE MATCHES < 15 WORDS