Connect the Dots

Ingwersen, Erik Wouter

2024

DOI (link to publisher)
10.5463/thesis.749

document version
Publisher's PDF, also known as Version of record

Connect the Dots

Towards a data-driven and multidisciplinary approach
for the management of anastomotic leakage

Erik Wouter Ingwersen

Colofon

VRIJE UNIVERSITEIT

Connect the Dots

Towards a data-driven and multidisciplinary approach
for the management of anastomotic leakage

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. J.J.G. Geurts,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geneeskunde
op vrijdag 1 november 2024 om 9.45 uur
in een bijeenkomst van de universiteit,
De Boelelaan 1105

door

Erik Wouter Ingwersen

geboren te Haarlem

promotor:	prof.dr. G. Kazemier prof.dr.
copromotor:	dr. F. Daams
promotiecommissie:	prof. dr. J. Berkhof prof.dr. E. Consten prof.dr. H.C. Van Santvoort dr. J.B. Tuynman dr. M. Lahaye

Table of contents

1

Chapter 1

General introduction and thesis outline

Anastomosis

Colorectal and pancreatic cancer continue to pose significant health challenges worldwide, with rising incidence rates in many countries. In the Netherlands, these cancers represent a substantial burden on public health, with colorectal cancer being the second leading cause of cancer related death and pancreatic cancer exhibiting dismal 5-year survival rates of around 8.5%.[1,2] To achieve potential curative outcomes, surgical resection plays a pivotal role in the management of both malignancies. Subsequent to bowel or pancreatic head resections, the establishment of an anastomosis is required for restoring gastrointestinal continuity. Colorectal cancer patients undergoing resection may undergo various types of anastomoses, including ileocolic, colo-colic, ileorectal, or colorectal anastomosis, depending on tumor site and the specific surgical procedure performed. For patients undergoing pancreatic head resection, reconstruction between the remnant pancreas and jejunum is typically created to ensure a drainage of secreted pancreatic juice into the bowel.

Anastomotic failure

Dehiscence of the anastomosis, also known as anastomotic leakage, is a potentially devastating complication that may arise after the construction of one of the abovementioned anastomoses. It stands as the single most critical determinant of short- and long-term outcomes following colorectal and pancreatic head surgery.[3,4] Anastomotic leakage leads to increased mortality rates, a higher need for re-interventions, prolonged length of hospital stay, and intensive care unit use.[4,5] Moreover, its impact extends beyond the immediate postoperative phase, potentially affecting oncological outcomes with a heightened risk of cancer recurrence.[6,7] This has underscored the importance of clinical vigilance regarding anastomotic leakage. Additionally, the financial burden associated with anastomotic leakage is substantial, with significantly higher health care costs compared to uncomplicated postoperative courses.[8]

The reported incidence of colorectal anastomotic leakage (CAL) varies considerably, ranging from 1% and 19%.[9] Such disparity is mainly attributed to the lack of a standardized definition for CAL.[10] The clinical severity of CAL spans a broad spectrum, ranging from minor, contained leaks with limited symptoms to widespread peritoneal contamination, severe sepsis, multiple organ failure, and even death.[9] Severe leakage of the pancreatojejunostomy, known as a clinically relevant postoperative pancreatic fistula (CR-POPF), is often classified following the International Study Group in Pancreatic Surgery (ISGPS) criteria. A CR-POPF a is defined as a drain output of any measurable volume of fluid with an amylase level >3 times the upper limit of institutional normal serum amylase activity, associated with a clinically relevant development/condition related directly to the postoperative pancreatic fistula.[11]

Risk factors

CAL is influenced by various risk factors, which have been extensively studies over the past decades. Identification of these risk factors holds great importance as it enables better prevention, early diagnosis, and treatment of this serious complication. Risk factors for CAL can be divided into pre- and intraoperative factors and modifiable and non-modifiable factors. Several preoperative factors have been linked to an increased risk of CAL. These include: male sex, the presence of comorbidities, tumor stage and site, administration of neoadjuvant radiotherapy, weight and nutrition status, and the American Society of Anesthesiologists score. Lifestyle such as alcohol consumption, drug use, and smoking have also been associated with an increased risk of anastomotic leakage.[9] During the surgical procedure, certain factors may also contribute to the occurrence of CAL. Intraoperative contamination, blood loss, and the chosen surgical approach are elements that can influence the development of CAL.[9] Moreover, recent research by Huisman et al. has highlighted the significance of specific perioperative modifiable risk factors, including low hemoglobin levels, contamination of the operative field, hyperglycemia, duration of surgery exceeding three hours, administration of vasopressors, inadequate timing of antibiotic prophylaxis, and application of epidural analgesia.[12] Identifying and managing these factors, if, may help to reduce the likelihood of CAL. Several risk factors have been associated with CR-POPF. These include male sex, body mass index greater than 25kg/m^2, pancreatic duct diameter less than 3mm and a soft pancreatic texture, and receiving a blood transfusion. Interestingly, the administration of neoadjuvant chemoradiotherapy has shown to decrease the risk of a CR-POPF.[13]

Benefit of accurate prediction

Despite numerous strategies and recent advancements in colorectal and pancreatic surgery, the Dutch National Audit reveals a consistent percentage of CAL and CR-POPF over the past decade.[14,15] This underscores the critical need for an accurate risk prediction model to facilitate improved prevention and treatment of these complications.[16]

In patients undergoing colorectal resection, the use of a diverting stoma in high-risk individuals has been considered as a measure to reduce the clinical impact of CAL.[17] However, it is essential to acknowledge that a diverting stoma negatively impacts patients' quality of life.[18] Presently, the decision to employ a diverting stoma is subjective and varies among surgeons, necessitating the development of an objective risk assessment tool based on a prediction model to guide clinical practice effectively. Similarly, in high-risk patients for CR-POPF, prophylactic somatostatin analogues have been administrated to reduce the incidence of this complication.[19] Additionally, in certain cases, a total pancreatectomy, where no pancreatojejunostomy is constructed, might be considered as an alternative approach.[20] However, despite the identification of various risk factors, accurately and timely predicting the occurrence of CAL or CR-POPF

for an individual patient remains challenging. This highlights the urgent need for the development of more sophisticated and reliable prediction models to enhance patient outcomes.

Artificial intelligence

As in many other medical fields, the emergence of 'Big Data' in surgery has been facilitated by technological advancements in data processing, storage, and analysis.[21] The application of artificial intelligence (AI) algorithms to process and analyze large quantities of data has already shown promising results in the prediction of postoperative complications.[22] Artificial intelligence, as wide branch of computer sciences, encompasses the ability to learn, adapt, and perform tasks at human-level intelligence, including cognitive functions. Machine learning, a subfield of AI, is employed to analyze structured data, such as patients' characteristics or tumor profiles, to develop prediction models. In imaging analysis, AI techniques, such as radiomic or deep learning features, can extract a vast number of features from medical images using data-characterization algorithms. These features have the potential to uncover new parameters that may not be visible to the naked eye.[23] The adoption of AI in surgery has gained considerable popularity in recent years and is poised to transform daily clinical practice. With the ability to analyze large amounts of data rapidly, AI can assist healthcare professionals in making more informed decision, optimizing patient care, and predicting and preventing potential complications after surgery.[24]

Despite the promising prospects of AI in surgery, there are inherent challenges and limitations that must be addressed. A fundamental issue in the implementation of AI for predictive modelling in clinical purposes is the lack of methodological quality in studies developing such models.[25] Poor methodological quality increases the risk of bias and strongly impedes the current application of machine learning or deep learning models in a clinical setting.[25] To ensure the reliability and validity of AI-generated predictions, rigorous validation and standardization of AI algorithms are essential. Moreover, AI models heavily rely on the quality and diversity of data used for training. If the data are incomplete, biased, or not representative of the target population, the AI model's performance and generalizability may be compromised. Addressing these data-related challenges is crucial to developing robust and clinically applicable AI models.

Aim of this thesis and thesis outline

Anastomotic failure after pancreatic and colorectal surgery remains a significant concern, despite extensive clinical research aimed at reducing its incidence. The high incidence of this complication requires accurate prediction for preventive treatment

and early detection. AI algorithms can effectively analyze large datasets and develop prediction models for anastomotic leakage. While several AI-based prediction models have been developed, the availability of well-established and clinically applicable models for anastomotic leakage remains limited. One of the main objectives of this thesis is to address this gap by conducting extensive research to identify advances and current pitfalls in AI implementation. Using machine learning machine learning techniques, prediction models for CAL and CR-POPF were developed and externally validated. Additionally, this thesis assesses the current state of data science in surgery and offers recommendations to enhance the clinical applicability of AI-based models. Throughout the thesis, interdisciplinary collaboration played a crucial role, involving experts from surgery, radiology, patients' associations, biomedical statistics, data science, and the IT department. Connecting these dots is essential to effectively use AI in surgery, ultimately improving the daily clinical practice.

Part I – Impact of anastomotic leakage and mitigation strategies

Since surgeons perform restorative colorectal, leakage of the anastomosis is the most dreaded complications for both surgeon and patient. To safeguard the primary anastomosis, the construction of a diverting stoma is often considered. However, the decision on when and in which patients to use a diverting stoma remains a topic of ongoing debate. In **Chapter 2**, pre- and intraoperative factors associated with a diverting stoma are identified. By analyzing these factors, insight is provided into the appropriate use of diverting stomas in specific patient populations. In **Chapter 3,** a comprehensive assessment of a decade of rectal cancer surgery is conducted, focusing on diverting stoma construction and CAL. Utilizing propensity score matching, the association between diverting stomas and the incidence as well as the severity of CAL is investigated. **Chapter 4** delves into the impact of the merger of two surgical teams, those of the the Academic Medical Center and Vrije Universiteit Medical Center into the Amsterdam University Medical Center on surgical outcomes after esophagastric and hepato-pancreato-biliary surgery. This investigation aims to understand the effects of organizational changes on patients' outcomes in these complex surgical procedures. Furthermore, in **Chapter 5,** the long-term consequences of CAL on oncological outcomes and the presence of stoma are scrutinized. It seeks to gain a deeper understanding of the lasting impact of CAL on patient prognosis and quality of life.

Part II – The use of artificial intelligence in predictive modelling

The increasing application of AI in predictive modelling has opened new possibilities in healthcare. However, the transition of these models into clinical practice is still limited, and several challenges must be addressed. In this part of the thesis, the current state of AI-driven predictive modelling is explored and its implementation is critically assessed. **Chapter 6** reviews the current literature on the use of machine learning

models in clinical practice for the prediction of postoperative complications after major abdominal surgery. Strengths and limitations are identified, providing a valuable insight for improving the clinical application. **Chapter 7** focuses on the application of radiomic and deep learning features on preoperative imaging to predict a CR-POPF. Through an in-depth review of current literature, the predictive performances of these models are analyzed. Additionally, the methodological quality of the included studies is critically assessed, enhancing the reliability and validity of the findings. In **Chapter 8** the current state of AI applications in the operating room are described. By exploring the various ways AI is used during surgical procedures, the potential benefits and challenges in real-time clinical decision-making are delineated.

Part III – Development of prediction models

Despite the growing application of AI in predictive modelling, there are no publicly available and externally validated prediction models for CAL and CR-POPF using AI. In this section the development and validation of novel prediction models using advanced AI techniques are presented. **Chapter 9** compares the predictive performance of a machine learning model with a conventional logistic regression model in predicting CR-POPF after a pancreatoduodenectomy. Leveraging data from the Dutch Pancreatic Cancer Audit, strengths and limitations of AI-based models compared to traditional approaches are identified. In **Chapter 10**, the Radiomics preoperative-Fistula Risk Score (RAD-FRS), a novel prediction model for pancreatic fistula is introduced. The RAD-FRS is the first externally validated and preoperatively available model, using radiomic features extracted from preoperative CT-scans. **Chapter 11** showcases another prediction model for a pancreatic fistula is developed: the Artificial Intelligence-based fully automatic prediction of postoperative pancreatic fistula (AI-FRS). This innovative model automates the segmentation of the pancreas, extracts relevant features, and generate a risk score. By streaming the process with AI, a more efficient and reliable prediction tool is provided. **Chapter 12** describes the study protocol of the A1Check study, a prospective study that externally validates a machine learning model using pre- and intraoperative data to predict CAL. This study aims to validate and optimize the performance of the AI-based model for clinical use.

References

1. Registry NKR. National Data [Available from: https://iknl.nl/nkr-cijfers.]
2. Latenstein AEJ, van der Geest LGM, Bonsing BA, et al. Nationwide trends in incidence, treatment and survival of pancreatic ductal adenocarcinoma. *Eur J Cancer*. 2020;125:83-93.
3. Chiarello MM, Fransvea P, Cariati M, et al. Anastomotic leakage in colorectal cancer surgery. *Surg Oncol*. 2022;40:101708.
4. Smits FJ, Henry AC, Besselink MG, et al. Algorithm-based care versus usual care for the early recognition and management of complications after pancreatic resection in the Netherlands: an open-label, nationwide, stepped-wedge cluster-randomised trial. *Lancet*. 2022;399(10338):1867-75.
5. Meyer J, Naiken S, Christou N, et al. Reducing anastomotic leak in colorectal surgery: The old dogmas and the new challenges. *World J Gastroenterol*. 2019;25(34):5017-25.
6. Koedam TWA, Bootsma BT, Deijen CL, et al. Oncological Outcomes After Anastomotic Leakage After Surgery for Colon or Rectal Cancer: Increased Risk of Local Recurrence. *Ann Surg*. 2022;275(2):e420-e7.
7. Nagai S, Fujii T, Kodera Y, et al. Recurrence pattern and prognosis of pancreatic cancer after pancreatic fistula. *Ann Surg Oncol*. 2011;18(8):2329-37.
8. La Regina D, Di Giuseppe M, Lucchelli M, et al. Financial Impact of Anastomotic Leakage in Colorectal Surgery. *J Gastrointest Surg*. 2019;23(3):580-6.
9. McDermott FD, Heeney A, Kelly ME, et al. Systematic review of preoperative, intraoperative and postoperative risk factors for colorectal anastomotic leaks. *BJS*. 2015;102(5):462-79.
10. Rahbari NN, Weitz J, Hohenberger W, Heald RJ, et al. Definition and grading of anastomotic leakage following anterior resection of the rectum: A proposal by the International Study Group of Rectal Cancer. *Surgery*. 2010;147(3):339-51
11. Bassi C, Marchegiani G, Dervenis C, et al. The 2016 update of the International Study Group (ISGPS) definition and grading of postoperative pancreatic fistula: 11 Years After. *Surgery*. 2017;161(3):584-91.
12. Huisman DE, Reudink M, van Rooijen SJ, et al. LekCheck: A Prospective Study to Identify Perioperative Modifiable Risk Factors for Anastomotic Leakage in Colorectal Surgery. *Ann Surg*. 2022;275(1):e189-e97.
13. Zhang B, Yuan Q, Li S, et al. Risk factors of clinically relevant postoperative pancreatic fistula after pancreaticoduodenectomy: A systematic review and meta-analysis. *Medicine*. 2022;101(26).
14. Arron MN, Greijdanus NG, Ten Broek RP, et al. Trends in risk factors of anastomotic leakage after colorectal cancer surgery (2011–2019): A Dutch population-based study. *Colorectal Dis*. 2021;23(12):3251-61.
15. van Dongen JC, van Dam JL, Besselink MG, et al. Fistula Risk Score for Auditing Pancreatoduodenectomy: The Auditing-FRS. *Ann Surg*. 2023;278(2):e272-e7.
16. Zarnescu EC, Zarnescu NO, Costea R. Updates of Risk Factors for Anastomotic Leakage after Colorectal Surgery. *Diagnostics (Basel)*. 2021;11(12).
17. Munshi E, Lydrup M-L, Buchwald P. Defunctioning stoma in anterior resection for rectal cancer does not impact anastomotic leakage: a national population-based cohort study. *BMC Surg*. 2023;23(1):167.
18. Herrle F, Sandra-Petrescu F, Weiss C, et al. Quality of Life and Timing of Stoma Closure in Patients With Rectal Cancer Undergoing Low Anterior Resection With Diverting Stoma: A Multicenter Longitudinal Observational Study. *Dis Colon Rectum*. 2016;59(4):281-90.

19. Li T, D'Cruz RT, Lim SY, et al. Somatostatin analogues and the risk of post-operative pancreatic fistulas after pancreatic resection - A systematic review & meta-analysis. *Pancreatology*. 2020;20(2):158-68.
20. Marchegiani G, Perri G, Burelli A, et al. High-risk Pancreatic Anastomosis Versus Total Pancreatectomy After Pancreatoduodenectomy: Postoperative Outcomes and Quality of Life Analysis. *Ann Surg*. 2022;276(6):e905-e13.
21. Prien C, Lincango EP, Holubar SD. Big Data in Surgery. *Surg Clin North Am*. 2023;103(2):219-32.
22. Hassan AM, Rajesh A, Asaad M, et al. Artificial Intelligence and Machine Learning in Prediction of Surgical Complications: Current State, Applications, and Implications. *Am Surg*. 2023;89(1):25-30.
23. van Timmeren JE, Cester D, Tanadini-Lang S, et al. Radiomics in medical imaging—"how-to" guide and critical reflection. *Insights Imaging*. 2020;11(1):91.
24. Kulkarni S, Seneviratne N, Baig MS, et al. Artificial Intelligence in Medicine: Where Are We Now? *Acad Radiol*. 2020;27(1):62-70.
25. Andaur Navarro CL, Damen JAA, Takada T, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. *BMJ*. 2021;375:n2281.

Part I

Impact of anastomotic leakage and mitigation strategies

2

Chapter 2

The Selective Use of a Diverting Stoma in Rectal Surgery

Daitlin E. Huisman
Erik W. Ingwersen
Muriël Reudink
Boukje T. Bootsma
Gerrit D. Slooter
Jan Willem T. Dekker
Freek Daams
On behalf of the LekCheck study group

Journal of Gastrointestinal Surgery 2022 Jul; 26 (7): 1509 -1512

Introduction

A diverting stoma is recommended by some as routine procedure to lower incidence of AL and mitigate its consequences. Others, however, state that the diverting stoma is overused, as they describe no differences in AL rate between patients with and without a diverting stoma.[1] Next to this, many patients experience stoma-related morbidity such as skin irritation, dehydration, stoma site complications, psychological distress and reversal surgery with potential complications. Ultimately, many patients never undergo reversal of their diverting stoma.[2] Between 2007 and 2019 89% of the patients who underwent rectal cancer surgery in Australia or New Zealand received a diverting stoma.[3] In the Netherlands, however, the incidence of a diverting stoma in rectal surgery is considerably low (40% in 2016) and reduced in the last decade.[4] The relatively low application rate is possibly due to a selection of patients. This study was designed to identify patient characteristics and intraoperative conditions related to the presence of a diverting stoma and the impact on anastomotic leakage.

Method

The data were used of a prospective, observational study from January 2016 to December 2019 from fourteen hospitals in four countries. This study was an additional sub analysis of the LekCheck study.[5] All patients undergoing rectum resection with primary anastomosis were included.

Results

A total of 351 patients were included for this sub-study. A diverting stoma was created in 97 patients (27.6%). The following seven variables were associated with a diverting stoma in univariate analysis: smoking status, neoadjuvant therapy, American Joint Committee on Cancer stage, tumor distance, fluid administration, blood loss and an intraoperative event. In the multivariate analysis the following variables were independently associated with a diverting stoma: tumor distance ($p< 0.001$), neoadjuvant therapy ($p< 0.001$), blood loos ($p= 0.003$), fluid administration ($p= 0.003$) and an intraoperative event ($p= 0.022$). Anastomotic leakage occurred in nine patients (9.3%) with a diverting stoma and in 34 patients (13.4%) without ($p= 0.297$). In patients with anastomotic leakage, less interventions were necessary when a diverting stoma was constructed ($p= 0.001$).

Discussion

This study found that the following factors were independently associated with a diverting stoma: tumor distance, neoadjuvant therapy, blood loss during surgery, fluid administration and an intraoperative event. The use of a diverting stoma in this study was relatively low (27.6%) and although the anastomotic leak rate was lower in patients with a diversion, this difference was not statistically significant. Other authors found that selective use of diverting stomas led to the same incidence of AL compared to policies in which diverting stoma were more routinely used.[6] Proper application of selective use would drastically lower the burden of the stoma, preventing stoma-related complications (e.g., parastomal hernias, dehydration, stoma prolapse), discomfort and costs, for many patients. On the other hand, it can potentially reduce complications in patients who are at high risk for AL, since less reintervention were necessary in patients with AL and a diverting stoma.

Conclusion

The study showed differences in patient characteristics and intraoperative variables in patients with and without a diverting stoma. A diverting stoma showed a protective effect as the impact of AL was less severe, resulting in less reinterventions. Selective use is therefore suggested, since it prevents unnecessary application while protecting patients. Current focus should be on techniques to identify patients with increased risk as soon as the rectum resection, in order to apply the protective stoma restrictively in patients at risk.

Table 1, patient characteristics of patients with and without diverting stomas, compared in a univariate and multvariate analysis.

					Univariate analysis		Multivariate analysis*	
Variable	**Diverting stoma (n=97)**	**No diverting stoma (n=254)**	**Total (n= 351)**	**Missing**	**OR (95% CI)**	**P value**	**OR (95% CI)**	**P value**
Age (mean in years)	64.6±11.0	65.8±12.2	65.4±11.9	*n= 2*		0.289		
Sex						0.174		
Female	35 (36.1%)	112 (44.1%)	147 (41.9%)					
Male	62 (63.9%)	142 (55.9%)	204 (58.1%)					
BMI (mean in kg/m^2)	26.3±4.4	26.7±4.4	26.6±4.4	*n= 5*		0.522		
ASA classification				*n= 3*		0.552		
<3	77 (80.2%)	209 (82.9%)	286 (82.2%)					
≥3	19 (19.8%)	43 (17.1%)	62 (17.8%)					
Diabetes mellitus				*n= 2*		0.109		
No	80 (82.5%)	224 (88.9%)	304 (87.1%)					
Yes	17 (17.5%)	28 (11.1%)	45 (12.9%)					
Current smoker						**0.035**		0.131
No	74 (80.4%)	228 (88.9%)	282 (87.5%)		1		1	
Yes	18 (18.6%)	26 (10.2%)	44 (12.5%)		2.00 (1.04 - 3.84)		1.81 (0.84 – 3.95)	
Pack years						0.603		
<15 pack years	52 (53.6%)	144 (56.7%)	196 (55.8%)					
≥15 pack years	45 (46.4%)	110 (43.3%)	155 (44.2%)					
Alcohol use						0.257		
<3 units per day	87 (89.7%)	216 (85.0%)	303 (86.3%)					
≥3 units per day	10 (10.3%)	38 (15.0%)	48 (13.7%)					
Disease				*n= 1*		0.224		

Malignant	92 (94.8%)	230 (90.6%)	322 (92.0%)					
Benign	5 (5.2%)	23 (9.4%)	28 (8.0%)					
Neoadjuvant therapy				*n= 4*		**<0.001**		**<0.001**
None	39 (40.6%)	179 (71.3)	218 (62.8%)				1	
5 x 5 radiotherapy	37 (38.5%)	53 (21.1%)	90 (25.9%)				3.86 (2.01 – 7.42)	
Chemotherapy	1 (1.0%)	2 (0.8%)	3 (0.9%)					
Chemoradiotherapy	19 (19.8%)	17 (6.8%)	36 (10.4%)					
AJCC stage				*n= 32*		**0.023**		0.613
I & II (T1-4N0M0)	28 (31.1)	103 (45.0%)	131(41.1%)		1		1	
III & IV (T1-4N1-2M0-1)	62 (68.9%)	126 (55.0%)	188 (58.9%)		1.81 (1.08 - 3.03)		0.84 (0.42 – 1.65)	
Tumour distance from AV						**<0.001**		**<0.001**
>10 cm	46 (47.4%)	158 (68.7%)	192 (61.0%)	*n=36*			1	
≤10 cm and >5cm	35 (41.2%)	61 (26.5%)	96 (30.5%)				2.46 (1.64 – 3.73)	
≤5cm	16 (16.8%)	11 (4.8%)	27 (8.6%)					

OR, odds ratio; CI, confidence-interval; BMI, body mass index; ASA, American Society of Anesthesiology score; AJCC, American Joint Committee of Cancer; TNM, Tumour Node Metastasis; AV, anal verge; cm; centimeter. Bold values are statistically significant ($p < 0.05$). Adjusted for: current smoker, neoadjuvant therapy, AJCC stage and tumour distance from AV.

Table 2, intraoperative factors of patients with and without diverting stomas, compared in a univariate and multivariate analysis

					Univariate analysis		Multivariate analysis*	
Variable	**Diverting stoma (n=97)**	**No diverting stoma (n=254)**	**Total (n= 351)**	**Missing**	**OR (95% CI)**	**P value**	**OR (95% CI)**	**P value**
Use of vasopressor						0.555		
No	67 (69.1%)	167 (65.7%)	234 (66.7%)					
Yes	30 (30.9%)	87 (34.3%)	117 (33.3%)					
Epidural use				*n= 11*		0.337		
No	55 (59.1%)	160 (64.8%)	215 (63.2%)					
Yes	38 (40.9%)	87 (35.2%)	125 (36.8%)					
Hemoglobin						0.363		
<6.0 mmol/L female or <6.5 mmol/L male	3 (3.1%)	4 (1.6%)	7 (2.0%)					
≥6.0 mmol/L female or ≥6.5 mmol/L male	94 (96.9%)	250 (98.4%)	344 (98.0%)					
Fluid administration						**<0.001**		**0.003**
<1000 ml	20 (20.6%)	104 (40.9%)	124 (35.3%)		1		1	
≥1000 ml	77 (79.4%)	150 (59.1%)	227 (64.7%)		2.67 (1.54 - 4.64)		2.50 (1.37 – 4.57)	
Blood loss						**<0.001**		**0.003**
<100 ml	28 (28.9%)	140 (55.1%)	168 (47.9%)		1		1	
≥100 ml	69 (71.1%)	114 (44.9%)	183 (52.1%)		3.03 (1.83 - 5.01)		2.34 (1.34 – 4.01)	
Intraoperative event				*n= 6*		**0.001**		**0.022**
No	78 (80.4%)	229 (92.3%)	307 (89.0%)		1		1	
Yes	19 (19.6%)	19 (7.7%)	38 (11.0%)		2.94 (1.48 - 5.83)		2.410 (1.13 – 5.13)	
Approach						0.797		
Open	9 (9.3%)	18 (7.3%)	27 (7.8%)	*n= 6*				
Laparoscopy	84 (86.6%)	218 (87.9%)	302 (87.5%)					
Laparoscopy with conversion	4 (4.1%)	12 (4.8%)	16 (4.6%)					

OR, odds ratio; CI, confidence-interval; MAP, mean arterial pressure. Bold values are statistically significant ($p < 0.05$). Adjusted for: current smoker, neoadjuvant therapy, AJCC stage and tumour distance from AV.

Table 3, patients with and without a diverting stoma and occurrence of anastomotic leakage, days until leakage was detected and severity. Reinterventions were scored when Clavien-Dindo was grade 3 or higher.

	Diverting stoma and AL (n= 9)	**No diverting stoma and AL (n= 34)**	**Missing**	**P value**
Anastomotic leakage	9	34		0.297*
Days until anastomotic leakage was detected			*n= 5*	
<7 days	4 (57.1%)	24 (77.4%)		0.257^
≥7 days	3 (42.9%)	7 (22.6%)		
Reintervention needed				**0.046^**
Yes	5 (55.6%)	30 (88.2%)		
No	4 (44.4%)	4 (11.8%)		
Death within 30 days postoperatively				0.370^
Yes	1 (11.1%)	1 (2.9%)		
No	8 (88.9%)	33 (97.1%)		

AL, anastomotic leakage. *X^2 test. ^Fisher's exact test.
Entries in boldface is due to the significance P value

References

1. Snijders HS, van den Broek CB, Wouters MW, Meershoek-Klein Kranenbarg E, Wiggers T, Rutten H, et al. An increasing use of defunctioning stomas after low anterior resection for rectal cancer. Is this the way to go? European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 2013;39(7):715-20.
2. Emmanuel A, Chohda E, Lapa C, Miles A, Haji A, Ellul J. Defunctioning Stomas Result in Significantly More Short-Term Complications Following Low Anterior Resection for Rectal Cancer. World J Surg. 2018;42(11):3755-64.
3. Grupa VEM, Kroon HM, Ozmen I, Bedrikovetski S, Dudi-Venkata NN, Hunter RA, et al. Current practice in Australia and New Zealand for defunctioning ileostomy after rectal cancer surgery with anastomosis: Analysis of the Binational Colorectal Cancer Audit. Colorectal disease : the official journal of the Association of Coloproctology of Great Britain and Ireland. 2021;23(6):1421-33.
4. de Neree Tot Babberich MPM, Detering R, Dekker JWT, Elferink MA, Tollenaar R, Wouters M, et al. Achievements in colorectal cancer care during 8 years of auditing in The Netherlands. European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 2018;44(9):1361-70.
5. Huisman DE, Reudink M, van Rooijen SJ, Bootsma BT, van de Brug T, Stens J, et al. LekCheck: A Prospective Study to Identify Perioperative Modifiable Risk Factors for Anastomotic Leakage in Colorectal Surgery. Annals of surgery. 2020.
6. Snijders HS, van Leersum NJ, Henneman D, de Vries AC, Tollenaar RA, Stiggelbout AM, et al. Optimal Treatment Strategy in Rectal Cancer Surgery: Should We Be Cowboys or Chickens? Annals of surgical oncology. 2015;22(11):3582-9.

3

Chapter 3

One decade of declining use of defunctioning stomas after rectal cancer surgery in the Netherlands: are we on the right track?

Erik W. Ingwersen
Paulien J.K. van der Beek
Jan Willem T. Dekker
Susan van Dieren
Freek Daams

Diseases of the Colon & Rectum 2023 Jul 1;66(7):1003 -1011

Abstract

Background: The beneficial effect of a defunctioning stoma in mitigating consequences of anastomotic leakage after rectal cancer surgery is still debated.

Objective: This study aims to reflect on a decade of rectal cancer surgery in terms of stoma construction and anastomotic leakage.

Study design: Retrospective observational study

Study setting: This study used data of the Dutch Colorectal Audit from 2011 to 2020.

Study population: Patients undergoing rectal cancer surgery with a primary anastomosis.

Outcome measures: Primary outcome was anastomotic leakage. Secondary outcomes were minor complications, admission to intensive care, length of stay, readmission and patient death.

Results: A total of 13,263 patients were included in this study. A defunctioning stoma was constructed in 7,106 (53.6%) patients. Patients with a defunctioning stoma were less likely to develop anastomotic leakage (7.9% vs 13.0%), and if anastomotic leakage occurred, less patients needed surgical reintervention (37.7% vs 81.1%). An annual decrease in the construction of a defunctioning stoma was seen (69.8% in 2011 vs 51.8% in 2015 vs 29.7% in 2020), accompanied with a 5% increase in anastomotic leakage (9.1% in 2011 vs 14.1% in 2020). A defunctioning stoma was associated with a higher occurrence of minor complications, admissions to intensive care unit, longer length of stay and more readmissions within 90 days.

Conclusion: The reduction in defunctioning stomas is paralleled with an increase in anastomotic leakage. Patients with a defunctioning stoma, however, also showed more minor complications, a prolonged length of stay, more ICU admissions and readmissions. In our opinion the trade-offs of selective use should be individually considered.

Introduction

Anastomotic leakage (AL) after rectal cancer surgery is a bête noire, resulting in increased morbidity, mortality and prolonged hospitalization.[1,2] Many surgeons regard a defunctioning stoma as a strategy to mitigate its consequences.[3,4] But when and in which patient a defunctioning stoma should be applied, remains questionable and has been heavily debated for several decades.[5] Some propose that a defunctioning stoma should be routinely used, regardless the patients' condition, since it reduces the chance of developing AL and reduces its severity.[6-8] Others think of the defunctioning stoma as overused.[9] It does not lower incidence of AL and possibly results in myriad impediments to patients, including skin irritation, reoperation and parastomal hernia.[1,10] Moreover, one-third of the patients never undergoes closure of their defunctioning stoma.[11]

In New Zealand and Australia surgeons seem to advocate the first strategy and defunctioning stomas are routinely performed. The binational colorectal cancer audit reported that between 2007 to 2019 89% of patients obtained a defunctioning stoma after rectal cancer surgery.[12]

In several other surgical communities, including the UK, the US, Japan and the Netherlands, selective use of the defunctioning stoma is propagated by some.[13,15] One study showed that highly selective use of a defunctioning stoma was safe and should be recommended. In this study, 76% of the patients never had a stoma after their lower anterior resection.[16] In 2015, Snijder et al. found in a retrospective cohort hospitals with a high tendency towards more stoma use did not have less AL then hospitals with a low tendency of defunctioning stoma use. This study led to a national decrease in stoma use.[17] Remarkably, a recent Dutch study showed an increase in incidence of AL after rectal cancer surgery in the last decade and suggested a relationship with the decrease in defunctioning stoma use.[18]

Albeit protagonists of both strategies continue to disagree, they do consent on one thing: high-risk anastomoses need to be diverted. However, when surgeons regard an anastomosis as high-risk is a conundrum, resulting in a large variety of stoma implementation between hospitals, even within the same country.[19]

This population-based study with data of the Dutch Colorectal Audit (DCRA) investigates one decade of defunctioning stoma use in the Netherlands in patients undergoing rectal cancer surgery with a primary anastomosis. It scrutinizes the trend in stoma use and its impact on the incidence and severity of AL.

Materials and Methods

Study design, setting and population

A retrospective observational population-based study collected data from a nationwide database from the DCRA. The DCRA annually collects data, from all Dutch hospitals providing colorectal care, of patients undergoing resection of colorectal cancer in the Netherlands. This is a mandatory registry; no ethical approval or informed consent was necessary under Dutch law.[20] The follow-up time was 90 days postoperatively. Data were collected from 2011 until 2020. All patients older than 18 years with rectal cancer undergoing a resection with a primary anastomosis with or without a defunctioning stoma were included. Patients were excluded if the construction of a defunctioning stoma was not specified.

Outcome, data collection and definitions

Primary outcome was AL. Secondary outcomes were minor complications, length of stay, admission to intensive care unit (ICU), readmission or patient death within 90 days.

The following preoperative variables and surgical factors were collected: age, gender, body mass index (BMI), comorbidities, American Joint Committee on Cancer stage (AJCC), American Society of Anesthesiologists (ASA) classification, administration of neoadjuvant therapy, tumor distance from the anorectal junction, preoperative tumor related complications, surgical procedure, laparoscopy/laparotomy, conversion, elective or urgent procedure, intraoperative radiological or chemotherapeutical intervention, and an additional or macroscopic irradical (R2) resection. The following postoperative outcomes were collected: AL, minor complications, need of blood transfusion, length of stay, admission to ICU and readmission or death of patient within 90 days from surgical procedure.

In the DCRA registry, from 2011 to 2018 AL was defined as clinically relevant AL requiring surgical, radiological or endoscopic reintervention (Clavien-Dindo grade 3a or 3b), diagnosed within 30 days postoperatively. Since 2018, the DCRA registered AL within 90 days postoperative and also registered patients with AL that did not received any type of reintervention . We defined AL as clinical relevant with a Clavien-Dindo grade of 3a, 3b or 4, diagnosed within 30 days postoperatively. Minor complications were defined as any deviation from the normal postoperatively course with a Clavien-Dindo grade of 1 or 2.[21,21] Since 2018, the DCRA registered data about comorbidities and reinterventions. These variables were analyzed in a subanalysis.

Imputation

Variables were excluded if missing data were above 50%. In variables with less than 50% missing data, there was imputed using predictive mean matching with 10 iterations. The pooled outcome of these 10 iterations was used for analysis.

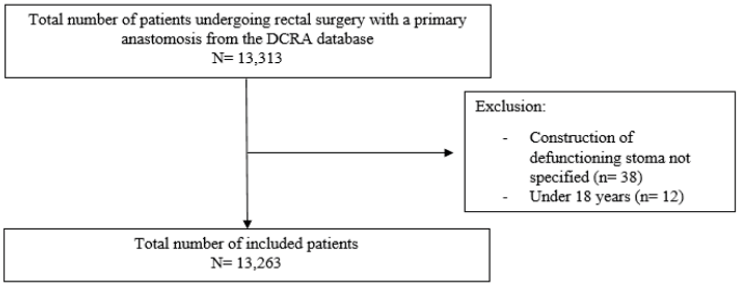
Statistics

Statistical Package for the Social Sciences (SPSS) and RStudio were used for analyzing the data. Continuous variables were reported as mean with a standard deviation and dichotomous variables as number and percentage. Differences between patient with and without a defunctioning stoma were compared using X^2 test in dichotomous variables and a student's T-test or Mann-Whitney U test (skew distribution) in continuous variables. Odds ratio (OR) and 95% confidence interval (CI) were reported for differences in dichotomous variables and mean differences (MD) or interquartile range (IQR) in continuous variables. Differences between the construction of a defunctioning stoma per year were analyzed with logistic regression with defunctioning stoma as dependent variable and year as an independent, continuous, variable in a univariate analysis. Postoperative outcomes were analyzed using a multivariate analysis and propensity score matching (PSM). In the multivariate analysis, differences in postoperative outcomes were analyzed with linear (continuous) or logistic regression (dichotomous) with the postoperative outcome as dependent variable. Potential confounders were determined via directed acyclic graph, to reduce bias.[23] Confounders were included in the multivariate analysis if the difference between the two measures of association (with and without confounder) was 10% or more. Every outcome of the multivariate analysis was time adjusted, so the outcomes were adjusted for year of surgical procedure. PSM was performed for age, BMI, sex, AJCC stage, distance to AV, surgical procedure and application of neoadjuvant therapy in a 1:1 ratio with a caliper of 0.1 between patients with and without a stoma to analyze postoperative outcomes. The within approach for multiple imputed sets was used to derive the score with the MatchThem package in Rstudio.[24] The outcome of the analyses in each imputation set was pooled using the Rubin's rule. A total of 4,147 patients in the stoma group were matched with 4,147 patients without a stoma. In Appendix I a balance table with the maximum standardized mean differences and Kolmogorov-Smirnov test of the covariates is found. A p value <0.05 was considered statistically significant.

Results

A total of 13,313 patients underwent rectal cancer surgery between 2011 – 2020. Fifty patients were excluded. A total of 13,263 patients were included (Figure 1).

Figure 1, flowchart



Participant flowchart. DCRA; Dutch colorectal Audit, N; number.

Participant flowchart. DCRA; Dutch colorectal Audit, N; number.

Baseline characteristics and surgical factors

Table 1 summarizes patient and tumor characteristics and surgical factors. Patients with a defunctioning stoma were younger (64.7 vs 65.6 years, p<0.001), more often male (66.3% vs 58.2%, p<0.001) and were less likely to have an ASA classification of 3 or more (12.8% vs 14.4%, p<0.001). BMI was not different for both groups (26.2 vs 26.1 kg/m^2, p<0.001). Patients with a defunctioning stoma had more advanced tumors (AJCC stage III or IV of 64.2% vs 41.8%, p<0.001), had neoadjuvant treatment more often (78.3% vs 42.7%, p<0.001) and more patients had preoperative tumor related complications (17.7% vs 15.8%, p=0.004). The mean tumor distance to the anorectal junction was 83mm in patients with and 95mm in patients without a defunctioning stoma (p<0.001). Data from 2018 to 2020 showed that patients with a defunctioning stoma more often had one or more comorbidities (58.3% vs 32.7%, p<0.001). Diverted patients more often underwent a total mesorectal excision compared to patients who were not diverted (96.9% vs 88.8%, p<0.001). A defunctioning stoma was associated with a laparotomy (p<0.001), and conversion rates were higher in diverted patients (10.8% vs 6.0%, p<0.001). Intraoperatively chemotherapeutical or radiological intervention (1.9% vs 0.6%, p<0.001) and an additional resection (4.1% vs 2.9%, p<0.001) were more frequent in patients with a defunctioning stoma. There were no differences between stoma construction and urgent surgery (0.6% vs 0.7%, p=0.187) or R2 resections (0.1% vs 0.2%, p=0.748). Table 2 displays the means and 95% CI after PSM matching of the variables used in the patients with and without a defunctioning stoma. It shows that the PSM matching was successful in balancing the data between the groups.

Table 2, patient characteristics and surgical factors

Variable	Defunctioning stoma (n= 7,106)	No defunctioning stoma (n= 6,157)	Total (n= 13,263)	OR/MD (95% CI)	P value
Age (mean in years)	64.7±10.1	65.6±10.1	65.1±10.1	-0.88 (-1.23 to -0.54)	**<0.001***
Sex					**<0.001^**
Female	2,398 (33.7%)	2,568 (41.7%)	4,966 (37.4%)	0.71 (0.66 - 0.76)	
Male	4,708 (66.3%)	3,589 (58.2%)	8,297 (62.6%)	1	
BMI (mean in kg/m^2)	26.2±4.1	26.1±4.2	26.1±4.1	0.07 (-0.01 to 0.27)	0.064*
ASA classification					
<3	6,194 (87.2%)	5,266 (85.6%)	11,460 (86.4%)	1.15 (1.04 – 1.27)	**0.006^**
≥3	912 (12.8%)	891 (14.4%)	1,803 (13.6%)	1	
Neoadjuvant therapy	5,562 (78.3%)	2,631 (42.7%)	8,193 (61.8%)	4.83 (4.29 – 5.21)	**<0.001^**
AJCC stage					
I & II (T1-4N0M0)	2,538 (35.7%)	3,588 (58.2%)	6,126 (46.1%)	0.40 (0.37 – 0.43)	**<0.001^**
III & IV (T1-4N1-2M0-1)	4,568 (64.2%)	2,569 (41.8%)	7,137 (53.9%)	1	
Tumor distance to anal junction (in mm)	83.2±41.8	94.8±41.5	88.6±42.0	-11.6 (-13.0 to -10.2)	**<0.001***
Preoperative tumor related complications	1,257 (17.7%)	973 (15.8%)	2,230 (16.8%)	1.14 (1.04 – 1.25)	**0.004^**
Surgical procedure					**<0.001^**
Total mesorectal excision	6,889 (96.9%)	5,465 (88.8%)	12,354 (93.1 %)	4.02 (3.44 – 4.69)	
Partial mesorectal excision	217 (3.1%)	692 (11.2%)	909 (6.9%)	1	
Approach					**<0.001^**
Abdominal laparotomy	1,579 (22.2%)	732 (11.9%)	2,311 (17.4%)	2.11 (1.92 – 2.32)	
Laparoscopic/transanal	5,527 (77.8%)	5,425 (88.1%)	10,952 (82.6%)	1	
Conversion	766 (10.8%)	371 (6.0%)	1,136 (8.6%)	1.87 (1.61 – 2.13)	**<0.001^**
Emergency surgery	40 (0.6%)	46 (0.7%)	86 (0.6%)	1.33 (0.87 – 2.03)	0.187^
Additional resection	295 (4.1%)	177 (2.9%)	472 (3.5%)	1.46 (1.20 – 1.76)	**<0.001^**
Macroscopic irradical	10 (0.1%)	10 (0.2%)	20 (0.2%)	0.87 (0.36 – 2.08)	0.748
HIPEC/IORT	133 (1.9%)	37 (0.6%)	170 (1.3%)	3.07 (2.14 – 4.41)	**<0.001^**

OR; odds ratio, MD; mean difference, BMI; body mass index, ASA; American Society of Anesthesiologists, AJCC; American Joint Committee on Cancer, HIPEC; hyperthermic intraperitoneal chemotherapy, IORT; intraoperative radiotherapy. *Student's T test. ^X^2 test.

Table 2, propensity score matching baseline table

Covariates used in PSM	Mean defunctioning stoma (n= 4,148)	95% CI	Mean no defunctioning stoma (n=4,148)	95% CI
Sex (male)	0.61	0.60 – 0.63	0.61	0.59 – 0.62
Age (years)	65.1	64.2 – 65.9	65.2	64.9 – 65.5
BMI (kg/m^2)	26.1	25.8 – 26.4	26.1	25.9 – 26.2
AJCC stage III & IV (T1-4N1-2M0-1)	0.53	0.52 – 0.55	0.53	0.51 – 0.54
Tumor distance to anal junction (in mm)	87.9	84.8 – 91.1	90.2	88.9 – 91.4
Surgical procedure (PME)	0.05	0.05 – 0.06	0.05	0.05 – 0.06
Neoadjuvant therapy	0.63	0.61 – 0.64	0.61	0.60 – 0.63

PSM; propensity score matching, CI; confidence interval, BMI; body mass index, AJCC; American Joint Committee on Cancer, PME: partial mesorectal excision.

Table 3 shows baseline characteristics per year for patients with a defunctioning stoma. An annual decrease in age (MD per year -0.15, p<0.001), application of neoadjuvant therapy (OR 0.85, p<0.001) and tumor distance from the anorectal junction (MD -4.0 mm per year, p<0.001) was found. Between 2011 and 2020 an increasing number of patients with a defunctioning stoma had an ASA classification of II or higher (OR 1.06, p<0.001), and an AJCC stage of III or IV (OR 1.03, p=0.010).

Table 3, patients and tumor characteristics of patients with a defunctioning stoma per year from 2011 - 2020

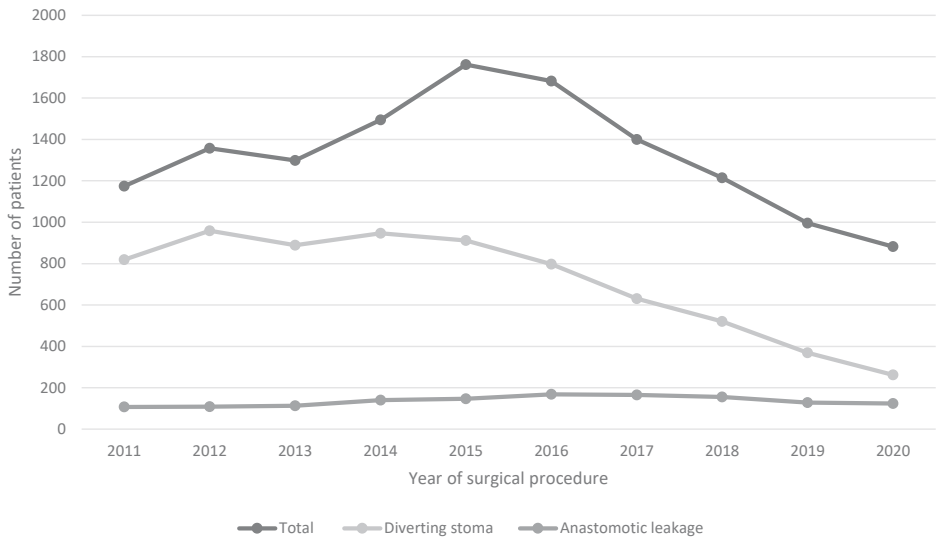
Variable	2011 (n= 819)	2012 (n= 959)	2013 (n= 889)	2014 (n= 947)	2015 (n= 912)		
Patients with a defunctioning stoma	819 (69.8%)	959 (70.6%)	889 (68.4%)	947 (63.3%)	912 (51.8%)		
Age (years)	66.1±11.1	65.3±10.8	66.4±10.3	66.6±10.3	66.4±9.2		
Sex (male)	536 (65.4%)	606 (63.2%)	583 (65.6%)	622 (65.7%)	613 (67.4%)		
Body mass index	26.0±3.9	25.9±3.9	26.2±4.2	26.1±4.1	26.4±3.9		
ASA classification III or IV	113 (13.8%)	101 (10.5%)	99 (11.1%)	100 (10.6%)	111 (12.2%)		
AJCC stage III or IV	484 (59.1%)	594 (61.9%)	596 (67.0%)	584 (61.7%)	605 (66.3%)		
Metastatic disease	34 (4.2%)	68 (7.1%)	60 (6.7%)	56 (5.9%)	59 (6.5%)		
Distance from tumor to anal junction (in mm)	91.9±41.8	89.7±32.8	93.8±65.8	88.0±35.5	88.5±36.1		
Neoadjuvant therapy	739 (90.2%)	959 (89.1%)	781 (87.9%)	679 (71.7%)	646 (70.8%)		
Developed AL	66 (8.1%)	62 (6.5%)	72 (8.1%)	80 (8.4%)	61 (6.7%)		
	2016 (n= 797)	**2017 (n= 631)**	**2018 (n= 521)**	**2019 (n= 369)**	**2020 (n= 262)**	**OR/MD (95% CI)**	**P value**
Patients with a defunctioning stoma	797 (47.4%)	631 (45.1%)	521 (42.9%)	369 (37%)	262 (29.7%)	0.82 (0.81 – 0.83)	**<0.001**
Age (years)	65.9±9.8	65.4±9.9	64.8±9.8	65.0±9.9	64.4±10.1	-0.150	**0.001**
Sex (male)	558 (70.0%)	426 (67.5%)	335 (64.3%)	244 (66.1%)	184 (70.2%)	0.98 (0.96 – 1.00)	0.051
Body mass index	26.4±4.1	26.3±4.3	26.0±4.0	26.5±4.2	26.4±4.3	0.047	**0.012**
ASA classification III or IV	105 (13.2%)	87 (13.8%)	80 (15.4%)	59 (16.0%)	59 (22.1%)	1.06 (1.03 – 1.09)	**<0.001**
AJCC stage III or IV	551 (69.1%)	407 (64.5%)	350 (67.2%)	237 (64.2%)	163 (62.2%)	1.03 (1.01 – 1.05)	**0.010**
Metastatic disease	55 (6.9%)	51 (8.3%)	41 (7.9%)	34 (9.2%)	21 (8.0%)	1.06 (1.02 – 1.10)	**0.002**
Distance from tumor to anal junction (in mm)	87.9±35.5	68.9±33.1	58.8±30.4	62.6±31.4	59.5±28.1	-4.02	**<0.001**
Neoadjuvant therapy	593 (74.4%)	451 (71.5%)	377 (72.4%)	264 (71.5%)	178 (67.9%)	0.85 (0.83 – 0.87)	**<0.001**
Developed AL	53 (6.6%)	58 (9.2%)	56 (10.7%)	28 (7.6%)	22 (8.4%)	1.02 (0.99 - 1.06)	0.171

OR; odds ratio, MD; mean difference, ASA; American Society of Anesthesiologists, AJCC; American Joint Committee on Cancer, AL; anastomotic leakage. The OR/MD an P value reflects the annual increase of decrease of the variable between 2011 and 2020.

Primary and secondary outcomes

A total of 7,106 (53.6%) patients were diverted. AL occurred in 1,356 (10.2%) patients. Patients with a defunctioning stoma developed AL less frequently (7.9%) than patients without (13.0%). The OR of developing AL with a defunctioning stoma was 0.57 in the univariate (p<0.001) and 0.47 in the multivariate analysis (p<0.001) compared to patients without a defunctioning stoma. The relative risk reduction of a defunctioning stoma on AL was 65%, the absolute risk reduction 5.1% and the number needed to treat (NNT) 20. Figure 2 shows the absolute and relative construction of a defunctioning stoma, incidence of AL and differences between both groups. A significant annual decrease in stoma construction was seen. In the univariate analysis the OR for every year from 2011 to 2020 of having a stoma was 0.82 pertaining to the previous year, with a 95% CI of 0.81 – 0.83 (p<0.001). In the multivariate analysis the OR was 0.87, with a CI of 0.85 – 0.89 (p<0.001). An annual increase in AL was seen (9.1% in 2011, 9.9% in 2016 and 14.1% in 2020).

Figure 2, Annual amount of patients undergoing rectal cancer surgery, construction of a defunctioning stoma and occurrence of anastomotic leakage



A decrease in stoma construction was seen between 2011 and 2020. The OR of having a stoma was 0.82 (95% CI 0.81 – 0.83, p<0.001) per year from 2011 – 2020 in the univariate analysis. In the multivariate analysis the outcome was adjusted for application of neoadjuvant therapy, tumor distance to anorectal junction and surgical procedure. The OR in the multivariate analysis was 0.87 (95% CI 0.81 – 0.83, p<0.001). The occurrence of AL increased with an OR of 1.07 (95% CI 1.05 –1.10, p<0.001) per year from 2011 – 2020. In the multivariate the outcome was adjusted for application of neoadjuvant therapy, tumor distance to anorectal junction, ASA classification, surgical procedure and the presence of a defunctioning stoma. The difference in the multivariate analysis was not significant (OR 0.99, 95% CI 0.97 –1.02, p=0.503). OR; odds ratio, CI; confidence interval, AL; anastomotic leakage, ASA; American Society of Anesthesiologists

Table 4 shows postoperative outcomes of the univariate, multivariate and PSM analysis. Patients with a defunctioning stoma developed more often minor complications (18.3% vs 9.6%, p<0.001). Admission to ICU (16.1% vs 12.3%, p<0.001) and readmission within 90 days was also seen more in patients with a defunctioning stoma (15.3% vs 10.7%, p<0.001). Patients with a defunctioning stoma had a median length of stay of 8 days and patients without a defunctioning stoma of 5 days (p<0.001). No significant difference in mortality within 90 days was found (1.1% vs 0.8%, p=0.294). Need of blood transfusion was not significant in the multivariate analysis (6.9% vs 5.3%, p=0.231), but was in the PSM analysis (p=0.012). Data from 2018 to 2020 showed that if AL occurred, patients with a defunctioning stoma needed a surgical intervention less frequently compared to those without (37.7% vs 81.1%, p<0.001)

There was an annual decrease from 2011- 2020 in the total cohort for minor complications (OR 0.96, 95% CI 0.95 – 0.98, p<0.001), length of stay (MD -0.58 days per year, 95% CI -0.68 to -.048, p<0.001), admission to ICU (OR 0.95, 95% CI 0.93 – 0.96, p<0.001), need of blood transfusion (OR 0.88, 95% CI 0.86 – 0.90, p<0.001) and patient death within 90 days (0.88, 95% CI 0.82 – 0.94, p<0.001). An increase was seen in the overall readmission rate within 90 days (OR 1.03, 95% CI 1.01 – 1.06, p=0.002).

Table 4, postoperative outcomes, univariate and multivariate analysis

Postoperative outcome	Defunctioning stoma (n= 7,106)	No defunctioning stoma (n= 6,157)	Total (n= 13,263)	OR / MD (95% CI)	P value	OR / MD (95% CI)	P value	OR / MD (95%CI)	P value
Anastomotic leakage	558 (7.9%)	798 (13.0%)	1,356 (10.2%)	0.57 (0.51 – 0.64)	**<0.001**	0.47 (0.42 – 0.54)	**<0.001**	0.46 (0.40 – 0.54)	**<0.001**
Minor complication	1,297 (18.3%)	594 (9.6%)	1,891 (14.3%)	2.10 (1.89 – 2.32)	**<0.001**	2.08 (1.87 - 2.31)	**<0.001**	1.84 (1.60 – 2.21)	**<0.001**
Need of blood transfusion	492 (6.9%)	324 (5.3%)	816 (6.2%)	1.34 (1.17 - 1.56)	**<0.001**	1.11 (0.94 – 1.31)	0.231	1.32 (1.06 – 1.63)	**0.012**
Admission to intensive care	1,142 (16.1%)	758 (12.3%)	1,900 (14.3%)	1.36 (1.24 – 1.51)	**<0.001**	1.26 (1.13 – 1.40)	**<0.001**	1.31 (1.14 – 1.51)	**<0.001**
Length of stay (IQR)	8 (3 - 12)	5 (6 - 3)	6 (12 - 5)	3.55 (3.02 – 4.08)	<**0.001**	2.96 (2.41 – 3.50)	**<0.001**	2.67 (1.97 – 3.38)	**<0.001**
Readmission within 90 days	1,084 (15.3%)	658 (10.7%)	1,742 (13.1%)	1.50 (1.36 – 1.67)	**<0.001**	1.54 (1.37 – 1.72)	**<0.001**	1.23 (1.07 – 1.41)	**0.003**
Patient death within 90 days	81 (1.1%)	51 (0.8%)	132 (1.0%)	1.38 (0.97 – 1.96)	0.071	1.27 (0.88 – 1.83)	0.207	1.24 (0.79 – 1.95)	0.350

In the multivariate analysis anastomotic leakage was adjusted for year of surgical procedure, application of neoadjuvant therapy, tumor distance to anal junction and surgical procedure. Minor complications were adjusted for year of surgical procedure. Need of blood transfusion was adjusted for year of surgical procedure, application of neoadjuvant therapy, surgical procedure, approach. Length of stay was adjusted for year of surgical procedure and surgical procedure. Admission to intensive care was adjusted for year of surgical procedure, BMI, age, ASA classification, AJCC stage and surgical procedure. Readmission within 90 days was adjusted for year of surgical procedure, application of neoadjuvant therapy, surgical procedure. Patient death within 90 days was adjusted for year of surgical procedure, age, surgical procedure and approach.

IQR; interquartile range, OR; odds ratio, MD; mean difference, CI; confidence interval, AJCC; American Joint Committee on Cancer, BMI; body mass index, ASA; American Society of Anesthesiologists.

Discussion

This study showed an annual decrease in the use of defunctioning stomas in the Netherlands over the last decade. Overall, a reduction in defunctioning stomas of 40% was found. This reduction was paralleled with a 5% overall increase in AL (9.1% in 2011, 14.1% in 2020).

The reduction in stoma use appears to reflect a shift towards more selective use, where a defunctioning stoma is often omitted and only constructed in patients with a high-risk anastomosis. Throughout the past decade, tumors were lower in the stoma group and there was an increase in ASA and AJCC classification. An overall decrease in neoadjuvant radiotherapy was seen after the implementation of a revised national guideline on colorectal cancer in 2014.[25] In conflict with Snijders et al., who argued that there was no difference in AL between hospitals routinely and selectively using a diverting stoma (8.4% in low stoma rate and 11.3% in high stoma rate hospitals), this study found that more selective use did lead to more AL.[17] In response to the results of Snijder et al., selective use was increasingly advocated and, the overall use of the defunctioning stoma in the Netherlands decreased. This could have led to a hasted decline in stoma use in hospitals used to routinely divert resulting in increase of AL. Talboom et al. also emphasized in their study, which had a highly selective diversion rate of 24% after a LAR, that implementation of a management protocol is a prerequisite to obtaining these results.[16] The increase in AL, however, could also be a consequence of a more strict and honest registration. In the last decade, there is a growing interest in quality of care and registration, accompanied by the need for uniformity of definitions of outcomes such as AL.[26-28]

Although a slightly lower AL-rate was seen in diverted patients, the defunctioning stoma is far from the holy grail and must be balanced against its negative outcomes. The defunctioning stoma led to a higher minor complication rate, admission to the ICU, readmission within 90 days and length of stay. Many patients with a defunctioning stoma suffer from stoma-related complications (e.g., stoma necrosis or stenosis, retraction, parastomal hernia dehydration).[1] In many cases, the prolonged length of stay was likely due to stoma management or so that the problem that initially required the stoma could be further managed.[29] It is likely that patients also had a longer length of stay due to the associated increased likelihood of minor surgical complications and vice versa.[30] Previous studies showed that stoma-related complications, especially dehydration, are also associated with readmission.[31] Furthermore, some patients will have a revision of their defunctioning stoma within 90 days and thus this is registered as a readmission. While this counts for a readmission in our analysis, it is in fact a planned admission, making diverting ileostomy seem more prone to unscheduled readmission

than it actually is. However, it is important to mention that up to one-third of stomas are never reversed.[12]

The question rises whether we are on a good track or not. The decrease of 40% of stoma use coincided with an increase in AL and like several other studies our data showed a protective effect of the defunctioning stoma on both the incidence and severity of AL.[4,7,33] The defunctioning stoma, however, also led to a longer length of stay, more admissions to the ICU and its decrease was accompanied by an overall annual decrease in mortality, length of stay, ICU admission and minor surgical related morbidity. Borstlap et al. showed that differences between AL in between patients with and without a defunctioning stoma diminish after a longer follow-up period. They hypothesize that AL will occur irrespective of a defunctioning stoma, but that diagnosis is delayed. Long-term data is missing in our study, causing a potential overestimation of the benefits of the defunctioning stoma for patients in preventing AL.[34] A trend towards more constructions of primary anastomosis in patients with rectal cancer surgery is seen in the Netherlands.[35] This has arguably resulted in more patients at risk for AL and could have altered the selection of patients with a defunctioning stoma.

Altogether, selective use of a defunctioning stoma has trade-offs. Considering a high percentage of stoma-related morbidity, longer length of stay, higher readmission rate, missing data about on late leaks and a NNT of 20 to prevent AL, the Netherlands are arguably on the right track. However, this track is only feasible when the trade-offs are individually balanced. Proper risk prediction for AL would be desirable, to predict which patients are at high-risk. Some tools already exist, but overall performance of these models remains suboptimal.[36,37] The preliminary results of machine learning models in predicting AL are promising and an internally and externally validated prediction model could assist the surgeon in the near future.[38] With these prediction models, surgeons can find out better how to select patients and how best to prevent AL. In the end, the surgeon decides after the anastomosis whether to divert or not, but these models help the surgeon to present patients with better figures and discuss treatment options via shared decision-making, leading to better patient comprehension.[39]

This nationwide cohort was strengthened by the large amount of included patients with rectal cancer. By analyzing patients during 2011 – 2020, this study was able to obtain trends in time and could analyze the accompanied risks and benefits of these trends. This retrospective study is susceptible for confounders by indications and there could be risk factors for AL and defunctioning stoma that were not regarded. Differences in postoperative outcomes could be due to the diverting stoma. However, it is still plausible that some differences that were caused by unforeseen confounders were not taken into account in the multivariate and PSM analysis. In the PSM analysis,

the results could be prone to selection bias due to a relatively low number of matched cases. Moreover, several secular changes, such as increased compliance e to enhanced recovery after surgery over time, might have influenced the findings. There was no data available about long-term outcomes, readmissions after 90 days and data on late leaks, to further define the impact of a defunctioning stoma. Hazen et al. showed that patients with a defunctioning stoma had significant number of re-

interventions and stoma-related complications (e.g. stoma-site hernias, high-stoma output, dehydration), even after one year.[40] Future studies on the topic should include long-term follow-up on patients with and without a defunctioning stoma, including endpoints as cancer recurrence, quality of life and stoma reversal rates.

Conclusion

This study showed a declining trend in the use of defunctioning stomas over the last decade. The reduction in defunctioning stomas of 40% is paralleled with a 5% increase in AL. Patients with a defunctioning stoma, however, also showed more minor complications, a prolonged length of stay, more ICU admissions and readmissions. Are we on the right track in the Netherlands? Considering a high percentage of stoma-related morbidity and a NNT of 20 to prevent one AL, arguably yes. However, this track is only feasible when the trade-offs are individually balanced.

References

1. Emmanuel A, Chohda E, Lapa C, Miles A, Haji A, Ellul J. Defunctioning Stomas Result in Significantly More Short-Term Complications Following Low Anterior Resection for Rectal Cancer. World J Surg. 2018;42(11):3755-64.
2. Huisman DE, Reudink M, van Rooijen SJ, Bootsma BT, van de Brug T, Stens J, et al. LekCheck: A Prospective Study to Identify Perioperative Modifiable Risk Factors for Anastomotic Leakage in Colorectal Surgery. Annals of surgery. 2020.
3. Gu W-l, Wu S-w. Meta-analysis of defunctioning stoma in low anterior resection with total mesorectal excision for rectal cancer: evidence based on thirteen studies. World journal of surgical oncology. 2015;13:9-.
4. Phan K, Oh L, Ctercteko G, Pathma-Nathan N, El Khoury T, Azam H, et al. Does a stoma reduce the risk of anastomotic leak and need for re-operation following low anterior resection for rectal cancer: systematic review and meta-analysis of randomized controlled trials. Journal of gastrointestinal oncology. 2019;10(2):179-87.
5. Snijders HS, van den Broek CB, Wouters MW, Meershoek-Klein Kranenbarg E, Wiggers T, Rutten H, et al. An increasing use of defunctioning stomas after low anterior resection for rectal cancer. Is this the way to go? European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 2013;39(7):715-20.
6. Lee BC, Lim SB, Lee JL, Kim CW, Yoon YS, Park IJ, et al. Defunctioning Protective Stoma Can Reduce the Rate of Anastomotic Leakage after Low Anterior Resection in Rectal Cancer Patients. Ann Coloproctol. 2020.
7. Matthiessen P, Hallbook O, Rutegard J, Simert G, Sjodahl R. Defunctioning stoma reduces symptomatic anastomotic leakage after low anterior resection of the rectum for cancer: a randomized multicenter trial. Annals of surgery. 2007;246(2):207-14.
8. Veenhof AA, van der Peet DL, Meijerink WJ, Cuesta MA. Defunctioning stoma reduces symptomatic anastomotic leakage after low anterior resection of the rectum for cancer: a randomized multicenter trial. Annals of surgery. 2008;247(4):718-9; author reply 9-20.
9. Niu L, Wang J, Zhang P, Zhao X. Protective ileostomy does not prevent anastomotic leakage after anterior resection of rectal cancer. J Int Med Res. 2020;48(8):300060520946520-.
10. Malik T, Lee MJ, Harikrishnan AB. The incidence of stoma related morbidity - a systematic review of randomised controlled trials. Ann R Coll Surg Engl. 2018;100(7):501-8.
11. Zeman M, Czarnecki M, Chmielarz A, Idasiak A, Grajek M, Czarniecka A. Assessment of the risk of permanent stoma after low anterior resection in rectal cancer patients. World journal of surgical oncology. 2020;18(1):207-.
12. Grupa VEM, Kroon HM, Ozmen I, Bedrikovetski S, Dudi-Venkata NN, Hunter RA, et al. Current practice in Australia and New Zealand for defunctioning ileostomy after rectal cancer surgery with anastomosis: Analysis of the Binational Colorectal Cancer Audit. Colorectal disease : the official journal of the Association of Coloproctology of Great Britain and Ireland. 2021;23(6):1421-33.
13. Emmanuel A, Chohda E, Lapa C, Miles A, Haji A, Ellul J. Defunctioning Stomas Result in Significantly More Short-Term Complications Following Low Anterior Resection for Rectal Cancer. World J Surg. 2018;42(11):3755-64.
14. Wang L, Hirano Y, Ishii T, Kondo H, Hara K, Obara N, et al. Diverting Stoma Versus No Diversion in Laparoscopic Low Anterior Resection: A Single-center Retrospective Study in Japan. In Vivo. 2019;33(6):2125-31.

15. Nurkin S, Kakarla VR, Ruiz DE, Cance WG, Tiszenkel HI. The role of faecal diversion in low rectal cancer: a review of 1791 patients having rectal resection with anastomosis for cancer, with and without a proximal stoma. Colorectal disease : the official journal of the Association of Coloproctology of Great Britain and Ireland. 2013;15(6):e309-16.
16. Talboom K, Vogel I, Blok RD, Roodbeen SX, Ponsioen CY, Bemelman WA, et al. Highly selective diversion with proactive leakage management after low anterior resection for rectal cancer. The British journal of surgery. 2021;108(6):609-12.
17. Snijders HS, van Leersum NJ, Henneman D, de Vries AC, Tollenaar RA, Stiggelbout AM, et al. Optimal Treatment Strategy in Rectal Cancer Surgery: Should We Be Cowboys or Chickens? Annals of surgical oncology. 2015;22(11):3582-9.
18. Arron MNN, Greijdanus NG, Ten Broek RPG, Dekker JWT, van Workum F, van Goor H, et al. Trends in risk factors of anastomotic leakage after colorectal cancer surgery (2011-2019): A Dutch population-based study. Colorectal disease : the official journal of the Association of Coloproctology of Great Britain and Ireland. 2021;23(12):3251-61.
19. Snijders HS, Kunneman M, Tollenaar RA, Boerma D, Pieterse AH, Wouters MJ, et al. Large variation in the use of defunctioning stomas after rectal cancer surgery. A lack of consensus. Acta oncologica (Stockholm, Sweden). 2016;55(4):509-15.
20. Van Leersum NJ, Snijders HS, Henneman D, Kolfschoten NE, Gooiker GA, ten Berge MG, et al. The Dutch Surgical Colorectal Audit. European Journal of Surgical Oncology (EJSO). 2013;39(10):1063-70.
21. Katayama H, Kurokawa Y, Nakamura K, Ito H, Kanemitsu Y, Masuda N, et al. Extended Clavien-Dindo classification of surgical complications: Japan Clinical Oncology Group postoperative complications criteria. Surg Today. 2016;46(6):668-85.
22. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. Annals of surgery. 2004;240(2):205-13.
23. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. BMC Medical Research Methodology. 2008;8(1):70.
24. Pishgar F, Greifer N, Leyrat C, Stuart E. MatchThem:: Matching and Weighting after Multiple Imputation: The R Journal (2021); 2020.
25. Gietelink L, Wouters M, Marijnen CAM, van Groningen J, van Leersum N, Beets-Tan RGH, et al. Changes in nationwide use of preoperative radiotherapy for rectal cancer after revision of the national colorectal cancer guideline. European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 2017;43(7):1297-303.
26. Frouws MA, Snijders HS, Malm SH, Liefers G-J, Van de Velde CJH, Neijenhuis PA, et al. Clinical Relevance of a Grading System for Anastomotic Leakage After Low Anterior Resection: Analysis From a National Cohort Database. Diseases of the Colon & Rectum. 2017;60(7):706-13.
27. van Gijn W, van de Velde CJ. Improving quality of cancer care through surgical audit. European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 2010;36 Suppl 1:S23-6.
28. Van Leersum NJ, Snijders HS, Henneman D, Kolfschoten NE, Gooiker GA, ten Berge MG, et al. The Dutch surgical colorectal audit. European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 2013;39(10):1063-70.

29. Cartmell MT, Jones OM, Moran BJ, Cecil TD. A defunctioning stoma significantly prolongs the length of stay in laparoscopic colorectal resection. Surgical Endoscopy. 2008;22(12):2643-7.
30. Bona S, Molteni M, Rosati R, Elmore U, Bagnoli P, Monzani R, et al. Introducing an enhanced recovery after surgery program in colorectal surgery: a single center experience. World J Gastroenterol. 2014;20(46):17578-87.
31. Li W, Stocchi L, Cherla D, Liu G, Agostinelli A, Delaney CP, et al. Factors associated with hospital readmission following diverting ileostomy creation. Tech Coloproctol. 2017;21(8):641-8.
32. Herrle F, Sandra-Petrescu F, Weiss C, Post S, Runkel N, Kienle P. Quality of Life and Timing of Stoma Closure in Patients With Rectal Cancer Undergoing Low Anterior Resection With Diverting Stoma: A Multicenter Longitudinal Observational Study. Diseases of the colon and rectum. 2016;59(4):281-90.
33. Hüser N, Michalski CW, Erkan M, Schuster T, Rosenberg R, Kleeff J, et al. Systematic review and meta-analysis of the role of defunctioning stoma in low rectal cancer surgery. Annals of surgery. 2008;248(1):52-60.
34. Borstlap WAA, Westerduin E, Aukema TS, Bemelman WA, Tanis PJ, Group DSR. Anastomotic Leakage and Chronic Presacral Sinus Formation After Low Anterior Resection: Results From a Large Cross-sectional Study. Annals of surgery. 2017;266(5):870-7.
35. de Neree Tot Babberich MPM, Detering R, Dekker JWT, Elferink MA, Tollenaar R, Wouters M, et al. Achievements in colorectal cancer care during 8 years of auditing in The Netherlands. European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 2018;44(9):1361-70.
36. Nikolian VC, Kamdar NS, Regenbogen SE, Morris AM, Byrn JC, Suwanabol PA, et al. Anastomotic leak after colorectal resection: A population-based study of risk factors and hospital variation. Surgery. 2017;161(6):1619-27.
37. Shiwakoti E, Song J, Li J, Wu S, Zhang Z. Prediction model for anastomotic leakage after laparoscopic rectal cancer resection. The Journal of international medical research. 2020;48(9):300060520957547.
38. Stam WT, Goedknegt LK, Ingwersen EW, Schoonmade LJ, Bruns ERJ, Daams F. The prediction of surgical complications using artificial intelligence in patients undergoing major abdominal surgery: A systematic review. Surgery. 2021.
39. de Mik SML, Stubenrouch FE, Balm R, Ubbink DT. Systematic review of shared decision-making in surgery. The British journal of surgery. 2018;105(13):1721-30.
40. Hazen SJA, Vogel I, Borstlap WAA, Dekker JWT, Tuynman JB, Tanis PJ, et al. Long-term stoma-related reinterventions after anterior resection for rectal cancer with or without anastomosis: population data from the Dutch snapshot study. Tech Coloproctol. 2021.

4

Chapter 4

Impact of merging two university hospitals on surgical outcome after esophagogastric and heapot-pancreato-biliary surgery: Results from a retrospective study

Erik W. Ingwersen
Wessel T. Stam
Lars J. van Kesteren
Ilse J.A. Wissink
Mark van Berge Henegouwen
Marc G. Besselink
Olivier R. Busch
Joris I. Erdmann
Wietse J. Eshuis
Suzanne S. Gisbertz
Geert Kazemier
Donald L. van der Peet
Rutger Jan Swijnenburg
Babs M. Zonderhuis

Surgery Open Science 2023 Jan 24:1-8

Abstract

Background: Due to centralization and super-specialization in medicine, hospital mergers are increasingly common. Their effect on postoperative outcomes in highly specialized surgical departments is unclear. As quality metrics often worsen after major organizational changes, preservation of quality of care during an hospital merge is of the utmost importance.

Objective: To evaluate the effect of a merger of two Dutch university hospitals on quality of surgical care, volume, and timeliness of care.

Methods: The upper gastro-intestinal and hepato-biliary-pancreatic sections merged on the 27th of January 2020 and the 31th of May 2021 respectively. Outcomes of all adult surgical patients were compared six months before and six months after the merger. Short-term quality metrics, volume, and timeliness of care were assessed.

Results: Overall, a cohort of 631 patients were included of whom 195 were upper gastro-intestinal (97 prior to the merger, 98 after the merger) and 436 (223 prior to the merger, 213 after) hepato-biliary-pancreatic patients. There were no differences in mortality, readmission, number and severity of complications, volume, and timeliness of care six months post-merger as compared to before merger.

Conclusion: This study shows that a hospital merger of two university hospitals can be performed without jeopardizing patient safety and while benefitting from centralization of highly specialized care and enhancement of medical research.

Key Message

This study investigated the impact of a merger of two Dutch university hospitals on quality of care, timeliness of care, and volume. It showed no deterioration in the evaluated short-term quality metrics, volume or timeliness for upper GI and HPB surgery, suggesting that a hospital merger of two university hospitals can be performed safely, while benefitting from centralization of highly specialized care and enhancement of medical research.

Introduction

Worldwide there is an increase in hospital mergers.[1] In the Netherlands, the introduction of managed competition in 2006 led to a wave of hospital mergers. From January 2006 almost 30 Dutch hospitals have merged. Potential positive effects of hospital mergers on quality of care (QoC) by centralization and super-specialization, include creation of synergies in clinical forces and material, effects on volume, timeliness of care, and increased efficiency and finance. Theoretically, negative effects include increased bureaucracy and a weakened level of competition resulting in decreased incentives for quality improvement in a market-based system.[2] The effects of general hospital mergers on QoC remain uncertain due to inconsistent findings[3,4], but they might be promising for university hospitals.[5,6] For university hospitals merging can be an opportunity to facilitate further concentration of typical academic high-complex-low-volume healthcare. Multiple studies have shown the benefits of increased volumes of high-complex healthcare on QoC, which is why volume norms for many complex procedures are generally accepted.[7,8] In 2017, the Dutch authority for consumers and markets conducted a major study in the Netherlands on the effect of general hospital mergers on the QoC and found no improvement in the QoC after the merger.[9] Recently, Wang et al. found a temporary increase in overall hospital mortality immediately after a merger, that recovered to a significant lower mortality two to three years later. Temporary decreased quality outcomes are common after major organization changes.[5] Therefore, the period after organization change is an important window, in which deterioration in QoC should be prevented, and acted upon if necessary.

A merger of two university and tertiary centers is unprecedented in the Netherlands and its effect on the short-term QoC, volume, and timeliness of care on highly specialized departments, such as upper gastrointestinal (GI) and the hepato-biliary-pancreatic (HPB), is more or less a conundrum. The aim of this study is to evaluate whether quality of surgical care for oncological HPB and upper GI procedures has been maintained six months after a Dutch university hospital merger. Furthermore, this study investigates the influence of the merger on the volume and timeliness of care. The hypothesize of this study was that the merger did not influence quality of surgical care, volume, or timeliness of care.

Methods

This study retrospectively collected QoC data from electronic health records (EHRs). Data concerning the volume and timeliness of surgical care were extracted from a prospectively collected operating room (OR) planning database. Data were evaluated

by two independent researchers (EWI and LJK) to check the correctness of the registered complications and supplement the extracted data in case of missing input. Any discrepancies in the data were assessed by a senior author (FD). The STROBE guidelines were followed to ensure correct reporting of the study methods and results.[10] Ethical approval for the data used for this study was waived by the local review board.

Setting

In 2018 the executive boards of the Academic Medical Center (AMC) and the Free University Medical Center (VUmc) conducted a horizontal merger into the Amsterdam University Medical Center (UMC), whereafter harmonization and integration processes was started.[11] To date, the legal merging for the hospitals is still ongoing, but both affiliated universities will remain different legal entities. Finalizing this requires a national political adjustment of legislation concerning the act on higher education. Currently this process has been set in motion. As a practical implication, EHRs cannot be exchanged between the two locations without consent of patients, for example. Prior to the merger, a director was appointed for every department managing both locations. Next to this, a multidisciplinary lateralization steering group was responsible for all the practical issues concerning the merger. Thirdly, a steering group was established for the harmonization and integration processes. The hospitals, both situated in Amsterdam, the Netherlands, have a physical distance of approximately 12 kilometers. The AMC was a tertiary care university hospital, facilitating 1,002 clinical beds with 8,720 employees (7,354 full-time equivalents [FTEs]).[12] The VUmc was a tertiary care university hospital, facilitating 733 clinical beds with 7,380 employees (6,156 FTEs).[13] Since the merger, half of the specializations are centered in the building of the former AMC hospital and the other half in the building of the former VUmc hospital. Both the upper GI and HPB surgical care were transferred to the building of the former VUmc hospital, with preservation of the operating room capacity. Before the merger, 3 Upper GI surgeons and 4 HPB surgeons worked in the AMC, and 3 upper GI surgeons and 3 HPB surgeons in the VUMC. After the merger, 5 upper GI surgeons (one surgeon quitted as an upper GI surgeon) and 7 HPB surgeons worked in the Amsterdam UMC. The AMC performed annually approximately 120 upper GI and 250 HPB surgical procedures and the VUmc approximately 70 upper GI and 150 HPB surgical procedures. The most important goal of the merger was: preservation and improvement of QoC for complex tertiary care, to excel in medical research, and to improve accessibility and efficiency.[14,15] The merger was executed as a phased process, in which clinical wards merge successively, to ensure preservation of high-quality and critical care. In so called waves, the joined departments were moved to their final location. For the upper GI group, the pre-merger period was July 27th, 2019, through January 26th, 2020, and the post-merger period was January 27th, 2020, through July 26th, 2020. For the HPB group, the pre-merger period was

December 1st, 2020, through May 30th, 2021, and the post-merger period was May 31th, 2021, through November 30th, 2021.

Patient selection and definitions

Adult patients undergoing upper GI or HPB surgical procedures were included. Both elective and urgent procedures were eligible for this study. Upper GI procedures were defined as non-traumatic oncological procedures under general anesthesia performed on the stomach, esophagus, or both. The HPB procedures were defined as any non-traumatic oncological procedures under general anesthesia performed on the liver, pancreas, gallbladder or biliary ducts. The Age Adjusted Charlson Comorbidity Index (AAC) was used in assessing surgical outcomes weighed by comorbid conditions.[16,17] The AAC utilizes 19 pre-determined groups, weighted between one and six points and is adjusted for one point for every decade over 40 years of age with a maximum score of 6 points.

Outcomes

This study investigated short-term QoC, volume, and timeliness of care. Quality of care was assessed using mortality within 90 days postoperatively, hospital readmission rate within 90 days postoperatively, and incidence and severity of postoperative complications as outcome measures. A complication was defined as any deviation from the normal postoperative course.[18] The severity of the complications was assessed with the Comprehensive Complication (CC) index and the Clavien-Dindo classification (CD). The CD was used to categorize adverse events whereas the CC index was used to accurately represent the impact of multiple complications.[18,19] The CC index was calculated based on the registered and corrected CD data extracted from the database. The volume was determined scrutinizing the number of surgical procedures per working day in the pre-merger and post-merger phase. A working day was defined as every midweek day, with the exemptions of national public holidays in the Netherlands (in which no elective procedures are performed). Timeliness of care is defined as the system's capacity to provide care quickly after a need is recognized.[20] In this study it was assessed by investigating differences in waiting time of the surgical procedure in the pre-merger and post-merger phase. Waiting time was defined as interval between the moment a surgical procedure was scheduled and the day of surgery.

Analysis of the impact of the COVID-19 pandemic

The coronavirus disease 2019 pandemic broke out during the inclusion period of the upper GI group with extensive national measures from March 2020, resulting in a general surgical scale down. However, the board of directors of the Amsterdam UMC decided not to postpone surgical procedures of patients undergoing HPB or upper GI procedures for cancer in COVID-19 negative patients, since this oncological care was nationally

considered as urgent. This was also recommended in a national guideline of the Dutch association of surgeons.[21] To analyze the impact on volume of these measures, applied from the 12th of March 2020, patients undergoing an upper GI surgical procedure before March 12th were compared with patients undergoing an upper GI from March 12th.

Statistical analysis

Statistical Package for the Social Sciences, version 26, was used for analyses. Continuous variables were reported as mean with a standard deviation or median with an interquartile range (IQR) in variables with a skewed distribution. Dichotomous variables were noted as number and percentage. Patients in the pre-merger period were compared with the post-merger period. Within the pre-merger period, patient characteristics of patients who underwent their surgical procedure in the AMC were compared with those in the VUmc. Differences between patient characteristics before and after the merger were compared using Pearson chi-square and Fisher's exact test (in case of small cell counts) in dichotomous variables and a student's T-test or Mann-Whitney U test (skew distribution) in continuous variables. The outcomes were assessed utilizing univariate- and multivariate analyses. Multivariate regression analyses were utilized to adjust for confounders with an impact of 10% or more on the unstandardized beta. To assess the relation between the two waves, odds Ratio (OR), mean differences (MD), and 95% confidence interval (95% CI) were used. A p-value <0.05 was considered statistically significant.

Results

Patient characteristics are shown in Table 1 and Table 2. Between July 27th 2019 and November 30th 2021, a total of 631 patients were included: 195 patients (30.9%) underwent an upper GI procedure and 436 (69.1%) a HPB procedure. The pre-wave group contained 320 patients (50.7%) and the post-wave 311 (49.3%). In the pre-merger period, 62 and 135 patients in the AMC, and 35 and 88 in the VUmc underwent an upper GI or HPB surgical procedure, respectively. The median age was 68 years (IQR: 61 – 74) in the upper GI and 66 (IQR: 57 – 73) in the HPB group. The majority of patients was male (71.0% vs 68.9%; *p=0.719* in the upper GI group, 54.3% vs 55.5%; *p=0.243* in the HPB group). In patients undergoing upper GI surgery, esophagectomies constituted the largest proportion of procedures (n=100, 51.3%). In the HPB group pancreatoduodenectomies were performed most often (n=142; 32.6%).

Table 1, patients characteristics, surgery characteristics, and patients history in the upper GI group

	Prior to merge (n=97)				After merge (n=98)	P-value
	AMC (n=62)	**VUmc (n=35)**	**P-value**	**Total pre-merger**		
Elective procedures				93 (95.9%)	97 (99.0%)	0.170
Number of working days	*	*		126	125	
Surgical procedures per working day [median, IQR]	*	*		1 [0-2]	1 [0-2]	0.893
Waiting time for elective surgery in days [IQR, median]	82 [31 – 125]	76 [25 – 97]	0.561	80 [28 – 124]	83 [16 – 105]	0.734
Patient characteristics						
Age in years [median, IQR]	68 [63-74]	69 [61-76]	0.550	68 [62-75]	69 [64-75]	0.493
Sex						
Male	60 (67.4%)	33 (78.6%)	0.189	72 (74.2%)	70 (71.4%)	0.661
Female	23 (32.6%)	9 (21.4%)		25 (25.8%)	28 (28.6%)	
BMI in kg/m^2 [median, IQR]	25 [23- 28]	25 [23 - 28]	0.400	25 [23 - 28]	24 [23 - 29]	0.839
Surgery characteristics						
Site of surgery			0.710			0.352
Gastric	29 (46.7%)	14 (40.0%)		43 (44.3%)	38 (38.8%)	
Esophageal	33 (53.2%)	21 (60.0%)		54 (55.6%)	60 (61.2%)	
Surgical procedure			**0.006**			0.125
Esophagectomy	28 (45.2%)	21 (60.0%)		49 (50.1%)	51 (52.0%)	
McKeown	5 (17.9%)	8 (38.1%)		13 (26.5%)	7 (13.7%)	
Ivor Lewis	18 (64.2%)	9 (42.8%)		27 (55.1%)	38 (74.5%)	
Other	5 (17.9%)	4 (19.0%)		9 (18.4%)	6 (11.8%)	
(Sub)total gastrectomy	19 (30.1%)	12 (34.2%)		31 (32.0%)	32 (32.7%)	
Other	15 (24.2%)	2 (2.1%)		17 (17.5%)	15 (15.3%)	
Patient history						
AAC [median, IQR]	4 [3-6]	4 [3-6]		4 [3-6]	4 [3-6]	0.502
0-1	4 (6.5%)	1 (2.9%)		5 (5.2%)	6 (6.1%)	
2-3	22 (35.5%)	2 (5.7%)		24 24.7%)	16 (16.3%)	
4-5	25 (40.3%)	16 (45.7%)		41 (42.3%)	40 (40.8%)	
≥6	11 (17.7%)	16 (45.7%)		27 (27.8%)	36 (36.7%)	
Abdominal or thoracic surgery in medical history	25 (40.3%)	12 (34.4%)	0.557	37 (38.1%)	40 (40.8%)	0.986

GI; Gastro-Intestinal *surgery, AMC; Academic Medical Center, VUmc; Free University Medical Center, IQR; interquartile range, BMI; body mass index, AAC; Age adjusted Charlson Comorbidity index. *Not applicable.*

Table 2, patients characteristics, surgery characteristics, and patients history in the HPB group

	Prior to the merge (n= 223)				After merge (n= 213)	P-value
	AMC (n=135)	**VUmc (n=88)**	**P-value**	**Total pre-merge**		
Elective procedures	122 (90.4%)	77 (87.5%)	0.499	199 (89.2%)	194 (92.8%)	0.194
Number of working day	*	*		122	131	
Surgical procedures per working day [median, IQR]	*	*		2 [1-2]	1 [1-2]	0.278
Waiting time for elective surgery in days [IQR, median]	22 [12 – 39]	22 [13 -33]	0.559	22 [12 – 36]	24 [13 – 43]	0.163
Patients' characteristics						
Age in years [median, IQR]	65 [57-72]	68 [58-77]	0.182	64 [56-72]	66 [57-73]	0.243
Sex						
Male	83 (61.5%)	38 (43.2%)	**0.007**	121 (54.3%)	116 (55.5%)	0.736
Female	52 (38.5%)	50 (56.8%)		102 (45.7%)	97 (44.5%)	0.719
BMI in kg/m^2 [median, IQR]	25.4 [22.9-29.0]	25.9 [22.9-28.7]	0.898	25.6 [23.0-28.7]	25.0 [22.5-28.0]	0.226
Surgery characteristics						
Site of surgery			0.347			0.205
Hepatic	32 (23.7%)	23 (26.1%)		55 (24.7%)	71 (33.3%)	
Pancreatic	68 (50.4%)	34 (38.6%)		102 (45.7%)	84 (39.4%)	
Biliary	29 (21.5%)	26 (29.5%)		55 (24.7%)	51 (23.9%)	
Other	6 (4.4%)	5 (5.7%)		11 (4.9%)	7 (3.3%)	
Surgical procedure			0.306			0.591
Pancreatoduodenectomy	46 (34.1%)	29 (33.0%)		75 (33.6%)	67 (31.5%)	
Partial hepatectomy	27 (20%)	19 (21.6%)		46 (20.6%)	55 (25.8%)	
Distal pancreatectomy	19 (14.1%)	5 (5.7%)		24 (10.8%)	16 (7.5%)	
Cholecystectomy	27 (20%)	24 (27.3%)		51 (22.9%)	47 (22.1%)	
Other	16 (11.9%)	11 9 (12.5%)		27 (12.1%)	28 (13.1%)	
Patient history						
AAC [median, IQR]	4 [3-6]	4 [3-6]	0.942	4 [3-6]	4 [3-5]	0.458
0-1	10 (7.4%)	8 (9.1%)		15 (11.5%)	13 (10.9%)	
2-3	33 (24.4%)	19 (21.4%)		32 (24.4%)	22 (18.5%)	
4-5	58 (43%)	38 (43.2%)		49 (37.4%)	42 (35.3%)	
≥6	34 (25.2%)	23 (26.1%)		35 (26.7%)	42 (35.3%)	
Abdominal or thoracic surgery in medical history	44 (32.6%)	40 (45.5%)	0.053	84 (37.7%)	80 (37.6%)	0.981

HPB; hepato- *biliary-pancreatic, AMC; Academic Medical Center, VUmc; Free University Medical Center, IQR; Interquartile range, BMI; body mass index, AAC; Age adjusted Charlson Comorbidity index*

Upper GI outcomes

The proportion of patients undergoing upper GI surgery who developed a postoperative complication in the pre-merger and post-merger phase was comparable (58.8% vs 58.2%; OR: 0.88 [95% CI: 0.58-1.88]; *p=0.880*), with a median number of 2 complications in both the pre-merger and post-merger period in patients with a complicated postoperative course (MD: 0.18 [95% CI: -0.59 – 0.95]; *p=0.643*). The median CC index score was 33.5 (IQR: 20.9 – 50.9) in the pre-merger and 33.7 in the post-merger (IQR: 20.9 – 46.0) group (MD: -2.54 [95% CI: 1.-11.21 – 6.12]; *p= 0.562*). A total of 37 patients (34.0%) in the pre-merger and 29 (28.6%) in the post-merger group had a CD of 3 or higher (OR: 0.87 [95% CI: 0.47 – 1.61]; *p=0.661*). Mortality (2.1% vs 1.0%), readmission within 90 days (13.4% vs 13.3%; *p=0.922*), and median length of stay (8 vs 9, *p=0.915*) were all not statistically significant. (Table 3)

Table 3, postoperative outcomes of patients undergoing upper GI surgery

	Prior to merge (n=97)				After merge (n=98)	Multivariate	
	AMC (n=62)	VUmc (n=35)	P-value	Total Pre-merge		OR (95% CI)	P-value
Patients with complication	33 (53.2%)	24 (68.6%)	0.140	57 (58.8%)	57 (58.2%)	1.05 (0.58 – 1.88)	0.880[a]
CD ≥ 3	19 (30.6%)	13 (37.1%)	0.513	33 (34.0%)	28 (28.6%)	0.87 (0.47 – 1.61)	0.661[b]
Mortality	2 (3.2%)	0		2 (2.1%)	1 (1.0%)	*	*
Readmission**	7 (11.3%)	6 (17.1%)	0.416	13 (13.4%)	13 (13.3%)	0.96 (0.412 – 2.23)	0.922[c]
						Multivariate	
						MD (95% CI)	P-value
Number of complications^	2 [1 – 5]	2 [1 – 3]	0.548	2 [1-3]	2 [1-4]	0.18 (-0.59 – 0.95)	0.643[d]
CC index^	33.5 [20.9 – 53.0]	33.7 [16.5 – 68.9]	0.689	33.5 [20.9-51.7]	29.6 [20.9-46.5]	-2.54 (-11.21 – 6.12)	0.562[e]
Length of stay	9 [5 – 14]	7 [6 – 10]	0.440	8 [6-14]	9 [6-14]	-0.08 (-1.52 – 1.37)	0.915[f]

GI; gastro-intestinal, *OR; odds ratio, 95% CI; 95% confidence interval, CD; Clavien-Dindo, CC; comprehensive complication, MD; mean Difference.*

[a] **Adjusted** *for: BMI, AAC*

[b] **Adjusted** *for: BMI*

[c] **Adjusted** *for: age, BMI*

[d] **Adjusted** *for: BMI, gender, elective procedure, type of surgery, AAC*

[e] **Adjusted** *for: age, BMI*

[f] **Adjusted** *for: no confounders found*

***Not applicable**

****Readmission within** *90 days postoperatively*

^In patients *with at least one postoperative complication*

HPB outcomes

The proportion of patients undergoing HPB surgery with postoperative complications in the pre-merger and post-merger phase was comparable (48.4% vs 45.1%; OR: 0.90 [95% CI: 0.59-1.36]; *p=0.606*), with a median number of complications per patient of 2 in both phases (MD: 0.33 [95% CI: -0.08 – 0.73]; *p=0.113*). The median CC index score was 26.2 (IQR: 20.9 – 40.7) in the pre-merger and 33.5 (IQR: 26.2 – 44.9) in the post-merger group (MD: 1.53 [95% CI -2.21 – 5.28]; *p=0.421*). A total of 53 patients (23.8%) in the pre-merger and 64 (30.0%) in the post-merger group had a CD of 3 or higher (OR: 1.44 [95% CI: 0.94 – 2.22]; *p=0.095*). Mortality (0.4% vs 0%; *p=0.995*), readmission within 90 days (11.7% vs 16.9%; *p=0.119*), and median length of stay (7 vs 6, *p=0.905*) were all statistically not significant. (Table 4)

Table 4, postoperative outcomes of patients undergoing HPB surgery

	Prior to merge (n=223)				After merge (n=213)	Multivariate	
	AMC (n=135)	**VUmc (n=88)**	**P- value**	**Total Pre-merger**		**OR (95% CI)**	**P-value**
Patients with complication	61 (45.2%)	47 (53.4%)	0.230	108 (48.4%)	96 (45.1%)	0.90 (0.59 – 1.36)	0.606 [a]
CD ≥ 3	34 (25.2%)	19 (21.6%)	0.538	53 (23.8%)	64 (30.0%)	1.44 (0.94 – 2.22)	0.095 [b]
Mortality	1 (0.7%)	0	0.418	1 (0.4%)	0	*	*
Readmission*	17 (12.6%)	9 (10.2%)	0.591	26 (11.7%)	36 (16.9%)	1.54 (0.90 – 2.65)	0.119 [c]
						Multivariate	
				Pre-wave Median [IQR]	**Post-wave Median [IQR]**	**MD (95% CI)**	**P-value**
Number of complications^	2 [1 – 4]	2 [1 – 4]	0.676	2 [1-4]	2 [1-5]	0.33 (-0.08 -0.73)	0.113 [d]
CC index^	29.5 [20.9 – 42.6]	22.6 [8.7 – 33.7]	0.584	26.2 [20.9-40.7]	33.5 [26.2-44.9]	1.53 (-2.21 - 5.28)	0.421 [e]
Length of stay	7 [4 – 11]	6 [4 -11]	0.672	7 [4-10]	6 [3-10]	-0.12 (-2.10 – 1.87)	0.905 [f]

HPB; hepato- *biliary-pancreatic, OR; odds ratio, 95% CI; 95% confidence interval, CD; Clavien-Dindo, MD; mean Difference, CC; comprehensive complication*

[a] **Adjusted** *for: age, elective procedure, type of surgery*

[b] **Adjusted** *for: age*

[c] **Adjusted** *for: no confounders found*

[d] **Adjusted** *for: age*

[e] **Adjusted** *for: age, elective procedure*

[f] **Adjusted** *for: age, BMI, elective procedure, site of surgery, type of surgery*

***Not applicable**

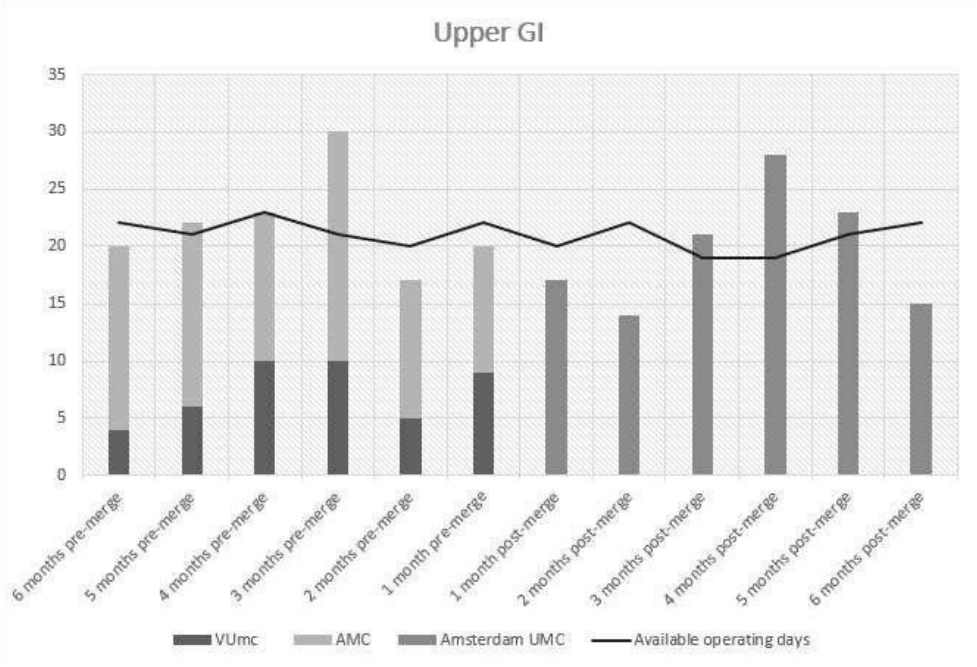
****Readmission within** *90 postoperatively*

^In patients *with at least one postoperative complications*

Impact of the hospital merger on number of surgical procedures and timeliness of care

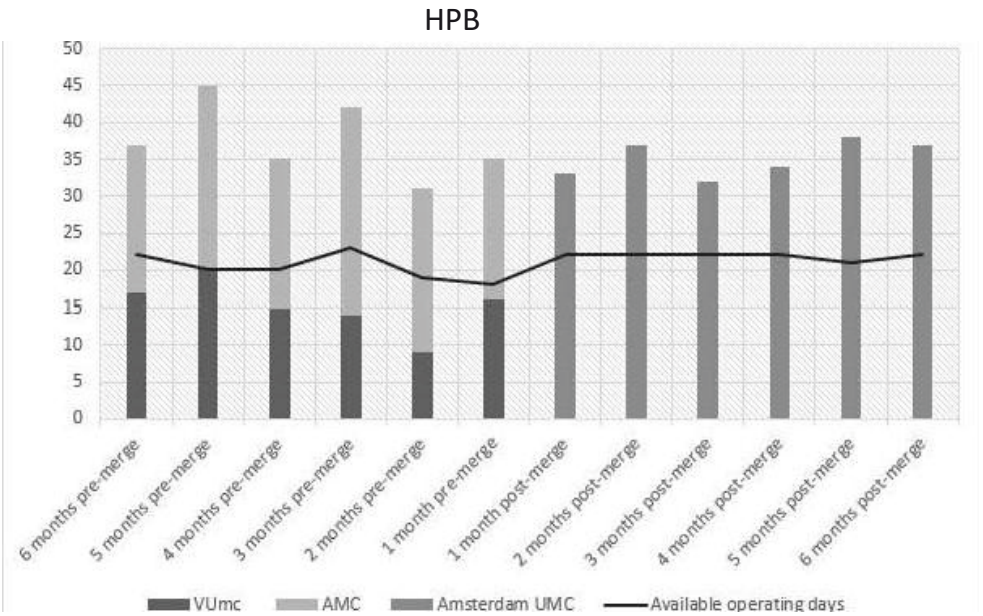
As shown in Table 1 and Figure 1, the median number of upper GI procedures per working day was 1 in both the pre-merger (IQR: 0-2) and the post-merger (IQR: 0-2) phase (*p=0.595*). The median waiting time for elective surgery was 80 days in the pre-merger period and 83 in the post-merger phase (p=0.893). As shown in Table 2 and Figure 2, the number of surgical procedures per working day in the HPB group was 2 in the pre-merger phase (IQR: 1-2) and 1 in the post-merger (IQR: 1-2) phase. This difference, however, was not statistically significant (*p= 0.278*). The median waiting time for elective surgery was 22 days in the pre-merger period and 24 in the post-merger phase (p=0.163).

Figure 1, number of surgical procedures performed in the upper GI group.



This figure shows the number of procedures in patients undergoing upper GI surgery in the six months before the merger on 27th June 2020, in the AMC (light colored bar, n=62) and VUmc hospital (dark colored bar, n=35) and in the six after the merger in the Amsterdam UMC (grey bar, n=98). The black line displays the number of working days in that month.
GI; gastro-intestinal, AMC; Academic Medical Center, VUmc; Free University Medical Center, Amsterdam UMC; Amsterdam University Medical Center.

Figure 2, number of procedures performed in the HPB group HPB



This figure shows the number of procedures in patients undergoing HPB surgery in the six months before the merger on 30th June 2021, in the AMC (light colored bar, n=135) and VUmc hospital (dark colored bar, n=88) and in the six after the merger in the Amsterdam UMC (grey bar, n=213). The black line displays the number of working days in that month.
HPB; hepato-biliary-pancreatic, VUmc; Vrije Universiteit Medisch Centrum, AMC; Academisch Medisch Centrum, Amsterdam UMC; Amsterdam University Medical Center.

Impact of the COVID-19 pandemic

There was no difference in the average number of surgical upper GI procedures during the COVID-19 pandemic. The median number of performed upper GI procedures per working day was 1 before (IQR: 0-1) and 1 (IQR: 0-2) during the pandemic (*p=0.331*). (Table 5)

Table 5, number of patients undergoing upper GI surgery during the COVID-19 pandemic

	Before the COVID-19 pandemic (n=121)	**During the COVID-19 pandemic (n=74)**	**Total (n=195)**	**P-value**
Elective procedures	117 (96.7%)	73 (98.6%)	190 (97.4%)	0.873
Working days	161	90	251	
Elective procedures per working day [median, IQR]	1 [0-1]	1 [0-2]	1 [0-2]	0.331

This table *compares the number of procedures per working day before en during the COVID-19 pandemic, with 12 March 2020 as reference date*
GI; gastro-intestinal, *COVID-19; coronavirus disease 2019, IQR; interquartile range..*

Discussion

The aim of this study was to evaluate the short-term effect of a merger of two university hospitals on quality of surgical care, volume, and timeliness of care in patients undergoing an oncological upper GI or HPB surgical procedure. There were no signs of short-term deterioration in the evaluated quality metrics, volume and timeliness of care after the merger. This suggests that, a hospital merger of two university hospitals can be performed without jeopardizing patient safety in the short-term of two highly-complex surgical disciplines, while profiting from the benefits of the merger. The putative long-term benefits hospital mergers in general are reduced duplication services, beneficial synergistic impacts, and increased market powers.[22] The benefits of this specific merger are, as mentioned in the predetermined goals, centralization of highly specialized tertiary care and enhancement of medical research.[14,15]

Some factors influencing QoC should be addressed in the light of these findings. Firstly, some nurses and paramedics remained at their initial location, according to their own individual preference. This may have led to a loss in expertise and subsequently to QoC, since many of these nurses and paramedics were experienced in the postoperative management of both upper GI and/or HPB patients. It is therefore possible that the QoC will improve in the near future, since new expertise will be gained over time. Secondly, the geographical distance between the merging hospitals might have played a facilitating role in the current merger. Several studies showed negative associations between the distance to the nearest hospital and QoC.[23,24] The travel time between both hospitals both with car and public transport is approximately 20-30 minutes. In this merger it is unlikely that geographical distance influenced the post-merger QoC. The findings of the current study are in contrast with two studies, that described temporary deterioration in QoC in terms of increased waiting times, decreased quantity of patient contacts and increased mortality rates, after university hospital mergers.[11,14] Arguably, the meticulous preparations, such as the foundation of steering groups assisting in the integration, harmonization, and lateralization processes, prior to the merger described in this study must have played an important role in the preservation of QoC.

Yet, the impact of a hospital merger is more than just the effect on QoC and a myriad of caveats should be borne in mind. The rub lies in alignment of the pre- and post-merger culture. The organizational culture of the two hospitals and the communication with employees are found to be dominant factors in determining success or failure of a merger, since it has a very large psychological effect on individual healthcare providers.[25] Next to this, the larger the cultural differences, the more arduous the integration. Gathering and evaluating pre-merger cultural data is essential to provide insight in differences and to formulate a post-merger culture.[26] The aforementioned steering

group were aware of these threats and were designed to tackle them during the whole process. Finally, the meticulous preparations prior to the merger potentially raised the awareness to provide top quality health care. In the long-term this awareness could drop and lead to a decrease in quality of care.[27] A future study is needed to elaborate on the effect of such merger on long-term quality of care.

This is the first study investigating the short-term impact of a merger of two highly specialized surgical departments. However, some limitations should be addressed. The number of included patients was relatively small, reducing the power of the study and increasing the margin of error. Especially in rare events (*e.g.,* patient death) it is usually difficult to assess differences. Patients compared in the pre-merger and post-merger phase underwent a surgical procedure in contrasting seasons. There is some contradictory evidence of the effect of seasonality on postoperative outcomes. Spencer et al. provided in a recently published systematic review tentative evidence for an increased risk of postoperative complications in the summer.[28] This putative effect of seasonality could have influenced the findings of this study. However, this effect is small and unlikely to significantly influence the findings of our relatively small cohort. Then, the COVID-19 pandemic broke out during the post-merger period of the upper GI cohort. This could have led to changes in the outcomes. However, although extensive measures within the hospital were applied, there was no decrease in number of upper GI procedures or increase in morbidity during the COVID-19 pandemic in this study. The national recommendations were followed and were in alignment with the results of Borgstein et al., who showed no increase in morbidity in four European tertiary esophageal cancer referral centers during the first wave of the COVID-19 pandemic.[29] Due to insufficient data available, no analyzes could be performed on patient satisfaction or nurse satisfaction score therefore patient and nursing experience and satisfaction were not regarded in this study. Linking short-term postoperative outcomes to patient experience data can reveal ways to optimize care.[30] Finally, the number of surgical procedures per day also depends on local policy decisions. It could be that OR time was divided unequal between different specialisms pre- and post-merger.

Conclusion

The merger of two Dutch university hospitals did not show any sign of deterioration in the evaluated short-term quality metrics, volume or timeliness for upper GI and HPB surgery. This suggest that a hospital merger of two university hospitals can be performed safely, and while benefitting from centralization of highly specialized care and enhancement of medical research. Future studies should investigate the long-term effect of the merger on quality of care.

References

1. Williams D, Jr., Reiter KL, Pink GH, Holmes GM, Song PH. Rural Hospital Mergers Increased Between 2005 and 2016-What Did Those Hospitals Look Like? Inquiry. 2020;57:1-15.
2. Roos AF. Mergers and competition in the Dutch healthcare sector. Dissertation [May, 2022].
3. Beaulieu ND, Dafny LS, Landon BE, Dalton JB, Kuye I, McWilliams JM. Changes in Quality of Care after Hospital Mergers and Acquisitions. New England Journal of Medicine. 2020;382(1):51-9.
4. Jiang HJ, Fingar KR, Liang L et al. Quality of Care Before and After Mergers and Acquisitions of Rural Hospitals. JAMA Network Open. 2021;4(9):e2124662.
5. Wang E, Arnold S, Jones S et al. Quality and Safety Outcomes of a Hospital Merger Following a Full Integration at a Safety Net Hospital. JAMA Network Open. 2022;5(1):e2142382.
6. Maile EJ, Mitra M, Ovseiko P et al. Merger and post-merger integration at Oxford University Hospitals: mixed-methods evaluation and lessons learned. J Health Organ Manag. 2022;ahead-of-print.
7. Sheetz KH, Dimick JB, Nathan H. Centralization of High-Risk Cancer Surgery Within Existing Hospital Systems. J of Clinical Oncology. 2019;37(34):3234-42.
8. Finks JF, Osborne NH, Birkmeyer JD. Trends in Hospital Volume and Operative Mortality for High-Risk Surgery. New English J Medicine.2011;364:2128-2137
9. Broers B, Kemp R. Dutch hospital mergers: no evidence for improvement of healthcare quality. https://www.acm.nl/sites/default/files/documents/cpi-broers-kemp.pdf [April 2017].
10. Elm Ev, Altman DG, Egger M et al. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. British Medical J. 2007;335(7624):806-8.
11. Heeringa J, Mutti A, Furukawa MF et al. Horizontal and Vertical Integration of Health Care Providers: A Framework for Understanding Various Provider Organizational Structures. Int J Integr Care. 2020;20(1):2.
12. Jaarverslaglegging 2020 Amsterdam UMC, locatie AMC. https://www.amsterdamumc.org/nl/organisatie/jaarverslagen.htm [June 2020].
13. Jaarverslaglegging 2020 Amsterdam UMC, locatie VUmc https://www.amsterdamumc.org/nl/organisatie/jaarverslagen.htm [June 2020].
14. VUMC A-. Naar een Alliantie AMC - VUmc. 2013.
15. VUMC AA-. Intentieverklaring | Visiedocument 2011.
16. Charlson M, Szatrowski TP, Peterson et al. Validation of a combined comorbidity index. J Clinical Epidemiol. 1994;47(11):1245-51.
17. Charlson ME, Pompei P, Ales KL et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Disease. 1987;40(5):373-83.
18. Dindo D, Demartines N, Clavien P-A. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. Annals of Surgery. 2004;240(2):205-13.
19. Slankamenac K, Graf R, Barkun J et al. The comprehensive complication index: a novel continuous scale to measure surgical morbidity. Annals of Surgery. 2013;258(1):1-7.
20. Brown T, Liu JX, Scheffler RM. Does the under- or overrepresentation of minority physicians across geographical areas affect the location decisions of minority physicians? Health Services Research. 2009;44(4):1290-308.

21. Heelkunde NVvH. Handvat voor chirurgische ingrepen tijdens Corona-crisis. https://heelkunde.nl/nieuws/nieuwsbericht?newsitemid=23658500 [June 2020].
22. Harris J, Ozgen H, Ozcan Y. Do mergers enhance the performance of hospital efficiency? J of the Operational Research Society. 2000;51(7):801-11.
23. Tracey E, McCaughan B, Badgery-Parker T et al. Survival of Australian lung cancer patients and the impact of distance from and attendance at a thoracic specialist centre: a data linkage study. Thorax. 2015;70(2):152-60.
24. Virgilsen LF, Hvidberg L, Vedsted P. Patient's travel distance to specialised cancer diagnostics and the association with the general practitioner's diagnostic strategy and satisfaction with the access to diagnostic procedures: an observational study in Denmark. BMC Primary Care. 2020;21(1):97.
25. Keane M, Farragher L. Barriers and facilitators to successful hospital mergers: A systematic review. BMJ Evidence-Based Medicine. 2018;23(Supplement 1):A29.
26. Chesley CG. Merging Cultures: Organizational Culture and Leadership in a Health System Merger. J of Healthcare Management. 2020;65(2):135-50.
27. Aravindakshan A, Naik PA. How does awareness evolve when advertising stops? The role of memory. Marketing Letters. 2011;22(3):315-26.
28. Spencer E, Berry M, Martin P et al. Seasonality in surgical outcome data: a systematic review and narrative synthesis. British J Anaesthesiology. 2022;128(2):321-32.
29. Borgstein ABJ, Brunner S, Hayami M et al. Safety of Esophageal Cancer Surgery During the First Wave of the COVID-19 Pandemic in Europe: A Multicenter Study. Annals of Surgery Oncology. 2021;28(9):4805-13.
30. de Vos MS, Hamming JF, Boosman H et al. The Association Between Complications, Incidents, and Patient Experience: Retrospective Linkage of Routine Patient Experience Surveys and Safety Data. J Patient Safety. 2021;17(2):e91-e7.

5

Chapter 5

Impact of anastomotic bowel leakage on oncological outcomes and stoma presence

Erik W. Ingwersen
Daitlin E. Huisman
Emma E. van der Meulen
Boukje T. Bootsma
Geert Kazemier
Hidde M. Kroon
Ishmam Murshed
Tarrik Sammour
Freek Daams
On behalf of the LekCheck study group

Submitted

Abstract

Importance: Literature on the long-term effects of anastomotic bowel leakage (ABL) with regards to postoperative oncological outcomes and stoma presence is equivocal. Understanding the association between ABL and long-term outcomes is crucial for patient follow-up, counseling, and tailored treatment decisions.

Objective: To investigate the impact of ABL on cancer recurrence and stoma presence in patients undergoing colon and rectal cancer surgery.

Design: This prospective cohort study was conducted in patients included in the LekCheck study. Follow-up period was 5 years.

Setting: A multicenter international study conducted in 14 hospitals in three countries.

Participants: Patients undergoing colon cancer surgery and rectal cancer surgery were included and analyzed separately. Patients with a primary anastomosis were included.

Main outcomes and measures: The outcomes assessed were cancer recurrence rate, overall survival, disease-free survival, and stoma presence at five years of follow-up, stratified based on ABL.

Results: A total of 1 042 patients were included. ABL was not significantly associated with cancer recurrence, overall survival, or disease-free survival after colon or rectal cancer surgery. However, for colon cancer surgery, ABL was associated with the presence of a stoma at five years of follow-up (18.8% vs 1.9%, odds ratio 16.96, 95% confidence interval 8.69 – 33.10, P<0.001). Similarly, in rectal cancer surgery, stoma presence at five years was significantly higher in patients after ABL (21.9% vs 1.9%, odds ratio 5.61, 95% confidence interval 1.82 - 17.27, P<0.001).

Conclusion and relevance: This study did not demonstrate a negative impact of ABL on oncological outcomes. However, ABL was strongly associated with an increased incidence of stoma presence after five years, significantly affecting the patients' quality of life after colon and rectal cancer surgery.

Introduction

Anastomotic bowel leakage (ABL) is a serious complication that can occur following colorectal surgery when an anastomosis is formed. It seriously affects short-term outcomes and leads to increased morbidity, mortality, prolonged hospitalization, and increased costs.[1, 2] The majority of patients who experience ABL will have to undergo surgical reintervention, often necessitating a defunctioning or permanent stoma.[3]

Although the negative impact of ABL on short-term outcomes is well established, its long-term implications remain a topic of debate. Firstly, existing literature shows an association between ABL and the non-reversal of a stoma. However, the findings regarding this association vary considerably across studies.[4-6] Secondly, it is unclear whether ABL has an influence on cancer recurrence. Some studies have confirmed the detrimental effect of ABL on recurrence.[7, 8] Other studies, on the other hand, failed to demonstrate this association.[9, 10] Three main hypotheses explain the pathophysiology of ABL leading to recurrence. The first hypothesis proposes that spillage of cancerous cells into the peritoneal cavity or local implantation of cancer cells at the anastomotic dehiscence increases the risk of recurrence.[11, 12] The second hypothesis suggests that ABL triggers an inflammatory response, leading to the release of cytokines and growth factors, promoting the growth of residual tumor cells.[13] Additionally, severe ABL can cause sepsis, which in turn leads to immunosuppression.[14] Existing data indicate that this immunosuppression fosters the proliferation of metastatic disease and, consequently, cancer recurrence.[15-18] The final hypothesis states that the higher recurrence rates might be caused by no receipt or delay in the administration of adjuvant therapy.[19, 20] The receipt of chemotherapy is at least partially associated with oncological outcomes, and a delay or absence could increase cancer recurrence.

Despite numerous surgical innovations and advancements in perioperative management, data of the national colorectal cancer surgery audit in the Netherlands showed no reduction of the incidence of ABL over the past decade.[21] Consequently, ABL is expected to continue to pose a significant threat to patients. Therefore, it is crucial to demystify the long-term impact on oncological and functional outcomes, as understanding such associations may have implications for follow-up, patient counseling, and in tailoring treatment strategies. To investigate this long-term impact, the current study used long-term follow-up data of the LekCheck study, an international multicenter prospective trial. The Lekcheck study identified perioperative modifiable risk factors for ABL in patients undergoing colorectal surgery.[19]

The objective of this study was to assess the long-term impact of ABL on colon or rectal cancer recurrence, overall survival, disease-free survival, and stoma presence.

Methods

This study utilized long-term data from LekCheck study.[22] This international, multicenter, and prospective study was conducted across fourteen hospitals: eleven in the Netherlands, one in Belgium, one in Italy, and one in Australia. It included patients undergoing elective colorectal surgery with the formation of a primary anastomosis for both malign and benign indications. Preoperative baseline characteristics and real-time intraoperative data using a multifactorial checklist were collected as part of the study. The presence of ABL was evaluated within 30 days postoperatively. ABL was defined according to the definition as proposed by Reisinger et al.: extra luminal presence of contrast fluid on contrast-enhanced computed tomography scans and/or leakage during relaparotomy, requiring re-intervention or other treatment.[23] Additional details regarding the LekCheck study can be found in the previously published study by Huisman et al.[22]

The protocol of the Lekcheck study and the amendment for this current study were approved by the Medical Ethics Review Committee of the participating centers of this study. Informed consent was obtained from all patients. The study adhered to the STROBE guidelines.[24]

Patient population

All patients included in the LekCheck study were eligible for inclusion in the current study. Patients with a benign indication for surgery were excluded. Follow-up duration was defined as the time between the surgical procedure and either death or the moment of data collection. Patients without available medical information after their initial hospital admission were considered lost to follow-up and were also excluded from the analysis. Due to differences in staging procedures, surgical approaches, (neo) adjuvant therapy, and metastatic patterns, colon and rectal cancer were considered as separate entities and analyzed accordingly.[25]

Outcomes and definitions

The primary outcome was the long-term impact of ABL, in terms of oncological outcomes and stoma presence, stratified based on the presence of ABL. Oncological outcomes included cancer recurrence, overall survival, and disease-free survival. For colon cancer, local recurrence was defined as malignant disease located near the primary tumor, port site, or wound sites. In addition, in rectal cancer pelvic recurrence was also classified as local recurrence. Distant recurrence encompassed lymphatic or hematogenous dissemination or invasive growth of the primary tumor to other organs.[26] Recurrent disease was confirmed through histopathology. Overall survival was defined as the time from primary surgery to the date of death from any cause.[26] Disease-free survival

was defined as time from primary surgery to the date of recurrence or death from any cause.[26] Stoma presence was calculated as the proportion of patients with a stoma at five years of follow-up. This was determined from 30 days postoperatively to ensure that all patients who developed ABL and subsequently received a stoma were included. Patients with a stoma reversal within the five years of follow-up were considered stoma-free after reversal.

Statistical analysis

Statistical analyses were performed using IBM SPSS Statistics, version 28.0.1.1. Continuous variables are presented as mean with standard deviation (SD), or as median with an interquartile range (IQR) for skewed variables. Dichotomous and nominal variables are presented as number with percentages. Baseline characteristics were compared between patients with and without ABL and patients with and without cancer recurrence using the Chi-Square test for dichotomous variables and Student t test for continuous variables. Cancer recurrence, overall survival, disease-free survival, and stoma presence were stratified by ABL and initially analyzed using univariate Cox regression models. Variables with a P-value <0.10 in the univariate analysis were subsequently included in a multivariate Cox regression model. Kaplan-Meier curves were used to visualize the effect of ABL on cancer recurrence, overall survival, and disease-free survival. Patients that did not complete five years of follow-up were censored at the time of data collection. A P-value of <0.05 indicated statistical significance.

Results

A total of 1562 patients were included in the LekCheck study. Among them, 286 patients with a benign indication for colorectal surgery were excluded and an additional 234 patients were lost to follow-up. The final analysis included 1042 patients. (Supplementary material 1) The median age of the included patients was 71 years (IQR 61 – 76), with 586 (56.2%) patients being male. The median BMI was 26 kg/m^2 (IQR 21 – 34). Colon cancer was present in 796 patients (76.4%) and rectal cancer in 246 (23.6%). The median follow-up time was 57 months (IQR: 51 – 60).

Baseline characteristics

Baseline characteristics of patients with and without ABL are presented in Table 1, and those with and without cancer recurrence are presented in Table 2. A total of 96 patients (9.2%) developed ABL, 64 (8.0%) of all patients undergoing surgery for colon cancer and 32 (13.0%) of all patients undergoing surgery for rectal cancer.

Table 1. Descriptive statistics of patients with and without anastomotic bowel leakage.

	COLON			RECTUM		
	No ABL (n=732)	**ABL (n=64)**	**P-value**	**No ABL (n=214)**	**ABL (n=32)**	**P-value**
Variable						
Sex			**<0.001**			0.412
Female	346 (47.3%)	17 (26.6%)		83 (38.8%)	10 (31.3%)	
Male	386 (52.7%)	47 (73.4%)		131 (61.2%)	22 (68.8%)	
Age (years)			0.205			0.530
< 70	323 (44.1%)	23 (35.9%)		128 (59.8%)	21 (65.6%)	
≥ 70	409 (55.9%)	41 (64.1%)		86 (40.2%)	11 (34.4%)	
BMI (kg/m^2)			**0.039**			0.280
< 30	546 (75.3%)	40 (63.5%)		165 (77.5%)	22 (68.8%)	
≥ 30	179 (24.7%)	23 (36.5%)		48 (22.5%)	10 (31.3%)	
ASA classification			**0.007**			0.327
< 3	497 (68.2%)	33 (51.6%)		176 (82.2%)	24 (75.0%)	
≥ 3	232 (31.8%)	31 (48.4%)		38 (17.8%)	8 (25.0%)	
Diabetes	111 (15.2%)	18 (28.1%)	**0.007**	35 (16.4%)	4 (12.5%)	0.571
Smoking						
Current smoker	84 (11.6%)	9 (14.1%)	0.564	26 (12.6%)	4 (12.5%)	0.985
Pack years ≥ 15	161 (26.7%)	18 (35.3%)	0.189	41 (25.0%)	11 (40.7%)	0.089
AJCC Stage						
I (T1-2N0M0)	251 (34.9%)	25 (39.7%)	Ref.	74 (35.4%)	13 (40.6%)	Ref.
II (T3-4N0M0)	217 (30.2%)	15 (23.8%)	0.282	60 (28.7%)	6 (18.8%)	0.282
III (T1-4N1-2M0)	202 (28.1%)	16 (25.4%)	0.493	65 (31.1%)	11 (34.4%)	0.933
IV (T1-4N1-2M1)	49 (6.8%)	7 (11.1%)	0.428	10 (4.8%)	2 (6.3%)	0.876
Neoadjuvant therapy						
None	696 (95.3%)	57 (91.9%)	Ref.	127 (59.3%)	18 (56.3%)	Ref.
Radiotherapy	9 (1.2%)	3 (4.8%_	0.838	55 (25.7%)	9 (28.1%)	0.231
Chemotherapy	16 (2.2%)	2 (3.2%)	0.891	3 (1.4%)	0 (0.0%)	0.897
Chemoradiotherapy	9 (1.2%)	0 (0.0%)	0.921	29 (13.6%)	5 (15.6%)	0.982
Adjuvant therapy						
None	590 (80.6%)	50 (79.4%)	Ref.	171 (81.4%)	29 (90.6%)	Ref.
Radiotherapy	9 (0.2%)	0 (0.0%)	0.921	3 (1.4%)	0 (0.0%)	0.412
Chemotherapy	132 (18.0%)	13 (20.6%)	0.631	36 (17.1%)	3 (9.4%)	0.667
Hormonal therapy	1 (0.1%)	0 (0.0%)	0.589	0 (0.0%)	0 (0.0%)	0.815
Surgical approach			**<0.001**			0.677
Open	97 (13.4%)	19 (30.6%)		25 (11.9%)	3 (9.4%)	
Laparoscopic	628 (86.6%)	43 (69.4%)		185 (88.1%)	29 (13.6%)	

Descriptive statistics *of patients with and without anastomotic bowel leakage.*

Bold values *indicate statistical significance.*

ABL; anastomotic *bowel leakage, BMI; body mass index, ASA; American Society of Anesthesiologists, AJCC; American Joint Committee on Cancer*

Table 2. *Descriptive statistics of patients with and without cancer recurrence.*

	COLON			RECTUM		
	No recurrence (n=679)	**Recurrence (n=115)**	**P-value**	**No recurrence (n=199)**	**Recurrence (n=46)**	**P-value**
Variable						
Sex			0.602			**0.004**
Female	313 (46.1%)	50 (43.5%)		84 (42.2%)	9 (19.6%)	
Male	366 (53.9%)	65 (56.5%)		115 (57.8%)	37 (80.4%)	
Age (years)			0.161			0.792
< 70	289 (42.6%)	57 (49.6%)		121 (60.8%)	27 (58.7%)	
≥ 70	390 (57.4%)	58 (50.4%)		78 (39.2%)	19 (41.3%)	
BMI (kg/m^2)			0.319			0.130
< 30	495 (73.7%)	89 (78.1%)		147 (74.2%)	39 (84.8%)	
≥ 30	177 (26.3%)	25 (21.9%)		51 (25.8%)	7 (15.2%)	
ASA classification			0.293			0.512
< 3	457 (67.6%)	72 (62.6%)		162 (82.4%)	36 (78.3%)	
≥ 3	219 (32.4%)	43 (37.4%)		35 (17.6%)	10 (21.7%)	
Diabetes	113 (16.7%)	16 (13.9%)	0.459	27 (13.6%)	11 (23.9%)	0.083
Smoking						0.803
Current smoker	82 (12.2%)	11 (9.6%)	0.429	23 (12.0%)	6 (13.3%)	
Pack years ≥ 15	159 (28.7%)	19 (19.6%)	0.063	41 (27.3%)	10 (25.0%)	
AJCC Stage						
I (T1-2N0M0)	261 (39.1%)	15 (13.3%)	Ref.	75 (38.5%)	12 (26.7%)	Ref.
II (T3-4N0M0)	206 (30.9%)	25 (22.1%)	**0.028**	59 (30.3%)	7 (15.6%)	0.555
III (T1-4N1-2M0)	1734 (25.9%)	44 (38.9%)	**<0.001**	57 (29.2%)	18 (40.0%)	0.099
IV (T1-4N1-2M1)	27 (4.0%)	29 (25.7%)	**<0.001**	4 (2.1%)	8 (17.8%)	**<0.001**
Neoadjuvant therapy						
None	656 (97.0%)	95 (83.3%)	Ref.	125 (63.3%)	19 (41.3%)	Ref.
Radiotherapy	7 (1.0%)	5 (4.4%)	**0.007**	50 (25.1%)	13 (28.3%)	0.700
Chemotherapy	9 (1.3%)	9 (7.9%)	**<0.001**	2 (1.0%)	1 (2.2%)	0.514
Chemoradiotherapy	4 (0.6%)	5 (34.4%)	**<0.001**	21 (10.6%)	13 (28.3%)	**0.002**
Adjuvant therapy						
None	656 (97.0%)	58 (50.4%)	Ref.	172 (87.3%)	28 (62.2%)	Ref.
Radiotherapy	7 (1.0%)	49 (42.6%)	**<0.001**	25 (12.7%)	14 (31.3%)	**<0.001**
Chemotherapy	9 (1.3%)	8 (7.0%)	**<0.001**	0 (0.0%)	3 (1.2%)	-
Hormonal therapy	4 (0.6%)	0 (0.0%)	-	0 (0.0%)	0 (0.0%)	-
Surgical approach			0.322			**0.049**
Open	95 (14.1%)	20 (17.7%)		19 (9.7%)	9 (19.6%)	
Laparoscopic	577 (85.9%)	93 (82.3%)		176 (90.3%)	37 (80.4%)	

Descriptive statistics *of patients with and without cancer recurrence.*

Bold values *indicate statistical significance.*

BMI; body *mass index, ASA; American Society of Anesthesiologists, AJCC; American Joint Committee on Cancer*

After colon cancer surgery, ABL occurred significantly more frequent in male patients, in patients with a BMI ≥30, those with ASA classification ≥3, those with diabetes, and in those who underwent an open procedure. There were no significant differences in baseline characteristics between patients with and without ABL after rectal cancer surgery.

Cancer recurrence

After colon cancer surgery, cancer recurrence was significantly higher in patients with an advanced tumor stage, and when neoadjuvant or adjuvant therapy was administrated. After five year follow-up, the cancer recurrence rate was 15.6% in patients with ABL and 14.3% of the patients without ABL. In both the univariate (P=0.93) and the multivariate analysis (HR 0.88, 95% CI 0.41 – 1.85, P=0.73), ABL was not significantly associated with cancer recurrence.

After rectal cancer surgery, cancer recurrence was significantly higher in male patients, those with an advanced tumor and in patients who underwent an open surgical procedure. After five year follow-up, the cancer recurrence rate was similar in both groups: 18.8% in patients with ABL and 18.7% in those without (P=0.891; multivariate analysis: HR 1.36, 95% CI 0.55 – 3.35, P=0.505). (Figure 1, Supplementary material 3).

Overall survival

For patients who underwent colon cancer surgery, overall survival after five year follow-up was 84.4% in patients with ABL and 86.2% in those without. This difference was not significant in the univariate (P=0.584) and multivariate analysis (HR 1.43, 95% CI 0.68 – 2.94, P=0.348).

After rectal cancer surgery the overall survival rate was 75.0% in patients with ABL and 86.4% in those without. This was not significant either in both the univariate (P=0.059) and multivariate analysis (HR 0.46, 95% CI 0.21 – 1.02, P=0.056). (Figure 2, Supplementary material 3)

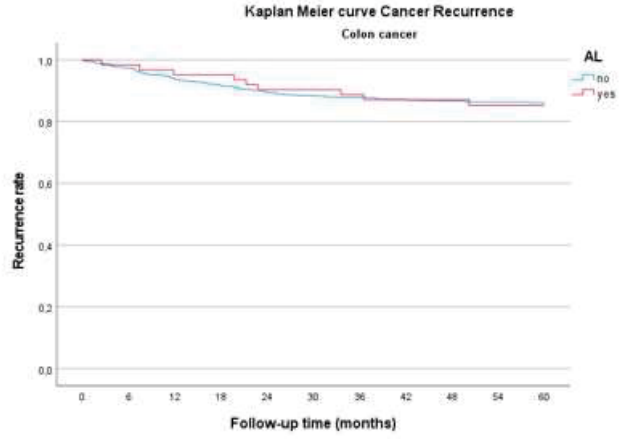
Disease-free survival

Disease-free survival rate after colon cancer surgery was 75.0% in patients with ABL and 76.6% in those without, with no significant difference found in both the univariate (P=0.668) and multivariate analysis (HR 1.79, 95% CI 0.88 – 3.57, P=0.110).

After rectal cancer surgery, 65.6% of the patients with ABL and 74.2% of patients without ABL had a disease-free survival at five year follow-up. This difference was also not significant in both the univariate (P=0.168) and multivariate analysis (HR 0.58, 95% CI 0.29 – 1.14, P=0.111). (Supplementary material 2&3)

Figure 1. *Cancer recurrence*

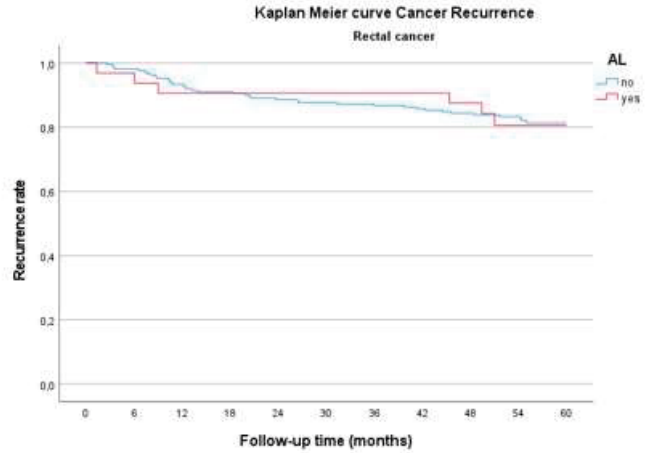
A.



No. At risk

Months after surgery	0	6	12	18	24	30	36	42	48	54	60
ABL	61	60	59	58	56	56	55	52	51	39	23
no ABL	726	708	682	666	651	643	639	628	606	450	273

B.



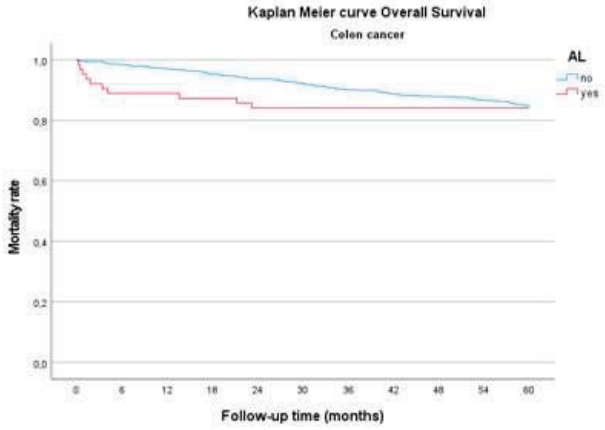
No. At risk

Months after surgery	0	6	12	18	24	30	36	42	48	54	60
ABL	31	30	29	29	29	29	29	29	28	16	5
no ABL	210	206	199	192	187	185	183	180	173	132	77

Cancer recurrence *at five year follow-up. A, cancer recurrence after colon resection. B, cancer recurrence after rectal resection. Red line indicates patients with ABL and the blue line patients without ABL.*
ABL; anastomotic *bowel leakage, no; number.*

Figure 2. *Overall survival*

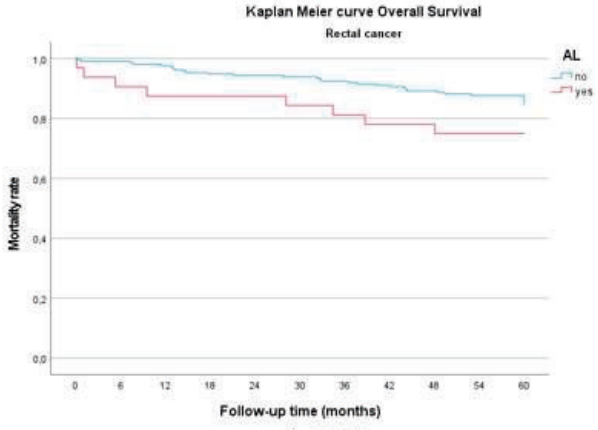
A.



No. At risk

Months after surgery	0	6	12	18	24	30	36	42	48	54	60
ABL	62	56	56	55	53	53	53	51	50	40	25
no ABL	723	713	703	690	678	668	652	637	607	449	269

B.



No. At risk

Months after surgery	0	6	12	18	24	30	36	42	48	54	60
ABL	62	56	56	55	53	53	53	51	50	40	25
no ABL	723	713	703	690	678	668	652	637	607	449	269

Overall survival *at five year follow-up. A, overall survival after colon resection. B, overall survival after rectal resection. Red line indicates patients with ABL and the blue line patients without ABL.*
ABL; anastomotic *bowel leakage, no; number.*

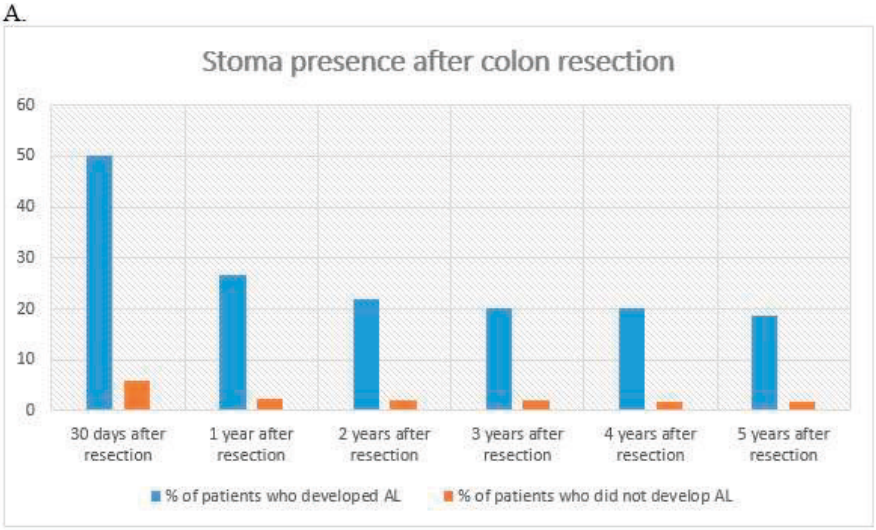
Stoma presence

After colon cancer surgery, 75 (9.4%) patients had a stoma 30 days postoperatively. Among patients who developed ABL, 50.0% had a stoma, while 5.9% of patients without ABL received a stoma. This difference was significant in both the univariate (P<0.001) and multivariate analysis (OR 15.79, 95% CI 7.72 – 32.30, P<0.001). In the multivariate analysis the following confounders were included: ASA classification, neoadjuvant therapy, and surgical approach. After five year follow-up, there was still a difference between stoma presence in patients with ABL (18.8%) and without (1.9%). This difference was significant in both the univariate (P<0.001) and multivariate analysis (OR 16.78, 95% CI 6.22 – 45.29, P<0.001). Age and ASA classification were included as confounders in the multivariate analysis.

After rectal surgery, 97 (39.4%) patients had a stoma 30 days postoperatively. Patients who developed ABL were more likely to have a stoma (90.6%) compared to patients without ABL (31.8%). Both in the univariate (P<0.001) and multivariate analysis (OR 16.77, 95 CI 4.62 – 60.80, P<0.001) significance was found. In the multivariate analysis the following confounders were included: sex, age, ASA classification and neoadjuvant therapy. After five year follow-up, 24.2% (21.9% of the total rectal cohort) of the patients who developed ABL and had a stoma after 30 days postoperatively still had a stoma. In patients without ABL, 4.8% (1.9% of the total cohort) still had a stoma present after five year follow-up. Both in the univariate (P<0.001) and multivariate analysis (OR 5.61, 95% CI 1.82 – 17.27, P<0.001) this difference was significant. In the multivariate analysis only sex showed significance as a confounder. (Figure 3, Supplementary Material 3)

Figure 3. *Stoma presence*

A.



B.

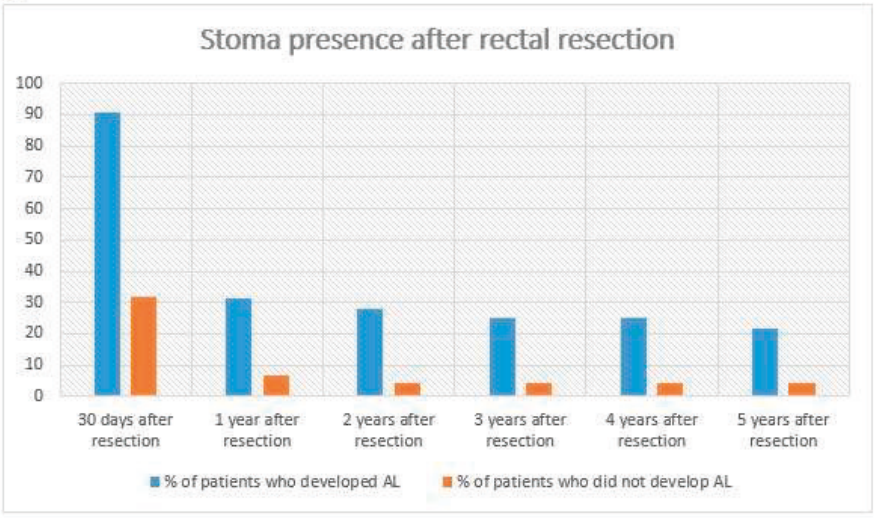


Figure legend *3. Stoma presence during five years of follow-up. A, stoma presence in patients after colon resection. B, stoma presence in patients after rectal resection. The blue bar indicates patients who developed ABL and the orange bar patients who did not develop ABL.*

Discussion

This study did not find an influence of ABL on oncological outcomes after colon and rectal cancer surgery. However, more patients had a stoma after five year of follow-up if they experienced ABL.

Several studies have examined the impact of ABL on cancer recurrence, yielding contrasting results. Some studies, consistent with the current findings, have not been able to show a significant influence of ABL on cancer recurrence.[9, 10, 27] Other studies, however, have demonstrated a positive correlation between ABL and cancer recurrence.[7, 8, 26, 28, 29] The variation in the reported incidence of ABL, ranging from 2.0% to 33.0%, may have contributed to the discrepancy between the results of these studies. It is important to note that the lack of a universally accepted definition for ABL complicates comparison of results.[30,31] Some studies consider only severe leaks requiring surgical reintervention as ABL, while others also include 'subclinical' leaks in their definition.[30, 31] Studies that exclusively report severe cases of leak as ABL are more likely to capture patients with severe infection and immunosuppression, which could potentially result in higher rates of cancer recurrence.[14-18] The study by Koedam et al., for example, who did demonstrate an association between ABL and cancer recurrence, reported a much lower incidence of ABL of 2.5% compared to the current study.[23] For future studies, it is recommended to establish a universally accepted and well described definition of ABL, incorporating a grading system based on severity. Changes in oncological management over the past decades, particularly the increased use of adjuvant and neoadjuvant therapy, may also account for differences in study results. The administration of (neo)adjuvant therapy reduces cancer recurrence rates and improves overall survival, potentially mitigating the effect of ABL.[32, 33]

The Kaplan-Meier curves of rectal cancer patients (Figure 2b and Supplementary material 2b) show that there was a notable reduction of more than 9% in both overall and disease-free survival after ABL. However, this difference was not statistically significant. Specifically, when comparing patients with ABL to those without, the most decrease in overall and disease-free survival is observed in the first year of follow-up, indicating the detrimental short-term effect of ABL after rectal surgery. A larger study, using data of 20,000 rectal cancer patients from the Netherlands Cancer Registry from 2008-2018, did demonstrate a negative impact of ABL on overall and disease-free survival in patients undergoing rectal cancer surgery.[27] Therefore, lack of statistical significance might be the result of the current study being underpowered.

Adjuvant and neoadjuvant therapy was significantly more often administered in patients with cancer recurrence after colon and rectal surgeries. This relation is likely

attributed to the more advanced preoperative tumor stages in these patients, which increases the risk of cancer dissemination. Interestingly, patients who underwent an open rectal resection had a significantly higher chance of cancer recurrence compared to those undergoing laparoscopic surgery. This observation is likely also due to patient characteristics, as laparoscopic surgery is the standard approach for rectal surgery in the predominantly Dutch patient population of this study, whereas open surgery is often only reserved for those with more advanced tumor stages.

The current study clearly highlights the negative impact of ABL on stoma presence after colon and rectal cancer surgery. The presence of a stoma, not only at 30 days postoperatively, but throughout the five years of follow-up, was significantly higher in patients who experienced ABL. Notably, the non-reversal rates in the five year follow-up period were higher in the patients who developed ABL compared to patients who did not (24.2% vs 4.8%). These findings align with previous studies, emphasizing the importance of preventing ABL.[34-36] Stoma-related problems, such as sexual problems, depression, constipation, and travel limitations, significantly impact the overall quality of life for patients.[37] Therefore, substantial benefits can be achieved by preventing ABL. One way to assist achieving this goal is making use of artificial intelligence and machine learning models that could help to obtains more accurate predictions.[38, 39] Preliminary results of a study utilizing a machine learning model to predict ABL intraoperatively using pre- and perioperative data have shown promising results.[40] This model is currently undergoing external validation in several hospitals in the Netherlands. A diverting stoma could be applied in patients who are at high-risk for ABL, since the diverting stoma reduces the severity of ABL and, as shown in this study, successful reversal rates of this stoma are high.[18]

The strengths of the current study include its multicentre design and the large patient cohort from the LekCheck study. Additionally, data were collected prospectively, which provided a reliable form of data collection. Finally, this study not only analyzed oncological outcomes, but also investigated stoma presence, providing valuable insights into the long-term impact of ABL. Several limitations should be acknowledged. Firstly, this study did not differentiate between local and distant cancer recurrence due to a low number of patients with local recurrence (n=36, 3.5%). Therefore, we combined both types of recurrence for the analysis. Secondly, some patients did not reach the full 60-month follow-up period and had to be censored. However, a significant portion of patients reached the 60-month follow-up period, and nearly all patients reached 36 months, minimizing the potential impact of missing data of the results. Lastly, although the outcomes were analyzed in a multivariate analysis, where confounders were included, it is possible that some factors that contributed to the outcome, such as cancer stage of individual patients were missed.

Conclusion

This study showed no significant difference in cancer recurrence, overall survival, or disease-free survival between patients with and without ABL after surgery for colon or rectal cancer. However, AL was strongly associated with increased stoma presence after five years of follow-up in this group of patients.

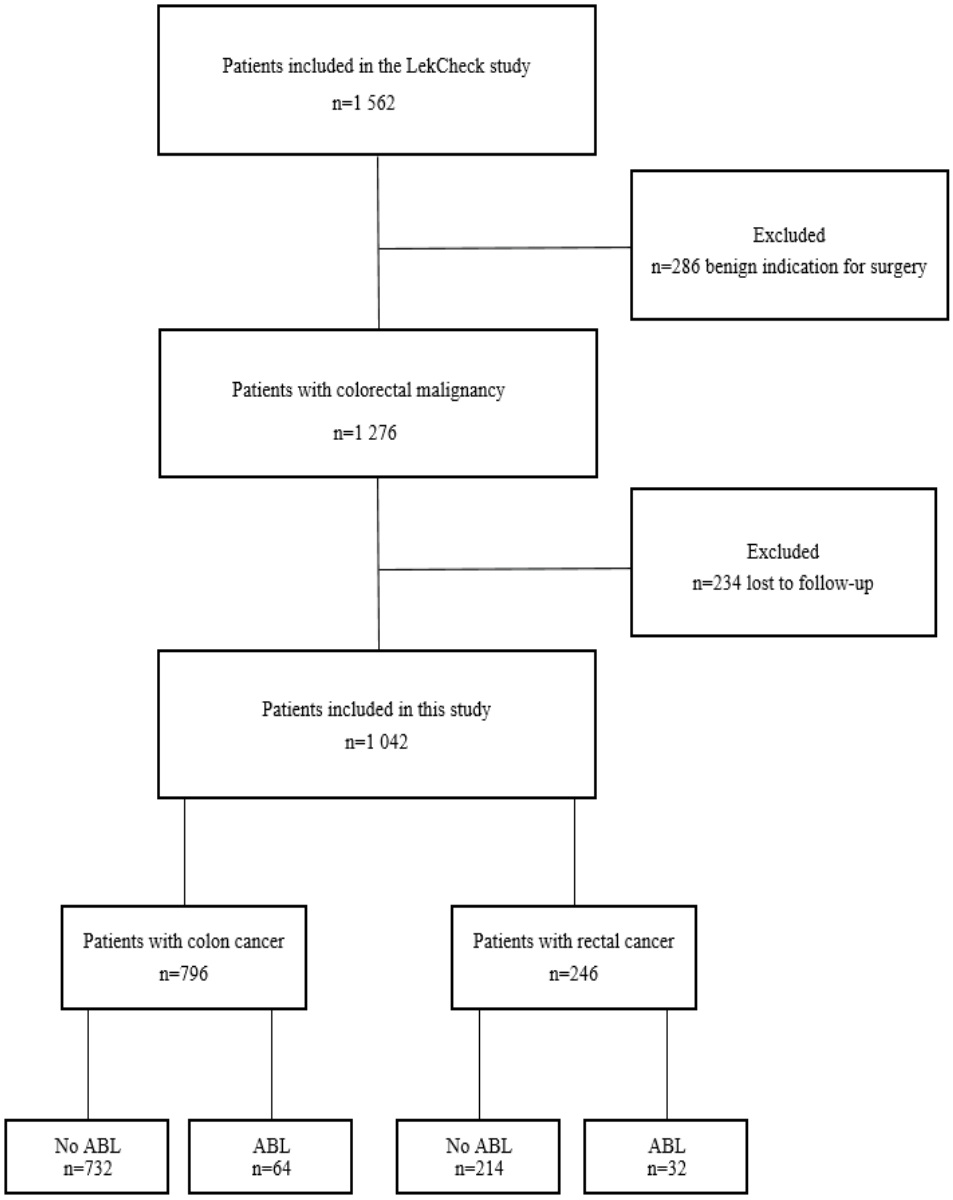
References

1. Chadi SA, Fingerhut A, Berho M, et al. Emerging Trends in the Etiology, Prevention, and Treatment of Gastrointestinal Anastomotic Leakage. *J Gastrointest Surg*. 2016;20(12):2035-51.
2. Chiarello MM, Fransvea P, Cariati M, et al. Anastomotic leakage in colorectal cancer surgery. *Surg Oncol*. 2022;40:101708.
3. van Workum F, Talboom K, Hannink G, et al. Treatment of anastomotic leakage after rectal cancer resection: The TENTACLE-Rectum study. *Colorectal Dis*. 2021;23(4):982-8.
4. Eriksen JD, Emmertsen KJ, Madsen AH, et al. Anastomotic leakage following restorative rectal cancer resection: treatment and impact on stoma presence 1 year after surgery—a population-based study. *Int J Colorectal Dis*. 2022;37(5):1161-72.
5. Holmgren K, Kverneng Hultberg D, Het al. High stoma prevalence and stoma reversal complications following anterior resection for rectal cancer: a population-based multicentre study. *Colorectal Dis*. 2017;19(12):1067-75.
6. Mak JCK, Foo DCC, Wei R, et al. Sphincter-Preserving Surgery for Low Rectal Cancers: Incidence and Risk Factors for Permanent Stoma. *World J Surg*. 2017;41(11):2912-22.
7. Wang S, Liu J, Wang S, et al. Adverse Effects of Anastomotic Leakage on Local Recurrence and Survival After Curative Anterior Resection for Rectal Cancer: A Systematic Review and Meta-analysis. *World J Surg*. 2017;41(1):277-84.
8. Mirnezami A, Mirnezami R, Chandrakumaran K, et al. Increased local recurrence and reduced survival from colorectal cancer following anastomotic leak: systematic review and meta-analysis. A*nn Surg*. 2011;253(5):890-9.
9. Espín E, Ciga MA, Pera M, et al. Oncological outcome following anastomotic leak in rectal surgery. *Br J Surg*. 2015;102(4):416-22.
10. Nachiappan S, Askari A, Malietzis G, et al. The impact of anastomotic leak and its treatment on cancer recurrence and survival following elective colorectal cancer resection. *World J Surg*. 2015;39(4):1052-8.
11. Gertsch P, Baer HU, Kraft R, et al. Malignant cells are collected on circular staplers. *Dis Colon Rectum*. 1992;35(3):238-41.
12. Fermor B, Umpleby HC, Lever JV, et al. Proliferative and metastatic potential of exfoliated colorectal cancer cells. *J Natl Cancer Inst*. 1986;76(2):347-9.
13. Miki C, Tanaka K, Inoue Y, et al. Perioperative host-tumor inflammatory interactions: a potential trigger for disease recurrence following a curative resection for colorectal cancer. *Surg Today*. 2008;38(7):579-84.
14. Liu D, Huang SY, Sun JH, et al. Sepsis-induced immunosuppression: mechanisms, diagnosis and current treatment options. *Mil Med Res*. 2022;9(1):56.
15. Crozier JE, McKee RF, McArdle CS, et al. The presence of a systemic inflammatory response predicts poorer survival in patients receiving adjuvant 5-FU chemotherapy following potentially curative resection for colorectal cancer. *Br J Cancer*. 2006;94(12):1833-6.
16. Mynster T, Christensen IJ, Moesgaard F, et al. Effects of the combination of blood transfusion and postoperative infectious complications on prognosis after surgery for colorectal cancer. Danish RANX05 Colorectal Cancer Study Group. *Br J Surg*. 2000;87(11):1553-62.
17. Law WL, Choi HK, Lee YM, et al. The impact of postoperative complications on long-term outcomes following curative resection for colorectal cancer. *Ann Surg Oncol*. 2007;14(9):2559-66.

18. Khoury W, Lavery IC, Kiran RP. Effects of chronic immunosuppression on long-term oncologic outcomes for colorectal cancer patients undergoing surgery. *Ann Surg*. 2011;253(2):323-7.
19. Fang C, Nie P, Jing P, et al. Effects of Adjuvant Therapy Compliance and Anastomotic Leakage on the Oncologic Outcomes of Patients With Rectal Cancer After Curative Resection. *Dis Colon Rectum*. 2021;64(6):689-96.
20. Kim IY, Kim BR, Kim YW. The impact of anastomotic leakage on oncologic outcomes and the receipt and timing of adjuvant chemotherapy after colorectal cancer surgery. *Int J Surg*. 2015;22:3-9.
21. Ingwersen EW, van der Beek PJK, Dekker JWT, et al. One Decade of Declining Use of Defunctioning Stomas After Rectal Cancer Surgery in the Netherlands: Are We on the Right Track? *Dis Colon Rectum*. 2023 Jul 1;66(7):1003-1011.
22. Huisman DE, Reudink M, van Rooijen SJ, et al. LekCheck: A Prospective Study to Identify Perioperative Modifiable Risk Factors for Anastomotic Leakage in Colorectal Surgery. *Ann Surg*. 2022;275(1):e189-e97.
23. Reisinger KW, Poeze M, Hulsewé KW, et al. Accurate prediction of anastomotic leakage after colorectal surgery using plasma markers for intestinal damage and inflammation. *J Am Coll Surg*. 2014;219(4):744-51.
24. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med*. 2007;4(10):e297.
25. Tamas K, Walenkamp AM, de Vries EG, et al. Rectal and colon cancer: Not just a different anatomic site. *Cancer Treat Rev*. 2015;41(8):671-9.
26. Koedam TWA, Bootsma BT, Deijen CL, et al. Oncological Outcomes After Anastomotic Leakage After Surgery for Colon or Rectal Cancer: Increased Risk of Local Recurrence. *Ann Surg*. 2022;275(2):e420-e7.
27. Arron MNN, Greijdanus NG, Bastiaans S, et al. Long-Term Oncological Outcomes After Colorectal Anastomotic Leakage: A Retrospective Dutch Population-based Study. *Ann Surg*. 2022;276(5):882-9.
28. Branagan G, Finnis D. Prognosis after anastomotic leakage in colorectal surgery. *Dis Colon Rectum*. 2005;48(5):1021-6.
29. Law WL, Choi HK, Lee YM, et al. Anastomotic leakage is associated with poor long-term outcome in patients after curative colorectal resection for malignancy. *J Gastrointest Surg*. 2007;11(1):8-15.
30. van Helsdingen CP, Jongen AC, de Jonge WJ, et al. Consensus on the definition of colorectal anastomotic leakage: A modified Delphi study. *World J Gastroenterol*. 2020;26(23):3293-303.
31. van Rooijen SJ, Jongen AC, Wu ZQ, et al. Definition of colorectal anastomotic leakage: A consensus survey among Dutch and Chinese colorectal surgeons. *World J Gastroenterol*. 2017;23(33):6172-80.
32. Glimelius B. Multidisciplinary treatment of patients with rectal cancer: Development during the past decades and plans for the future. *Ups J Med Sci*. 2012;117(2):225-36.
33. Lemmens V, van Steenbergen L, Janssen-Heijnen M, et al. Trends in colorectal cancer in the south of the Netherlands 1975-2007: rectal cancer survival levels with colon cancer survival. *Acta Oncol*. 2010;49(6):784-96.
34. Jutesten H, Draus J, Frey J, et al. High risk of permanent stoma after anastomotic leakage in anterior resection for rectal cancer. *Colorectal Dis*. 2019;21(2):174-82.

35. Lindgren R, Hallböök O, Rutegård J, et al. What is the risk for a permanent stoma after low anterior resection of the rectum for cancer? A six-year follow-up of a multicenter trial. *Dis Colon Rectum*. 2011;54(1):41-7.
36. Hüser N, Michalski CW, Erkan M, et al. Systematic review and meta-analysis of the role of defunctioning stoma in low rectal cancer surgery. *Ann Surg*. 2008;248(1):52-60.
37. Vonk-Klaassen SM, de Vocht HM, den Ouden ME, et al. Ostomy-related problems and their impact on quality of life of colorectal cancer ostomates: a systematic review. *Qual Life Res*. 2016;25(1):125-33.
38. Wen R, Zheng K, Zhang Q, et al. Machine learning-based random forest predicts anastomotic leakage after anterior resection for rectal cancer. *J Gastrointest Oncol*. 2021;12(3):921-32.
39. Stam WT, Goedknegt LK, Ingwersen EW, et al. The prediction of surgical complications using artificial intelligence in patients undergoing major abdominal surgery: A systematic review. *Surgery*. 2022;171(4):1014-21.
40. Ingweren EW, Stam, W.T., Daams, F. et al. A1Check: the External Validation of a Machine Learning Model Predicting Colorectal Anastomotic Leakage (A1Check). Protocol on *Clinical trial.gov.* 2023.

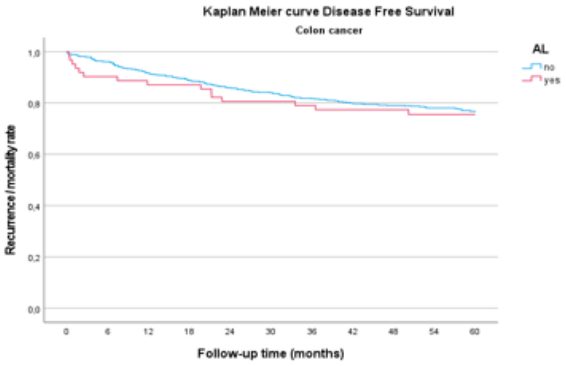
Supplementary material *1. Flowchart.*



Flowchart. A *total of 1 562 patients were included in the LekCheck study and were found eligible for this study. Some 286 patients were excluded because of a benign indication for surgery and 234 patients were excluded due to lost to follow-up. A total of 1 042 patients were ultimately included. ABL; anastomotic bowel leakage.*

Supplementary Material *2.*

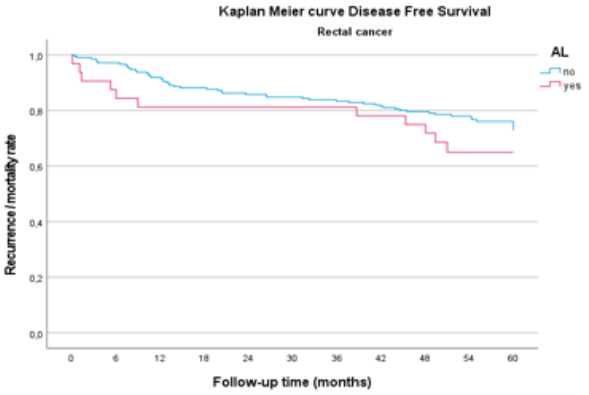
A.



No. At risk

Months after surgery	0	6	12	18	24	30	36	42	48	54	60
ABL	62	56	56	55	53	53	53	51	50	40	25
no ABL	723	713	703	690	678	668	652	637	607	449	269

B.



No. At risk

Months after surgery	0	6	12	18	24	30	36	42	48	54	60
ABL	62	56	56	55	53	53	53	51	50	40	25
no ABL	723	713	703	690	678	668	652	637	607	449	269

Disease-free survival *at five year follow-up. A, disease-free survival after colon resection. B, disease-free survival after rectal resection. Red line indicates patients with ABL and the blue line patients without ABL.*
ABL; anastomotic *bowel leakage, no; number.*

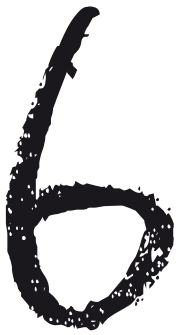
Supplementary material *3. Results of uni- and multivariate analysis for oncological outcomes and stoma presence at 5-years follow-up*

	COLON						RECTUM					
Variable	**ABL (n=64)**	**No ABL (n=732)**	**Univariate P-value**	**Multivariate P-value**	**HR/ OR**	**95%-Confidence interval**	**ABL (n=32)**	**No ABL (n=214)**	**Univariate P-value**	**Multivariate P-value**	**HR/ OR**	**95%-Confidence interval**
Cancer recurrence	10 (15.6%)	105 (14.3%)	0.93	0.73	0.88	0.41 – 1.85	6 (18.8%)	40 (18.7%)	0.891	0.505	1.36	0.55 – 3.35
Overall survival	54 (84.4%)	623 (86.2%)	0.584	0.348	1.43	0.68 – 2.94	24 (75.0%)	185 (86.4%)	0.059	0.056	0.46	0.21 – 1.02
Disease-free survival	48 (75.0%)	561 (76.6%)	0.668	0.110	1.79	0.88 – 3.57	21 (65.6%)	159 (74.2%)	0.168	0.111	0.58	0.29 – 1.14
Stoma presence	12 (18.8%)	14 (1.9%)	<0.001	<0.001	16.78	6.22 – 45.29	7 (21.9%)	9 (4.2%)	<0.001	<0.001	5.61	1.82 – 17.27

Univariate and *multivariate analysis for oncological outcomes and stoma presence after five years of follow-up. The hazard ratio was calculated for cancer recurrence, overall survival, and disease-free survival. The odds ratio was calculated for stoma presence.*
ABL; anastomotic *bowel leakage, HR; hazard ratio, OR; odds ratio.*

Part II

The use of artificial intelligence in predictive modelling



Chapter 6

Machine learning models in clinical practice for the prediction of postoperative complications after major abdominal surgery

Wessel T. Stam
Erik W. Ingwersen
Mahsoem Ali
Jorik T. Spijkerman
Geert Kazemier
Emma R.J. Bruns
Freek Daams

Surgery Today 2023 Oct 53:1209-1215

Abstract

Complications after surgery have major impact on short and long term outcome. Decades of technological advancement did not lead to eradication of its risk. Accurate prediction of complications, recently enhanced by development of machine learning algorithms, has the potential to completely reshape surgical patient management. In this paper we reflect on multiple issues from the development to the actual implementation of machine learning models in daily clinical practice, providing suggestions when using machine learning models for the prediction of postoperative complication after major abdominal surgery.

Introduction

Surgical patients are inevitably at risk of postoperative complications, despite decades of scientific and technological advancement. Accurate prediction of individual outcome has the potential to completely reshape the future of postoperative management. Accurate prediction would enable shared clinical decision making, individual perioperative care and postoperative management. In the last few years an abundance of prediction models have been developed using machine learning (ML) models. These models offer the opportunity to develop a more individualized approach allowing for data-driven individualized medicine.[1-3] However, clinical implementation and acceptance is cumbersome as it is often hampered by non-compliance to the necessary guidelines. To achieve transparent, safe, and applicable implementation of ML-models in the prediction of postoperative outcome, we propose uniform selection, training, and guideline compliance.

Barriers and solutions

There is a large gap between promising and comprehensive research on the potential of AI in the field of medicine and the actual implementation in daily clinical practice.[4] Several authors have tried to use ML-models to optimize postoperative management by predicting postoperative complications. Unfortunately, many conclude that the implementation of these models are far from being clinically applicable even though most ML-models achieve reasonable performance.[5-15] Cao et al.[5], Weller et al.[12], and Van den Bosch et al.[15] conclude that no practical implementation was achieved for ML-models due to too low a predictive value to clinically implement. Grass et al. predicted surgical site infection in patients after colorectal surgery and initially found that ML outperformed conventional logistic regression. However, after external validation the practical applicability dropped due to low predictive performance.[7] Table 1 summarizes the outcomes of the largest most recent articles on postoperative complication prediction with ML in major abdominal surgery. The shift to clinical implementation will depend on five main improvement categories: technology, policy, medical and economic impact, transparency, and reporting.[4, 16, 17]

Table 1. *Description of recent large studies on ML and complication prediction and their key findings N; number, ML; machine learning, LR; logistic regression*

First author (year)	Study design	Data source	Population	Prediction	Methods	Outcome
Cao (2019)	Retrospective	National database	N: 44,061 Surgical specialty: bariatric surgery number of patients with complications: 1,408 (3.2%)	30-days severe postoperative complications	Comparison of 29 ML-models Internal and external validation Overfitting prevention	No achievement of practical applicability due to low performance
Chen (2018)	Retrospective	Institutional database	N: 13,399 Surgical specialty: colorectal surgery number of patients with complications: 1,680 (12.5%)	7-days postoperative bleeding	Comparison of an ML-model to conventional logistic regression Internal validation, no external validation Overfitting prevention	ML-models outperformed conventional LR. They potentially allow for targeted surveillance
Grass (2020)	Retrospective	National database	N: 2,376 Surgical specialty: colorectal surgery number of patients with complications: 108 (4.6%)	Surgical site infection	Comparison of an ML-model to conventional logistic regression Internal validation, no external validation	ML-models outperformed conventional LR. However external validation showed low performance. Institutional data was advised.
Han (2020)	Retrospective	Institutional database	N: 1,769 Surgical specialty: pancreaticoduodenectomy number of patients with complications: 221 (12.5%)	Postoperative pancreatic fistula	Two ML-models tested Internal validation, no external validation	ML-models outperformed conventional LR.
Merath (2020)	Retrospective	National database	N: 15,657 Surgical specialty: liver (685 [4.4%]), pancreatic (6012 [38.4%]) and colorectal (8960 [57.2%]) surgery number of patients with complications: 6,073 (38.8%)	30-days postoperative complications	Comparison of an ML-model to conventional risk scores Internal validation, no external validation	ML-models outperformed conventional LR.

Nudel (2021)	Retrospective	National database	N: 436,807 Surgical specialty: bariatric surgery number of patients with complications: 3,068 (0.70%) developed anastomotic leakage, 2012 (0.46%) developed venous thromboembolism	30-days postoperative anastomotic leakage and venous thromboembolism	Comparison of two ML-models to conventional logistic regression Internal and external Overfitting prevention	ML-models outperformed conventional LR regarding anastomotic leakage LR-models outperformed ML regarding venous thromboembolism
Pera (2022)	Retrospective	National database	N: 3,182 Surgical specialty: gastric surgery Number of patients with complications: 178 (5.6%)	90-days postoperative mortality	Comparison of three ML-models to conventional logistic regression Internal validation, no external validation Overfitting prevention	Good ML-model performance, however due to no external validation and an imbalanced dataset outcomes may not be extrapolable.
Shi (2012)	Retrospective	National database	N: 22,926 Surgical specialty: liver number of patients with complications: 619 (2.7%)	In-hospital mortality	Comparison of two ML-models to conventional logistic regression Internal and external validation	ML-models outperformed conventional LR.
Van den Bosch (2022)	Retrospective	National database	N: 62,501 Surgical specialty: colorectal surgery Number of patients with complications: 1693 (2.7%)	30-days postoperative mortality	Comparison of four ML-methods Internal validation, no external validation Overfitting prevention	ML-models outperformed conventional LR. However, there was no achievement of practical applicability due to low performance
Weller (2018)	Retrospective	Institutional database	N: 9,598 Surgical specialty: colorectal surgery	Postoperative complications	Comparison of four ML-methods Internal and external validation	No achievement of practical applicability due to low performance
Wise (2019)	Retrospective	National database	N: 101,721 Surgical specialty: bariatric surgery number of patients with complications: 3,853 (3.8%)	30-days postoperative morbidity and mortality	Comparison of an ML-model to conventional logistic regression Internal validation, no external validation	LR-models outperformed ML

Selection of a model

Model selection in conventional statistics is well defined due to strict requirements, as opposed to ML models, in which it typically depends on several factors. Even experienced data scientists have difficulty in selecting the optimal model.[18] Model selection in ML, the highest over optimism-corrected performance in the metric and output of choice, is performed post hoc by exploration of different models. Factors of consideration are; quality and size of the data, questions which are required to be answered, the time available for running the model, and finally the type of desired output. Depending on the hypothesis, different metrics can be used to measure this performance. For example, in the prediction of events with a relatively low incidence (e.g. anastomotic leakages after hemicolectomy) imbalanced data may occur. In these cases a single metric as accuracy is less effective. Therefore, precision, recall, and the *f*1-score are more explanatory. Although it is advisable to use multiple metrics to measure performance more broadly, most studies published so far on artificial intelligence (AI) in complication prediction did not adhere to this principle. The majority of prior studies only focused on sensitivity, specificity, and the area under the receiver operating characteristic curve (AUROC).[19, 20]

The complexity of ML models potentially leads to challenges in the understanding of its mechanisms.[21] Many argue that it is highly unfavorable to rely on outcomes of 'black-box'-systems in the decision-making process as this would ignore the moral responsibilities of the medical professional. However, at the same time we regularly prescribe medication such as acetaminophen without fully understanding its mechanism.[22] It is argued that without pursuing the 'explainability' of AI tools, better outcomes for patients cannot be provided.[23-25] Furthermore, many state that carrying out the decision-making process by a 'black box'-system entails dangers.[26-28] However, these dangers are of different importance based on the ethical burden of the decisions that depend on it.[29] For example, in the prediction of postoperative mortality, model explainability might be of more value compared to the prediction of postoperative delayed gastric emptying after distal gastrectomy. As Aristotle stated over two millennia ago, 'the ability to verify results by empirical means are more important than to explain the etiology of these results' This is particularly important in a field in which knowledge on causality is often incomplete, as in postoperative sequalae.[21] Proving that a complication can be predicted, while having the ability to reproduce these results, might be more important than how this prediction is made. Therefore, a more complex and less explanatory model could be accepted in the prediction of complications.[30] However, to maximize the model interpretability, one could use either individual conditional expectations (ICE), local interpretable model-agnostic explanations (LIME) or Shapley Additive Explanations (SHAP).[31] These techniques aim to increase the comprehensibility of the rationale behind the model's prediction by visualizing the contributing impact of different variables. This offers the possibility to approach the mechanisms within black-box systems in a way which would empower the clinician to trust the results produced by these models.

Training, validating, and testing

Generalizability of an ML model depends on the extent and quality of its training, validating, and testing. A very complex model for the prediction of postoperative mortality after pancreatic surgery might perform near perfect while training the model, but it might not be able to properly predict its risk prospectively. This phenomenon is called overfitting. This occurs when a model is incapable of capturing a relationship between the input variables and the target output values. A way to estimate the extent of overfitting is through repeated cross-validation or preferentially via bootstrap resampling. This is where the entire modelling process is repeated in each bootstrap replicate. Quantifying the degree of overoptimism in the model's performance also enables observation of bias-corrected model performance estimates.[20, 32] Bootstrap resampling with preferentially 200 to 1000 bootstrap replications would give stable and accurate overoptimism-corrected performance estimates and is therefore the gold standard for internal validation.[33, 34] Contrarily, when there is a high error rate in both the training and testing data due to high bias and low variance, the model may be underfit. The balance between underfitting and overfitting is called the bias-variance trade-off.[35] This trade-off shows that when the complexity of a model increases, variance also increases and bias falls.

Infrastructure and transparency

Adaptation of facility infrastructure to enable safe, real-time interaction between the patient file and ML models is both time-consuming and expensive.[36] The implementation of working models in other facilities can be equally difficult since models are often facility specific.[10, 12, 36, 37] It is therefore of paramount importance that ML models are trained and validated within multiple healthcare facilities with adequate sample sizes to ensure generalizability as well as preventing substantial harm to the patient.[38-40] Next to this, the use of a uniform classification system, or international consensus, is a prerequisite for clinical applicability and generalizability.[37]

While abundant literature exists on methodology and the reporting quality of models using conventional statistics (i.e., logistic regression), there is an increasing concern about the transparency of studies using ML models. [17, 41] Suboptimal transparency on model development makes ML models hard to interpret which thus impedes their implementation. It is regarded as the main reason for the limited application of ML models in daily clinical practice.[42, 43] In addition to the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis statement (TRIPOD), a protocol for ML-specific prediction was recently published to optimize transparency of prediction models.[17, 44] Adherence to this statement will improve the interpretability, reproducibility, risk of bias assessment and ultimately, its applicability in the clinical practice.[45] The completeness of this checklist is generally poor as only 38.7% of the 152 articles published using ML models for prediction, adhered to the TRIPOD

items.[16] Model specification alongside model performance, which are both essential in transparency reports, were rarely reported.[16]

Data and interpretability

The large effect of the data quality on the final results is an often-mentioned pitfall in AI-implementation and is called the Garbage-In-Garbage-Out principle.[46] Inaccurate data directly causes unreliable results. The World Health Organization stated that proper data quality is multidimensional and should be accurate, available, complete, and valid.[46-48] This quality should always be accurate for ML purposes. The type of data used to feed the model depends on the moment a prediction has to be made. Xue et al.[49] evaluated the use of pre- and intraoperative in the prediction of postoperative complications. They conclude that having a combination of pre- and intraoperative data had a slightly better performance that only analyzing with one of the two types. Moreover, a combination of structured and unstructured data is advisable. Especially in unstructured data, for example from electronic health records and CT- or MRI-scans, ML models showed their superiority.[50] It is advisable to use a combination of both data types when possible in order to help formulate a risk score based on as many contributing static and modifiable risk factors allowing for early intervention. When the risk scores are created, it is essential that several experts in the field of surgery, nursery, or data science discuss the favorable cut-off value together. The previously discussed ICE, LIME, and SHAP can contribute to the interpretability of the model's output.

Acceptance and ethical considerations

AI shows a great potential for healthcare professionals to support or augment clinical decision-making. Multiple studies suggest that this must play a critical role in future surgical decision-making.[51] However, even when a model has been tested and validated correctly, reticence by both clinicians and the patients can greatly influence its successful implementation. Patients correlate AI with science fiction which creates a fear of machines and computers taking over the making of decisions. The majority of them thus prefer healthcare provider's supervision over AI.[52] The fear of clinicians replaced by AI leads to mistrust by patients. Equally, in a different study on acceptance of AI amongst clinicians, only 25% of radiologists had confidence in the results of the diagnostic from AI algorithms.[53]

Therefore, medical tools using AI should be used in an assistive manner as opposed to being ultimately responsible for the main decision-making.[54] The 'doctor in the loop' is responsible and this responsibility is classified into the following: accountability, liability, and culpability.[54, 55] To tackle this problem it is important to have proper patient education to reassure that the AI systems are not replacing the decision-making of the professional and are merely acting in a complementary manner to them.

Before the predictions of ML models can be used in daily clinical practice successfully, healthcare providers need to gain trust in these techniques. Surgeons deciding to restore colonic continuity after resection of colon cancer based on an intraoperative ML model predicting a high risk of anastomotic leakage, would want to rely on algorithms that have been validated at their own institute. A prospective simulation study, in which the predictive performance of a model is tested in addition to the regular local care without affecting its course, would enable correct calibration of the ML model. A calibration curve shows whether the predicted chance matches the actual population-based chance of developing a postoperative complication. This approach would allow surgeons to gain trust in the effectiveness and predictive performance of ML models and use them in their clinical practice.[56, 57]

The introduction of ML models led to an unprecedented amount of ethical issues and guidelines regarding these ethical considerations are still sparse.[58] Currently available frameworks for governance were discussed in a recently published review of the literature.[59] This study included 21 guidelines for gold-standard societal values, such as sustainability, freedom, and fairness. Although these guidelines appeared to be insufficient when analyzed separately, it was stated that ideal rules for ethical considerations should harmonize interests, offering benefit to the clinicians, patients and hospital.[59] A governance model for the application of ML models in healthcare based on the abovementioned concept was developed recently.[60] In our opinion it is of utmost importance to adhere to such governance models to guarantee acceptance as well as ethical and legal well-doing.

Correct prediction of postoperative complications using ML has the potential to dramatically improve the outcome of everyday clinical surgical care. Its implications for the patients should be considered before implementing prediction models in clinical practice. For example, predicting anastomotic leakage after colorectal surgery may lead to more stomas or earlier discharge when leakage is not expected to occur. This could also lead to collateral over- and under treatment. This change of paradigm in clinical practice must be accepted by all healthcare providers to ensure full benefit from these techniques. Therefore, it is advisable to obtain solid prospective validation from an external source in different centres with adequate sample sizes. All while adhering to the transparency and ethical guidelines to overcome potential distrust of ML from clinicians and patients.

Declaration of interests

All authors declare no competing interests

References

1. Beaulieu-Jones BK, Yuan W, Brat GA, Beam AL, Weber G, Ruffin M, et al. Machine learning for patient risk stratification: standing on, or looking over, the shoulders of clinicians? NPJ Digit Med. 2021;4(1):62.
2. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56.
3. Beam AL, Kohane IS. Translating Artificial Intelligence Into Clinical Care. Jama. 2016;316(22):2368-9.
4. Wolff J, Pauling J, Keck A, Baumbach J. Success Factors of Artificial Intelligence Implementation in Healthcare. Front Digit Health. 2021;3:594971.
5. Cao Y, Fang X, Ottosson J, Näslund E, Stenberg E. A Comparative Study of Machine Learning Algorithms in Predicting Severe Complications after Bariatric Surgery. Journal of clinical medicine. 2019;8(5).
6. Chen D, Afzal N, Sohn S, Habermann EB, Naessens JM, Larson DW, et al. Postoperative bleeding risk prediction for patients undergoing colorectal surgery. Surgery. 2018;164(6):1209-16.
7. Grass F, Storlie CB, Mathis KL, Bergquist JR, Asai S, Boughey JC, et al. Challenges of Modeling Outcomes for Surgical Infections: A Word of Caution. Surgical infections. 2020.
8. Han IW, Cho K, Ryu Y, Shin SH, Heo JS, Choi DW, et al. Risk prediction platform for pancreatic fistula after pancreatoduodenectomy using artificial intelligence. World journal of gastroenterology. 2020;26(30):4453-64.
9. Merath K, Hyer JM, Mehta R, Farooq A, Bagante F, Sahara K, et al. Use of Machine Learning for Prediction of Patient Risk of Postoperative Complications After Liver, Pancreatic, and Colorectal Surgery. Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract. 2020;24(8):1843-51.
10. Nudel J, Bishara AM, de Geus SWL, Patil P, Srinivasan J, Hess DT, et al. Development and validation of machine learning models to predict gastrointestinal leak and venous thromboembolism after weight loss surgery: an analysis of the MBSAQIP database. Surg Endosc. 2021;35(1):182-91.
11. Shi HY, Lee KT, Lee HH, Ho WH, Sun DP, Wang JJ, et al. Comparison of artificial neural network and logistic regression models for predicting in-hospital mortality after primary liver cancer surgery. PloS one. 2012;7(4):e35781.
12. Weller GB, Lovely J, Larson DW, Earnshaw BA, Huebner M. Leveraging electronic health records for predictive modeling of post-surgical complications. Stat Methods Med Res. 2018;27(11):3271-85.
13. Wise ES, Amateau SK, Ikramuddin S, Leslie DB. Prediction of thirty-day morbidity and mortality after laparoscopic sleeve gastrectomy: data from an artificial neural network. Surg Endosc. 2020;34(8):3590-6.
14. Pera M, Gibert J, Gimeno M, Garsot E, Eizaguirre E, Miró M, et al. Machine Learning Risk Prediction Model of 90-day Mortality after Gastrectomy for Cancer. Ann Surg. 2022.
15. van den Bosch T, Warps AK, de Nerée Tot Babberich MPM, Stamm C, Geerts BF, Vermeulen L, et al. Predictors of 30-Day Mortality Among Dutch Patients Undergoing Colorectal Cancer Surgery, 2011-2016. JAMA Netw Open. 2021;4(4):e217737.
16. Andaur Navarro CL, Damen JAA, Takada T, Nijman SWJ, Dhiman P, Ma J, et al. Completeness of reporting of clinical prediction models developed using supervised machine learning: a systematic review. BMC Med Res Methodol. 2022;22(1):12.

17. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Bmj. 2015;350:g7594.
18. Sarker IH. Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective.
19. Stam WT, Goedknegt LK, Ingwersen EW, Schoonmade LJ, Bruns ERJ, Daams F. The prediction of surgical complications using artificial intelligence in patients undergoing major abdominal surgery: A systematic review. Surgery. 2021.
20. Lever J, Krzywinski M, Altman N. Model selection and overfitting. Nature Methods. 2016;13(9):703-4.
21. A J. Artificial Intellgence and Black-box Medical Decisions: Accuracy versus Explainability. The Hastings Center Report. 2019;49(1):15 - 21.
22. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-44.
23. Li X, Yang L, Yuan Z, Lou J, Fan Y, Shi A, et al. Multi-institutional development and external validation of machine learning-based models to predict relapse risk of pancreatic ductal adenocarcinoma after radical resection. J Transl Med. 2021;19(1):281.
24. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17(1):195.
25. Ahuja AS. The impact of artificial intelligence in medicine on the future role of the physician. PeerJ. 2019;7:e7702.
26. Mazaki J, Katsumata K, Ohno Y, Udo R, Tago T, Kasahara K, et al. A Novel Predictive Model for Anastomotic Leakage in Colorectal Cancer Using Auto-artificial Intelligence. Anticancer Res. 2021;41(11):5821-5.
27. Habli I, Lawton T, Porter Z. Artificial intelligence in health care: accountability and safety. Bull World Health Organ. 2020;98(4):251-6.
28. Durán JM, Jongsma KR. Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. J Med Ethics. 2021.
29. Anis Mihdi Afnan ML, Y; Conitzer, V; Rudin, C; Mishra, A; Savulescu, J; Afnan, M. Interpretable, not black-box, artificial intelligence should be used for embryo selection. Human Reproduction Open. 2021;2021(4).
30. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Health. 2021;3(11):e745-e50.
31. Lundberg CL, S-I. A unified approach to interpreting model predictions. arXiv. 2017;1.
32. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. J Clin Epidemiol. 2003;56(5):441-7.
33. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating: Springer International Publishing; 2019.
34. Harrell FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis: Springer International Publishing; 2016.
35. Belkin M, Hsu D, Ma S, Mandal S. Reconciling modern machine-learning practice and the classical bias–variance trade-off. Proceedings of the National Academy of Sciences. 2019;116(32):15849-54.
36. Cao Y, Montgomery S, Ottosson J, Näslund E, Stenberg E. Deep Learning Neural Networks to Predict Serious Complications After Bariatric Surgery: Analysis of Scandinavian Obesity Surgery Registry Data. JMIR Med Inform. 2020;8(5):e15992.

37. Grass F, Storlie CB, Mathis KL, Bergquist JR, Asai S, Boughey JC, et al. Challenges of Modeling Outcomes for Surgical Infections: A Word of Caution. Surg Infect (Larchmt). 2021;22(5):523-31.
38. Azimi K, Honaker MD, Chalil Madathil S, Khasawneh MT. Post-Operative Infection Prediction and Risk Factor Analysis in Colorectal Surgery Using Data Mining Techniques: A Pilot Study. Surg Infect (Larchmt). 2020;21(9):784-92.
39. Gulati G, Upshaw J, Wessler BS, Brazil RJ, Nelson J, van Klaveren D, et al. Generalizability of Cardiovascular Disease Clinical Prediction Models: 158 Independent External Validations of 104 Unique Models. Circ Cardiovasc Qual Outcomes. 2022:101161circoutcomes121008487.
40. Riley RD, Ensor J, Snell KIE, Harrell FE, Jr., Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. Bmj. 2020;368:m441.
41. Zamanipoor Najafabadi AH, Ramspek CL, Dekker FW, Heus P, Hooft L, Moons KGM, et al. TRIPOD statement: a preliminary pre-post analysis of reporting and methods of prediction models. BMJ Open. 2020;10(9):e041537.
42. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol. 2019;110:12-22.
43. Gravesteijn BY, Nieboer D, Ercole A, Lingsma HF, Nelson D, van Calster B, et al. Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury. J Clin Epidemiol. 2020;122:95-107.
44. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open. 2021;11(7):e048008.
45. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. Lancet. 2014;383(9913):267-76.
46. Kilkenny MF, Robinson KM. Data quality: "Garbage in - garbage out". Health Inf Manag. 2018;47(3):103-5.
47. Poksinska BDJA, M. The state of ISO 9000 certification: a study of Swedisch organizations. The TQM Magazine. 2002;14:297 - 306.
48. World Health O. Guide to the health facility data quality report card. 2018. p. 1.
49. Xue B, Li D, Lu C, King CR, Wildes T, Avidan MS, et al. Use of Machine Learning to Develop and Evaluate Models Using Preoperative and Intraoperative Data to Identify Risks of Postoperative Complications. JAMA Netw Open. 2021;4(3):e212240.
50. Zhang D, Yin C, Zeng J, Yuan X, Zhang P. Combining structured and unstructured data for predictive models: a deep learning approach. BMC Med Inform Decis Mak. 2020;20(1):280.
51. Loftus TJ, Tighe PJ, Filiberto AC, Efron PA, Brakenridge SC, Mohr AM, et al. Artificial Intelligence and Surgical Decision-making. JAMA Surg. 2020;155(2):148-58.
52. Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. Lancet Digit Health. 2021;3(9):e599-e611.
53. Jungmann F, Jorg T, Hahn F, Pinto Dos Santos D, Jungmann SM, Düber C, et al. Attitudes Toward Artificial Intelligence Among Radiologists, IT Specialists, and Industry. Acad Radiol. 2021;28(6):834-40.
54. O'Sullivan S, Nevejans N, Allen C, Blyth A, Leonard S, Pagallo U, et al. Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and

autonomous robotic surgery. The International Journal of Medical Robotics and Computer Assisted Surgery. 2019;15(1):e1968.
55. Ho CWL, Soon D, Caals K, Kapur J. Governance of automated image analysis and artificial intelligence analytics in healthcare. Clinical Radiology. 2019;74(5):329-37.
56. Mu W, Liu C, Gao F, Qi Y, Lu H, Liu Z, et al. Prediction of clinically relevant Pancreatico-enteric Anastomotic Fistulas after Pancreatoduodenectomy using deep learning of Preoperative Computed Tomography. Theranostics. 2020;10(21):9779-88.
57. Adams K, Papagrigoriadis S. Creation of an effective colorectal anastomotic leak early detection tool using an artificial neural network. Int J Colorectal Dis. 2014;29(4):437-43.
58. Cave S, Nyrup R, Vold K, Weller A. Motivations and risks of machine ethics. Proceedings of the IEEE. 2018;107(3):562-74.
59. de Almeida PGR, dos Santos CD, Farias JS. Artificial Intelligence Regulation: a framework for governance. Ethics and Information Technology. 2021;23(3):505-25.
60. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. Journal of the American Medical Informatics Association. 2019;27(3):491-7.

7

Chapter 7

Radiomics for the prediction of a postoperative pancreatic fistula following a pancreatoduodenectomy: a systematic review and radiomic score quality assessment

Erik W. Ingwersen
Pieter M.W. Rijssenbeek
Henk A. Marquering
Geert Kazemier
Freek Daams

Pancreatology 2024 Mar 24(2):1003-1011

Abstract

Background: Postoperative pancreatic fistula (POPF) is a severe complication following a pancreatoduodenectomy**.** An accurate prediction of POPF could assist the surgeon in offering tailor-made treatment decisions. The use of radiomic features has been introduced to predict POPF. A systematic review was conducted to evaluate the performance of models predicting POPF using radiomic features and to systematically evaluate the methodological quality.

Methods: Studies with patients undergoing a pancreatoduodenectomy and radiomics analysis on computed tomography or magnetic resonance imaging were included. Methodological quality was assessed using the Radiomics Quality Score (RQS) and Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement.

Results: Seven studies were included in this systematic review, comprising 1 300 patients, of whom 364 patients (28%) developed POPF. The area under the curve (AUC) of the included studies ranged from 0.76 to 0.95. Only one study externally validated the model, showing an AUC of 0.89 on this dataset. Overall adherence to the RQS (31%) and TRIPOD guidelines (54%) was poor.

Conclusion: This systematic review showed that high predictive power was reported of studies using radiomic features to predict POPF. However, the quality of most studies was poor. Future studies need to standardize the methodology.

Introduction

Postoperative pancreatic fistula (POPF) is a common complication following a pancreatoduodenectomy, impairing short-term and long-term outcomes and cancer recurrence rates[1,2]. Moreover, POPF is a substantial burden for the health care resource utilization and health expenditure[3].

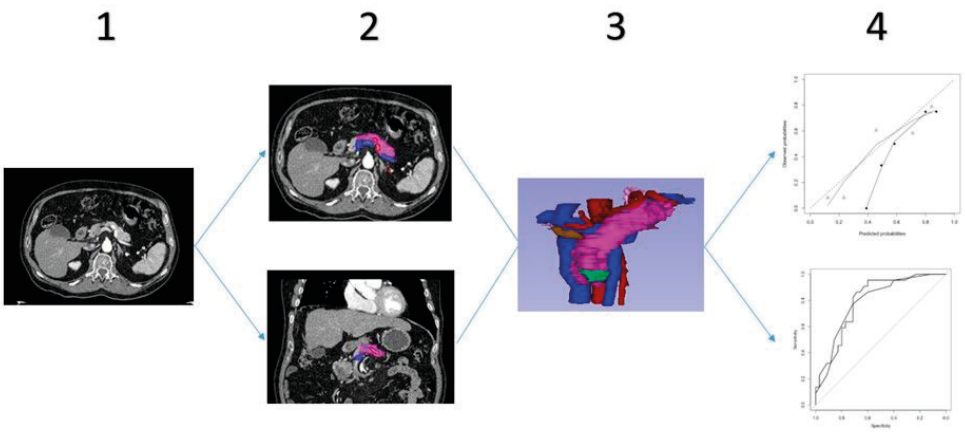
An accurate prediction model for POPF would enhance personalized treatment decisions based on individual patient risk. Several prediction models for POPF have been proposed in the past. However, most of these models rely on subjective measurements (e.g. intra-operative assessment texture of the pancreas) and are not able to provide a prediction prior to surgery.

Features extracted from the preoperative computed tomography (CT) and magnetic resonance imaging (MRI) scan can be used to facilitate preoperative prediction of POPF development. These features can be semantic or quantitative[4]. Semantic features are radiologist-defined features consisting of common and visible characteristics of a tumor or disease, such as pancreatic duct size and pancreatic volume[5, 6]. Use of these semantic features, however, is limited due to suboptimal inter- and intra-observer reliability (k-0.50–0.70)[7]. Quantitative features, on the other hand, can be computer-derived or mathematically extracted[8]. Radiomic features allow for large-scale collection of these quantitative and complex visual patterns from CT-scans and MRI[9]. Radiomic features can uncover valuable parameters for preoperative POPF prediction, some invisible to the human eye. The use of radiomic features has emerged to be included in tools for diagnosis, prognosis, patient's outcome in malignancies, such as liver lung, breast, and pancreatic cancer[10-13].

A radiomic feature workflow consists of four steps: 1) image acquisition, 2) image segmentation, 3) feature extraction, and 4) analysis[14]. (Figure 1) The first step is image acquisition. This includes the collection of data of a certain imaging modality (e.g. MRI/CT/ultrasound). Once a dataset is collected, the primary outcome of interest and the region of interest (ROI) or volume of interest (VOI) must be defined[16]. In image segmentation, the VOI must be segmented. Both manual and automated segmentations are possible. After segmentation, radiomic features are extracted from the VOI. This image features characterize the VOI. Specialized software is used to derive these features, generating more than thousand variables within a single VOI. These features are quantitative and are used to perform analysis on the VOI. The relationship between radiomic features and the described outcome are investigated. An important process in this step is feature selection: a technique in which the most valuable predictors are identified. Also, dimensionality reduction is used to reduce the number of features used as predictors in order to reduce the chance of overfitting[15]. Finally, the selected radiomic features can be analyzed to predict the outcome of interest. Machine learning models are the most common approach to analyse the radiomic features[16].

Recently, radiomic features are also used to predict POPF[17]. Despite the theoretical advantages of a radiomic features based prediction model, its use has been limited in context of research, and it is unclear what the added value of such prediction models may be in daily clinical practice. Therefore, a systematic review was conducted to evaluate the predictive performance of radiomic features in predicting POPF in patients undergoing a pancreatoduodenectomy. Furthermore, it systematically evaluate the methodology quality, reporting transparency, and risk of bias of the included studies.

Figure 1, visualization of the four steps of the radiomics workflow



The radiomics *workflow consists out of steps. 1) Image acquisition, 2) Image segmentation, 3) Feature extraction, and 4) analysis.*
VOI; volume *of interest.*

Methods

A comprehensive systematic review of the literature was performed according to the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) statement[18]. The literature review captured studies from PubMed, Embase, and Web of Science Core Collection from inception to January 29, 2023. The full search strategy, including controlled terms (e.g. Mesh, Emtree) as well as free-text words. The following terms were used: "pancreatoduodenectomy", "pancreaticoduodenectomy", "radiomic features", "computed tomography", "magnetic resonance imaging", and "pancreatic fistula" as well as synonyms and related words. There were no restrictions regarding publication date or language. The references of relevant articles were further crosschecked to identify additional articles. The final search strategy is shown in the Supplementary appendix 1 to 3.

Inclusion and exclusion criteria

Inclusion criteria included the following: studies with (1) patients undergoing a pancreatoduodenectomy with the construction of a pancreatojejunostomy and (2) radiomics analysis on CT- or MRI-scan. Exclusion criteria were patients undergoing a distal pancreatectomy or any pancreatectomy without pancreatojejunostomy, non-English language article, and systematic reviews and meta-analyses.

Screening

After elimination of duplicate articles, two reviewers (EI and PR) independently reviewed all titles and abstracts of the identified articles based on the pre-defined eligibility criteria. Articles meeting the eligibility criteria were subsequently assessed in full text. Discrepancies between both reviewers were assessed by a third author (FD).

Data extraction

The following data were collected from the included studies: first author, study design, number of included patients, imaging modality, information about the VOI, the primary outcome of interest, number of extracted features, preprocessing steps (e.g. data split, sampling techniques, feature selection), type of machine learning model used for analysis, methods of validation, type of features extracted, and most important feature to predict POPF.

Quality assessment

The Radiomic Quality Score (RQS) was independently assessed by two reviewers (EI and PR). The RQS is a relevant and convenient quantitative scoring system to critically assess methodology quality with the aim to reward and penalize the methodology, analysis, and reporting of a study, to encourage best scientific practice[19]. The checklist comprises 16 components, with a score ranging from 0 to 36. Since the included studies developed prediction models, adherence to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement was assessed[20]. Finally, the Risk Of Bias In Non-Randomized Studies – of interventions (ROBINS-I) was used to identify potential sources of bias[21]. Any discrepancies in the RQS, TRIPOD, or ROBINS-I between both reviewers were resolved by consensus.

Results

The PRISMA flow chart of study inclusion is summarized in Figure 2. Fifteen studies were assessed for eligibility and screened on full text. One article involved the prediction of POPF with radiomic features, but the full text was in Chinese.[22] The authors were contacted but did not respond to our request to obtain the full text in English. Therefore, this study was excluded from the systematic review. Finally, seven studies were included in this systematic review[17, 23-28]. A total number of 1 300 patients were included in the seven studies of whom 364 patients (28%) developed POPF. The median number of patients used to develop the prediction models per study was 117, ranging from 62 to 513 patients. The studies were published between 2018 and 2022. Study characteristics, demographics, and predictive performances of the included studies are summarized in Table 1. All seven studies developed a prediction model on a retrospective cohort. The

study of Mu et al. was the only to externally validate the prediction model and did this validation on a prospective cohort. Primary outcome in all studies was clinically relevant POPF. Most of the studies defined this as POPF grade B and C, according to the ISGPS criteria[29].

Figure 2, flowchart of study inclusion following the PRISMA statement

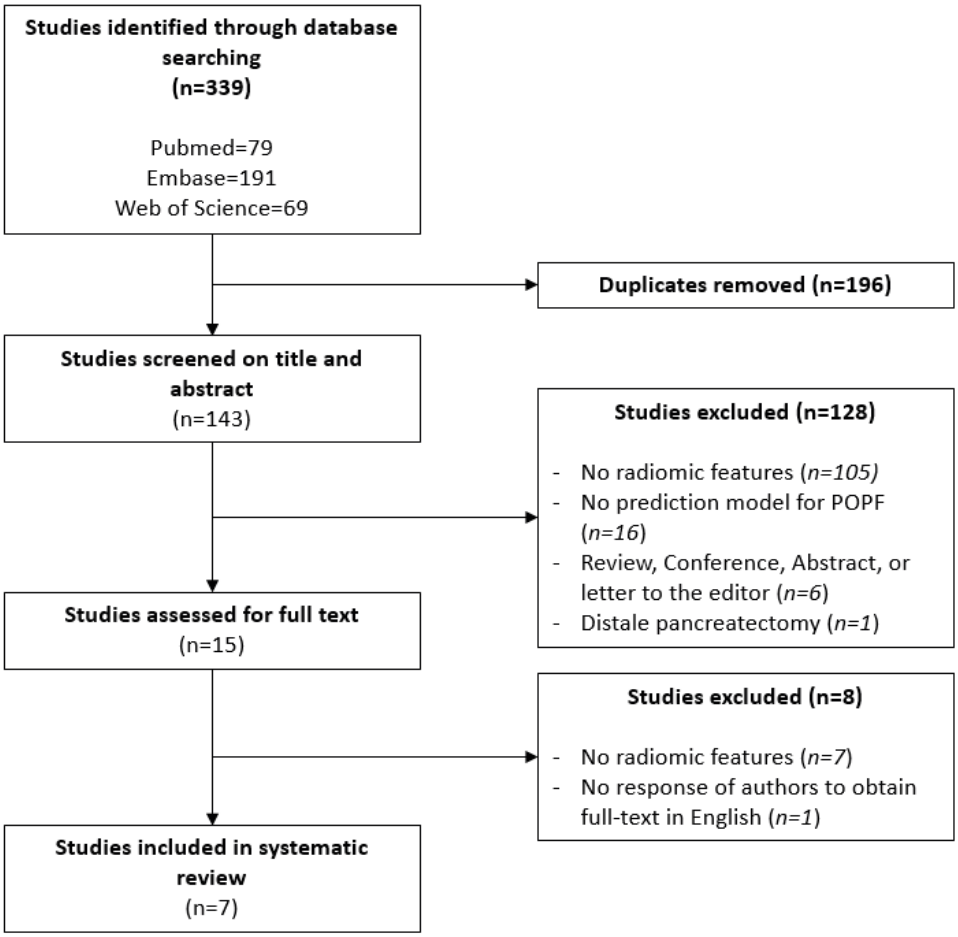


Table 1, study characteristics, demographics, and predictive performances.

First author (year)	Study design	Study population	Primary outcome	Compared with reference test	Predictive performance in AUC (95% CI)
Bhasker (2022) (23)	Retrospective, monocenter	148 patients between 2011 and 2019 48 developing POPF	Prediction of POPF grade B and C	NP	0.81 (0.73 – 0.87)
Capretti (2022) (24)	Retrospective, monocenter	100 patients between 2011 and 2019 20 developing POPF	Prediction of POPF grade B and C	NP	0.81 (0.50 – 1.00)
Kambakamba (2020) (17)	Retrospective, monocenter	110 patients between 2008 to 2018 55 developing POPF	Prediction of POPF grade B and C	FRS and a-FRS (30, 31)	0.95 (NR)
Lin (2021) (25)	Retrospective, monocenter	250 patients between 2013 and 2019 53 developing POPF	Prediction of POPF grade B and C	FRS (30)	0.79 (0.68 – 0.91)
Mu (2020) (26)	Retrospective for model development and prospective for model validation, multicenter	Development: 513 patients between 2006 and 2018 80 developing POPF Validation cohort: 70 patients between 2018 and 2019 15 developing POPF	Prediction of POPF grade B and C	FRS (30)	Internal validation set 0.81 (0.72 – 0.89) External test set 0.89 (0.76 – 0.96)
Skawran (2021) (27)	Retrospective, monocenter	62 patients between 2008 and 2018 45 developing POPF	Prediction of POPF grade B and C	NP	0.82 (0.74 – 0.89)
Zhang (2018) (28)	Retrospective, monocenter	117 patients, time period of inclusion NR 48 developing POPF	Prediction of POPF grade B and C	FRS (30)	0.76 (0.61 – 0.92)

The AUC *of the best performing prediction model in the validation or test set is reported.*
AUC; area *under the curve. NP; not performed. FRS; Fistula Risk Score, a-FRS: alternative Fistula Risk Score.*

Radiomics workflow

Table 2 shows information about the four steps in the radiomics workflow and an assessment following the RQS. Six studies used the preoperative CT-scan and one study used MRI as imaging modality. Four studies segmented in the portal venous phase, two in the non-contrast-enhanced phase, and one did not report on the used scan phase. The VOI was the pancreatic body and tail in five studies, in two studies the VOI comprised the entire pancreas, including the pathology. In most studies, scans were manually segmented by a radiologist. Lin et al. used a semiautomatic segmentation, which was then checked by a radiologist. The VOI was the pancreatic body and tail in most of the

studies. All studies used feature selection to select predictors with the highest correlation to POPF, with the least absolute shrinkage and selection operator (LASSO) regression being used most. All studies used at least one machine learning model to predict POPF and three studies used multiple. Table 3 provides a detailed overview of the radiomic features extracted in the studies included in our analysis. Among these studies, three explicitly disclosed the program utilized for radiomic feature extraction, with all three employing PyRadiomics—a widely recognized software for this purpose. The extracted features span across eight distinct categories. The employed subcategories encompass first-order statistics (Histogram), shape based features, Gray Level Co-occurrence Matrix (GLOM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), and Gray Level Dependence Matrix (GLDM). These subcategories collectively capture information pertaining to the distribution of voxel intensities within the image, the volume, size, density, shape of the Volume of Interest (VOI), and the quantity and distribution of connected voxels that share the same gray level intensity. Interestingly, the consensus among six studies, highlighting features related to pancreatic texture as consistently emerging as the most predictive for POPF. Additionally, Capretti et al. identified the feature associated with fat/muscle ratio as the most significant predictor in their study.

Table 2, information about radiomic workflow and calculated RQS.

First author (year)	Data acquisition	Medical imaging	Feature extraction	Modelling	RQS
Bhasker (2022) (23)	A portal venous phase CT-scan.	Manual segmentation performed by radiologist and supervised by a second radiologist Slice thickness: 1.0mm. VOI: PP including pathology, PV, SA, BD, PD	Features extracted: 1706 Feature selection: selected four features with highest variance Dimensionality reduction: NP No sampling technique	ML models used: LASSO, elastic-net Models validated using 10-fold cross validation	11 (31%)
Capretti (2022) (24)	A portal venous phase CT-scan	Manual segmentation performed by two radiologists Slice thickness: NR VOI: pancreatic body and tail, skeletal muscle tissue	Features extracted: NR Feature selection: LASSO Dimensionality reduction: NP Sampling technique: SMOTE	ML models used: L1 regularized LR, RF Models validated using 10-fold cross validation	7 (19%)

Kambakamba (2020) (17)	A non-contrasted-enhanced CT scan	Manual segmentation performed by two radiologists Slice thickness: 3.75mm VOI: PP including pathology	Features extracted: 278 features Feature selection: features with low interclass correlation were discarded Dimensionality reduction: PCA No sampling technique	ML models used: RF, DT, k-nearest neighbor, SVM ANN Models validated using 10-fold cross validation	11 (31%)
Lin (2021) (25)	NR	Semiautomatic segmentation checked by a radiologist Slice thickness ≤1.25mm VOI: pancreatic body and tail	Features extracted: 102 features Feature selection: Pearson Correlation Coefficient Dimensionality reduction: NP No sampling technique	ML model used: AdaBoost Model validated using 3/10 of dataset	12 (33%)
Mu (2020) (26)	A portal venous CT-scan	Manual segmentation performed by four radiologists Slice thickness≤3.0mm VOI: pancreatic body and tail	Features extracted: NR Feature selection: SMOTE Dimensionality reduction: NP Width/height shirt, horizontal/vertical flip to augment data	ML model used: ANN Model validated using 5-fold cross validation and externally validated in prospective cohort	16 (44%)
Skawran (2021) (27)	T1 and T2 no contrast enhanced MRI-scan	Manually segmentation Slice thickness Slice thickness: 3.0mm VOI: pancreatic body and tail in a single slice	Features extracted: 45 features Feature selection: Boruta feature selection Dimensionality reduction: NP MWMOTE to augment data	ML model used: GBT Models validated using 3/10 of dataset	11 (31%)
Zhang (2018) (28)	A portal venous CT-scan	NR Slice thickness: 1.0mm VOI: pancreatic body and tail	Features extracted: 1,219 features Feature selection: LASSO Dimensionality reduction: NP No sampling technique	ML model used: NR Models validated using 1/3 of dataset	9 (25%)

RQS; Radiomics *Quality Score, CT; computed tomography, VOI: volume of interest, PP; pancreatic parenchyma, PV: portal vein, SA: splenic artery, BD; bile duct, PD; pancreatic duct, NP; not performed, ML; machine learning, LASSO; Least Absolute Shrinkage and Selection Operator, NR: not reported, SMOTE; Synthetic Minority Oversampling Technique, LR; logistic regression, RF; random forest, PCA; principal component analysis, DT; decision tree, SVM; support vector machine, ANN: artificial neural network, GBT; gradient boosted tree*

Table 3, type of features used as predictors for POPF

Study	Program or package used to extract features	Type of features extracted as predictors	Most important feature to predict POPF
Bhasker (2022)[23]	NR	1. Morphological 2. Texture 3. Combined segment 4. Intensity volume	Radiomic texture feature associated with pancreatic texture
Capretti (2022)[24]	PyRadiomics	1. Fat / muscle ratio 2. Shape of pancreas 3. Position of PD	Radiomic feature associated with fat/ muscle ratio
Kambakamba (2020)[17]	NR	1. GLCM 2. GLRM 3. Histogram 4. Absolute gradient 5. Autoregressive model 6. Wavelet transform	Radiomic texture feature associated with pancreatic texture
Lin (2021)[25]	NR	1. Shape 2. GLRM 3. GLDM	Radiomic texture feature associated with pancreatic texture
Mu (2020)[26]	NR	Features correlated with: 1. Less fibrosis 2. Acinar atrophy 3. Soft pancreas	Deep learning feature associated with pancreatic texture
Skawran (2021)[27]	PyRadiomics	1. GLCM 2. GLRLM 3. GLDM 4. GLSZM 5. Histogram	Possibly radiomic feature that reflects pancreatic texture
Zhang (2018)[28]	PyRadiomics	1. GLCM 2. GLRM 3. GLDM 4. GLZSM	Radiomic features associated with pancreatic texture

POPF; postoperative *pancreatic fistula, NR; not reported, PD; pancreatic duct, GLCM; Gray Level co-Occurrence matrix. GLRM; Gray Level Run Length Matrix, GLDM; Gray Level Dependence Matrix, GLSZM; Gray Level Size Zone Matrix.*

Prediction performances

All studies measured an area under the curve (AUC) to evaluate the discriminative performance of their prediction model. Three studies only tested their model on a validation cohort using k-fold cross validation and reported these results. Three studies created a test set of 30-33% of the data and evaluated model performances on this test set. One study performed internal validation and subsequently externally validated the model.

The lowest AUC of the prediction model was determined in the study of Zhang et al. with a value of 0.76 (95% CI 0.61 – 0.92). The highest AUC was found in the study of Kambakamba, with a value of 0.95 (95% CI not reported). The AUC of the model developed by Mu et al. was 0.89 (95% CI 0.76 – 0.96) on the external validation set, which was higher than the performance on the internal dataset (AUC: 0.81 (95% CI 0.72 – 0.89)). The median AUC was 0.81 in the studies using CT as imaging modality and 0.82 in the study using MRI.

Four studies compared the results of their prediction model to the Fistula Risk Score (FRS), a previously validated clinical risk score for predicting POPF[30]. Kambakamba et al. also compared the developed prediction model to the alternative Fistula Risk Score (a-FRS), a more recently developed and validated clinical risk score[31]. In all studies, the radiomics-based prediction model outperformed the FRS and, if applicable, the a-FRS. Three studies compared the AUCs of their model to the FRS using the Delong test[25, 26, 28]. They all showed a significant difference in favor of the radiomics model. Mu et al. also compared the AUC on the external test cohort with the FRS and showed that their model significantly outperformed the FRS when tested on this cohort. Bhasker et al. and Lin et al. developed a combination model, combining radiomic features wit clinical variables (e.g. BMI, age, sex). Both models showed better predictive performances if clinical variables were added to the radiomics model, with Bhasker et al. showing a AUC of 0.81 (95% CI 0.73 – 0.88) and Lin et al. an AUC of 0.87 (95% CI 0.78 – 0.96).

Quality assessment and study validity

The median RQS of the seven studies was 11 (31%) and the range was 7 – 16. The most common limitation of the studies was the lack of a phantom study on all scanners, calibration statistics, registration of a prospective study, validation, cost effectiveness analysis, and publicly available models. The quality adherence was highest for feature reduction or adjustment for multiple testing, discrimination statistics, and potential clinical utility. The study of Mu et al. scored highest with 16 points, primarily because a validation cohort was incorporated and compared to a reference test for predicting POPF. Supplementary appendix 4 shows the adherence of the included studies to the TRIPOD guidelines. The overall adherence rate was 54% (107/200 points). None of the studies scored points in Title (item 1), Abstract (item 2), blind assessment of the outcome (item 6b), sample size (item 8), missing data (item 9) and characteristics of the participants (item 13b). Overall adherence was highest in the introduction (100%) and discussion (86%) and lowest in the title and abstract (0%) and the methods (33%) section. In the Supplementary Appendix 5, the risk of bias was evaluated using the ROBINS-I tool. Overall risk of bias was low.

Discussion

This systematic review shows that radiomic features can be used to accurately predict POPF after pancreatoduodenectomy prior to surgery, with AUCs exceeding 0.80. Although these results were promising, most studies were limited by single-center study design, small study populations, and a lack of external validation. Furthermore, the overall adherence to the TRIPOD guidelines (54%) and the RQS (31%) was low. Only one study externally validated the developed model, but failed to report eligibility criteria for participants, a clear definition of the outcome of interest, and a calibration curve to estimate the absolute risk of POPF in the total population. Moreover, the external validation cohort consisted of 70 patients and only 15 events. An external validation with a larger sample size is required to validate the clinical application of the model. Despite the fact that both CT and MRI were used as imaging modality to segment the VOI, it is currently unclear which is preferable, given that only one study used MRI.

Well-known risk factors for POPF are, among others, a soft pancreatic texture and a small pancreatic duct diameter[30-32]. In the included studies, the radiomic features concerning pancreatic texture and pancreatic duct size were clearly associated with POPF[17, 23, 25, 26, 28]. As shown in Table 2, consensus emerges among all studies, highlighting pancreatic texture as the foremost radiomic feature category associated with POPF. This consistent and recurrent observation underscores the significant correlation between radiomic features and clinical risk factors, particularly in the context of pancreatic texture. The collective emphasis on this specific category across various studies adds robustness to the evidence supporting the relevance of radiomics in predicting POPF. Bhasker et al. also showed that these radiomic features were comparable to the intraoperative assessment of these risk factors.[23] Radiomics were able to reliably and objectively assess texture, which are not necessarily discernable to the human eye.[33] This allows for a preoperative prediction to potentially improve patient prognosis, assisting surgeons in making tailored treatment decisions and counseling patients. Interestingly, discriminative performances of the prediction models improved after combining the extracted radiomic features with clinical variables[23, 25]. Both studies included BMI and sex as clinical variables in the combination model, which are also known risk factors for POPF[30-32].

Most included studies compared the developed radiomics model to previous validated clinical risk score for POPF, the FRS or the a-FRS[17, 25, 26, 28]. Although most of these studies showed a significant difference in AUC in favor of the developed radiomics model, little value could be derived from this finding. Most of the studies did not perform external validation, thus comparing their model to the FRS or a-FRS using the same database on

which the model was developed. This will be in favor of the radiomics models, especially because they use machine learning models, which are prone for overfitting the data[34].

To date, several systematic reviews have summarized the use of radiomics in pancreatic diseases, including staging pancreatic cancer and neuro-endocrine tumours, stratification of cystic lesions, and prediction of the severity of acute pancreatitis [35-39]. Although none of these reviews summarized the ability of radiomics in predicting POPF, they all also concluded that prior to clinical application, methodological implementation needs to be optimized and further validations are warranted. To take radiomics further and to ascertain compliance with good practice, future studies need to standardize methodology, externally validate the prediction model in a multicenter design, and make the model publicly available. Adherence to quality assessment tools, such as the TRIPOD guidelines and RQS, are also critical for approval, and journals should consider demanding these guidelines for radiomics studies.

An important technical aspect must be overcome to consider radiomics ready for use in daily clinical practice. Most of the studies manually segmented the VOI and an auto-segmentation, in which the pancreatic parenchyma is automatically segmented, is pivotal prior to implementation of radiomics into daily clinical practice. This would drastically reduce the time-consuming workflow and provide a more consistent delineation of the VOI.

This review was strengthened by providing an overview over the current situation regarding the prediction of POFP using radiomic features. It also assessed the quality of the included studies using scores or guidelines to objectively appraise methodology and reporting of the radiomics models. Some limitations of the present study should also be addressed. First, we did not perform a meta-analysis to further analyze the use of radiomics in predicting POPF. Although data was intended for meta-analysis, it was not admissible to do so due to a high heterogeneity within the included studies. Second, the TRIPOD guidelines were used to assess quality of the included studies. The TRIPOD guidelines were developed to assess quality of multivariable prediction models and they are not perfectly suitable to assess radiomics studies. However, there is still a clear overlap. The TRIPOD-artificial intelligence tool is currently being developed, and will hopefully be a checklist to specifically assess quality of studies using artificial intelligence[40].

Conclusion

This systematic review showed that radiomic features can be used to accurately predict POPF after pancreatoduodenectomy before surgery, with AUCs exceeding 0.80. Although these results are promising, evidence supporting clinical application was poor, external validation was lacking, and overall adherence to the TRIPOD guidelines and the RQS was low.

References

1. Grego A, Friziero A, Serafini S, Belluzzi A, Moletta L, Saadeh LM, et al. Does Pancreatic Fistula Affect Long-Term Survival after Resection for Pancreatic Cancer? A Systematic Review and Meta-Analysis. Cancers (Basel). 2021;13(22).
2. Hackert T, Werner J, Büchler MW. Postoperative pancreatic fistula. Surgeon. 2011;9(4):211-7.
3. Wang J, Ma R, Churilov L, Eleftheriou P, Nikfarjam M, Christophi C, et al. The cost of perioperative complications following pancreaticoduodenectomy: A systematic review. Pancreatology. 2018;18(2):208-20.
4. Wu G, Jochems A, Refaee T, Ibrahim A, Yan C, Sanduleanu S, et al. Structural and functional radiomics for lung cancer. Eur J Nucl Med Mol Imaging. 2021;48(12):3961-74.
5. Maqueda González R, Di Martino M, Galán González I, Rodríguez Carnero P, Martín-Pérez E. Development of a prediction model of pancreatic fistula after duodenopancreatectomy and soft pancreas by assessing the preoperative image. Langenbecks Arch Surg. 2022;407(6):2363-72.
6. Paul R, Schabath M, Balagurunathan Y, Liu Y, Li Q, Gillies R, et al. Explaining Deep Features Using Radiologist-Defined Semantic Features and Traditional Quantitative Features. Tomography. 2019;5(1):192-200.
7. Fowler KJ, Tang A, Santillan C, Bhargavan-Chatfield M, Heiken J, Jha RC, et al. Interreader Reliability of LI-RADS Version 2014 Algorithm and Imaging Features for Diagnosis of Hepatocellular Carcinoma: A Large International Multireader Study. Radiology. 2018;286(1):173-85.
8. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2016;278(2):563-77.
9. Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, et al. Introduction to Radiomics. J Nucl Med. 2020;61(4):488-95.
10. Harding-Theobald E, Louissaint J, Maraj B, Cuaresma E, Townsend W, Mendiratta-Lala M, et al. Systematic review: radiomics for the diagnosis and prognosis of hepatocellular carcinoma. Aliment Pharmacol Ther. 2021;54(7):890-901.
11. Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell Lung Cancer. Radiology. 2016;281(3):947-57.
12. Li H, Zhu Y, Burnside ES, Drukker K, Hoadley KA, Fan C, et al. MR Imaging Radiomics Signatures for Predicting the Risk of Breast Cancer Recurrence as Given by Research Versions of MammaPrint, Oncotype DX, and PAM50 Gene Assays. Radiology. 2016;281(2):382-91.
13. Toyama Y, Hotta M, Motoi F, Takanami K, Minamimoto R, Takase K. Prognostic value of FDG-PET radiomics with machine learning in pancreatic cancer. Sci Rep. 2020;10(1):17024.
14. Avery E, Sanelli PC, Aboian M, Payabvash S. Radiomics: A Primer on Processing Workflow and Analysis. Seminars in Ultrasound, CT and MRI. 2022;43(2):142-6.
15. Clarke R, Ressom HW, Wang A, Xuan J, Liu MC, Gehan EA, et al. The properties of high-dimensional data spaces: implications for exploring gene and protein expression data. Nat Rev Cancer. 2008;8(1):37-49.
16. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts H. Machine Learning methods for Quantitative Radiomic Biomarkers. Sci Rep. 2015;5:13087.

17. Kambakamba P, Mannil M, Herrera PE, Müller PC, Kuemmerli C, Linecker M, et al. The potential of machine learning to predict postoperative pancreatic fistula based on preoperative, non-contrast-enhanced CT: A proof-of-principle study. Surgery. 2020;167(2):448-54.
18. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Systematic Reviews. 2021;10(1):89.
19. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14(12):749-62.
20. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Medicine. 2015;13(1):1.
21. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919.
22. Hu X, Ge Y, Wei W. Radiomics model based on enhanced CT for predicting pancreatic fistula after pancreaticoduodenectomy. Chinese Journal of Medical Imaging Technology. 2021;37(10):1496-500.
23. Bhasker N, Kolbinger FR, Skorobohach N, Zwanenburg A, Löck S, Weitz J, et al. Prediction of clinically relevant postoperative pancreatic fistula using radiomic features and preoperative data. medRxiv. 2022:2022.10.22.22281403.
24. Capretti G, Bonifacio C, De Palma C, Nebbia M, Giannitto C, Cancian P, et al. A machine learning risk model based on preoperative computed tomography scan to predict postoperative outcomes after pancreatoduodenectomy. Updates Surg. 2022;74(1):235-43.
25. Lin Z, Tang B, Cai J, Wang X, Li C, Tian X, et al. Preoperative prediction of clinically relevant postoperative pancreatic fistula after pancreaticoduodenectomy. Eur J Radiol. 2021;139:109693.
26. Mu W, Liu C, Gao F, Qi Y, Lu H, Liu Z, et al. Prediction of clinically relevant Pancreatico-enteric Anastomotic Fistulas after Pancreatoduodenectomy using deep learning of Preoperative Computed Tomography. Theranostics. 2020;10(21):9779-88.
27. Skawran SM, Kambakamba P, Baessler B, von Spiczak J, Kupka M, Müller PC, et al. Can magnetic resonance imaging radiomics of the pancreas predict postoperative pancreatic fistula? Eur J Radiol. 2021;140:109733.
28. Zhang W, Cai W, He B, Xiang N, Fang C, Jia F. A radiomics-based formula for the preoperative prediction of postoperative pancreatic fistula in patients with pancreaticoduodenectomy. Cancer Manag Res. 2018;10:6469-78.
29. Bassi C, Marchegiani G, Dervenis C, Sarr M, Abu Hilal M, Adham M, et al. The 2016 update of the International Study Group (ISGPS) definition and grading of postoperative pancreatic fistula: 11 Years After. Surgery. 2017;161(3):584-91.
30. Callery MP, Pratt WB, Kent TS, Chaikof EL, Vollmer CM, Jr. A prospectively validated clinical risk score accurately predicts pancreatic fistula after pancreatoduodenectomy. J Am Coll Surg. 2013;216(1):1-14.
31. Mungroop TH, van Rijssen LB, van Klaveren D, Smits FJ, van Woerden V, Linnemann RJ, et al. Alternative Fistula Risk Score for Pancreatoduodenectomy (a-FRS): Design and International External Validation. Annals of Surgery. 2019;269(5).

32. Mungroop TH, Klompmaker S, Wellner UF, Steyerberg EW, Coratti A, D'Hondt M, et al. Updated Alternative Fistula Risk Score (ua-FRS) to Include Minimally Invasive Pancreatoduodenectomy: Pan-European Validation. Annals of Surgery. 2021;273(2).
33. Mannil M, Spiczak J, Manka R, Alkadhi H. Texture Analysis and Machine Learning for Detecting Myocardial Infarction in Noncontrast Low-Dose Computed Tomography: Unveiling the Invisible. Investigative Radiology. 2018;53:1.
34. Subramanian J, Simon R. Overfitting in prediction models – Is it a problem only in high dimensions? Contemporary Clinical Trials. 2013;36(2):636-41.
35. Bezzi C, Mapelli P, Presotto L, Neri I, Scifo P, Savi A, et al. Radiomics in pancreatic neuroendocrine tumors: methodological issues and clinical significance. Eur J Nucl Med Mol Imaging. 2021;48(12):4002-15.
36. Dalal V, Carmicheal J, Dhaliwal A, Jain M, Kaur S, Batra SK. Radiomics in stratification of pancreatic cystic lesions: Machine learning in action. Cancer Lett. 2020;469:228-37.
37. Staal FCR, Aalbersberg EA, van der Velden D, Wilthagen EA, Tesselaar MET, Beets-Tan RGH, et al. GEP-NET radiomics: a systematic review and radiomics quality score assessment. Eur Radiol. 2022;32(10):7278-94.
38. Virarkar M, Wong VK, Morani AC, Tamm EP, Bhosale P. Update on quantitative radiomics of pancreatic tumors. Abdom Radiol (NY). 2022;47(9):3118-60.
39. Zhong J, Hu Y, Xing Y, Ge X, Ding D, Zhang H, et al. A systematic review of radiomics in pancreatitis: applying the evidence level rating tool for promoting clinical transferability. Insights Imaging. 2022;13(1):139.
40. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open. 2021;11(7):e048008.

Supplementary appendix

Supplementary appendix I.

Search strategy: PubMed		
Search	**PubMed query - January 29, 2023**	**Results**
#7	#1 AND #2 AND #3 AND #6	76
#6	#4 OR #5	1 450 551
#5	"Tomography, X-Ray Computed"[Mesh] OR "Multidetector Computed Tomography"[Mesh] OR "Four-Dimensional Computed Tomography"[Mesh] OR "Single Photon Emission Computed Tomography Computed Tomography"[Mesh] OR "Spiral Cone-Beam Computed Tomography"[Mesh] OR "Cone-Beam Computed Tomography"[Mesh] OR "CT"[tiab] OR "computer tomography"[tiab] OR "computed tomography"[tiab] OR "computed*"[tiab]	901 311
#4	"Magnetic Resonance Imaging"[Mesh] OR "Diffusion Magnetic Resonance Imaging"[Mesh] OR "MRI"[tiab] OR "magnetic resonance imaging"[tiab]	699 618
#3	"radiomics"[tiab] OR "radiomic feature*"[tiab] OR "radiomic*" OR "texture*"[tiab] OR "characteristic*"[tiab] OR "computer analysis"[tiab]	1 786 137
#2	"Pancreatic Fistula"[Mesh] OR "pancreatic fistula"[tiab] OR "pancreas fistula"[tiab] OR "POPF"[tiab] OR "CR-POPF"[tiab]	6 237
#1	"Pancreaticoduodenectomy"[Mesh] OR "Pancreatoduodenectomy"[tiab] OR "pancreaticoduodenectomy"[tiab] OR "pancreatic surgery"[tiab] OR "whipple"[tiab] OR "pancreas surgery"[tiab]	20 444

Supplementary appendix II.

	Search strategy: Embase	
Search	**Embase query - January 29, 2023**	**Results**
#7	#1 AND #2 AND #3 AND #6	191
#6	#4 OR #5	2 585 706
#5	'computer assisted tomography'/exp OR 'ct':ti,ab,kw OR 'computer tomography':ti,ab,kw OR 'computed tomography':ti,ab,kw OR 'computed*':ti,ab,kw	1 735 559
#4	'nuclear magnetic resonance imaging'/exp OR 'diffusion weighted imaging'/exp OR 'mri':ti,ab,kw OR 'magnetic resonance imaging':ti,ab,kw	1 245 413
#3	'radiomics'/exp OR 'radiomics score'/exp OR 'radiomics quality score'/exp OR 'radiomic feature'/exp OR 'radiomics model'/exp OR 'radiomics':ti,ab,kw OR 'radiomic features':ti,ab,kw OR 'radiomic*':ti,ab,kw OR 'texture*':ti,ab,kw OR 'characteristic*':ti,ab,kw OR 'computer analysis':ti,ab,kw OR 'computer*':ti,ab,kw	2 715 001
#2	'pancreas fistula'/exp OR 'pancreatic fistula':ti,ab,kw OR 'pancreas fistula':ti,ab,kw OR popf:ti,ab,kw OR 'cr-popf':ti,ab,kw	12 698
#1	'pancreaticoduodenectomy'/exp OR 'pancreatoduodenectomy':ti,ab,kw OR 'pancreaticoduodenectomy':ti,ab,kw OR 'pancreatic surgery':ti,ab,kw OR 'whipple':ti,ab,kw OR 'pancreas surgery':ti,ab,kw	37 317

Supplementary appendix III

Search strategy: Web of Science Core Collection		
Search	**Web of Science Core Collection query - January 29, 2023**	**Results**
#7	#1 AND #2 AND #3 AND #6	69
#6	#4 OR #5	1 820 602
#5	AB='CT' OR AB='computer tomography' OR AB='computer*' OR AB='computed tomography' OR AB='compute*'	1 475 494
#4	AB='MRI' OR AB='magnetic resonance imaging'	432 628
#3	AB='radiomic*' OR AB='radiomic feature* OR AB='texture*' OR AB='texture analysis' OR AB='computer analysis' OR AB=characteristic*'	3 295 335
#2	AB='pancreatic fistula' OR AB='pancreas fistula' OR AB='POPF' OR AB='CR-POPF'	5 254
#1	AB='pancreatoduodenectomy' OR AB='pancreaticoduodenectomy' OR AB='pancreatic surgery' OR AB='whipple' OR AB='pancreas surgery'	26 123

Supplementary appendix IV, adherence to TRIPOD Guidelines.

Section/Topic	Item		Checklist item	Study, *n (%)*
Title	1	D;V	Identify the study as developing and/or validating a prediction model, the target population, and the primary outcome of interest	0 (0)
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusion	0 (0)
Introduction				
Background and objectives	3a	D;V	Explain the medical context and rationale for developing the prediction model	7 (100)
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both	7 (100)
Methods				
Source of data	4a	D;V	Describe the study design source of data	6 (86)
	4b	D;V	Specify the key study dates	6 (86)
Participants	5a	D;V	Specify key elements of the study design including number and location of centers	4 (57)
	5b	D;V	Describe eligibility criteria for participants	6 (86)
	5c	D;V	Give details of treatments received, if relevant	NA
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed	5 (71)
	6b	D;V	Report any actions to blind assessment of the outcome of the predicted	0 (0)
Predictors	7a	D;V	Clearly define all predictors used in developing or validation the prediction model	6 (86)
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors	3 (43)
Sample size	8	D;V	Explain how the study size was arrived at	0 (0)
Missing data	9	D;V	Describe how missing data were handled with details of any imputation method	0 (0)

Statistical data analysis methods	10a	D	Describe how predictors were handled in the analysis	7 (100)
	10b	D	Specify type of model, all model-building procedures, any method for internal validation	5 (71)
	10c	V	For validation, describe how the prediction were calculated	1 (100)
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models	1 (14)
	10e	V	Describe any model updating arising from the validation, if done	NA
Risk groups	11	D;V	Provide details on how risk groups were created, if done	NA
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors	0 (0)
Results				
Participants	13a	D;V	Describe the flow of participants trough the study, including the number of participants with and without outcome and, if applicable, a summary of the follow-up time.	5 (71)
	13b	D;V	Describe the characteristics of the participants, including the number of participants with missing data for predictors and outcome	0 (0)
	13c	V	For validation, show a comparison with the development data of the distribution of important variables	1 (100)
Model development	14a	D	Specify the number of participants and outcome events in each analysis	2 (29)
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome	NA
Model specification	15a	D	Present the full prediction model to allow predictions for individuals	4 (57)
	15b	D	Explain how to use the prediction model	1 (14)
Model performance	16	D;V	Report performance measures (with Cis) for the prediction model	1 (14)
Model-updating	17	V	If done, report the results from any model updating	NA
Discussion				
Limitations	18	D;V	Discuss any limitations of the study	7 (100)
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data	0 (0)
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	7 (100)
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research	5 (71)
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, web calculator, and data sets	6 (86)
Funding	22	D;V	Give the source of funding and the role of the funders for the present study	4 (57)

Items relevant *only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V.*

TRIPOD: transparent *reporting of a multivariable prediction model for individual prognosis or diagnosis. NA; not applicable*

Supplementary appendix V, quality assessment of the included studies using the Review of the Development of the risk of bias tool for nonrandomized studies for interventions (ROBINS-I) tool.

Study	Bias due to confounding	Bias in the selection of participants into the study	Bias in classification of interventions	Bias due to deviations from intended interventions	Bias due to missing data	Bias in measurement of outcomes	Bias in selection of the reported results
Bhasker (2022) (23)							
Capretti (2022) (24)							
Kambakamba (2020) (17)							
Lin (2021) (25)							
Mu (2020) (26)							
Skawran (2021) (27)							
Zhang (2018) (28)							

8

Chapter 8

Artificial Intelligence: application for the operating room

Erik W. Ingwersen*
Boukje T. Bootsma*
Freek Daams
Donald L. van der Peet

Nederlands Tijdschrift voor de Geneeskunde 2023 Aug 23;167
Originally published in Dutch: Artificiële intelligentie: toepassingen voor de operatiekamer
**These authors contributed equally to this manuscript*

Summary

The operating room is nowadays a data-rich environment where artificial intelligence (AI) can play a significant role. Current AI applications are aimed at supporting perioperative decision-making and enhancing surgical skills and safety. To promote the implementation of AI, several steps need to be taken. Studies focusing on external validation and data standardization, as well as monitoring of implementation, are required. Additionally, consensus is needed regarding the ethical and legal aspects. Overall, AI is expected to have a substantial impact on making surgical care more efficient and safe.

Article

Artificial intelligence (AI) is on the rise. It also has the potential to be a valuable tool in and around surgeries. Despite numerous studies on AI, there is a question about how concrete its application currently is. AI is sometimes viewed with skepticism as a hype that receives excessive attention. In this article, we will discuss the reality behind this and the benefits that AI can offer to surgical specialties.

What is AI?

There is much discussion about AI in healthcare. For example, the Intensive Care unit at Amsterdam UMC has recently started using AI to determine whether a patient can be discharged to the ward.[1] AI is a collective term that encompasses the development of computer systems capable of performing tasks that normally require human intelligence. This includes understanding natural language, recognizing images, making decisions, and problem-solving. AI relies on various components such as machine learning, deep learning, neural networks, or computer vision.

The more information ('big data') the computer system is provided with, the more accurate the model's outcomes. With the introduction of the electronic health record, minimally invasive surgery, and robot-assisted surgery, more and more big data is being generated. This can include patient characteristics, imaging studies, quantitative data from surgical equipment, and anesthesiological parameters. These data are often unstructured and exist in various formats. Unlike traditional data analysis techniques, AI can analyze such data. This offers several possibilities for the operating room. The applications can be divided into preoperative, intraoperative, and postoperative (see Table 1).

Table 1, the possibilities of artificial intelligence for surgical specialties

Artificial Intelligence applications
Preoperative
- Predictive modelling
- 'Radiomics'
- Personalized care
Preoperative
- Recognition of anatomical structures
- 3D view
- Real-time analysis and prediction of risks
Postoperative
- Benchmarking of surgical skills
- Evaluation of working processes

Preoperative

Preoperatively, AI can contribute to personalized patient care and support clinical decision-making. Current scoring systems that support surgical decision-making are often based on post-hoc analyses and are complex. Additionally, they do not always function optimally and are time-consuming because data often need to be collected and input manually. AI models do not have to have this disadvantage since they can use live streaming of data from the electronic health record. It has been demonstrated that AI, based on preoperative data, can predict the risk of anastomotic leakage after a 'low anterior' resection.[2] In the Netherlands, an AI model has been developed that can predict the risk of anastomotic leakage during a colorectal resection. This model is currently being externally validated in various hospitals.[3]

Another rapidly growing field of research is the use of 'radiomics'. In this approach, conventional medical images, such as CT, PET, or MRI scans, are processed into large amounts of quantitative data using AI. An externally validated study in patients with pancreatic cancer showed that a model based on preoperative CT radiomics and patient characteristics could better predict the risk of a pancreatic fistula than current scoring systems.[4]

In both examples, the AI model can serve as an aid in decision-making and be used to provide the patient with more tailored information about the risk of complications before surgery. Additionally, the surgeon can decide, for example, not to create an anastomosis at all or to take additional measures during surgery to reduce the severity of a leakage for high-risk patients identified by AI.

Intraoperative

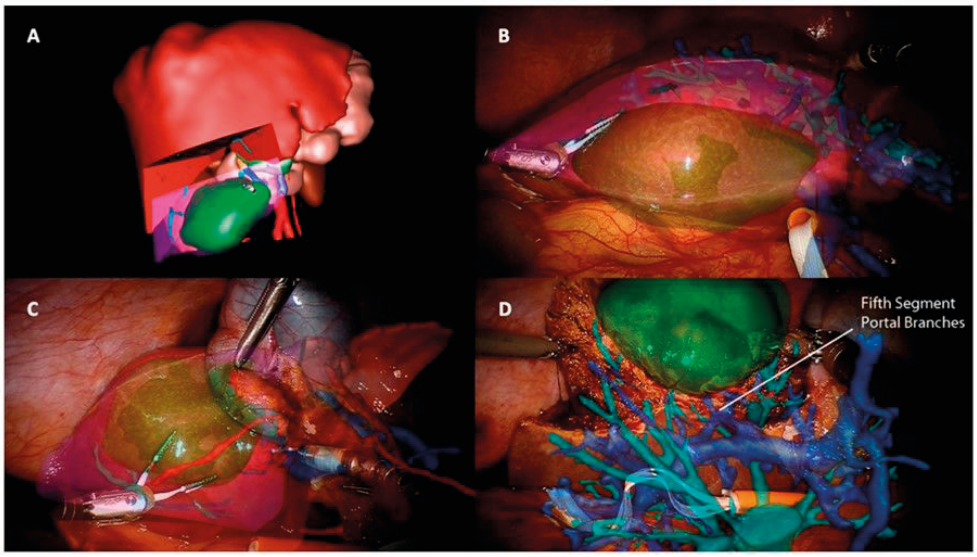
Recognition of Anatomy

Intraoperatively, AI can analyze image and video material through computer vision (CV). CV is used in research to recognize anatomy during minimally invasive intra-thoracic and intra-abdominal surgery.[5] Vital structures such as the aorta, bile ducts, and lymphatic vessels are identified by an AI model. Several models have been developed that show valuable outcomes.[6] For instance, it has been demonstrated that during a laparoscopic cholecystectomy, important anatomical structures, such as the Calot's triangle, can be successfully identified. This could potentially lead to a reduction in bile duct injuries and improved postoperative outcomes.[7]

3D Visualization

A step beyond recognizing anatomy is 3D visualization. It has been shown that AI can accurately depict abnormalities and vascularization in the liver and can adequately delineate the resection margin for tumor tissue (see Figure 1).[8,9] In vascular surgery, AI navigation software is now available that projects the preoperative CT scan onto the X-ray screen. The images are automatically aligned with the patient's position, allowing the surgeon to continuously see a detailed 3D image of the blood vessels during the procedure.

Figure 1, an example of 3D visualization during liver surgery



a) Representation of the tumor (green) and an optional resection plane (pink-red). (b-d). Vascular and biliary structures can be projected during different surgical phases. Figure derived from previous publication.[9]

Further applications are in development for assessing a completion angiogram at the end of the procedure. A completion angiogram contains information about the position of the placed stent, the patency of the stent and arteries, and can visualize leaks ('endoleaks'). Currently, these images are assessed by the naked eye by the surgical team. AI can assess these images at the pixel level, making it possibly more accurate in recognizing endoleaks.[10] The future scenario is that the surgeon is alerted by the algorithm in the case of an endoleak, and the surgeon can then decide to take additional steps during surgery to prevent re-intervention.

These examples of 3D applications are very complex and essentially straddle the boundary between AI and augmented reality. Displaying deformable structures like these requires significant computing power. AI navigation software is already standard in use in one hospital in the Netherlands for vascular surgeries. In other centers, these 3D techniques are still only applied in a research context.

Predicting Blood Pressure Drops

In the field of anesthesiology, AI also appears to have measurable benefits. An American technology company developed a machine learning algorithm that can predict a drop in blood pressure during surgery through highly accurate analysis of the arterial blood pressure curve.[11] This warning system triggers an alert when a patient is at risk of developing hypotension, even before the patient actually experiences it. This is advantageous because an intervention in response to hypotension typically comes a few minutes too late, which can lead to severe complications. Amsterdam UMC tested this AI algorithm in a research setting during elective surgeries and demonstrated a 70% decrease in the incidence of hypotension.[12] Currently, the system is undergoing further testing within the cardiothoracic surgical population.

Postoperative

In the postoperative phase, AI offers many opportunities for evaluating workflow processes in the operating room and for the training of surgeons. A good example is benchmarking surgical performance based on deep learning analysis of surgical videos. One of the current shortcomings of the existing evaluation process is that it is time-consuming, often making it unfeasible to analyze the entire video. Additionally, there is significant variability among evaluators.

AI analysis does not have these shortcomings. It provides an objective assessment and efficiently manages time. It can provide insight into someone's technical performance over an extended period, document it in a personal portfolio, and specify learning objectives.[13] Furthermore, it can be used to assess quality in the training of surgeons and in the evaluation of already certified surgeons. A large American consensus study among surgeons, urologists, gynecologists, supervisors, data scientists, and lawyers demonstrates support for these applications.[14]

Future Outlook

An important caveat regarding the described AI applications is that they are based on results from relatively small studies. Most models are specifically designed and tuned to the hospital where the model is used. Little can be said about the generalizability of the

models because it is still unknown whether the model from Hospital X is also applicable in Hospital Y.

Headlines like 'Computer can detect cancer better than a pathologist' and 'Google says AI can detect lung cancer a year earlier than doctors' contribute to the assumption that AI should be implemented as soon as possible.[15,16] However, what is suggested here is far from reality. While the results of perioperative AI applications are promising, few have actually been implemented. This is due in part to the lack of external validation studies. Additionally, it is important to optimize collaboration among different parties, and the support among healthcare providers and patients is still limited.[17]

How Can We Promote Implementation?

To promote the integration of AI into surgical care, certain aspects deserve attention. First, it is important to take a nuanced view of existing AI applications. The focus should be on developing applications that provide a solution to an existing problem, not on creating gadgets for non-existent problems. Additionally, it must be proven that AI can actually solve the problem better than regular healthcare. Transparency plays a role in this. Publications about AI models should provide detailed descriptions of the application or model and be open about the results and the development of the models.[18]

Regarding the use of AI in medical decision-making, it is important to keep in mind that AI should always play a supplementary role and not a replacement role. The fear of autonomous decision-making by AI can make patients wary.[19]

Finally, clear legal and ethical frameworks must be established for the use of AI, which includes algorithmic bias and liability for errors. These frameworks are subject to ongoing debate and societal developments. For example, the question may arise about who is responsible if an AI model makes an incorrect prediction during a procedure, leading to harm to the patient: the technology company, the researchers who developed the model, or the surgeon? The Dutch AI Coalition and the Ministry of Health, Welfare, and Sport's "Valuable AI" program hope to provide answers to these questions.[20]

Conclusion

AI is already being used by surgical specialties and will continue to be applied more extensively. To promote implementation, the focus must be on standardizing data, clinically validating AI systems, carefully evaluating the implementation process, and establishing ethical and legal frameworks. Overall, there are high expectations for AI in making surgical care more efficient and safer.

References

1. Van Lonkhuyzen L. De ontslagsoftware zegt dat de patiënt van de IC mag. NRC, 24 augustus 2022.
2. Wen R, Zheng K, Zhang Q, et al. Machine learning-based random forest predicts anastomotic leakage after anterior resection for rectal cancer. J Gastrointest Oncol. 2021;12:921-32. doi:10.21037/jgo-20-436. Medline
3. Dutch Society of Colorectal Surgery 'A1Check studie': https://dscrs.nl/portfolio_page/a1check-studie/, geraadpleegd op 5 augustus 2023.
4. Shi Y, Gao F, Qi Y, et al. Computed tomography-adjusted fistula risk score for predicting clinically relevant postoperative pancreatic fistula after pancreatoduodenectomy: Training and external validation of model upgrade. EBioMedicine. 2020;62:103096. doi:10.1016/j.ebiom.2020.103096. Medline
5. den Boer RB, de Jongh C, Huijbers WTE, et al. Computer-aided anatomy recognition in intrathoracic and -abdominal surgery: a systematic review. Surg Endosc. 2022;36:8737-52. doi:10.1007/s00464-022-09421-5. Medline
6. Mascagni P, Alapatt D, Sestini L, et al. Computer vision in surgery: from potential to clinical value. NPJ Digit Med. 2022;5:163. doi:10.1038/s41746-022-00707-5. Medline
7. Mascagni P, Vardazaryan A, Alapatt D, et al. Artificial Intelligence for Surgical Safety: Automatic Assessment of the Critical View of Safety in Laparoscopic Cholecystectomy Using Deep Learning. Ann Surg. 2022;275:955-61. doi:10.1097/SLA.0000000000004351. Published online November 16, 2020. Medline
8. Bari H, Wadhwani S, Dasari BVM. Role of artificial intelligence in hepatobiliary and pancreatic surgery. World J Gastrointest Surg. 2021;13:7-18. doi:10.4240/wjgs.v13.i1.7. Medline
9. Giannone F, Felli E, Cherkaoui Z, Mascagni P, Pessaux P. Augmented Reality and Image-Guided Robotic Liver Surgery. Cancers (Basel). 2021;13:6268. doi:10.3390/cancers13246268. Medline
10. Kappe KO, Smorenburg SPM, Hoksbergen AWJ, Wolterink JM, Yeung KK. Deep Learning-Based Intraoperative Stent Graft Segmentation on Completion Digital Subtraction Angiography During Endovascular Aneurysm Repair. J Endovasc Ther. 2022;0:15266028221105840. doi:10.1177/15266028221105840. Medline
11. Hatib F, Jian Z, Buddi S, et al. Machine-learning Algorithm to Predict Hypotension Based on High-fidelity Arterial Pressure Waveform Analysis. Anesthesiology. 2018;129:663-74. doi:10.1097/ALN.0000000000002300. Medline
12. Wijnberge M, Geerts BF, Hol L, et al. Effect of a Machine Learning-Derived Early Warning System for Intraoperative Hypotension vs Standard Care on Depth and Duration of Intraoperative Hypotension During Elective Noncardiac Surgery: The HYPE Randomized Clinical Trial. JAMA. 2020;323:1052-60. doi:10.1001/jama.2020.0592. Medline
13. Khalid S, Goldenberg M, Grantcharov T, Taati B, Rudzicz F. Evaluation of Deep Learning Models for Identifying Surgical Actions and Measuring Performance. JAMA Netw Open. 2020;3:e201664. doi:10.1001/jamanetworkopen.2020.1664. Medline
14. Vedula SS, Ghazi A, Collins JW, et al; Collaborative for Advanced Assessment of Robotic Surgical Skills. Artificial Intelligence Methods and Artificial Intelligence-Enabled Metrics for Surgical Education: A Multidisciplinary Consensus. J Am Coll Surg. 2022;234:1181-92. doi:10.1097/XCS.0000000000000190. Published online May 11, 2022. Medline
15. Murphy M. Google says its AI can spot lung cancer a year before doctors. The Telegraph, 7 mei 2019.

16. Computer kan kanker beter herkennen dan patholoog. NOS Nieuws, 3 maart 2017.
17. Starke G, Ienca M. Misplaced Trust and Distrust: How Not to Engage with Medical Artificial Intelligence. Camb Q Healthc Ethics. 20 oktober 2022 (epub). doi:10.1017/S0963180122000445. Medline
18. Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. BMJ. 2020;368:m689. doi:10.1136/bmj.m689.
19. Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. Lancet Digit Health. 2021;3:e599-611. doi:10.1016/S2589-7500(21)00132-1. Medline
20. Onderzoeksrapport 'Artificiële Intelligentie en Passende Zorg'. Diemen: Zorginstituut Nederland; 2022.

Part III

Development of prediction models

9

Chapter 9

Machine learning versus logistic regression for the prediction of complications after pancreatoduodenectomy

Erik W. Ingwersen
Wessel T. Stam
Bono J.V. Meijs
Joran Roor
Marc G Besselink
Bas Groot Koerkamp
Ignace H.J.T. de Hingh
Hjalmar C. van Santvoort
Martijn W. Stommel
Freek Daams
On behalf of the Dutch Pancreatic Cancer Group

Surgery 2023 Sep; 174(3):435-440

Abstract

Background: Machine learning is increasingly advocated to develop prediction models for postoperative complications. It is, however, unclear if machine learning is superior to logistic regression when using structured clinical data. Postoperative pancreatic fistula and delayed gastric emptying are the two most common complications with the biggest impact on patient condition and length of hospital stay after pancreatoduodenectomy. The aim of this study was to compare the performance of machine learning and logistic regression in predicting pancreatic fistula and delayed gastric emptying after pancreatoduodenectomy.

Methods: This retrospective observational study used nationwide data from 16 centers in the Dutch Pancreatic Cancer Audit between January 2014 and January 2021. The area under the curve of a machine learning and logistic regression model for clinically relevant postoperative pancreatic fistula and delayed gastric emptying were compared.

Results: Overall, 799 (16.3%) patients developed postoperative pancreatic fistula and 943 (19.2%) delayed gastric emptying. For postoperative pancreatic fistula, the area under the curve of the machine learning model was 0.74 and the area under the curve of the logistic regression model was 0.73. For delayed gastric emptying, the area under the curve of both the machine learning model and logistic regression was 0.59.

Conclusion: Machine learning did not outperform logistic regression modelling in predicting postoperative complications after pancreatoduodenectomy.

Introduction

Pancreatoduodenectomy is a complex surgical procedure with considerable morbidity and a negative influence on short-term quality of life.[1] It is a substantial burden for the health care resource utilization and health expenditure, especially in those with a complicated postoperative recovery.[2,3] Postoperative pancreatic fistula (POPF) and delayed gastric emptying (DGE) are the two most common, high-impact complications after pancreatoduodenectomy, with sizable effects on resource utilization and prolonged length of stay.[4,5]

Both patients and clinicians would benefit from an accurate prediction of POPF and DGE, and, if available, a proper risk assessment would be adopted widely.[6] In the preoperative setting, individual risk assessment could aid clinical-decision making. Peri-operatively, this could lead to early awareness of complications, allowing for selective preventive and timely therapeutic measures. The majority of these models are based on traditional logistic regression models, in which the probability of an outcome is related to a certain number of predictors. These models are relatively easy to use and interpret, but have some drawbacks.[7] First, the usual assumption for logistic regression is that there is a linear relation between the independent and dependent variables. Second, predictors are usually chosen using backward selection. This technique has some problems (e.g. once a variables has been eliminated, it cannot be re-entered).[8]

Machine learning algorithms are increasingly advocated as they are less prone to the abovementioned problems.[7,9] Machine learning algorithms are able to detect non-linear relationship between independent and dependent variables and can incorporate a large number of variables.[10,11] These models are, however, more susceptible to overfitting and have the so-called black-box phenomenon, meaning that the models are not interpretable for humans. This can undermine the use of these model in daily clinical practice.[12]

Although both techniques have been used to develop risk models for postoperative complications, it is unclear if machine learning is superior to logistic regression when using structured data.[13-16] The aim of this study was to compare machine learning to logistic regression in predicting POPF and DGE after pancreatoduodenectomy using structured clinical data from a nationwide audit.

Methods

Study design and population

This retrospective cohort study used data from the nationwide Dutch Pancreatic Cancer Audit (DPCA) including data from 16 centers (2014-2020). In the Netherlands, all patients undergoing a pancreatic resection are prospectively registered in the DPCA. This type of study does not require approval from an ethics committee. The study protocol was approved by the scientific committee of the Dutch Pancreatic Cancer Group.[17] The DPCA data has been verified regarding its completeness and accuracy of postoperative complication registration, showing that the data is representative for the Dutch pancreatic cancer population.[18]

All patients undergoing elective, open and minimally invasive pancreatoduodenectomy were included. Patients with no information about the outcome and patients below the age of 18 were excluded. The STROBE guidelines were followed to ensure correct reporting of the study methods and results.[19]

Outcome and definition

The primary outcomes were POPF and DGE, both defined by the International Study Group of Pancreatic Surgery.[20,21] Only grade B/C POPF and DGE were included. The area under the curve – receiver operating characteristic (AUROC) was calculated to evaluate the discriminative power of the logistic regression and machine learning model for prediction of POPF and DGE. Four different machine learning models were examined: random forest, a neural network, support vector machine, and gradient boosting. An AUROC between 0.50 and 0.60 was defined as bad, between 0.60 and 0.70 as poor, between 0.70 and 0.80 as moderate, and between 0.80 and 0.90 as good.[22-24]

Variables

The following preoperative patient characteristics and preoperative data were collected: age, sex, body mass index (BMI), weight loss per week, Eastern Cooperative Oncology Group (ECOG) performance score, and the latest measurement of preoperative: serum bilirubin, hemoglobin, albumin, IgG4, CEA and CA19.9. American Society of Anesthesiologists (ASA) classification, American Joint Committee of Cancer (AJCC) stage, age-adjusted Charlson Comorbidity Index (CCIa),[25] tumor location (i.e. pancreatic body, head, peri-ampullary, duodenum), tumor size, vessel involvement, tumor involvement in other organs, lymph nodes bigger than 10 millimeter, remote metastases, the sum of worrisome features,[26] histological diagnosis, application of neoadjuvent therapy, biliary drainage, type of stent, complications and/or re-intervention after biliary drainage, and the response evaluation criteria in solid tumors criteria (RECIST).

The following intraoperative data were collected: type of surgical procedure (i.e. pylorus preserving pancreaticoduodenectomy (PPPD) or pylorus resecting pancreaticoduodenectomy (PRPD)), pancreatic duct diameter, perioperative evidence of residual macroscopic tumor, arterial vessel resection, venous vessel resection, additional resections (including spleen, transversal mesocolon, colon segment resection, hemicolectomy, gastric resection or other resections), texture of the pancreas, single row anastomosis or non-single row anastomosis, other perioperative measures (i.e. drains, stents, probes, somatostatins), length of surgical procedure, and blood loss.

Statistical analysis

The analysis was performed using SAS VIYA version 2021.2.3. Baseline pre- and intraoperative characteristics were described. Normally distributed continuous variables were presented as mean with a standard deviation (SD) and continuous variables with skew distribution as median with an interquartile range (IQR). Dichotomous variables were noted as number and percentage. Differences between patients with or without either a POPF and DGE were compared using Pearson chi-square test in dichotomous variables and a student's T-test (normally distributed) or Mann-Whitney U test (skew distribution) in continuous variables. A p-value <0.05 was considered as statistically significant.

Logistic regression model

Variables were excluded if more than 80% of the data were missing. Data were imputed via predictive mean matching with 10 iterations.[27] The pooled outcome of the 10 iterations was used. The logistic regression model was developed using backward selection, including all variables with at least a p-value <0.1 in at least six of the ten imputed datasets. The association between the predictors and outcome were displayed an odds ratio (OR), 95% confidence interval (95% CI), and p-values. The models were validated using bootstrapping, with 250 bootstrapping samples. In order to evaluate and compare the discriminative power of the models, the median AUROC with an IQR and the brier score of the 250 bootstrap samples was calculated. The brier score is a method to verify the accuracy of a probability forecast. A brier score of 0 means perfect accuracy and a score of 1 means perfect inaccuracy.[28]

Machine learning model

Variables were excluded if more than 80% of the data were missing. Continuous variables were imputed following median imputation and categorical variables following count imputation. The hyperparameters of the best machine learnings model were determined and chosen using the autotuning tool: this tool seeks the best hyperparameters, while keeping the risk of overfitting low. This hyperparameter optimization tool is integrated in SAS VIYA. The models were validated using bootstrapping, with 250 bootstrapping samples. Supplementary appendices I and II show how many times variables were

included in the 250 different bootstrapped of the best performing machine learning model. The tables also show the mean importance of the variable and the total importance, which is the mean importance multiplied by the number inclusion in the model. The mean importance of the variables was determined using the Gini Importance, which calculates each feature importance as the sum over the number of splits that include the features, proportionally to the number of samples it splits.[29] In order to evaluate and compare the discriminative power of the models, the median AUROC of the 250 bootstrap samples with an IQR was calculated.

Results

Between January 2014 and December 2020, 4 972 patients after pancreatoduodenectomy were registered in the DPCA. In total, 60 patients were excluded because of missing information on one of the outcomes. None of the included variables exceeded 80% of missing data. Among the 4 912 included patients, 799 (16.3%) patients developed a POPF, 943 (19.2%) patients DGE. Patients had a mean age of 66.7 (SD:10.5), the majority was male (55.4%) and underwent a procedure via laparotomy (82.9%). All pre- and intraoperative characteristics per postoperative outcomes are displayed in Table 1.

Table 1. Pre- and intraoperative characteristics

Preoperative characteristics	Total (N=4 912)	POPF B/C (n= 799)	DGE B/C (n=943)	Missing (%)
Age (mean in years)	66.7±10.5	67.1±10.3	67.6±10.2	8 (0.2%)
Sex (male)	2 720 (55.4%)	486 (60.8%)	551 (58.4%)	2 (<0.1%)
BMI (mean in kg/m^2)	25.3±4.3	26.6±4.3	25.3±4.3	166 (3.4%)
ASA classification ≥3	1 335 (27.6%)	224 (28.6%)	268 (29.0%)	71 (1.4%)
ECOG score ≥2	370 (8.8%)	43 (6.3%)	72 (8.7%)	688 (14.0%)
Hemoglobin (mean in mmol/l)	7.8±1.1	7.9±1.1	7.8±1.1	2 095 (42.7%)
Albumin (mean in mmol/l)	36.5±10.2	36.8±7.7	36.2±7.8	2 812 (57.2%)
Weight loss per week in kg (median)	0.4 [0-1]	0.1 [0 – 0.88)	0.2 [0 – 1.0]	1 567 (31.9%)
Bilirubin (median in mmol/L)	17 [7 – 66]	11.5 (7.0 – 29.8)	10.0 [6.0 – 38.5]	2 972 (60.5%)
CA19.9 (median in U/ml)	23 [0 – 160]	24.5 (9.0 – 99.4)	27.0[16.5 – 327]	347 (7.1%)
CEA (median in µg/l)	3.1 [1.9 – 5.0]	2.7 (1.7 – 4.7)	3.0 [2.0 – 5.5]	2 383 (48.5%)
IgG4 (median in g/l)	0 [0 – 0.29]	0 (0 – 0.4)	0.1 [0 – 0.42]	674 (13.7%)
Diameter of tumour on CT-scan (mean in mm)	27.7±22.2	28.2±27.0	28.5±28.1	1 924 (39.2%)
Vessel involvement	1 213 (25.9%)	116 (15.3%)	200 (22.3%)	233 (4.7%)
Involvement of structures	457 (9.8%)	67 (8.8%)	107 (12.0%)	234 (4.8%)
Lymph node involvement	583 (12.4%)	101 (13.2%)	115 (12.8%)	229 (4.7%)

Biliary drainage	2526 (53.7%)	362 (45.3%)	442 (48.4%)	207 (4.2%)
Application of neoadjuvant therapy	426 (14.8%)	20 (4.0%)	74 (13.5%)	2 028 (41.3%)
CCIa ≥4	2 940 (59.9%)	497 (62.2%)	589 (62.5%)	0
Surgical procedure				
PRPD	2 396 (48.8%)	435 (54.4%)	502 (53.2%)	
PPPD	2 516 (51.2%)	364 (45.6%)	441 (46.8%)	
Approach				87 (1.8%)
Open	4 001 (82.9%)	591 (74.0%)	741 (80.5%)	
Laparoscopic	298 (6.2%)	69 (8.8%)	60 (6.5%)	
Robotic	526 (10.9%)	126 (16.0%)	119 (12.9%)	
Location of tumor				1 671 (34.0%)
Duodenum	264 (8.1%)	76 (14.6%)	72 (11.8%)	
Head of pancreas	2 414 (74.5%)	287 (55.3%)	406 (66.8%)	
Peri-ampullary	563 (17.4%)	156 (30.1%)	130 (21.4%)	
Arterial resection	74 (1.5%)	15 (1.9%)	21 (2.2%)	46 (0.9%)
Venous resection	710 (14.6%)	57 (7.2%)	116 (12.4%)	45 (0.9%)
Additional resection	464 (9.8%)	94 (12.2%)	135 (14.3%)	172 (3.5%)
Length of surgical procedure in minutes	340±115	342±113	341.6±110.7	3 664 (74.6%)
Blood loss in ml (median in ml)	450 [200 – 845]	450 [200 – 800]	500 [200 – 800]	3,545 (72.2%)
Soft pancreas	2 755 (61.7%)	606 (83.6%)	577 (67.2%)	450 (9.1%)
Diameter pancreatic duct	4.8±4.6	3.5±5.2	4.2±4.5	1,442 (29.4%)
Single row pancreatic anastomosis	1 208 (24.6%)	145 (18.1%)	225 (23.9%)	494 (10.1%)
Intra-abdominal drains	4 725 (96.2%)	772 (97.1%)	901 (96.9%)	18 (0.4%)
Administration of somatostatin analogue	2 990 (61.1%)	528 (66.4%)	589 (62.7%)	18 (0.4%)
Stent in pancreas anastomosis	1 619 (33.1%)	355 (44.7%)	349 (37.0%)	18 (0.4%)
Nasojejunal feeding tube	987 (20.2%)	106 (13.3%)	168 (17.9%)	18 (0.4%)

POPF; postoperative *pancreatic fistula, DGE; delayed gastric emptying, BMI; body mass index, ASA; American Society of Anesthesiologists, ECOG; Eastern Cooperative Oncology Group, CA19.9; cancer-antigen 19.9, CEA; carcinoembryonic antigen, IgG4; immunoglobulin G4, CT; computer tomography, CCIa; adjusted comprehensive complication index, PRPD; pylorus resecting pancreaticoduodenectomy, PPPD; pylorus preserving pancreaticoduodenectomy.*

Postoperative pancreatic fistula

The AUROC of the logistic regression for POPF after bootstrapping was 0.73 with a pooled brier score of 0.12. Predictors in the logistic regression model were: soft pancreatic texture (OR: 2.70, p<0.001), non-single row pancreatic anastomosis (OR: 1.41, p<0.002),

male sex (OR: 1.38, p<0.001), plastic stent for biliary drainage (OR: 1.12, p=0.043), BMI (OR per kg/m^2 increase: 1.08, p<0.001), bilirubin (per µmol/L increase: 0.99, p=0.053), pancreatic duct size (OR per mm increase: 0.91, p<0.001), weight loss (OR per kg weight loss: 0.89, p=0.012), venous resection (OR: 0.74, p=0.003), nasojejunal feeding tube (OR: 0.61, p<0.001), localization of tumor, and surgical approach. (see Table 2) The mean AUROC of the best performing machine learning model (gradient boosting) for POPF after bootstrapping was 0.74 (IQR 0.73 - 0.74), with a pooled brier score of 0.13. The ten variables with the highest sum of the relative importance were: BMI, diameter of pancreatic duct, pancreatic texture, CA19.9, hemoglobin, age, localization of tumor, CEA, diameter of tumor, and non-single row pancreatic anastomosis. (Table 3). Supplementary appendix I shows the relative importance of all pre- and intraoperative variables.

Table 2. Predictors in the multivariate logistic regression model for POPF grade B/C

Variables	OR (95% CI))	P - value
Soft pancreas	2.70 (2.19 – 3.33)	<0.001
Non-single row anastomosis*	1.41 (1.40 – 1.76)	0.002
Male sex	1.38 (1.17 – 1.62)	<0.001
Plastic stent for biliary drainage	1.12 (1.00 – 1.26)	0.043
BMI per kg/m^2 increase	1.08 (1.06 – 1.10)	<0.001
Bilirubin per µmol/L increase	0.99 (0.99 – 1.00)	0.053
Pancreatic duct size, per mm increase	0.91 (0.88 – 0.95)	<0.001
Weight loss per kg loss	0.89 (0.81 – 0.97)	0.012
Venous resection	0.74 (0.61 – 0.91)	0.003
Nasojejunal feeding tube	0.61 (0.48 – 0.78)	<0.001
Localization of tumor	-	-
Head of pancreas	Reference	-
Body of pancreas	1.15 (0.65 – 2.02)	0.640
Peri-ampullary	1.74 (1.38 – 2.20)	<0.001
Duodenum	1.61 (1.18 – 2.20)	0.003
Approach	-	-
Open surgery	Reference	-
Laparoscopic	1.84 (1.37 – 2.49)	<0.001
Robotic	1.28 (1.01 – 1.62)	0.045

***This included** *the following techniques for PJ: Blumgart, duct-to-mucosa, and dunking/intussusception.*
POPF; postoperative *pancreatic fistula, OR; odds ratio, CI; confidence interval, BMI; body mass index.*

Table 3. The ten variables in the gradient boosting model with the highest sum of the mean relative importance in predicting POPF grade B/C

Variables	Included in prediction model	Mean relative importance	Sum
BMI	250	5.66	1414.39
Diameter pancreatic duct	250	5.19	1297.37
Soft pancreas	250	4.40	1099.60
CA19.9	250	2.85	711.56
Hemoglobin	250	2.67	668.17
Age	250	2.44	609.17
Tumor location	250	2.27	566.33
CEA	250	1.97	491.36
Diameter of tumor	250	1.70	425.00
Non-single row anastomosis	250	1.62	404.51

Table 4 *shows the 10 most important variables in the gradient boosting model for predicting POPF. The entire model with all included variables and their importance is shown in Supplementary Appendix I.*
***This included** the following techniques for PJ: Blumgart, duct-to-mucosa, and dunking/intussusception.*
POPF; postoperative *pancreatic fistula, BMI; body mass index, CA19.9; cancer-antigen 19.9, CEA; carcinoembryonic antigen.*

Delayed gastric emptying

The AUROC of the logistic regression for DGE after bootstrapping was 0.59 with a pooled brier score of 0.15. Predictors in the logistic regression model were: soft pancreatic texture (OR: 1.32, p<0.001), a PRPD (OR: 1.33, p<0.001), the application of a somatostatin analogue (OR: 1.18, p=0.002), plastic stent (OR: 1.15, p=0.010), age (OR per year increase: 1.01, p=0.005), bilirubin (OR per µmol/L increase: 0.99, p=0.007), nasojejunal feeding tube (OR 0.84, p=0.070), and localization of tumor. (see Table 4) The mean AUROC of the best performing machine learning model (gradient boosting) for DGE after bootstrapping was 0.59 (IQR 0.58 – 0.60), with a pooled brier score of 0.14. The ten variables with the highest sum of the relative importance were: BMI, age, CA19.9, diameter of pancreatic duct, CEA, hemoglobin lever, diameter of tumor, albumin, weight loss, and non-single row anastomosis. (Table 5) Supplementary appendix II shows the relative importance of the pre- and intraoperative variable.

Table 4. Predictors in the multivariate model for DGE grade B/C

Variables	OR (95% CI)	P - value
Soft pancreas	1.32 (1.13 -1.55)	<0.001
Type surgical procedure - PRPD	1.33 (1.14 – 1.56)	<0.001
Somatostatin analogue	1.18 (1.07 – 1.31)	0.002
Plastic stent	1.15 (1.03 – 1.27)	0.010
Age per year increase	1.01 (1.00 – 1.02)	0.005
Bilirubin per µmol/L increase	0.99 (0.99 – 1.00)	0.007
Nasojejunal feeding tube	0.84 (0.70 – 1.01)	0.070
Localization of tumor	-	-
Head of pancreas	Reference	-
Body of pancreas	0.88 (0.48 - 1.59)	0.623
Peri-ampullary	1.27 (1.03 – 1.57)	0.029
Duodenum	1.34 (1.03 – 1.77)	0.038

DGE; delayed *gastric emptying, OR; odds ratio, CI; confidence interval, PRPD; pylorus resecting pancreatoduodenectomy.*

Table 5. The ten variables in the gradient boosting model with the highest sum of the mean relative importance in predicting DGE

Variables	Included in prediction model	Mean relative importance	Sum
BMI	250	3.62	904.32
Age	250	3.02	755.54
CA19.9	250	2.69	672.80
Diameter pancreatic duct	250	2.34	584.68
CEA	250	2.12	530.58
Hemoglobin	250	2.11	527.92
Diameter of tumor	250	2.01	503.36
Albumin	250	1.69	422.43
Weight loss	250	1.62	404.62
Non-single row anastomosis	250	1.58	394.09

Table 5 *shows the 10 most important variables in the gradient boosting model for predicting DGE. The entire model with all included variables and their importance is shown in Supplementary Appendix II.*
***This included** the following techniques for PJ: Blumgart, duct-to-mucosa, and dunking/intussusception.*
DGE; delayed *gastric emptying, BMI; body mass index, CA19.9; cancer-antigen 19.9, CEA; carcinoembryonic antigen.*

Discussion

This first multicenter audit-based study to compare machine learning and logistic regression for prediction models found no superiority of the machine learning model for the prediction of POPF and DGE after pancreatoduodenectomy using structured pre- and intraoperative variables. For both the machine learning and logistic regression prediction model, the predictive performance of the POPF model was moderate and for DGE bad.

An accurate prediction of a POPF would be extremely helpful as it would support clinicians in identifying patients at risk, thereby improving peri-operative decision-making.[30,31] However, given the performance scores, the models developed in this study are not useful for daily clinical practice. Especially, concerning the fact that there are several prediction models available with higher AUROCs and less input variables. For example, the Fistula Risk Score, developed by Callery et al. (2013) This logistic regression model was based on four common POPF related risk factors (small duct, soft pancreas, high-risk pathology and excessive blood loss). After analysis and internal validation on 233 patients, the Fistula Risk Score performed moderate with an AUROC of 0.72.[32] Six years later, Mungroop et al. (2019) developed an alternative Fistula Risk Score based on only three risk factors (small duct, soft pancreas and high BMI). The group designed the model based on 1 924 patients and the alternative score was externally validated and compared to the original fistula risk score based on 926 patients. The models achieved an AUROC of 0.78 for the alternative and 0.75 for the original fistula risk score in the external validation.[33] The models are easy to understand and convenient to use, since they use only three or four predictors.

There are several reasons why machine learning did not outperform logistic regression, as was hypothesized in this study. As was mentioned earlier, machine learning models are able to take into account the nonlinear interaction between predictors and the outcomes.[34] The majority of the variables in this study were discrete values, while the potency of machine learning models lies in continuous data. These variables lie on a higher-dimensional space and the relationship of the predictors with the outcomes is more complex. The number of variables in this database may have been too small for machine learning to show its benefit. Moreover, only structured data was used. Machine learning showed its superiority over logistic regression when unstructured data was added.[35] Radiomic features derived from CT or MRI images, for instance, might be able to provide a lot of unstructured data. The results from studies using radiomics to predict a POPF are promising, with an AUROC varying from 0.80 to 0.90 in several monocenter retrospective studies.[36-39] Future studies should further investigate its use and externally validate it, preferably using both radiomics features and structured clinical data. It

should also be acknowledged that not everything is predictable and a perfect predictive performance is an utopia.[40] Although the etiology and the accompanying risk factors of POPF and DGE are partly understood, there might still be unknown and unrecognized factors (both in the pre- and peri-operative phase) playing a role of importance that are lacking in our current audit.

Despite the findings of this study, the use of ML models in comparison to logistic regression may have has some advantages.[41,42] First, they are data driven and have the ability to continuously improve, since both algorithms can see patterns or features itself without human intervention.[42] Second, each factor is individually reassessed where the complex interrelationship per patient between the different factors and outcomes are also taken into account. This is important in an individualized prediction since these interrelationships are different for each person, making some factors more relevant than others. Third, in the future, machine learning can easily be added to electronic health records, which makes it possible to easily and automatically predict complications based on the pre- and intra operative information, and unstructured available at that moment.[42] All in all, accurate prediction of individual outcome has the potential to reshape the future of postoperative management and offers the opportunity to develop a more individualized approach allowing for data-driven individualized medicine.[43-45]

The results of this study should be interpreted in light of several limitations. First, the proportion of missing data was large for some variables (up to 75%), potentially causing bias and reduction in efficiency. Because complete case analysis would drastically reduce the sample size available, multiple imputation was performed. Madley-Dowd et al. showed that imputation reduced bias even when the proportion of missingness was large (up to 90%) and missing data was at random (as was the case in this study).[46] Variables were therefore only excluded if missingness exceeded 80%. Second, this study defined both POPF grade B and C as outcome. While grade B POPF is a complication that can have a significant impact on quality of life, it is not likely that the complication will determine whether a patient will undergo a pancreatoduodenectomy. A future prediction model should only predict a grade C POPF in order to accurately determine which patients are at high-risk for this life-threatening complication. Third, both pre- and intraoperative data were used and in order to improve the predictability of the outcomes. However, clinically it may be more helpful if the model uses only preoperative predictors. Finally, this study used multicenter data from a prospectively maintained audit. However, external and prospective validation is necessary to increase the generalizability of the findings. Strengths of this study include the large study population and its multicenter nationwide study design. Moreover, this study compared machine learning with logistic regression within the same database.

Conclusion

This study showed no difference in predictive performance of machine learning compared to logistic regression in predicting POPF and DGE. Performances of both models are suboptimal and not useful for daily clinical practice. Future research should focus on adding radiomics to these models, in order to objectively determine risks preoperatively.

Funding/Support

The author received no funding or support for this work.

COI/disclosures

The authors have no relevant financial disclosures.

References

1. Nahm CB, Connor SJ, Samra JS, Mittal A. Postoperative pancreatic fistula: a review of traditional and emerging concepts. Clin Exp Gastroenterol. 2018;11:105-18
2. Wang J, Ma R, Churilov L, Eleftheriou P, Nikfarjam M, Christophi C, et al. The cost of perioperative complications following pancreaticoduodenectomy: A systematic review. Pancreatology. 2018;18(2):208-20.
3. Wang J, Ma R, Eleftheriou P, Churilov L, Debono D, Robbins R, et al. Health economic implications of complications associated with pancreaticoduodenectomy at a University Hospital: a retrospective cohort cost study. HPB (Oxford). 2018;20(5):423-31.
4. Mirrielees JA, Weber SM, Abbott DE, Greenberg CC, Minter RM, Scarborough JE. Pancreatic Fistula and Delayed Gastric Emptying Are the Highest-Impact Complications After Whipple. J Surg Res. 2020;250:80-7.
5. Smits FJ, Verweij ME, Daamen LA, van Werkhoven CH, Goense L, Besselink MG, et al. Impact of Complications After Pancreatoduodenectomy on Mortality, Organ Failure, Hospital Stay, and Readmission: Analysis of a Nationwide Audit. Ann Surg. 2022;275(1):e222-e8.
6. Marchegiani G, Bassi C. Prevention, prediction, and mitigation of postoperative pancreatic fistula. British Journal of Surgery. 2021;108(6):602-4.
7. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. Journal of Clinical Epidemiology. 1996;49(11):1225-31.
8. Chowdhury MZI, Turin TC. Variable selection strategies and its importance in clinical prediction modelling. Fam Med Community Health. 2020;8(1):e000262.
9. Deo RC. Machine Learning in Medicine. Circulation. 2015;132(20):1920-30.
10. Kuhn M, Johnson K. Applied predictive modeling: Springer; 2013.
11. LeCun Y, Bengio Y, Hinton G. Deep learning. nature. 2015;521(7553):436-44.
12. Guidi JL, Clark K, Upton MT, Faust H, Umscheid CA, Lane-Fall MB, et al. Clinician perception of the effectiveness of an automated early warning and response system for sepsis in an academic medical center. Annals of the American Thoracic Society. 2015;12(10):1514-9.
13. Callery MP, Pratt WB, Kent TS, Chaikof EL, Vollmer CM, Jr. A prospectively validated clinical risk score accurately predicts pancreatic fistula after pancreatoduodenectomy. J Am Coll Surg. 2013;216(1):1-14.
14. Han IW, Cho K, Ryu Y, Shin SH, Heo JS, Choi DW, et al. Risk prediction platform for pancreatic fistula after pancreatoduodenectomy using artificial intelligence. World J Gastroenterol. 2020;26(30):4453-64.
15. Merath K, Hyer JM, Mehta R, Farooq A, Bagante F, Sahara K, et al. Use of Machine Learning for Prediction of Patient Risk of Postoperative Complications After Liver, Pancreatic, and Colorectal Surgery. Journal of Gastrointestinal Surgery. 2020;24(8):1843-51.
16. Stam WT, Goedknegt LK, Ingwersen EW, Schoonmade LJ, Bruns ERJ, Daams F. The prediction of surgical complications using artificial intelligence in patients undergoing major abdominal surgery: A systematic review. Surgery. 2022;171(4):1014-21.
17. Strijker M, Mackay TM, Bonsing BA, Bruno MJ, van Eijck CHJ, de Hingh I, et al. Establishing and Coordinating a Nationwide Multidisciplinary Study Group: Lessons Learned by the Dutch Pancreatic Cancer Group. Annals of surgery. 2020;271(4):e102-e4.
18. van Rijssen LB, Koerkamp BG, Zwart MJ, Bonsing BA, Bosscha K, van Dam RM, et al. Nationwide prospective audit of pancreatic surgery: design, accuracy, and outcomes of the Dutch Pancreatic Cancer Audit. HPB (Oxford). 2017;19(10):919-26.

19. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Bmj. 2007;335(7624):806-8.
20. Bassi C, Marchegiani G, Dervenis C, Sarr M, Abu Hilal M, Adham M, et al. The 2016 update of the International Study Group (ISGPS) definition and grading of postoperative pancreatic fistula: 11 Years After. Surgery. 2017;161(3):584-91.
21. Welsch T, Borm M, Degrate L, Hinz U, Büchler MW, Wente MN. Evaluation of the International Study Group of Pancreatic Surgery definition of delayed gastric emptying after pancreatoduodenectomy in a high-volume centre. Br J Surg. 2010;97(7):1043-50.
22. Tape TG. Interpretation of Diagnostic Tests. Annals of Internal Medicine. 2001.
23. Trevethan R. Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. Frontiers in public health. 2017;5:307.
24. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. Indian journal of ophthalmology. 2008;56(1):45-50.
25. de Groot V, Beckerman H, Lankhorst GJ, Bouter LM. How to measure comorbidity. a critical review of available methods. J Clin Epidemiol. 2003;56(3):221-9.
26. Hipp J, Mohamed S, Pott J, Sick O, Makowiec F, Hopt UT, et al. Management and outcomes of intraductal papillary mucinous neoplasms. BJS Open. 2019;3(4):490-9.
27. Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. BMC Medical Research Methodology. 2014;14(1):75.
28. Brier GW. Verification of forecasts expressed in terms of probability. Monthly weather review. 1950;78(1):1-3.
29. Nembrini S, König IR, Wright MN. The revival of the Gini importance? Bioinformatics. 2018;34(21):3711-8.
30. Bootsma BT, Plat VD, van de Brug T, Huisman DE, Botti M, van den Boezem PB, et al. Somatostatin analogues for the prevention of pancreatic fistula after open pancreatoduodenectomy: A nationwide analysis. Pancreatology. 2022;22(3):421-6.
31. Smits FJ, Henry AC, Besselink MG, Busch OR, van Eijck CH, Arntz M, et al. Algorithm-based care versus usual care for the early recognition and management of complications after pancreatic resection in the Netherlands: an open-label, nationwide, stepped-wedge cluster-randomised trial. Lancet. 2022;399(10338):1867-75.
32. Miller BC, Christein JD, Behrman SW, et al. A multi-institutional external validation of the fistula risk score for pancreatoduodenectomy. *J Gastrointest Surg*. 2014 Jan;18(1):172-79; discussion 179-80.
33. Mungroop TH, van Rijssen LB, van Klaveren D, Smits FJ, van Woerden V, Linnemann RJ, et al. Alternative Fistula Risk Score for Pancreatoduodenectomy (a-FRS): Design and International External Validation. Ann Surg. 2019;269(5):937-43.
34. Gravesteijn BY, Nieboer D, Ercole A, Lingsma HF, Nelson D, van Calster B, et al. Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury. J Clin Epidemiol. 2020;122:95-107.
35. Zhang D, Yin C, Zeng J, Yuan X, Zhang P. Combining structured and unstructured data for predictive models: a deep learning approach. BMC Med Inform Decis Mak. 2020;20(1):280.
36. Capretti G, Bonifacio C, De Palma C, Nebbia M, Giannitto C, Cancian P, et al. A machine learning risk model based on preoperative computed tomography scan to predict postoperative outcomes after pancreatoduodenectomy. Updates Surg. 2022;74(1):235-43.

37. Lin Z, Tang B, Cai J, Wang X, Li C, Tian X, et al. Preoperative prediction of clinically relevant postoperative pancreatic fistula after pancreaticoduodenectomy. Eur J Radiol. 2021;139:109693.
38. Skawran SM, Kambakamba P, Baessler B, von Spiczak J, Kupka M, Müller PC, et al. Can magnetic resonance imaging radiomics of the pancreas predict postoperative pancreatic fistula? Eur J Radiol. 2021;140:109733.
39. Zhang W, Cai W, He B, Xiang N, Fang C, Jia F. A radiomics-based formula for the preoperative prediction of postoperative pancreatic fistula in patients with pancreaticoduodenectomy. Cancer Manag Res. 2018;10:6469-78.
40. Coggon DIW, Martyn CN. Time and chance: the stochastic nature of disease causation. The Lancet. 2005;365(9468):1434-7.
41. Bertsimas D, Dunn J, Velmahos GC, Kaafarani HMA. Surgical Risk Is Not Linear: Derivation and Validation of a Novel, User-friendly, and Machine-learning-based Predictive OpTimal Trees in Emergency Surgery Risk (POTTER) Calculator. Ann Surg. 2018;268(4):574-83.
42. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial Intelligence in Surgery: Promises and Perils. Ann Surg. 2018;268(1):70-6.
43. Beaulieu-Jones BK, Yuan W, Brat GA, Beam AL, Weber G, Ruffin M, et al. Machine learning for patient risk stratification: standing on, or looking over, the shoulders of clinicians? NPJ Digit Med. 2021;4(1):62.
44. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56.
45. Beam AL, Kohane IS. Translating Artificial Intelligence Into Clinical Care. Jama. 2016;316(22):2368-9.
46. Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. Journal of Clinical Epidemiology. 2019;110:63-73.

Supplementary appendix 1: importance and number of included variables of the gradient boosting model predicting POPF

Variables	Included in prediction model	Mean relative importance	Sum
BMI	250	5.66	1414.39
Diameter pancreatic duct	250	5.19	1297.37
Texture pancreas	250	4.40	1099.60
CA19.9	250	2.85	711.56
Hemoglobin	250	2.67	668.17
Age	250	2.44	609.17
Tumor location	250	2.27	566.33
CEA	250	1.97	491.36
Diameter of tumor	250	1.70	425.00
Non-single row anastomosis	250	1.62	404.51
Bilirubin	250	1.61	403.32
Weight loss	250	1.51	376.64
IgG4	250	1.12	280.30
Albumin	250	1.11	276.93
Approach	250	1.04	258.98
Stent in pancreatic anastomosis	250	1.03	257.00
Blood loss	250	1.02	254.95
Neoadjuvant therapy	249	0.78	193.09
Nasojejunal feeding tube	250	0.73	183.21
Type of surgical procedure	250	0.68	170.54
Length of surgical procedure	250	0.68	169.20
ECOG score	250	0.63	157.34
Preoperative pathology	250	0.60	151.16
Pancreatic anastomosis	250	0.60	149.98
Biliary drainage	250	0.59	147.89
Sex	250	0.59	147.25
CCIa	250	0.57	141.96
Tumor visible on CT/MRI	250	0.49	122.71
Comorbidity - diabetes	249	0.49	122.63
ASA classification	249	0.46	115.66
Vessel involvement PV / MSV	245	0.42	103.69
TNM classification	249	0.40	100.38
Venous resection - PV/MSV	247	0.40	99.50
Type of somatostatin analogue	249	0.36	88.87
Type of stent	244	0.32	78.06
Somatostatin analogue	248	0.30	74.12
Additional resection - n/s	241	0.25	60.54

Tumor visible on EUS	243	0.24	59.31
Preoperative laparoscopy	239	0.24	57.13
Use of standard CT	241	0.20	47.43
EUS performed	238	0.20	46.80
Lymph node involvement – hepatoduodenal ligament	222	0.21	46.71
Connective tissue disease	228	0.20	45.94
Comorbidity - peripheral arterial disease	225	0.20	45.12
Worrisome features	215	0.17	37.49
Comorbidity - myocardial infarction	228	0.16	37.33
Comorbidity - chronic lung disease	220	0.17	36.78
Lymph node involvement - peripancreatic	221	0.15	33.28
Comorbidity - ulcerative disease	215	0.15	32.06
Comorbidity - chronic kidney disease	216	0.14	29.63
RECIST criteria	165	0.18	29.08
Malignancy as comorbidity	212	0.14	28.74
Comorbidity - CVA/TIA	204	0.13	27.44
Comorbidity - heart failure	193	0.13	25.62
Stent in biliary digestive anastomosis	186	0.13	24.94
Lymph node involvement – para-aortic	162	0.13	20.53
Comorbidity - chronic pancreatitis	179	0.10	18.50
Vessel involvement – hepatic artery	167	0.09	14.25
Intra-abdominal drain	174	0.08	13.82
Vessel involvement - MSA	140	0.10	13.71
Tumor involvement peripancreatic fat tissue	156	0.09	13.51
Tumor involvement – small intestine	159	0.08	13.24
Type of reintervention after biliary drainage	143	0.09	13.14
Complication biliary drainage – cholangitis	88	0.15	12.82
Nasojejunal feeding tube	123	0.08	9.92
Tumor involvement – CBD	104	0.08	8.33
Complication biliary drainage - pancreatitis	97	0.07	6.43

Additional resection – mesocolon transversum	95	0.06	5.52
Additional resection - right hemicolectomy	2	0.20	0.40
Comorbidity - liver disease	1	0.10	0.10
Complication biliary drainage – occlusion	2	0.05	0.09

POPF; postoperative *pancreatic fistula, BMI: body mass index, CA19.9; cancer-antigen 19.9, CEA: carcinoembryonic antigen, IgG4; immunoglobulin G4, ECOG; Eastern Cooperative Oncology Group, ASA; American society of anesthesiologists, CCIa; age adjusted Charlson Comorbidity Index, PV; portal vein, MSV; mesenteric superior vein, EUS; endoscopic ultrasound, RECIST; response evaluation criteria in solid tumors, MSA; mesenteric superior artery, CBD; common bile duct*

Supplementary appendix II, importance and number of included variables of the gradient boosting model predicting DGE

Variables	Included in prediction model	Mean relative importance	Sum
BMI	250	3.62	904.32
Age	250	3.02	755.54
CA19.9	250	2.69	672.80
Diameter pancreatic duct	250	2.34	584.68
CEA	250	2.12	530.58
Hemoglobin	250	2.11	527.92
Diameter of tumor	250	2.01	503.36
Albumin	250	1.69	422.43
Weight loss	250	1.62	404.62
Non-single row anastomosis	250	1.58	394.09
IgG4	250	1.39	347.09
Blood loss	250	1.31	326.67
Bilirubin	250	1.30	326.22
Surgical procedure	250	1.05	261.75
Length of surgical procedure	250	0.93	233.03
Additional resection - n/s	250	0.83	208.12
Tumor location	250	0.83	208.09
Preoperative pathology	250	0.78	195.39
Biliary drainage	250	0.77	192.26
Somatostatin analogue	250	0.71	177.07
ECOG score	250	0.70	175.67
Stent in pancreatic anastomosis	250	0.66	165.55
CCIa	250	0.63	156.44
Texture pancreas	250	0.54	134.89

Approach	249	0.51	125.77
Nasojejunal feeding tube	249	0.50	124.36
Sex	250	0.49	122.54
Tumor visible on CT/MRI	250	0.47	117.41
ASA classification	250	0.40	99.82
Type of stent	248	0.38	95.25
TNM classification	247	0.36	89.83
Comorbidity - chronic kidney disease	239	0.37	88.59
Tumor visible on EUS	246	0.33	81.08
Comorbidity - diabetes	248	0.28	69.20
Nasojejunal feeding tube	204	0.33	67.73
Tumor involvement – small intestine	219	0.30	65.11
Worrisome features	234	0.27	64.17
Use of standard CT	241	0.26	62.55
Vessel involvement PV / MSV	240	0.25	61.04
Venous resection - PV/MSV	243	0.24	59.08
Preoperative laparoscopy	243	0.23	56.96
EUS performed	239	0.23	54.48
Comorbidity - chronic lung disease	237	0.21	49.71
Comorbidity - peripheral arterial disease	233	0.19	45.21
Comorbidity - myocardial infarction	229	0.20	44.77
Neoadjuvant therapy	226	0.19	42.81
Additional resection – mesocolon transversum	187	0.22	40.23
Lymph node involvement - peripancreatic	225	0.16	36.38
Malignancy as comorbidity	205	0.16	33.70
Lymph node involvement – hepatoduodenal ligament	210	0.13	28.04
Type of reintervention after biliary drainage	172	0.15	26.34
Comorbidity - ulcerative disease	192	0.12	23.62
RECIST criteria	181	0.11	20.70
Stent in biliary digestive anastomosis	177	0.11	20.33
Vessel involvement - MSA	174	0.12	20.32
Intra-abdominal drain	173	0.11	19.47
Comorbidity - chronic pancreatitis	177	0.11	19.20

Comorbidity - CVA/TIA	177	0.10	17.95
Connective tissue disease	166	0.11	17.85
Comorbidity - heart failure	157	0.11	16.67
Tumor involvement peripancreatic fat tissue	149	0.11	16.43
Lymph node involvement – para-aortic	137	0.11	15.58
Tumor involvement – CBD	142	0.10	14.30
Complication biliary drainage - pancreatitis	122	0.10	12.34
Vessel involvement – hepatic artery	119	0.08	9.83
Complication biliary drainage – cholangitis	52	0.08	4.12
Complication biliary drainage – occlusion	4	0.10	0.40
Additional resection - right hemicolectomy	2	0.17	0.35

DGE; delayed *gastric emptying, BMI; body mass index, CA19.9; cancer-antigen 19.9, CEA; carcinoembryonic antigen, IgG4; immunoglobulin G4, ECOG; Eastern Cooperative Oncology group, CCIa; age adjusted Charlson Comorbidity Index, ASA; American society of anesthesiologists, EUS; endoscopic ultrasound, PV; portal vein, MSV; mesenteric superior vein, RECIST; response evaluation criteria in solid tumors, MSA; mesenteric superior artery, CBD; common bile duct.*

10

Chapter 10

Radiomics preoperative-Fistula Risk Score (RAD-FRS) for pancreatoduodenectomy: development and external validation

Erik W. Ingwersen*
Jacqueline I. Bereska*
Alberto Balduzzi
Boris V. Janssen
Marc G. Besselink
Geert Kazemier
Giovanni Marchegiani
Giuseppe Malleo
Henk A. Marquering
C. Yung Nio
Riccardo de Robertis
Roberto Salvia
Ewout W. Steyerberg
Jaap Stoker
Femke Struik
Inez M. Verpalen#
Freek Daams#
On behalf of the Pancreatobiliary and Hepatic Artificial Intelligence Research (PHAIR) consortium

BJS open 2023 Sep 5; 7(5)
**These authors contributed equally to this manuscript*
#Shared last authorship

Structured abstract

Background: Accurately predicting the risk of clinically relevant postoperative pancreatic fistula (CR-POPF) after pancreatoduodenectomy prior to surgery may assist surgeons in making more informed treatment decisions and improved patient counseling. The aim was to evaluate the predictive accuracy of a radiomics-based preoperative-Fistula Risk Score (RAD-FRS) for CR-POPF.

Methods: Radiomic features were derived from preoperative CT scans from adult patients after pancreatoduodenectomy at a single center in the Netherlands (Amsterdam, 2013 -2018) to develop the RAD-FRS. Extracted radiomic features were analyzed with four machine learning classifiers. The model was externally validated in a single center in Italy (Verona, 2020 – 2021). The RAD-FRS was compared to the Fistula Risk Score (FRS) and the updated alternative Fistula Risk Score (ua-FRS).

Results: Overall, 359 patients underwent a pancreatoduodenectomy, of whom 89 (25%) developed a CR-POPF. The RAD-FRS model was developed using CT scans of 118 patients, of which three radiomic features where included in the random forest model, and externally validated in 57 patients. The model performed well with an area under the curve (AUC) of 0.90 (95% CI: 0.71 – 0.99) and 0.81 (95% CI: 0.69 – 0.92) in the Amsterdam test set and Verona dataset, respectively. The RAD-FRS performed similarly to the FRS (AUC 0.79) and ua-FRS (AUC 0.79).

Conclusion: The RAD-FRS, which uses only preoperative CT features, is a new and promising radiomics-based score that has the potential to be integrated with hospital CT report systems and improved patient counseling preoperatively. The model with underlying code is readily available via www.pancreascalculator.com and www.github.com/PHAIR-Consortium/POPF-predictor.

Introduction

Clinically relevant postoperative pancreatic fistula (CR-POPF) is a feared complication following pancreatoduodenectomy that negatively impacts short- and long-term outcomes.[1,2] Accurate risk stratification of CR-POPF in the preoperative setting can assist in determining the best surgical approach for high-risk or frail patients. In the perioperative period, high-risk patients for CR-POPF may be candidates for prophylactic treatment such as with somatostatin analogs.[3] In patients with cystic lesions of the pancreatic head, accurate risk stratification of CR-POPF can help decide whether to proceed with a pancreatoduodenectomy or consider alternative approaches.[4] A recent study demonstrated improved postoperative outcomes for high-risk patients who underwent a total pancreatectomy compared to those who underwent a pancreatoduodenectomy.[5]

Previous research has introduced several CR-POPF prediction models for pancreatoduodenectomy, including the Fistula Risk Score (FRS) and the updated alternative Fistula Risk Score (ua-FRS).[6,7] These risk models are commonly used but have limitations, including their reliance on subjective intraoperative assessments (e.g., the texture of the pancreas) and their inability to provide predictions before surgery.

Radiomics is an approach for extracting features from medical images. It enables objective approaches for texture analysis and can uncover new parameters, some invisible to the human eye.[8] Previous studies have investigated the use of computed tomography (CT)-based-radiomics models to predict CR-POPF.[9-13] However, radiomics-based models which are publicly available and externally validated are currently lacking. The objective of this study was to develop and externally validate a publicly available radiomics preoperative-Fistula Risk Score (RAD-FRS) in patients undergoing pancreatoduodenectomy using radiomic features from preoperative CT scans. Such a model could lead to automatic CR-POPF risk prediction in CT reports.

Methods

The Medical Ethics Review Committee of the Amsterdam UMC approved this study protocol and waived the need for informed consent. The Ethics Committee of Verona and Rovigo Provinces approved the utilization of data from the validation cohort (PAD-R, 1101CESC). The study adhered to the STROBE guidelines.[14] All patients were managed per institutional practices. At the Amsterdam UMC, it was a standard procedure to insert an abdominal drain after resection, and the routine administration of somatostatin analogues was not performed. The levels of drain fluid amylase were measured on

postoperative day 1 and 3. At Verona University Hospital, an assessment of the risk associated with CR-POPF was carried out intraoperatively using the FRS. In instances where a drain was employed, the levels of drain fluid amylase were measured on both postoperative day 1 and day 5.

Patients

For model design, adult patients after pancreatoduodenectomy in one of the two locations of the Amsterdam UMC (Vrije Universiteit Medical Center) were included from the Dutch Pancreatic Cancer Audit (January 2013-December 2018). Exclusion criteria were patients with a poor quality CT scan or with CTs with slice thickness >3.0mm. A poor quality CT scan was defined as a poor scan due to artifacts (e.g. respiratory motion artifacts). For external validation of the model, adult patients after pancreatoduodenectomy in the Verona University Hospital were included (January 2020-January 2021). The same eligibility criteria were applied. These datasets will be referred to as the Amsterdam and Verona dataset, respectively.

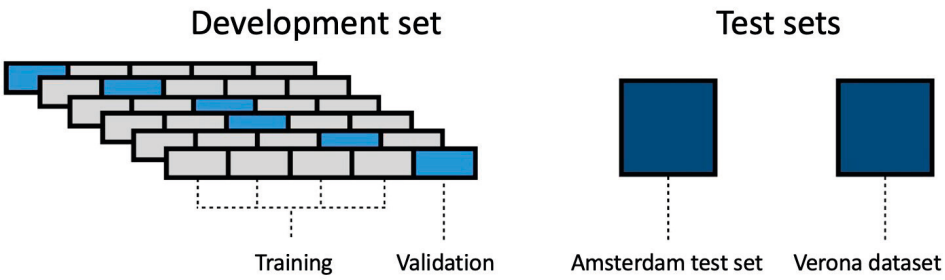
Data acquisition and outcome

The following patient demographics and tumor characteristics were retrospectively obtained from electronic health records: age (years), sex, BMI (kg/m^2), ASA classification, pre-existing diabetes, application of neoadjuvant therapy, surgical approach (i.e., open/laparoscopic/robotic surgery), pancreatic duct size, pancreatic texture, intraoperative blood loss, length of surgical procedure, pathology. The pancreatic duct size was measured intraoperatively with a ruler. The pancreatic texture was assessed subjectively by surgeon feeling intraoperatively and in minimally invasive surgery, the texture was determined at the resected pancreatoduodenectomy specimen. Contrast enhanced CT scans were obtained from the picture archiving and communication systems. The primary outcome was a clinically relevant POPF, defined as grade B or C according to the 2016 ISGPS definition.[15]

Datasets

Two full datasets, Amsterdam and Verona datasets, were used. Amsterdam dataset was further divided into three subdatasets as follows. First, the Amsterdam dataset was divided into two sets: the Amsterdam development set, comprising 90% of the Amsterdam dataset, and the Amsterdam test set, comprising the remaining 10% of the data. Within the Amsterdam development set, a further division was performed, creating Amsterdam training and validation sets using five-fold cross-validation. The models were trained on the Amsterdam development set and evaluated on both the Amsterdam test set and the Verona dataset. The data splits of the Amsterdam and Verona datasets are visualized in Figure 1.

Figure 1



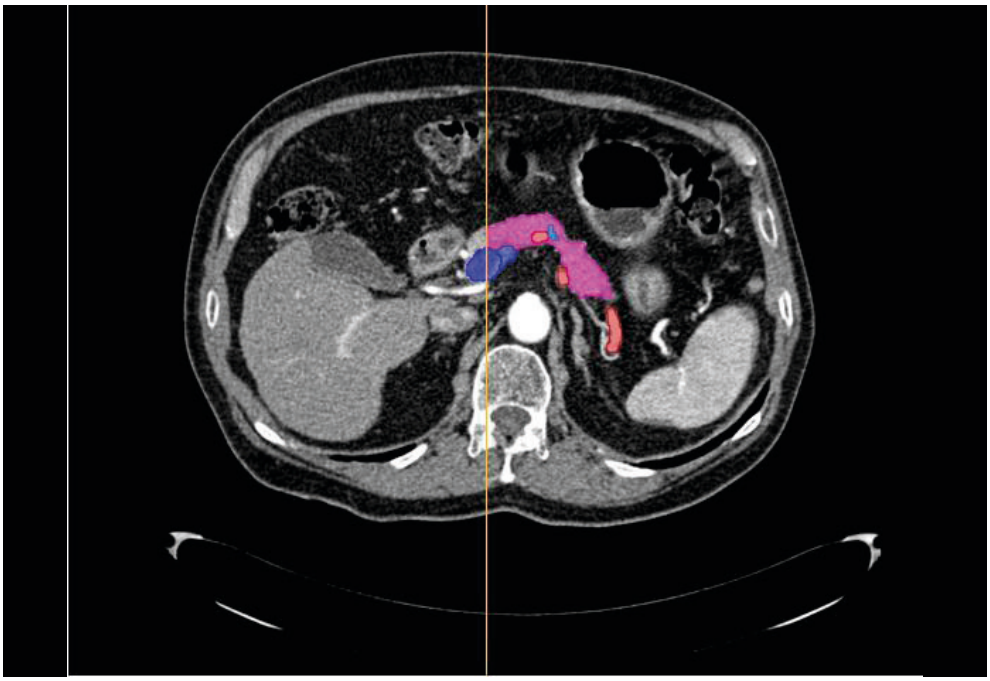
The Amsterdam *dataset (n=118) was split into Amsterdam development (90%, n=106) and test set (10%, n=12). The Amsterdam development set was further split into Amsterdam training and validation set using 5-fold cross validation. The best out of 25 models with regards to the AUROC on the Amsterdam test set was evaluated on the Verona dataset (n=57).*
AUROC; area *under the curve – receiver operating characteristic*

Image segmentation and radiomic analysis

We constructed a radiomics workflow consisting of the segmentation of the volume of interest (VOI) and radiomic feature extraction, feature selection and model construction on the Amsterdam development set, and performance evaluation on the Amsterdam test set and Verona dataset.

A PhD candidate specializing in the field of radiology and surgery (EI) and a resident in radiology (IV) manually pre-segmented the VOI in either the (late) arterial or portal venous phase using 3D slicer version 4.11.20210226 (www.slicer.org).[16] Subsequently, an experienced abdominal radiologist (YN or FS in the Amsterdam dataset and RR in the Verona dataset), who was blinded to the patient's outcome, finalized the segmentation. The VOI represented the pancreatic remnant, segmented distal/lateral (i.e., towards the tail) from the midline of the superior mesenteric vein, considering that all patients underwent a pancreatoduodenectomy. The segmentation encompassed both pancreatic tissue and the pancreatic duct. The surrounding vessels (i.e., splenic artery, splenic vein, superior mesenteric vein, and portal vein) were segmented to prevent their inclusion in the pancreatic segmentation mask but were not part of the final VOI (Figure 2). If both the arterial and portal venous phases were available, the pancreas was segmented in the (late) arterial phase. If the arterial phase was unavailable, the pancreas was segmented in the portal venous phase.

Figure 2.



An example *of a contrast-enhanced CT scan in the early arterial phase of the abdomen, showing the contoured pancreatic tissue (pink), pancreatic duct (light blue), splenic artery (red), and the superior mesenteric vein (dark blue) in an axial image. The vertical yellow line indicates the midline of the superior mesenteric vein, with the pancreas annotated on the left side. The volume of interest consisted of the pancreatic tissue and pancreatic duct*
CT; computed *tomography.*

Radiomic features were extracted from each VOI of patients in the Amsterdam dataset using PyRadiomics version 3.0 (http://github.com/radiomics/pyrdiomics).[17] Subsequently, 25 Amsterdam development test splits were created for final model selection. For each split, 90% of the Amsterdam dataset was allocated to training and 10% to testing. The features of each Amsterdam development and test set were normalized independently using the MinMax scaler. Two feature reduction methods were serially applied to the Amsterdam development sets to select the most predictive features for each split: removing of features with a variance of less than 0.001 and with an importance near zero using LASSO feature selection.

Four machine learning (ML) classifiers were fitted to the Amsterdam development sets: support vector machine, logistic regression, k-nearest neighbor, and random forest. Before training, the minority classes in the Amsterdam development sets were oversampled with a sampling strategy of 0.9, and, subsequently, the majority classes were undersampled. To

create more equally sized groups based on the presence or absence of CR-POPF, the data was manually undersampled. Finally, random Gaussian noise with a mean of zero and a variance of 0.2 was added to the Amsterdam development sets. A 5-fold cross-validation grid search optimizing AUROC was used to find the hyperparameters and fit all four ML classifiers on the Amsterdam development sets. All four ML classifiers were optimized, fitted, and evaluated on all 25 Amsterdam development-test splits. The best-performing ML classifier with regards to AUROC on the Amsterdam test set was validated on the Verona dataset. The performance of this model on both the Amsterdam test set and Verona dataset was evaluated using the AUROC, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and calibration plots.

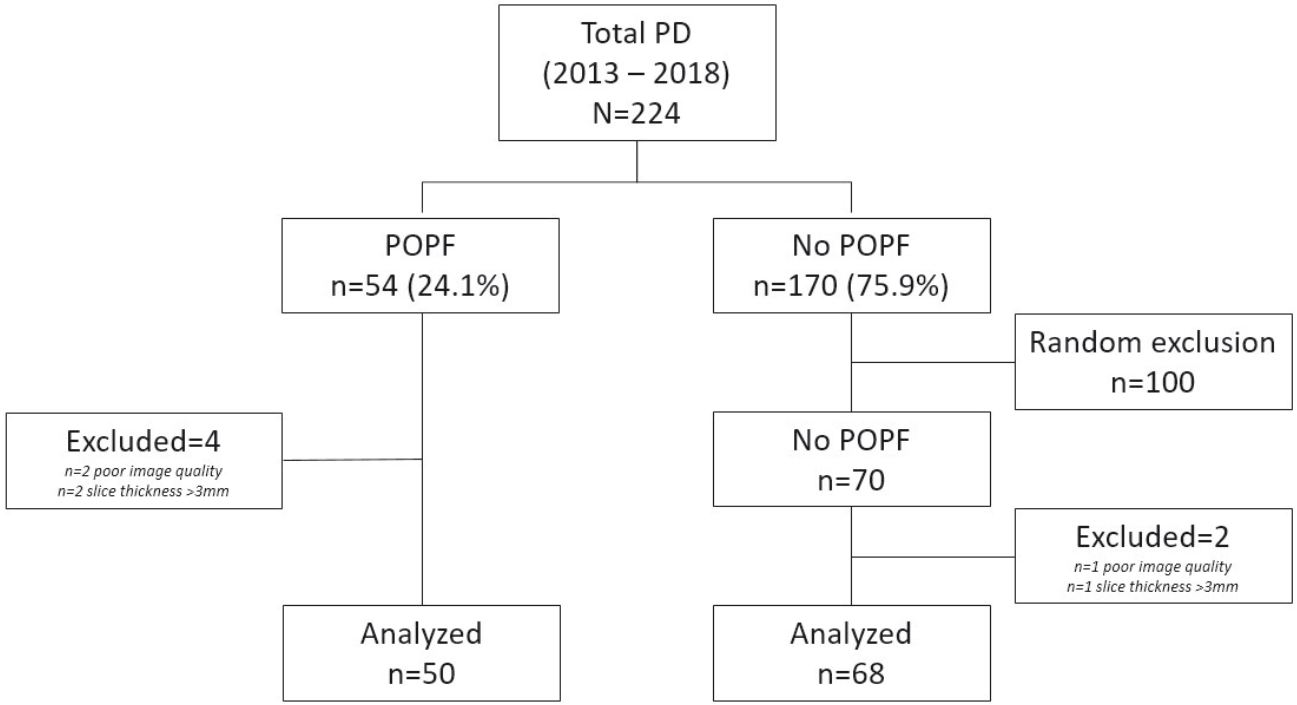
Statistical analysis

Statistical analysis was conducted using Python version 3.7 (Python Software Foundation, Wilmington, DE) and R-Studio version 2022.07.1 (R Studio Team (2018). RStudio: Integrated Development for R. RStudio Inc., Vienna, Austria). The AUROC of the RAD-FRS, FRS, and ua-FRS were reported with a 95% confidence interval (95% CI). The maximum Youden's *J* value from the AUROC was used to identify the cut probability where prediction discrimination was maximum according to sensitivity and specificity. The AUROC of the RAD-FRS was compared to the AUROC of the FRS and ua-FRS in the Verona dataset using DeLong's test. Calibration curves were used to compare the observed and estimated probability of the models. A p-value <0.05 was considered statistically significant. Continuous variables were reported as mean with a standard deviation (SD) or median with an interquartile range (IQR) if the distribution was skewed. Dichotomous, ordinal, and nominal variables were presented as numbers and percentages.

Results

Overall, 224 patients underwent a panceatoduodenectomy at UMC Amsterdam in the Netherlands and 135 at the Verona University Hospital in Italy. A total of 89 (25%) developed CR-POPF. In the Amsterdam dataset, a total of 106 patients were excluded: 100 patients without CR-POPF were randomly excluded for the analysis as a result of undersampling, 3 patients were excluded because of a poor quality CT scan, and 3 did not have slices ≤3.0 mm, resulting in 118 patients that were included in the Amsterdam dataset. Of those, 50 developed CR-POPF (Figure 3). In the Verona dataset, 135 patients underwent a pancreatoduodenectomy between January 2020 and January 2021, with CR-POPF occurring in 22 patients (Figure 4). As a result of undersampling, 60 patients without CR-POPF were randomly excluded from the analysis. Another 18 patients were excluded, 6 due to poor CT scan quality, and 12 did not have slices ≤3.0mm. In total, 57 patients were included in the Verona dataset, of whom 22 developed CR-POPF.

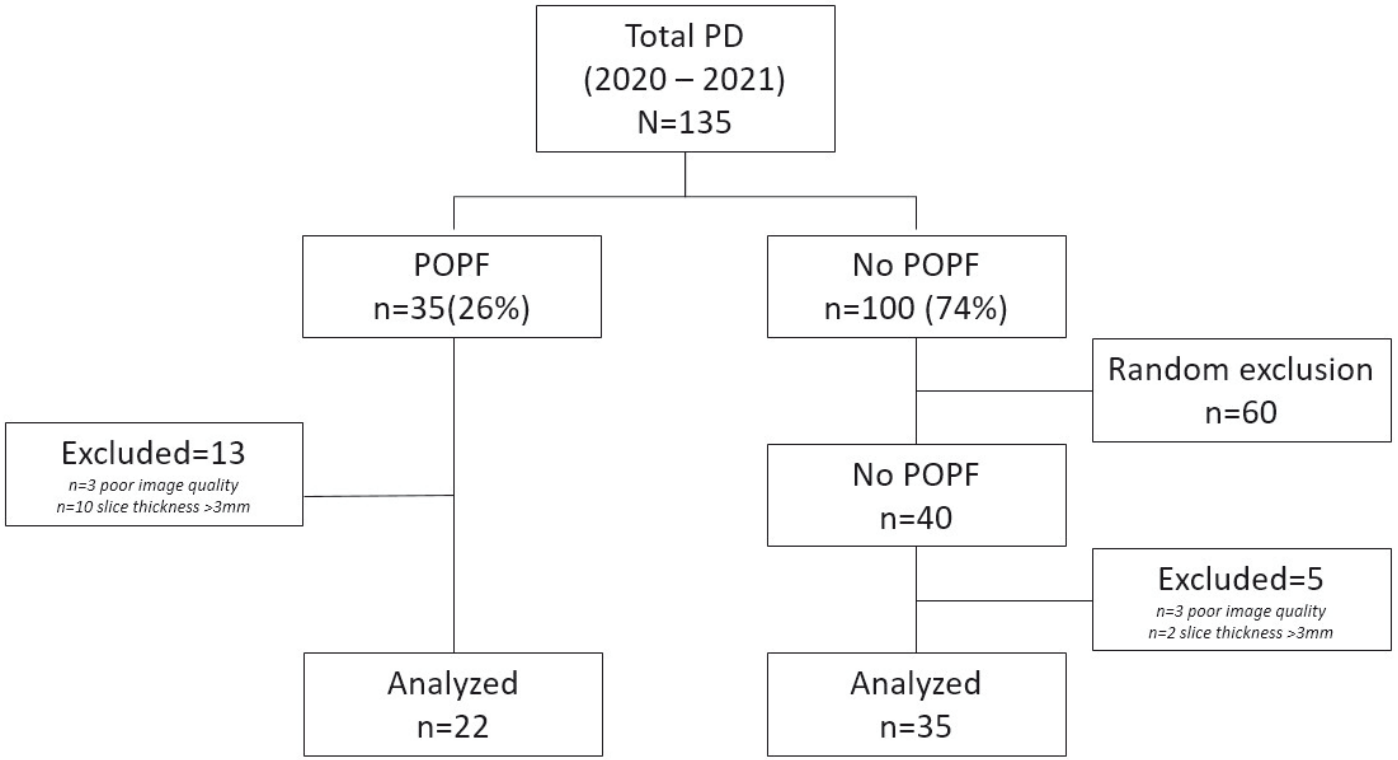
Figure 3. Flowchart of the Amsterdam dataset



Flowchart showing *the selection process of the Amsterdam dataset. A total of 100 patients were excluded in the group without CR-POPF to create more equally sized groups based on the presence or absence of CR-POPF. Further exclusion criteria were: poor image quality (n=3) and slice thickness above 3 mm (n=3). A total of 50 patients with CR-POPF and 68 without CR-POPF were found eligible for analysis.*

PD: pancreatoduodenectomy, *N; number, CR-POPF; clinically relevant postoperative pancreatic fistula, CT; computer tomography.*

Figure 4. Flowchart of the Verona dataset



Flowchart showing *the selection process of the Verona dataset. A total of 60 patients were excluded in the group without CR-POPF to create more equally sized groups based on the presence or absence of CR-POPF. Further exclusion criteria were: poor image quality (n=6) and slice thickness above 3 mm (n=12). A total of 22 patients with CR-POPF and 35 without CR-POPF were found eligible for analysis.*

PD: pancreatoduodenectomy, *N; number, CR-POPF; clinically relevant postoperative pancreatic fistula, CT; computer tomography.*

Characteristics of the Amsterdam and Verona datasets

The clinical characteristics of both the Amsterdam and Verona datasets are shown in Table 1. The dataset differed in sex (36% female in the Amsterdam dataset versus 45% in the Verona dataset), the application of neoadjuvant therapy (14% in the Amsterdam dataset versus 58% in the Verona dataset), surgical approach (15% of patients in the Amsterdam dataset underwent a laparoscopic pancreatoduodenectomy versus 0% in the Verona dataset), and application of a prophylactic somatostatin analogue (15.3% in the Amsterdam dataset versus 31.2% in the Verona dataset). The Amsterdam dataset comprised 118 patients with a median age of 68 years (IQR: 59 -79 years) and a mean BMI of 25.7 (SD: 3.9). The Verona dataset included 57 patients with a median age of 68 years (IQR: 61 – 75) and a mean BMI of 25 kg/m2 (SD: 4.1 kg/m^2). Reconstruction and acquisition parameters used to obtain the CT scans of the Amsterdam and Verona datasets are listed in Supplementary Table 1.

Table 1. Baseline characteristics of the Amsterdam and Verona datasets.

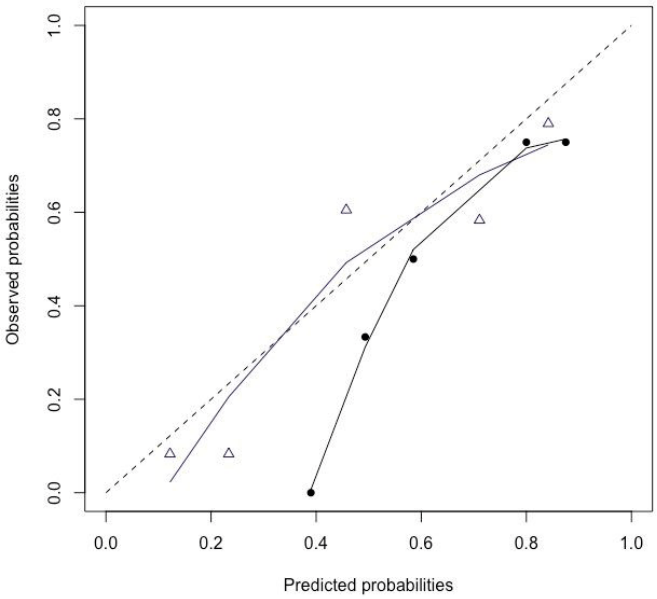
	Amsterdam dataset (n= 118)	**Verona dataset (n= 57)**
Age, median (IQR)	68 (59 – 79)	68 (61 – 75)
Sex (female), no. (%)	43 (36%)	25 (45%)
BMI, mean (SD)	26±3.9	25±4.1
ASA classification, no. (%)		
1 & 2	91 (77%)	42 (74%)
3 & 4	27 (23%)	15 (26%)
Diabetes Mellitus, no. (%)	20 (17%)	10 (18%)
Neoadjuvant therapy, no. (%)	16 (14%)	33 (58%)
Preoperative biliary drainage, no (%)	35 (29.6%)	17 (29.8%)
Laparoscopic procedure, no (%)	18 (15%)	0
Pancreatic duct size in mm, median (IQR)	3 (2 – 4)	4 (3 – 5)
Soft pancreas, no. (%)	76 (64%)	32 (55%)
Intraoperative blood loss in ml, median (IQR)	400 (380 – 420)	480 (350 – 792)
Length of surgical procedure (minutes), median (IQR)	325 (233 – 379)	390 (327 – 465)
Pathology		
PDAC	56 (48%)	38 (67%)
Distal cholangiocarcinoma	22 (19%)	3 (5.3%)
Ampullary carcinoma	13 (11%)	3 (5.3%)
Duodenal carcinoma	10 (8.5%)	4 (7.0%)
Cystic neoplasm (IPMN, serous cysts, MCN)	7 (5.9%)	3 (5.3%)
Inflammation / pancreatitis	5 (4.2%)	0
NET	4 (3.4%)	6 (10.5%)
Chronic pancreatitis	1 (0.8%)	0
Administration of prophylactic somatostatin analogues, no (%)	18 (15.3%)	18 (31.2%)

IQR; interquartile *range, BMI; body mass index, ASA; American Society of Anesthesiologists, SD; standard deviation, PDAC; pancreatic ductal adenocarcinoma, IPMN; intraductal papillary mucinous neoplasm, MCN; mucinous cystic neoplasm, NET; neuro-endocrine tumour.*

Amsterdam dataset

In the Amsterdam dataset, 103 (87%) CT scans were segmented in the (late) arterial phase and 15 (13%) in the portal venous phase. A total of 120 radiomic features were extracted. These features are listed in Supplementary Table 2. The final RAD-FRS model included three radiomic features: one gray level dependence matrix feature (gray level non uniformity) and two 3D shape based features (voxel volume and minor axis length). The gray level non uniformity feature quantifies gray level dependencies in an image and measures the variability in intensity values in the image. This feature corresponds to pancreatic texture. The 3D shape based features includes the descriptors of the three-dimensional size and shape of the VOI. The voxel volume feature corresponds to the volume of the VOI and thus the remnant of the pancreas volume. The minor axis length yields the second largest axis length of the VOI and could correlate to the pancreatic thickness. The random forest model performed best with an AUC of 0.90 (95% CI: 0.71 – 0.99) in the Amsterdam test set (n = 12). This model had a sensitivity of 1.00, specificity of 0.67, PPV of 0.71, and NPV of 1.00 in the Amsterdam test set. For the calibration curve of the Amsterdam test set, see Figure 5. The results of the other three machine learning models are listed in Table 2.

Figure 5.



A calibration *plot of the RAD-FRS in the Amsterdam test set and Verona dataset. The black dots represent the quintiles of the observed probabilities by quintiles of the predicted probabilities of Amsterdam test set, while the white triangles represent the same for the Verona dataset. The dashed line represents the ideal performance of the score.*

RAD-FRS; Radiomics *preoperative-Fistula Risk Score.*

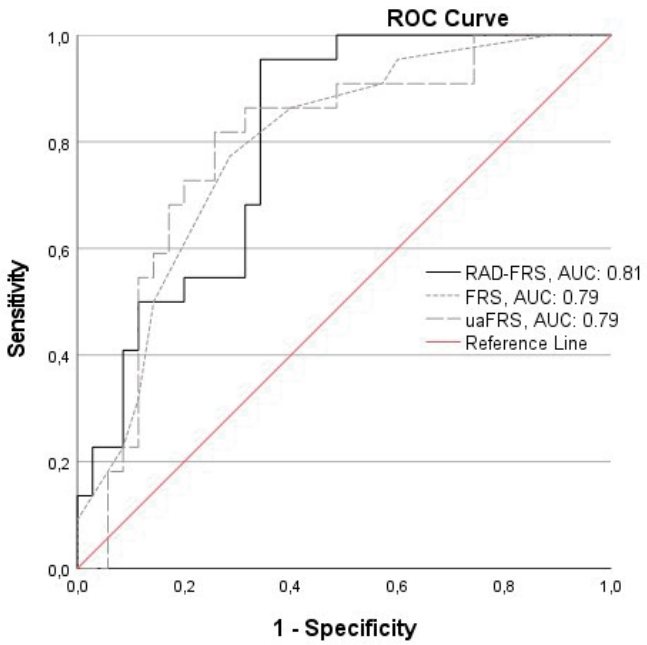
Verona dataset

In the Verona dataset, 52 (91%) CT scans were segmented in the (late) arterial phase and 5 (9.1%) in the portal venous phase. The AUC of the random forest model on the Verona dataset (n = 57) was 0.81 (95% CI: 0.69 – 0.92), with a sensitivity of 0.96, specificity of 0.66, PPV of 0.64, and NPV of 0.96. The calibration curve indicated that the model's performance was consistent with the observed data in Verona dataset and demonstrated an improvement in the consistency between estimated and observed probabilities compared to the Amsterdam test set. (Figure 5).

Comparison with the FRS and ua-FRS

The AUC of the RAD-FRS was 0.81 (95% CI: 0.60 – 0.92) on the Verona dataset, which was comparable to that of the FRS (AUC 0.79, 95% CI: 0.67 – 0.91, DeLong test: p = 0.85) and ua-FRS (AUC 0.79, 95% CI: 0.66 – 0.91, DeLong test: p = 0.87) (Figure 6).

Figure 6.



Comparison of *the ROC curves for RAD-FRS, ua-FRS, and FRS in predicting CR-POPF in the Verona dataset. The reference line represents the performance of a random guess.*
ROC; receiver *operating characteristics, RAD-FRS; Radiomics preoperative-Fistula Risk Score, ua-FRS; updated alternative Fistula Risk Score, FRS; Fistula Risk Score, CR-POPF; clinically relevant postoperative pancreatic fistula.*

Table 2. Predictive performances of the four machine learning models on the Amsterdam test set.

	AUC	Sensitivity	Specificity
Random Forest	0.90	0.99	0.67
Logistic regression	0.86	0.90	0.62
Support Vector Machine	0.81	0.98	0.53
K-nearest neighbor	0.80	0.80	0.75

AUROC; area *under the curve – receiver operating characteristics*

Discussion

The RAD-FRS is the first externally validated and publicly available prediction model for CR-POPF using radiomic features of preoperative CT scans, which performed well with similar external validity to two commonly used risk models for CR-POPF (AUC 0.81), the FRS (AUC 0.79) and ua-FRS (AUC 0.79).[6,7] Such a radiomics-based model could assist surgeons in counseling patients and making more informed treatment decisions in the preoperative setting already. Moreover, accurate risk stratification of CR-POPF in the preoperative setting can assist in determining the best surgical approach for high-risk or frail patients.

Four previous studies used radiomic features from the preoperative CT scan to predict CR-POPF. [9-13] These studies included a comparable number of patients after pancreatoduodenectomy, ranging from 100 to 250. In these studies, the AUC, sensitivity, and specificity were also comparable to the present study, with the AUC ranging from 0.76 to 0.81 in the Amsterdam test set (Table 3). However, these studies lacked external validation and did not include patients who received neoadjuvant or induction therapy patients. The present study did include patients with neoadjuvant therapy. However, its use was still relatively limited during the examined study period, especially in the Amsterdam dataset (14% and 58% of all patients and 27% and 75% of patients with pancreatic cancer, in the Amsterdam dataset and Verona dataset respectively).

Table 3. Comparison of the performance of the RAD-FRS model with other published studies that have used radiomic features derived from preoperative CT scans to predict the occurrence of CR-POPF.

All Risk	AUROC	Sensitivity	Specificity
RAD-FRS			
Amsterdam test set	0.90	0.99	0.67
Verona dataset	0.81	0.96	0.66
Capretti et al.[9]			
Internal test set	0.81	0.82	0.57
External test set	Missing	Missing	Missing
Kambakamba et al.[10]			
Internal test set	0.78	0.76	0.64
External test set	Missing	Missing	Missing
Lin et al.[11]			
Internal test set	0.79	0.63	0.78
External test set	Missing	Missing	Missing
Zhang et al.[12]			
Internal test set	0.76	Missing	Missing
External test set	Missing	Missing	Missing

CR-POPF; clinically *relevant postoperative pancreatic fistula, CT; computer tomography, RAD-FRS; Radiomics preoperative-Fistula Risk Score, AUROC; area under the curve – receiver operating characteristics.*

Although the RAD-FRS has the advantage over the FRS and ua-FRS in that it predicts CR-POPF preoperatively, it currently is less convenient to use than the FRS and ua-FRS due to the need for time-consuming manual segmentation of the pancreatic parenchyma. To make the RAD-FRS clinically applicable, auto-segmentation is required. Future work will involve developing an auto-segmentation model using deep learning techniques to fully automate the segmentation of the pancreatic parenchyma. Such a fully automated CR-POPF prediction model has the potential to be integrated into commonly used hospital radiology systems and CT scanners allowing for predictions at scanning time. These risk assessments could facilitate personalized treatment decisions by aiding surgeons in deciding on mitigation strategies and whether to proceed with a pancreatoduodenectomy. The initial results of a fully automatic CR-POPF prediction model developed by our group are promising. Another area of future work should involve the combination of radiomic features with preoperative clinical data, as demonstrated in a model developed by Lin et al., which exhibited excellent predictive performance.[11] Bhasker et al. also incorporated mesh-based volume to this combination model and demonstrated a significant improvement of the predictive performance

with this addition.[13] However, both the combination models developed by Lin et al. and Bhasker et al. require external validation and are currently not publicly available.

Several limitations of this study should be addressed. First, patients included in this study were managed following institutional practices. Therefore, differences in the peri-operative management, including drain management and administration of somatostatin analog, between both centers were conceivable. Despite these potential differences, the RAD-FRS performed well on the Verona dataset, showing its potential across heterogeneous populations. Second, this is a retrospective study with all its inherent limitations. A prospective study is needed to investigate the true value of the tool and to assess whether it can influence clinical decision-making. Third, most patients in both datasets underwent open surgery, and a subgroup analysis of open and minimally invasive pancreatic surgery is therefore lacking. As minimally invasive pancreatoduodenectomy is more commonly performed, future studies will need to validate the usefulness of the model in such patients for future applications.[18] Fourth, manual undersampling was used to create more equally sized groups based on the presence or absence of CR-POPF. However, manually annotating patients without CR-POPF would be futile, considering that undersampling would be one of the first preprocessing steps. Fifth, we did report calibration, but for clinical use, this would require correction for the undersampling. The discriminative ability of the RAD-FRS was promising as evaluated on the Verona dataset. Future work should enable absolute risk estimations in the total population.

The RAD-FRS can potentially improve patient prognosis and could eventually assist surgeons in making tailored treatment decisions. The externally validated RAD-RFS model and the code used for training, validation, and evaluation, are provided at www.pancreascalculator.com and www.github.com/PHAIR-Consortium/POPF-predictor.

References

1. Hackert T, Werner J, Büchler MW. Postoperative pancreatic fistula. *Surgeon*. 2011 Aug;9(4):211-7.
2. Grego A, Friziero A, Serafini S, Belluzzi A, Moletta L, Saadeh LM, Sperti C. Does Pancreatic Fistula Affect Long-Term Survival after Resection for Pancreatic Cancer? A Systematic Review and Meta-Analysis. *Cancers (Basel)*. 2021 Nov 19;13(22):5803.
3. Bootsma BT, Plat VD, van de Brug T, et al. Somatostatin analogues for the prevention of pancreatic fistula after open pancreatoduodenectomy: A nationwide analysis. *Pancreatology*. 2022 Apr;22(3):421-426.
4. Marchegiani G, Crippa S, Perri G, et al. Surgery for Intraductal Papillary Mucinous Neoplasms of the Pancreas: Preoperative Factors Tipping the Scale of Decision-Making. *Ann Surg Oncol*. 2022 May;29(5):3206-3214.
5. Marchegiani G, Perri G, Burelli A, et al. High-risk Pancreatic Anastomosis Versus Total Pancreatectomy After Pancreatoduodenectomy: Postoperative Outcomes and Quality of Life Analysis. *Ann Surg*. 2022 Dec 1;276(6):e905-e913.
6. Callery MP, Pratt WB, Kent TS, et al. A prospectively validated clinical risk score accurately predicts pancreatic fistula after pancreatoduodenectomy. *J Am Coll Surg*. 2013 Jan;216(1):1-14.
7. Mungroop TH, Klompmaker S, Wellner UF, et al. Updated Alternative Fistula Risk Score (ua-FRS) to Include Minimally Invasive Pancreatoduodenectomy: Pan-European Validation. *Ann Surg*. 2021 Feb 1;273(2):334-340.
8. Mayerhoefer ME, Materka A, Langs G, et al. Introduction to Radiomics. *J Nucl Med*. 2020 Apr;61(4):488-495.
9. Capretti G, Bonifacio C, De Palma C, et al. A machine learning risk model based on preoperative computed tomography scan to predict postoperative outcomes after pancreatoduodenectomy. *Updates Surg*. 2022 Feb;74(1):235-243.
10. Kambakamba P, Mannil M, Herrera PE, et al. The potential of machine learning to predict postoperative pancreatic fistula based on preoperative, non-contrast-enhanced CT: A proof-of-principle study. *Surgery*. 2020 Feb;167(2):448-454
11. Lin Z, Tang B, Cai J, et al. Preoperative prediction of clinically relevant postoperative pancreatic fistula after pancreaticoduodenectomy. *Eur J Radiol*. 2021 Jun;139:109693.
12. Zhang W, Cai W, He B, et al. A radiomics-based formula for the preoperative prediction of postoperative pancreatic fistula in patients with pancreaticoduodenectomy. *Cancer Manag Res*. 2018 Nov 28;10:6469-6478.
13. Bhasker N, Kolbinger FR, Skorobohach N, et al. Prediction of clinically relevant postoperative pancreatic fistula using radiomic features and preoperative data. *Sci Rep*. 2023 May 9;13(1):7506.
14. Vandenbroucke JP, von Elm E, Altman DG, Get al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med*. 2007 Oct 16;4(10):e297
15. Bassi C, Marchegiani G, Dervenis C, et al. The 2016 update of the International Study Group (ISGPS) definition and grading of postoperative pancreatic fistula: 11 Years After. *Surgery*. 2017 Mar;161(3):584-591.
16. Fedorov A., Beichel R., Kalpathy-Cramer J, et al. 3D Slicer as an Image Computing Platform for the Quantitative Imaging Network. *Magn Reson Imaging.* 2012 Nov;30(9):1323-41.

17. van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G. H., Fillon-Robin, J. C., Pieper, S., Aerts, H. J. W. L. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* 2017;77(21), e104–e107.
18. Van der Heijde N, Vissers FL, Manzoni A, et al. Use and outcome of minimally invasive pancreatic surgery in the European E-MIPS registry. *HBP*. 2022.

Supplementary Table 1. Acquisition and reconstruction parameters of the CT scans of the Amsterdam and Verona datasets.

Parameter	Amsterdam dataset	Verona test set
Pitch	0.6 – 1.388	0.5 – 0.891
X-ray tube current	17 – 183 mA	200 – 250mA
kVp	90 – 120 kVp	90 – 120 kVp
Tube rotation speed	0.4 – 0.7 ms	0.4 – 0.75 ms
Slice thickness	1.0 – 3.0mm	1.0 – 2.0 mm
Width	400 - 934	300 – 400
Height	300 - 975	300 - 400
Machine manufacturer	SIEMENS, GE Healthcare	Philips Brilliance 64 Siemens Somatom Definition Edge 128

Supplementary Table *1. Upper value is the maximum and lower value is the minimum of all values present in the dataset for that specific parameter.*
CT; computed *tomography, kVp; kilovoltage peak*

Supplementary Table 2. List of radiomic features extracted from Pyradiomics.

Feature class	Number of features
First order statistics	19 features
Shape-based (3D)	16 features
Shape-based (2D)	10 features
Gray Level Co-occurrence Matrix	24 features
Gray Level Run Length Matrix	16 features
Gray Level Size Zone Matrix	16 features
Neighboring Gray Tone Difference Matrix	5 features
Gray Level Dependence Matrix	14 features

Supplementary Table *2 lists all the extracted radiomic features. These features are extracted using Pyradiomics. A description of each feature can be found on: https://pyradiomics.readthedocs.io/en/latest/features.html*

11

Chapter 11

Artificial Intelligence-based fully automatic prediction of post pancreatic fistula from pre-operative CT scans of patients undergoing pancreatoduodenectomy

Jacqueline I. Bereska*
Erik W. Ingwersen*
Alberto Balduzzi
Boris V. Janssen
Marc G. Besselink
Geert Kazemier
Giovanni Marchegiani
Giuseppe Malleo
Henk A. Marquering
C. Yung Nio
Riccardo de Robertis
Roberto Salvia
Ewout W. Steyerberg
Jaap Stoker
Femke Struik
Freek Daams#
Inez M. Verpalen#
On behalf of the Pancreatobiliary and Hepatic Artificial Intelligence Research (PHAIR) consortium

Submitted

**These authors contributed equally to this manuscript*
#Shared last authorship

Abstract

Purpose: To develop and externally validate a pre-operative Artificial Intelligence (AI)-based Fistula Risk Score (AI-FRS) for predicting post-operative pancreatic fistula (POPF) in patients undergoing pancreatoduodenectomy.

Materials and Methods: In this retrospective study, we developed an AI model to segment the pancreas in contrast-enhanced CTs using 613 CTs of 467 patients with pancreatic abnormalities and 50 control patients. We developed the AI-FRS to predict POPF using a random forest model and radiomic features extracted from the pancreas segmentations in 118 pre-operative CT scans of 118 adult patients scheduled for pancreatoduodenectomy at a single center in the Netherlands (2013-2018). We evaluated the AI-FRS on an internal test set comprising 15 patients from the center in the Netherlands and an external test set comprising 50 patients from a single center in Italy (2020-2021) using the area under the receiver operating characteristic curve (AUROC). We compared the performance of the AI-FRS to three existing fistula risk scores (FRS, ua-FRS, RAD-FRS).

Results: In the internal data set, 50 (42%) patients developed POPF. In the external test set, 20 (40%) patients developed POPF. The AUROC of the AI-FRS was 0.86 (95% CI: 0.71 – 1.00) on the internal test set and 0.83 (95% CI: 0.72 – 0.94) on the external test set. The AI-FRS reached a higher AUROC than the three existing fistula risk scores. However, this difference was not statistically significant.

Conclusions: The AI-FRS accurately predicts the risk of POPF after pancreatoduodenectomy based on radiomic features from automatic pancreas segmentations. After further validation, this model may allow for automatic pre-operative prediction of POPF.

Introduction

Post pancreatic fistula (POPF), the leading complication following pancreato-duodenectomy, is associated with further morbidity, prolonged hospital stay, increased resource utilization, mortality, and worse oncologic outcome.[1,2] An accurate prediction could aid surgeons in making more informed treatment decisions, including the implementation of perioperative mitigation strategies, forgoing surgery in patients with low-risk pancreatic cysts, and considering alternative surgical approaches, such as total pancreatectomy in patients with high-risk pancreatic anastomoses.[3]

Several prediction models for POPF have been developed with good predictive performances. However, most of these models rely on intra-, or post-operatively extracted features, limiting their clinical relevance.[4-11] In contrast, prediction models based on radiomic features allow pre-operative predictions and have shown promising results.[9-12] Our research group recently developed a prediction model using radiomic features from manually segmented computed tomography (CT) scans, the radiomic-based pre-operative Fistula Risk Score (RAD-FRS). Nevertheless, these methods rely on time-consuming manual segmentations, assessments, or corrections performed by radiologists.[7,13,14] A fully automated segmentation model could make manual segmentations obsolete and is an essential step toward implementing reliable radiomics-based prediction models in daily clinical practice.

This study aimed to develop and externally validate an artificial intelligence (AI)-based Fistula Risk Score (AI-FRS) in patients undergoing pancreatoduodenectomy using radiomic features extracted from AI-generated segmentations of the pancreas on pre-operative CT scans. In addition, this study compared the predictive performance of the AI-FRS to several existing fistula risk scores, including the RAD-FRS constructed with manual pancreas segmentations, the Fistula Risk Score (FRS), and the updated alternative Fistula Risk Score (ua-FRS).[15,16]

Material and Methods

This retrospective study obtained delegated research ethics approval from the Medical Ethics Review Committee of the Amsterdam UMC (references: 2022.0241, CPB/aj/345654, MdV/ss/METC/68052). The Ethics Committee of Verona and Rovigo Provinces approved using data from the validation cohorts (PAD-R, 1101CESC). General informed consent was obtained from all patients. All research was performed in accordance with relevant guidelines and/or regulations.

Datasets

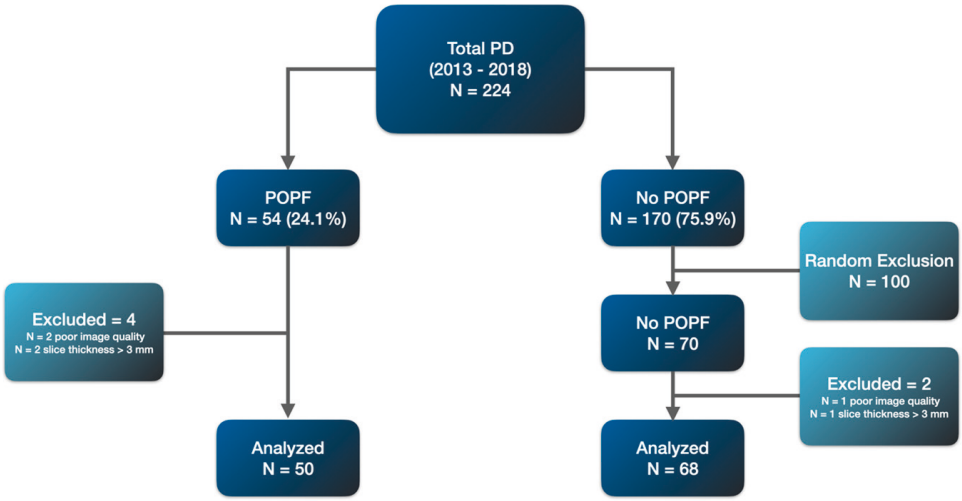
We used three datasets for this study containing 783 CTs of 642 patients with pancreatic abnormalities and 50 CTs of 50 control patients obtained between 2013 and 2021. Dataset *A* consisted of four sub-datasets (*A.1*, *A.2*, *A.3*, and *A.4*) and was used to train the AI segmentation model. The characteristics of dataset A are described in Table 1. Dataset *B* was an internal dataset used for developing the prediction model, while dataset *C* was an external dataset used for validating the model. For the construction of internal dataset *B*, adult patients who underwent a pancreatoduodenectomy between January 2013 to 2018 in the Amsterdam University Medical Centers (UMC) were screened for the occurrence of POPF (Figure 1). For the construction of the external dataset *C*, adult patients from Verona University Hospital who underwent a pancreatoduodenectomy between 2020 and 2021 were screened for the occurrence of POPF (Figure 2).

Table 1: *Characteristics for datasets A.1, A.2, A.3, and A.4.*

Characteristic	Dataset A.1	Dataset A.2	Dataset A.3	Dataset A.4
Number of scans	232	50	50	281
Number of patients	136	50	50	281
Timeframe	2013 – 2020	2019 – 2021	2013 – 2017	Not available
CT phase	Late arterial phase	Late arterial phase	Late arterial phase	Portal venous phase
Pancreatic abnormalities	Resectable and borderline resectable PDAC	Locally advanced PDAC	No pancreatic abnormalities	PDAC, intraductal mucinous neoplasms, pancreatic neuroendocrine tumors
Medical center	Amsterdam UMC	Amsterdam UMC	Amsterdam UMC	Memorial Sloan Kettering Cancer Center

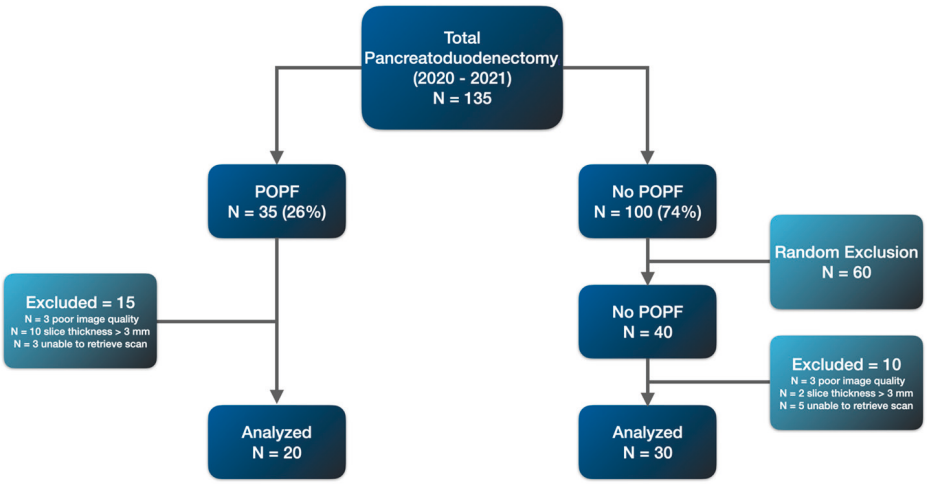
Legend: PDAC = *pancreatic ductal adenocarcinoma.*

Figure 1: *Flowchart showing the selection process of the internal dataset (dataset B).*



Legend: A *total of 100 patients were excluded in the group without POPF to create more equally sized groups based on the presence or absence of POPF. Further exclusion criteria were poor image quality (n=3) and slice thickness larger than 3 mm (n=3). A total of 50 patients with POPF and 68 patients without POPF were found eligible for analysis.*
PD = *pancreatoduodenectomy, N = number, POPF = post pancreatic fistula, CT = computer tomography.*

Figure 2: *Flowchart showing the selection process of the external dataset (dataset C).*



Legend: A *total of 60 patients were excluded in the group without POPF to create more equally sized groups based on the presence or absence of POPF. Further exclusion criteria were poor image quality (n=3) and slice thickness larger than 3 mm (n=3). A total of 20 patients with POPF and 30 patients without POPF were found eligible for analysis.*
N = *number, POPF = post pancreatic fistula, CT = computer tomography.*

We excluded patients in whom no pancreatojejunostomy could be constructed, CT slices of ≤3.0 mm, poor scan quality, or without an available CT scan within 60 days prior to surgery. From the screened patients without POPF in dataset *B*, we further randomly excluded 100 patients to create more equally sized groups based on the presence or absence of POPF. After exclusion, dataset *B* consisted of 50 patients with POPF and 68 patients without POPF, while dataset *C* consisted of 20 patients with POPF and 30 patients without POPF.

Data acquisition and outcome

The study collected patient demographics, tumor characteristics, intraoperative information from electronic health records, and contrast-enhanced CT scans from the picture archiving and communication system. The primary outcome was a clinically relevant POPF, defined as grade B or C according to the 2016 International Study Group of Pancreatic Fistula definition.[17]

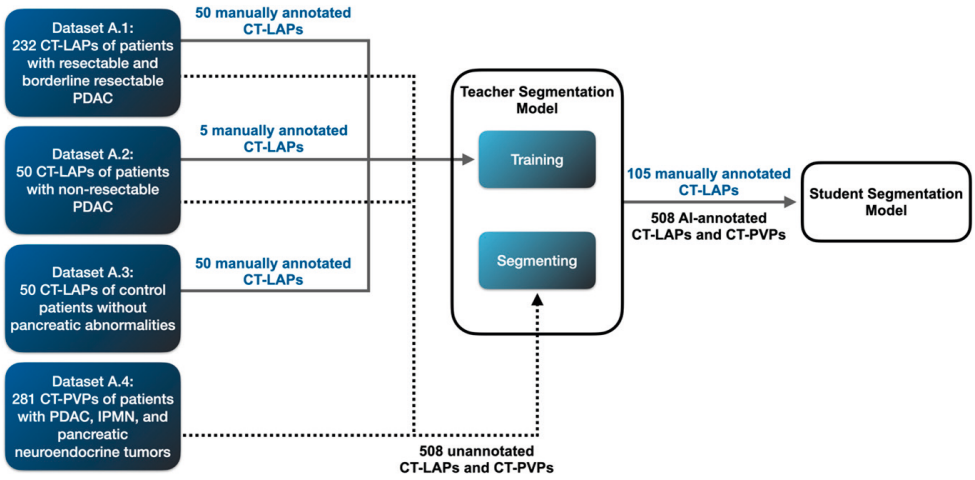
Model implementation

The overall workflow of our framework consisted of 1) automatically segmenting the pancreas and surrounding abdominal structures, 2) extracting radiomic features from the mask of the pancreas, and 3) predicting the occurrence of POPF.

Segmenting the Pancreas

To develop the pancreas segmentation model, we used a self-learning-based framework consisting of a teacher and a student model; see Figure 3 for a schematic representation. We trained the teacher model with manual segmentations of the pancreas and surrounding anatomical structures in 105 late arterial phase CT scans (CT-LAPs) from 55 patients from dataset *A.1* and *A.2* and 50 control patients from dataset *A.3* (see appendix S1 for details) (18). Subsequently, the trained teacher model segmented the remaining data consisting of 508 CTs of patients from datasets A.1, A.2., and A.4. We then trained the student model using the resulting 508 AI-generated segmentations generated by the teacher model along with the 105 manual segmentations. Finally, the student model segmented the CTs in both the internal dataset B and external dataset C. We trained the teacher and student model using a two-stage 5-fold 3D U-Net cascade requiring approximately one day per fold on an NVIDIA A100 GPU (NVIDIA, Santa Clara, California). [19,20]

Figure 3: *The proposed learning framework for pancreas segmentation on contrast-enhanced CT scans.*



Legend: CT-LAP = *late arterial phase CT scan, CT-PVP = portal venous phase CT scan.*

Extracting Radiomic Features

We extracted radiomic features from each AI-segmented pancreas in the internal dataset *B using Pyradiomics version 3.0* (http://github.com/radiomics/pyradiomics).[16] Subsequently, we randomly split the patients into 25 training (90%) and test sets (10%). Finally, we normalized, reduced, and augmented the radiomic features of the training and test sets (see Appendix S1 for details).

Predicting the Occurrence of POPF

We optimized four machine learning (ML) classifiers - support vector machine, logistic regression, k-nearest neighbor, and random forest - and trained these on the radiomic features of the training sets. We used a 10-fold cross-validation grid search optimizing area under the receiver operating characteristic curve (AUROC) to find the hyperparameters and fit all four ML classifiers on the 25 training set splits. We selected the best-performing model with regard to AUROC on the internal test set as our final model.

Performance Assessment and Statistical Analysis

We performed statistical analysis using Python version 3.7 (Python Software Foundation, Wilmington, DE) and R-Studio version 2022.07.1 (Posit PBC, Boston, Massachusetts). We assessed the performance of our final model on both the internal test set (a subset of dataset *B)* and the external test set *C* using the AUROC, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). We compared the performance of this model to the FRS, ua-FRS, and a model using radiomic features extracted from manual segmentation masks of the same dataset (RAD-FRS)[6,7] A p-value <0.05 was considered statistically significant.

Results

Patient characteristics

Overall, 783 CTs of 642 patients were included for segmentation and prediction model development and validation. The internal dataset (dataset B) comprised 118 patients, with 50 (42%) developing POPF. The median age in this cohort was 68 years (IQR: 20, Q1: 59, Q2: 79 years), the mean BMI was 25.7 (SD: 3.9) kg/m^2, and 16 (14%) patients received neoadjuvant therapy. The external test set (dataset C) included 50 patients, with 20 (40%) developing POPF. The median age was 66 years, the mean BMI was 24.9 kg/m2 (SD: 3.9 kg/m2), and 28 (56%) patients received neoadjuvant therapy. The clinical characteristics for datasets *B* and *C* are presented in Table 2. Table 3 details the reconstruction and acquisition parameters used to obtain the CTs from all three datasets, *A*, *B*, and *C*.

Table 2: *Clinical characteristics of patients in datasets B and C.*

Characteristic	Dataset B N=118	Dataset C N=50
Sex, n(%), female	43 (36.4)	22 (22.0)
Median age at diagnosis, years	68	66
BMI, mean (dev)	25.7 (3.9)	24.9 (3.9)
ASA classification, n(%)		
1 & 2	91 (77.1)	36 (72.0)
3 & 4	27 (22.9)	14 (28.0)
Diabetes Mellitus, n(%)	20 (17.1)	9 (18.0)
Neoadjuvant therapy, n(%)	16 (13.6)	28 (56.0)
Laparoscopic procedure, n(%)	18 (15.3)	0
Pancreatic duct size in mm, median (IQR)	3 (2-4)	4 (3-5)
Soft pancreas, n(%)	76 (64.4)	29 (58.0)
Intraoperative blood loss in ml, median (IQR)	400 (380 – 420)	500 (350 – 793)
Length of surgical procedure (minutes), median (IQR)	325 (233 – 379)	390 (319 – 463)
Pathology		
Pancreatic ductal adenocarcinoma	56 (48%)	32 (64%)
Distal cholangiocarcinoma	22 (19%)	3 (6%)
Ampullary carcinoma	13 (11%)	3 (6%)
Duodenal carcinoma	10 (8.5%)	3 (6%)
Cystic neoplasm (IPMN, serous cysts, MCN)	7 (5.9%)	3 (6%)
Inflammation / pancreatitis	5 (4.2%)	0
Neuro-endocrine tumor	4 (3.4%)	6 (12%)
Other	1 (0.8%)	0

Legend: IQR = *interquartile range, BMI = body mass index, ASA = American Society of Anesthesiologists, dev = standard deviation, IPMN = intraductal papillary mucinous neoplasm, MCN = mucinous cystic neoplasm*

Table 3: *Reconstruction and acquisition parameters used to obtain the CTs in datasets A, B, and C.*

Parameter	Dataset A N=613	Dataset B N=118	Dataset C N=50
Pitch/table speed	0.5 – 1.9	0.6 – 1.388	0.5 – 0.891
X-Ray tube current	50 – 2551 mA	17 – 183 mA	200 – 250 mA
kVp	70 – 120 kVp	90 – 120 kVp	90 – 120 kVp
Tube rotation speed	0.4 – 0.8 ms	0.4 – 0.7 ms	0.4 – 0.75 ms
Slice thickness	0.625 – 5.0 mm	1.0 – 3.0 mm	1.0 – 2.0 mm
Width X Height	512 – 1189 X 512 – 1189	400 - 934 X 300 - 975	300 – 400 X 300 - 400
Machine Manufacturer and Type	Siemens: Somatom Force/Definition AS+/ Edge/Flash, Aquilion ONE, iCT 256, Sensation64, Brilliance64 GE Healthcare: LightSpeed16, BrightSpeed Philips: Optima CT660	Siemens: Somatom Force/ Definition AS+/Edge/ Flash, Aquilion ONE, iCT 256, Sensation64, Brilliance64, Biograph 40 GE Healthcare: Discovery CT750 HD	Philips: Brilliance 64 Siemens: Somatom Definition Edge 128

Performance on the internal test set

In the internal dataset (dataset B), the AI segmentation model segmented the pancreas in 103 (87%) CT-LAPs and 15 (13%) portal venous phase CT scans (CT-PVPs). For each patient, a total of 108 radiomics features were extracted, of which five were ultimately used after feature reduction. The best-performing model with regard to AUROC on the internal test was a random forest model with an AUROC of 0.86 (95% CI: 0.71 – 1.00). This model had a sensitivity of 0.90, specificity of 0.77, PPV of 0.40, and NPV of 0.90 in the internal test set.

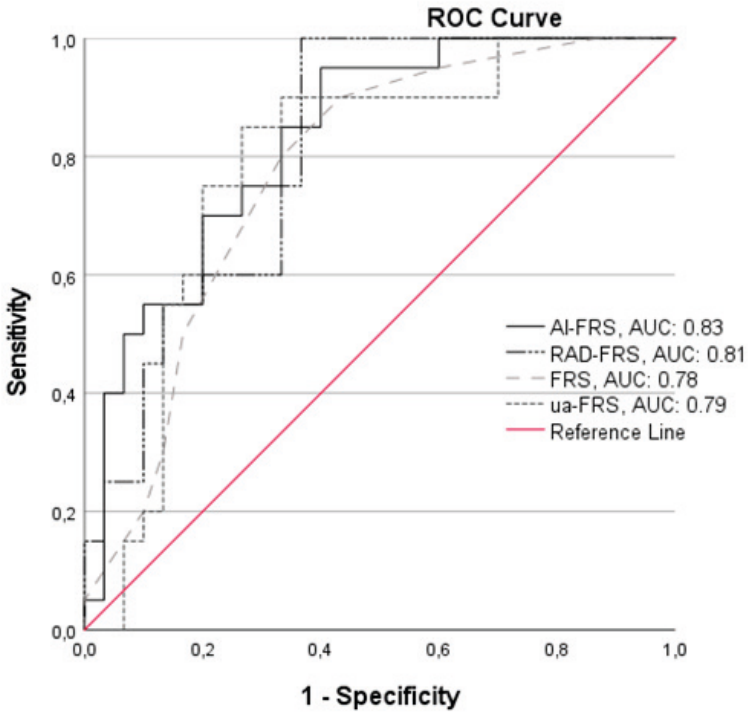
Performance on the external test set

In the external test set (dataset C), the AI segmentation model segmented the pancreas in 46 (92%) CT-LAPs and 4 (8%) CT-PVPs. The AUROC of the random forest model on the external test set was 0.83 (95% CI: 0.72 – 0.94), with a sensitivity of 0.85, specificity of 0.67, PPV of 0.63, and NPV of 0.87.

Comparison with the RAD-FRS, the FRS, and ua-FRS

The AUROC of the AI-FRS using the random forest model was comparable to that of the RAD-FRS (AUC 0.81, 95% CI: 0.70 – 0.93, DeLong test: p=0.83), FRS (AUC 0.78, 95% CI: 0.65 – 0.91, DeLong test, p=0.52), and ua-FRS (AUC 0.79, 95% CI: 0.65 – 0.92, DeLong test: p=0.60). (Figure 4, Table 4)

Figure 4: *Comparison of the ROC curves for AI-FRS, RAD-FRS, ua-FRS, and FRS in predicting POPF in the external test set (dataset C).*



Legend: The *reference line represents the performance of a random guess. ROC = receiver operating characteristics, AI-FRS = Artificial Intelligence-based Fistula Risk Score, RAD-FRS = Radiomics-based pre-operative-Fistula Risk Score, ua-FRS = updated alternative Fistula Risk Score, FRS = Fistula Risk Score, POPF = post pancreatic fistula.*

Table 4: *Performances of the RAD-FRS, the FRS, and ua-FRS in the external test set (dataset C) as compared to AI-FRS.*

Model	AUROC (95% CI)	Sensitivity	Specificity
AI-FRS	0.83 (0.72 – 0.94)	0.85	0.67
RAD-FRS	0.81 (0.70 – 0.93)	0.95	0.63
FRS	0.78 (0.65 – 0.91)	0.80	0.67
ua-FRS	0.79 (0.65 – 0.92)	0.90	0.67

Legend: AUROC *= area under the curve receiver operating characteristic curve, CI = confidence interval, AI-FRS = Artificial Intelligence-based Fistula Risk Score, RAD-FRS = radiomic-based pre-operative-Fistula Risk Score, FRS = Fistula Risk Score, ua-FRS = updated alternative Fistula Risk Score.*

Discussion

The AI-FRS, based on 168 pre-operative CT scans from two centers, is the first fully automatic and externally validated AI-based framework for predicting POPF after pancreatoduodenectomy. The results of our analysis demonstrate that radiomic features extracted from AI-segmented scans can be used to predict POPF with high predictive performance on both the internal and external test sets. Due to the fully automatic nature of our method, the AI-FRS has the potential to be integrated into commonly used hospital radiology systems allowing for predictions at scanning time. Accurate predictions could assist surgeons in counseling and making treatment decisions for high-risk or frail patients.

To date, at least four studies using radiomic features extracted from pre-operative CT scans to predict POPF have been published.[9,10,12,14] These studies included a comparable number of patients after pancreatoduodenectomy, ranging from 100 to 250. The AUC, sensitivity, and specificity reported in these studies were lower, with AUC ranging from 0.76 to 0.81 in the respective internal test sets. Only one study employed a segmentation model to segment the pancreas for radiomic feature extraction.[14] However, experienced abdominal radiologists manually checked and, if necessary, corrected the resulting AI-generated segmentations, preventing fully automatic predictions. In addition, despite promising initial results on their respective internal test set, the authors did not validate their model's performance on an external test set.

Previously, a radiomics-based Fistula Risk Score (RAD-FRS) using a similar prediction model with radiomic features extracted from manual segmentations of the pancreas was proposed. We found that the AI-FRS based on AI-generated segmentations performed similarly to the RAD-FRS on the same previously withheld external test set (AUC = 0.81). Both the AI-FRS and the RAD-FRS were trained and evaluated on the same internal and external datasets differing only in the segmentations used to extract the radiomic features. However, the AI-FRS used automatic segmentations to extract the radiomic features, which allowed for a more standardized and less time-consuming process, while the RAD-FRS relied on manual segmentations that required expert input and additional time for precise delineation. This fundamental difference in the segmentation process underscores the potential for AI-FRS to enhance efficiency without sacrificing predictive accuracy, offering a distinct advantage in clinical settings where timely risk assessment is critical.

Several limitations of this study should be addressed. First, while the AI-FRS forgoes time-consuming manual annotations, additional computational resources, at least 8 GB GPU, are needed to create the AI-generated segmentations, potentially limiting its

application in less equipped medical centers. Second, we did not report calibration, which would require correction for the undersampling performed during the preprocessing step. The discriminative ability of the AI-FRS was promising as evaluated on the external test set. Thus, future work should enable absolute risk estimations in the total population. Third, this is a retrospective study with all its known limitations. A prospective study is deemed necessary before clinical application to investigate the safety and effectiveness of the AI-FRS in a large-scale, comparative trial.[9-11] The objective of this study should be to provide more substantial evidence for translating the AI-FRS from a research application into a clinically relevant tool. Third, we only included radiomic features in the construction of our model. Future work should evaluate multi-modal models constructed with additional feature types, such as pre-operative clinical data and deep learning features.

The AI-FRS is the first fully automatic and externally validated AI-based framework to predict POPF using pre-operative CT scans, allowing for predictions at scanning time. The AI-FRS achieved a similar AUROC to the three existing fistula risk scores. After further validation, this model may allow for automatic pre-operative prediction of POPF.

Acknowledgements

We thank Anonymous MSc, Anonymous MSc, and Anonymous MSc, for their time and commitment to creating high-quality segmentations of abdominal organs and vessels within the scope of their Master's thesis which greatly assisted our research.

References

1. Bonaroti JW, Zenati MS, Al-abbas AI, Rieser CJ, Zureikat AH, Hogg ME, et al. Impact of post-operative pancreatic fistula on long-term oncologic outcomes after pancreatic resection. HPB. 2021 Aug 1;23(8):1269–76.
2. Marchegiani G, Andrianello S, Salvia R, Bassi C. Current Definition of and Controversial Issues Regarding Postoperative Pancreatic Fistulas. Gut Liver. 2019 Mar;13(2):149–53.
3. Grego A, Friziero A, Serafini S, Belluzzi A, Moletta L, Saadeh LM, et al. Does Pancreatic Fistula Affect Long-Term Survival after Resection for Pancreatic Cancer? A Systematic Review and Meta-Analysis. Cancers. 2021 Jan;13(22):5803.
4. Han IW, Cho K, Ryu Y, Shin SH, Heo JS, Choi DW, et al. Risk prediction platform for pancreatic fistula after pancreatoduodenectomy using artificial intelligence. World J Gastroenterol. 2020 Aug 14;26(30):4453–64.
5. Shi Y, Gao F, Qi Y, Lu H, Ai F, Hou Y, et al. Computed tomography-adjusted fistula risk score for predicting clinically relevant postoperative pancreatic fistula after pancreatoduodenectomy: Training and external validation of model upgrade. eBioMedicine. 2020 Dec 1;62:103096.
6. Yoon SJ, Kwon W, Lee OJ, Jung JH, Shin YC, Lim CS, et al. External validation of risk prediction platforms for pancreatic fistula after pancreatoduodenectomy using nomograms and artificial intelligence. Ann Surg Treat Res. 2022 Mar 4;102(3):147–52.
7. Kambakamba P, Mannil M, Herrera P, Linecker M, Müller P, Kümmerli C, et al. Machine learning based texture analysis predicts postoperative pancreatic fistula in preoperative non-contrast enhanced computed tomography. HPB. 2020 Jan 1;22:S384.
8. Xingjun G, Feng Z, Meiwen Y, Jianxin J, Zheng H, Jun G, et al. A score model based on pancreatic steatosis and fibrosis and pancreatic duct diameter to predict postoperative pancreatic fistula after Pancreatoduodenectomy. BMC Surgery. 2019 Jul 3;19(1):75.
9. Zhang W, Cai W, He B, Xiang N, Fang C, Jia F. A radiomics-based formula for the preoperative prediction of postoperative pancreatic fistula in patients with pancreaticoduodenectomy. Cancer Manag Res. 2018 Nov 28;10:6469–78.
10. Capretti G, Bonifacio C, De Palma C, Nebbia M, Giannitto C, Cancian P, et al. A machine learning risk model based on preoperative computed tomography scan to predict postoperative outcomes after pancreatoduodenectomy. Updates Surg. 2022 Feb 1;74(1):235–43.
11. Lin Z, Tang B, Cai J, Wang X, Li C, Tian X, et al. Preoperative prediction of clinically relevant postoperative pancreatic fistula after pancreaticoduodenectomy. European Journal of Radiology. 2021 Jun 1;139:109693.
12. Kambakamba P, Mannil M, Herrera P, Linecker M, Müller P, Kümmerli C, et al. Machine learning based texture analysis predicts postoperative pancreatic fistula in preoperative non-contrast enhanced computed tomography. HPB. 2020 Jan 1;22:S384.
13. Liu KL, Wu T, Chen PT, Tsai YM, Roth H, Wu MS, et al. Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: a retrospective study with cross-racial external validation. The Lancet Digital Health. 2020;2(6):e303–13.
14. Lin Z, Tang B, Cai J, Wang X, Li C, Tian X, et al. Preoperative prediction of clinically relevant postoperative pancreatic fistula after pancreaticoduodenectomy. European Journal of Radiology. 2021 Jun 1;139:109693.
15. Mungroop TH, Klompmaker S, Wellner UF, Steyerberg EW, Coratti A, D'Hondt M, et al. Updated Alternative Fistula Risk Score (ua-FRS) to Include Minimally Invasive Pancreatoduodenectomy: Pan-European Validation. Annals of Surgery. 2021 Feb;273(2):334.

16. Callery MP, Pratt WB, Kent TS, Chaikof EL, Vollmer CM. A Prospectively Validated Clinical Risk Score Accurately Predicts Pancreatic Fistula after Pancreatoduodenectomy. Journal of the American College of Surgeons. 2013 Jan 1;216(1):1–14.
17. Bassi C, Marchegiani G, Dervenis C, Sarr M, Abu Hilal M, Adham M, et al. The 2016 update of the International Study Group (ISGPS) definition and grading of postoperative pancreatic fistula: 11 Years After. Surgery. 2017 Mar 1;161(3):584–91.
18. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magnetic resonance imaging. 2012;30(9):1323–41.
19. Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods. 2021;18(2):203–11.
20. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer; 2015. p. 234–41.
21. Alves N, Schuurmans M, Litjens G, Bosma JS, Hermans J, Huisman H. Fully Automatic Deep Learning Framework for Pancreatic Ductal Adenocarcinoma Detection on Computed Tomography. Cancers. 2022;14(2):376.

APPENDIX S1

METHODS

Segmentation Model Training Data Preparation

Seven trained observers manually segmented the pancreas and surrounding anatomical structures in 105 late arterial phase CT scans from 105 patients from dataset *A* using 3D slicer version 4.11.20210226 (18). Including segmentations of surrounding abdominal structures during segmentation model training has been shown to increase performance, thus the observers also segmented the pancreatic tumor (if present), duodenum, liver, gallbladder, bile duct (and, if present, bile duct stents), celiac trunk, hepatic artery, superior mesenteric artery, superior mesenteric vein, portal vein, splenic artery, splenic vein, aorta, and inferior vena cava (21). The pancreatic tumors were segmented by one of a team of three abdominal radiologists (YN: 27 years' experience, FS: 3 years' experience, and MK: 6 years' experience) subspecialized in pancreatic pathologies at the Amsterdam UMC. After the pancreatic tumor was segmented, the pancreas and the remaining structures were segmented by a Ph.D. student (JB) or one of three master's students (AT, FR, and KV). These segmentations were performed under the instruction and supervision of a radiology resident (IV), who also checked the final segmentations. Figure S1 depicts a segmentation example.

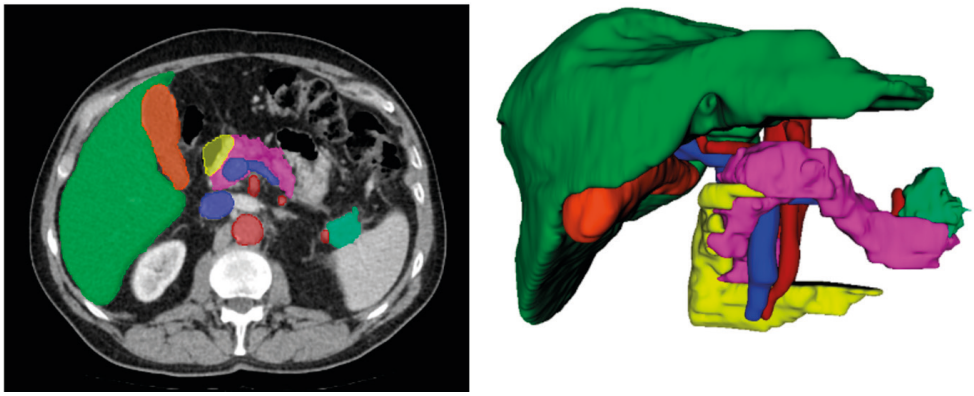
Normalization and Feature Reduction

We independently normalized the radiomic features of the training and the test sets using the MinMax scaler. Further, we serially applied two feature reduction methods to the training sets to select the most predictive features for each split: removal of features with a variance of less than 0.001 and discarding of features with an importance of zero using LASSO feature selection. Before training, we oversampled the minority classes in the training sets with a sampling strategy of 0.9 and subsequently undersampled the majority classes. Finally, we added random Gaussian noise with a mean of zero and a variance of 0.2.

Statistical Analysis

We reported the AUROC of the AI-FRS, RAD-FRS, FRS, and ua-FRS with a 95% confidence interval (95% CI). We employed the maximum Youden's J value from the AUROC to identify the cut probability where prediction discrimination was optimized between sensitivity and specificity.

Figure S1, *Example of manual segmentations of anatomical structures within a late arterial phase computed tomography scan.*



Legend: *Turquoise = PDAC tumor, red = arteries (aorta, celiac trunk, hepatic artery, splenic artery, and superior mesenteric artery), dark blue = veins (inferior vena cava, portal vein, splenic vein, and superior mesenteric vein), pink = pancreas, yellow = duodenum, green = liver, orange = gallbladder*

11

12

Chapter 12

A1Check: the External Validation of a Machine Learning Model Predicting Colorectal Anastomotic Leakage (A1Check) – study protocol (NCT05810207)

Erik W. Ingwersen
Wessel T. Stam
Freek Daams

ClinicalTrials.gov. Study: NCT05810207

Study Overview

Brief Summary:

Anastomotic leakage is a severe complication that can arise following a colorectal resection. It impairs both the short- and long term outcome, and negatively influences cancer recurrence rates.[1-4] Its detrimental effects resound in healthcare costs of a patient after anastomotic leakage, 71.978,- EUR, versus patients with an uncomplicated course, 17.647,- EUR.[5] Despite multiple innovation within the field of colorectal surgery, incidence of colorectal anastomotic leakage did not reduce in the past decade.[6] Mitigation strategies as prehabilitation, intraoperative optimization, selective bowel decontamination and reconstruction techniques are promising but do not completely eliminate the risk of leakage.[7-9] The only true prevention of colorectal anastomotic leakage is the omission of an anastomosis and implies an ostomy, which in itself has a negative impact on quality of life. A stoma is associated with stoma-related morbidity and should therefore be avoided in patient who do not need it. Predicting anastomotic leakage intra-operatively, just before the construction of the anastomosis, may offer a solution. A stoma will then only be constructed in those at high risk of anastomotic leakage. Currently, there are prediction models for anastomotic leakage base on conventional multivariate logistic regression analysis, however these are not useful for clinical practice due to suboptimal results. Machine learning algorithms on the other hand, take well into account the multifactorial nature of complications and might thus be able to predict anastomotic leakage more accurately.[10] The machine learning models we created proved to be well capable of making these predictions. This model was created based on the a database containing both pre- and intra-operative data from 2,483 patients. Before these models can be used in daily practice, external validation is essential. Our models should be tested on unseen data from patients treated in centers that were not previously involved in the database that was used to train the model in order to achieve high reproducibility. Our hypothesis is that with our model we can accurately predict anastomotic leakage intra-operatively during colorectal surgery.

Detailed Description:

Study procedure

The aim of the study is to externally validate a machine learning model predicting anastomotic leakage. The prediction model that will be externally validated, is developed on a prospective database.[11] This database contained data of 2,483 colorectal cancer patients who were operated between January 2016 and April 2021 in 14 hospitals, both rural and academic in 4 different countries (the Netherlands, Italy, Belgium, Australia). Some 189 patients (7.6%) developed colorectal anastomotic leakage. The models predicted risk of colorectal anastomotic leakage intraoperatively, just prior to the

construction of the anastomosis, using a total of 31 variables. These variables contain both preoperatively available data and the variables regarding the intraoperative condition of the patient. The models were internally validated using 10-fold cross validation and subsequently tested on 20% of unseen data of the database. The area under the curve – receiver operating characteristics (AUROC) of the best performing ML model on the test set was 0.84, with a sensitivity of 0.86, specificity of 0.78, a positive predictive value of 0.24 and a negative predictive value of 0.99. This means that this model have near perfect prediction of "no anastomotic leakage". Although the positive predictive value is low (24%) and could lead to overtreatment (i.e. unnecessary stomas) when adopted in every patient, it enables focused postoperative observation in a limited amount of patients.

During this prospective simulation study there are no direct benefits or risks for participating patients. This prospective simulation study will be non-interventional, the prediction models do not alter the original daily practice and in this phase, it is not intended to be used as a diagnostic device. Intraoperatively, just prior to the construction of the anastomosis machine learning model will predict, using the patient, tumor, and intraoperatively variables (listed in the Data Dictionary paragraph), the probability of anastomotic leakage to occur. SAS VIYA is used for deployment of the machine learning model. During the prospective simulation study, the scores of these predictions are only available to the principle and research investigators), and thus unknown to the participating hospitals or operating surgeons in order to prevent any influence on current daily practice in this stage of the research. Thirty days postoperatively, data of the patients regarding the occurrence of anastomotic leakage will be collected. AUC-ROC, sensitivity, specificity, and accuracy then will be calculated based on the number of patients assessed as true positive, true negative, false positive or false negative. After a minimum of 100 events and 100 non-events, the external validation is completed and the final AUC-ROC, sensitivity and specificity scores will be presented.

Quality assurance plan, data checks, source data verification

Data will be handled confidentially and anonymously. Data will be pseudo-anonymized for the principal investigator and the research investigators. Pseudo-anonymized data are entered in a Castor database.[12] A data dictionary is attached to the original dataset with metadata to describe the data. All participating hospitals have a Data Sharing Agreement to share data of included patients with the principal investigator and the research investigators. A data management plan will be created according to our institute's polices with the assistance of a data management expert, along with the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines.[13] The characteristics of the collected and generated data is clinical data extracted from the electronic health records. This contains continuous,

nominal, and dichotomous variables. Data will not be reused or coupled to existing data. Informed consent of patients is necessary to predict the outcome using the developed model. Privacy policies and laws are applicable to this project. The research is laid down legally by Dutch law concerning scientific medical research, according to the law: Wet Medischwetenschappelijk Onderzoek met mensen (WMO). The project will also comply with all data protection principles as is defined in the General Data Protection Regulation. The anonymized dataset can be accessed via a Castor database. Long term data will be saved in the Amsterdam UMC repository with help of the research data management (RDM) department. The data will be saved for five years after the project has ended. The costs for data management are provided in the project budget.

Data dictionary

The following variables will be collected:

i. Patient and tumor characteristics: Age; sex; body mass index; American Society of Anesthesiologists (ASA) classification; intoxications (smoking and/or alcohol consumption); medical history of diabetes; steroid use (not nasal); hemoglobin; benign or malignant disease. If there is malignant disease: TNM-stage, tumor distance from anal verge, neoadjuvant treatment.
ii. Perioperative characteristics: Surgical procedure, surgical approach; conversion; occurrence of intraoperative event (hypoxic events, hypercarbia, bradycardia, hypotension, embolism, reanimation, more extensive resection than planned, serosa lesions, bladder and ureteral injuries, intraoperative bleeding, splenectomy)
iii. Characteristics just prior to the creation of the anastomosis: Patient temperature; time of antibiotic administration; administration of vasopressors; blood loss; O2 saturation; mean arterial pressure; fluid administration; urine production; presence of fecal contamination; subjective assessment of local perfusion; epidural analgesia; dosing movements; time from incision until the creation of the anastomosis, intention to create stoma.
iv. Postoperative characteristics: Colorectal anastomotic leakage within 30 days and length of hospital stay.

Standard Operating procedures: Patients eligible for inclusion are detected in the first multidisciplinary team meeting. If eligible, the surgeon will inform and discuss this study with the patient in the preoperative consultation for surgery. If the patient consents to participation, written informed consent is required. The patient may withdraw this consent at any time.

Sample size calculation: In the participating hospitals, around 100 to 400 colorectal resections are performed annually, with an approximate incidence of anastomotic leakage of 5 to 15%. Multiple studies demonstrated a minimum of 100 events and 100 nonevents as an appropriate sample size for external validation.[14,15] With an expected

total of 1,200 patients included annually and a leakage percentage around 10%, including 100 events takes approximately one to two years.

Handling missing data: The machine learning model will make a prediction in patients with more than 80% of the required data available. Missing data are imputed using predictive mean matching with ten iterations.

Statistical analysis plan: The external validation will be performed on at least 100 events (anastomotic leakage) and 100 non-events (no anastomotic leakage). The machine learning model with the best predictive performance in terms of AUROC will be used as the implementation model. Colorectal anastomotic leakage rate will be compared in a multivariate logistic regression model. All analyses will be carried out under the supervision of a clinical epidemiologist.

Eligibility Criteria:

Inclusion Criteria:
- patients undergoing colorectal resection with the construction of an anastomosis
- patients with the age of 18 years or older
- patients able to give informed consent

Exclusion Criteria:
- when more than 25% of the target variables are missing on which the machine learning model bases the prediction
- non-elective surgeries

Study Plan

This section provides details of the study plan, including how the study is designed and what the study is measuring.

Design details

Observational Model: Cohort
Time Perspective: Prospective
Target Follow-up Duration: 90 Days
Estimated enrollment: 1000 patients

Study start: 2022-02-01
Estimated primary completion: 2024-07-01

Cohorts and interventions

Group/Cohort	Intervention/Treatment
Adult patients undergoing colorectal resection with the construction of an anastomosis. The exposure of interest in the current study regards the occurrence of anastomotic leakage in patients undergoing colorectal resection with the construction of an anastomosis. Information on the exposure of interest is gained by obtaining data from the patient files.	Procedure: Colorectal resection · Patients undergoing a colorectal resection with the construction of a primary anastomosis

Primary outcomes measures

Outcome Measure	Measure Description	Time Frame
Colorectal anastomotic leakage	Colorectal anastomotic leakage is defined according to Reisinger et al.[16]: "clinically relevant anastomotic leakage is defined as extra luminal presence of contrast fluid on contrast-enhanced computed tomography scans and/or leakage when relaparotomy was performed, requiring reintervention or treatment."	The occurrence of the primary outcome is assessed after 30 days postoperatively.

Secondary outcomes measures

Outcome Measure	Measure Description	Time Frame
Length of Hospital Stay	To evaluate the impact of anastomotic leakage on length of stay, the duration of a patient's hospital stay after undergoing a colorectal resection is investigated.	The length of hospital stay is assessed 90 days postoperatively
The predictive performance of the prediction model	The predicted probability of anastomotic leakage for every participant will be evaluated with the actual outcome after 30 days postoperatively. The discrimination of the external validation cohort is reported with the area under the curve - receiver operating characteristic, the sensitivity, specificity, and accuracy. Calibration of the prediction model will be visualized in a calibration plot.	30 days postoperatively

13

Chapter 13

Discussion and future perspectives

This thesis introduces novel artificial intelligence (AI) based prediction models designed for the prediction of colorectal anastomotic leakage (CAL) and postoperative pancreatic fistula (POPF). These AI models represent groundbreaking innovations that stand to revolutionize perioperative decision-making in surgical practice.

An AI model capable of accurately predicting CAL serves as a transformative decision support tool, empowering surgeons to make more informed choices. It aids in discerning when to employ stomas for high-risk patients and when it is judicious to forego them for those at lower risk. Additionally, this model's precision provides the assurance of low risk of CAL, thus facilitating earlier patient discharge. The forthcoming A1Check study, detailed in Chapter 12, plays a pivotal role in addressing this pressing need. Successful external validation of this model holds the potential to usher in its clinical implementation, thereby becoming a cornerstone in patient selection for early discharge, regular postoperative care, and stoma construction. The CHASE study has demonstrated the feasibility and safety of including colorectal cancer patients with an American Society of Anesthesiologists (ASA) score of 1 or 2 who underwent elective laparoscopic resection within a 23-hour accelerated enhanced recovery after surgery program.[1] However, it is imperative to acknowledge that the CHASE population exhibited a higher readmission rate than the control group. The AI model developed in the A1Check study could address this by accurately selecting patients for early discharge while minimizing the risk of readmission.

In parallel, the AI-FRS model, as presented in Chapter 11, carries significant implications for the future of pancreatic surgery. Similar to the CAL model, it plays a crucial role in surgical decision-making, especially for high-risk or frail patients. Beyond preventive applications, the PORSCH study underscores the potential for early recognition facilitated by accurate prediction, leading to substantial enhancements in clinical outcomes.[2]

While the shift toward AI-driven algorithms is undeniable in this data-driven era, the actual integration of AI models into daily clinical practice remains limited. Additionally, several challenges emerge from the studies presented in this thesis that necessitate careful consideration before successful implementation can be achieved. To further enhance the seamless integration of AI models for predicting complications following abdominal surgery, this discussion section addresses these challenges and provides actionable recommendations. A proposed framework, rooted in these recommendations, is presented to guide researchers and clinicians in future endeavors within the field of predictive modeling

Assess clinical relevancy

Assessing the clinical relevancy of a predictive model is a crucial preliminary step in the development and implementation of AI-models. In this era of data-driven healthcare, the ability to predict various outcomes has expanded significantly. However, it is essential to carefully consider how accurate predictions can genuinely benefit daily clinical practice. The process of creating and integrating a prediction model is resource-intensive, so it is imperative to evaluate the potential advantages of the intended model. In essence, a prediction model that justifies implementation should align with one or more of the following criteria:

1. Improved patient outcomes. A predictive model should have the potential to substantially enhance patient outcomes. This may involve the early identification of complications, better treatment decisions, or improved postoperative care strategies
2. Enhanced Care Efficacy: The model should contribute to the overall effectiveness of healthcare delivery. This can manifest in several ways, such as aiding in more accurate diagnosis, treatment planning, or patient management.
3. Reduced Hospital Costs. Efficient predictive models can play a vital role in reducing healthcare costs. By optimizing resource allocation, streamlining interventions, and potentially preventing adverse events, these models can contribute to cost savings for healthcare facilities and patients.

Define primary outcome

The primary outcome in a prediction model should closely align with the objective of the study. Precision in defining this outcome is important. As illustrated in Chapter 7, certain studies developed prediction models for postoperative pancreatic fistula (POPF) without providing a clear definition for this complication. Similarly, as explored in Chapter 5, the wide array of definitions for colorectal anastomotic leakage (CAL) has led to divergent study outcomes.

The importance of a standardized definition cannot be emphasized enough, as it fosters the applicability of prediction models across different healthcare settings and regions. Efforts have been made to seek consensus through two Delphi studies, yet the challenge of achieving unanimity regarding the definition, diagnosis, and management of CAL remains difficult.[3,4] To address this challenge, future Delphi studies must engage a globally diverse panel of experts. Their collective expertise has the potential to culminate in a universally accepted definition, which can then be seamlessly integrated

into both clinical practice and research. This harmonization will significantly enhance the comparability of research findings.

Using the right type of data

Selecting the right datasets for the development of AI prediction models is another perquisite. For instance, in the context of colorectal anastomotic leakage (CAL), the A1Check study has demonstrated that a combination of pre- and intraoperative structured data yields a highly accurate prediction model. A review of previous research reveals that many machine learning algorithms predominantly rely on preoperative data for CAL prediction. One notable example is the study conducted by Hoek et al., which aimed to construct a preoperative prediction model using a vast dataset from the Dutch Colorectal Audit, comprising 13,175 patients, making it the largest study of its kind in the Netherlands.[5] Despite its impressive scale, this model exhibited suboptimal predictive performance, as evidenced by a modest concordance index of 0.658 following cross-validation. Such findings underscore the crucial need for a model that also integrates intraoperative data. Indeed, the LekCheck study by Huisman et al.[6] identified certain intraoperative risk factors for CAL, and incorporating these factors into a CAL prediction model is likely to enhance its predictive performance.

However, the development of a prediction model for postoperative complications does not consistently yield high predictive performance when using structured data. Chapter 9 illustrates that the use of both pre- and intraoperative structured data did not result in a highly predictive model. Various factors can explain this phenomenon, and in this instance, two main reasons emerge. Either the quality of the data is suboptimal, attributed to missing data or the absence of crucial variables that would serve as significant predictors, or the primary outcome of interest cannot be adequately explained by structured independent variables. In such cases, it becomes pertinent to explore other types of data.

The realm of imaging analysis has gained substantial attention, exemplified by the utilization of radiomic and deep learning features to extract parameters from imaging modalities.[7,8] This approach offers the potential to uncover novel predictive parameters, especially relevant in the context of POPF, where imaging-derived parameters such as pancreatic texture and pancreatic duct diameter hold significance. Chapter 7's review reaffirms the promise of radiomic features in POPF prediction, although it highlights methodological limitations and a lack of external validation. Chapter 10 addresses this gap by presenting an externally validated and publicly available POPF prediction model based on radiomic features extracted from preoperative CT scans.

Assess the quality of a prediction model

Once a prediction model is developed, it is crucial to assess its performance using various methods and metrics. Traditional measures for binary outcomes provide essential insights and include the Brier score, the Area Under the Curve – Receiver Operating Characteristic (AUROC), sensitivity, specificity, and calibration.[9] Calibration, in particular, is a metric that is sometimes overlooked, yet it can carry profound significance. Neglecting calibration can be perilous because poorly calibrated models have the potential to mislead and, in certain cases, prove detrimental to clinical decision-making.[10] It is important to emphasize that calibration should not be underestimated. A poorly calibrated model, despite achieving a high AUROC, can be less clinically valuable than a competing prediction model with a lower AUROC but exceptional calibration.[11] The significance of calibration becomes even more evident when considering the practical implications of using a prediction model to inform both patients and healthcare professionals. Inaccurate risk estimates arising from poor calibration can lead to false expectations among patients and clinicians. Patients may make life-altering decisions based on an anticipated event, or its absence, only to realize that their decisions were based on misguided information.[10]

Be as transparent as possible

Artificial intelligence is a powerful force poised to revolutionize various domains, including surgical procedures. In Chapter 8, we explore the current implementation of AI applications in the operating room, with the expectation that their prevalence will continue to grow. Given this trajectory, it is crucial to emphasize the importance of transparent reporting and rigorous evaluation of AI models. In the realm of predictive modeling, the Transparent Reporting of multivariable prediction model of Individual Or Diagnosis Artificial Intelligence (TRIPOD-AI) initiative is currently in progress.[12] This initiative is tailored to prediction models that leverage machine learning techniques. TRIPOD-AI provides researchers with a comprehensive reporting guideline that empowers them to furnish essential details, enabling readers to assess the study's quality and meaningfully interpret its findings. The adoption of such a tool has the potential to reduce research inefficiency and minimize wasted resources.[12] The hope is that this valuable resource will gain widespread acceptance and catalyze improvements in research methodology, ensuring that the evolution of AI in healthcare is guided by a commitment to transparency and the highest standards of research integrity.

Establishing risk categories and define treatment consequences

To effectively support surgeons in their perioperative decision-making using a prediction model for postoperative complications, the creation of risk categories alongside individual risk assessments is pivotal. Risk categories enable the grouping of individual risks for a comprehensive evaluation and tailored response to each category. Typically, risk categories are categorized as low, intermediate, and high risk, with specific recommendations provided for each category.

For instance, in the case of postoperative pancreatic fistula (POPF), two widely used prediction models, the Fistula Risk Score (FRS) and the updated Alternative Fistula Risk Score (ua-FRS), advise considering preventive measures for patients at high risk of developing a fistula.[13,14] These measures may include the use of anastomotic stents or prophylactic ocreotide. However, in the prediction of CAL, such well-defined risk groups have not yet been established. To address this gap, a Delphi study should be conducted to achieve consensus on the specific cut-off values for risk categories in the prediction of CAL and the subsequent management of complications following colorectal resection. This consensus will provide actionable guidance for managing patients within each risk category, rendering the prediction model more practical and conducive to seamless implementation in clinical settings.

Collaborate

Collaboration stands as a cornerstone in the arena of predictive modeling using AI. Every phase of the process, from conceptualization to development, validation, and eventual integration of a prediction model, necessitates the establishment of a well-coordinated team. It is advisable to assemble these teams at the outset of the project to ensure the smooth progression of the work. In making these collaborative AI projects clinically applicable, it is imperative to establish a robust chain of communication among multifaceted teams. These teams should ideally include doctors, data scientists, biomedical researchers, and IT experts. Such complex problems require a mixture of domain-specific expertise, and this collaboration is the glue that binds these diverse skill sets together. Regular meetings among these teams are crucial. They provide a forum to discuss the current status of the project, share insights, and identify potential challenges. Each team can be thought of as a distinct point or "dot", and it is the process of collaboration that connects these dots. This interconnectedness is vital for the seamless integration of a prediction model into existing healthcare systems, ensuring that the technology serves its intended purpose effectively.

References

1. Tweed, T.T.T., Sier, M.A.T., Daher, I. et al. Accelerated 23-h enhanced recovery protocol for colon surgery: the CHASE-study. *Sci Rep*. 12, 20707 (2022).
2. Smits FJ, Henry AC, Besselink MG, et al. Algorithm-based care versus usual care for the early recognition and management of complications after pancreatic resection in the Netherlands: an open-label, nationwide, stepped-wedge cluster-randomised trial. *Lancet*. 2022;399(10338):1867-75.
3. Spinelli A, Anania G, Arezzo A, et al. Italian multi-society modified Delphi consensus on the definition and management of anastomotic leakage in colorectal surgery. *Updates Surg*. 2020;72(3):781-92.
4. van Helsdingen CP, Jongen AC, de Jonge WJ, Bouvy ND, Derikx JP. Consensus on the definition of colorectal anastomotic leakage: A modified Delphi study. *World J Gastroenterol*. 2020;26(23):3293-303.
5. Hoek VT, Buettner S, Sparreboom CL, et al. A preoperative prediction model for anastomotic leakage after rectal cancer resection based on 13.175 patients. *Eur J Surg Oncol*. 2022;48(12):2495-501.
6. Huisman DE, Reudink M, van Rooijen SJ, et al. LekCheck: A Prospective Study to Identify Perioperative Modifiable Risk Factors for Anastomotic Leakage in Colorectal Surgery. *Ann Surg*. 2022;275(1):e189-e97.
7. Agostini A, Borgheresi A, Bruno F, et al. New advances in CT imaging of pancreas diseases: a narrative review. *Gland Surg*. 2020;9(6):2283-94.
8. van Timmeren JE, Cester D, Tanadini-Lang S, et al. Radiomics in medical imaging—"how-to" guide and critical reflection. *Insights Imaging*. 2020;11(1):91.
9. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010 Jan;21(1):128-38.
10. Van Calster, B., McLernon, D.J., van Smeden, M. et al. Calibration: the Achilles heel of predictive analytics. *BMC Me.d* 17, 230 (2019).
11. Van Calster B, Vickers AJ. Calibration of risk prediction models: impact on decision-analytic performance. *Med Decis Making*. 2015 Feb;35(2):162-9.
12. Collins GS, Dhiman P, Navarro CLA, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open*. 2021;11(7):e048008.
13. Callery MP, Pratt WB, Kent TS, Chaikof EL, Vollmer CM Jr. A prospectively validated clinical risk score accurately predicts pancreatic fistula after pancreatoduodenectomy. J Am Coll *Surg*. 2013 Jan;216(1):1-14.
14. Mungroop TH, Klompmaker S, Wellner UF, Steyerberg EW, Coratti A, D'Hondt M, de Pastena M, Dokmak S, Khatkov I, Saint-Marc O, Wittel U, Abu Hilal M, Fuks D, Poves I, Keck T, Boggi U, Besselink MG; European Consortium on Minimally Invasive Pancreatic Surgery (E-MIPS). Updated Alternative Fistula Risk Score (ua-FRS) to Include Minimally Invasive Pancreatoduodenectomy: Pan-European Validation. *Ann Surg*. 2021 Feb 1;273(2):334-340.

13

14

Chapter 14

English summary
Nederlandse samenvatting

English summary

Part I: Impact of anastomotic leakage and mitigation strategies

An anastomosis is created to restore gastrointestinal continuity after a sick part within this this track is removed. Failure of this anastomosis is called colorectal anastomotic leakage (CAL) in colorectal surgery and a postoperative pancreatic fistula (POPF) in pancreatic surgery. Both a CAL and POPF are major impact complications on both the short- and long-term outcome, as well as oncological outcomes. Several mitigations strategies have been proposed to diminish the incidence or severity of anastomotic leakage. A diverting stoma is a commonly used mitigation strategy in colorectal surgery. However, it is not fully understood which patients benefit most from a diverting stoma. A diverting stoma is a commonly employed mitigation strategy in colorectal surgery. However, it remains unclear which patients benefit the most from a diverting stoma. **Chapter 2** identified patient characteristics and intraoperative factors associated with a diverting stoma, revealing statistically significant differences between patients with and without one. The study suggested a protective effect of a diverting stoma against CAL, advocating for selective use.

In **chapter 3**, data from the Netherlands' nationwide audit over the past decade on rectal surgery were analyzed. Out of 13,263 patients, 7,106 had a diverting stoma. The reduction in diverting stomas was correlated with an increase in CAL. Interestingly, if CAL occurred, fewer patients required surgical reintervention. However, patients with a diverting stoma experienced more readmissions to the hospital. Therefore, the trade-off of selective use should be carefully considered.

As hospital mergers become more common due to centralization in medicine, their impact on postoperative outcomes such as CAL and POPF remains unclear. **Chapter 4** evaluated the effect of a merger between two Dutch university hospitals on the quality of surgical care, volume, and timeliness. The study demonstrated that a hospital merger could be conducted without compromising patient safety, while benefiting from the centralization of highly specialized care and promoting medical research.

Literature on the long-term effects of CAL concerning postoperative oncological outcomes and stoma presence is inconclusive. Understanding the association between CAL and long-term outcomes is crucial for patient follow-up, counseling, and tailored treatment decisions. **Chapter 5** did not reveal a negative impact of CAL on oncological outcomes. However, CAL was strongly associated with an increased incidence of stoma

presence after five years, significantly affecting the quality of life for patients after colon and rectal cancer surgery.

Part II: The use of artificial intelligence in predictive modelling

The integration of AI is becoming increasingly prevalent in the development of prediction models within the healthcare sector. In **Chapter 6**, numerous challenges associated with the implementation of AI were thoroughly examined, encompassing aspects from initial development to practical application. The chapter offered insightful recommendations concerning the utilization of AI for predicting postoperative complications following major abdominal surgery.

The ability to POPF accurately holds significant potential in assisting surgeons with making individualized treatment decisions. **Chapter 7** delved into a systematic review, assessing the predictive capabilities of radiomic features in anticipating POPF. The evaluation encompassed considerations of methodological quality, reporting transparency, and the risk of bias. The findings underscored the promising potential of radiomic features in accurately predicting POPF. However, it was noted that the existing evidence supporting their clinical application was still lacking.

Moreover, the operating room stands out as a data-rich environment where AI can play a pivotal role. **Chapter 8** provided a comprehensive review of current AI applications in perioperative care, highlighting their focus on supporting decision-making during the perioperative phase and enhancing surgical skills and safety. This section emphasized the paramount importance of standardizing data, clinically validating AI systems, meticulously evaluating the implementation process, and establishing ethical and legal frameworks.

Part III: Development of prediction models

Within the realm of AI, various types of algorithms are used for the development of prediction models. Tree-based models, a subset of AI, have gained prominence in this regard. However, their superiority over logistic regression, especially in handling structured data, remains uncertain. **Chapter 9**, drawing on nationwide data from 16 centers in the Dutch Pancreatic Cancer Audit involving 799 patients, investigated this question. The findings revealed that, in the prediction of postoperative complications after pancreatic surgery, tree-based models did not outperform logistic regression.

In **chapter 10**, a groundbreaking model, the Radiomics preoperative-Fistula Risk Score (RAD-FRS), was introduced for predicting postoperative pancreatic fistula (POPF). This novel model derived radiomic features from preoperative CT scans of patients following pancreatoduodenectomy. Successfully developed and internally validated

in the Netherlands, it underwent external validation in Italy, demonstrating a robust performance with an area under the curve of 0.81 in the external validation cohort. The potential integration of this model with hospital CT reporting systems could enhance patient counseling before surgery.

Chapter 11 showcased yet another model for predicting POPF—the Artificial Intelligence-based Fistula Risk Score (AI-FRS). This model, relying on radiomic features from automatic pancreas segmentations, predicted the risk of POPF after pancreatoduodenectomy. Pending further validation, this model holds promise for automatic preoperative prediction of POPF.

In **chapter 12**, the A1Check study's study protocol was presented. This prospective study aims to externally validate AI algorithms predicting CAL. These algorithms utilize pre- and intraoperative data from patients undergoing colorectal resection with a primary anastomosis. The study contributes to the ongoing exploration and validation of AI's role in predicting postoperative complications.

Nederlandse Samenvatting

Deel I: Impact van naadlekkage en risico beperkende maatregelen

Een naad of anastomose wordt gecreëerd om de gastro-intestinale continuïteit te herstellen nadat een ziek stuk van de darm of alvleesklier is verwijderd. Wanneer deze naad niet goed heelt, wordt dit colorectale naadlekkage (CAL) genoemd bij colorectale chirurgie en een pancreasfistel (POPF) bij pancreaschirurgie. Zowel CAL als POPF hebben een grote impact op zowel de korte- als langetermijnresultaten, evenals op oncologische resultaten. Verschillende risico beperkende maatregelen zijn voorgesteld om de incidentie of ernst van naadlekkage te verminderen. Een ontlastend stoma is een veelgebruikte strategie bij colorectale chirurgie. Het is echter niet volledig begrepen welke patiënten het meest baat hebben bij een deviërend stoma. **Hoofdstuk 2** toonde aan welke patiëntkenmerken en intraoperatieve factoren geassocieerd waren met een ontlastend stoma. De studie toonde duidelijk statistisch significante verschillen aan tussen patiënten met en zonder ontlastend stoma. Een ontlastend stoma vertoonde een beschermend effect tegen CAL, en het selectief gebruik ervan werd aangeraden.

In **hoofdstuk 3** werden gegevens van de landelijke audit van rectale chirurgie in Nederland van het afgelopen decennium geanalyseerd. In totaal werden 13.263 patiënten geanalyseerd, waarvan 7.106 een ontlastend stoma hadden. Er werd vastgesteld dat de afname van ontlastend stoma's gepaard gingmet een toename van

CAL en als CAL optrad, hadden minder patiënten een chirurgische re-interventie nodig. Patiënten met een ontlastend stoma vertoonden echter ook meer ziekenhuisopnames. Dus, de afweging van selectief gebruik moet individueel worden overwogen.

Vanwege centralisatie en supercentralisatie in de geneeskunde komen ziekenhuisfusies steeds vaker voor. Het effect hiervan op postoperatieve uitkomsten, zoals CAL en POPF, is echter onduidelijk. In **hoofdstuk 4** werd het effect van een fusie van twee Nederlandse universitaire ziekenhuizen op de kwaliteit van chirurgische zorg, volume en tijdigheid van zorg geëvalueerd. De studie toonde aan dat een ziekenhuisfusie kan worden uitgevoerd zonder de veiligheid van patiënten in gevaar te brengen, en daarbij profiteert van de centralisatie van hooggespecialiseerde zorg en de bevordering van medisch onderzoek.

Literatuur over de langetermijneffecten van CAL met betrekking tot postoperatieve oncologische resultaten en stomapresentie is tegenstrijdig. Begrip van de associatie tussen CAL en lange termijn resultaten is cruciaal voor de follow-up van patiënten, counseling en op maat gemaakte behandelbeslissingen. **Hoofdstuk 5** toonde echter geen negatieve invloed van CAL op oncologische resultaten aan. Naadlekkage was echter sterk geassocieerd met een verhoogde aanwezigheid van een ontlastend of eindstandig stoma na vijf jaar, wat de kwaliteit van leven van patiënten na chirurgie voor colorectale kanker aanzienlijk beïnvloedde.

Deel II: Het gebruik van kunstmatige intelligentie in voorspellende modellering

Artificiële intelligentie (AI) wordt steeds vaker gebruikt om voorspellingsmodellen in de gezondheidszorg te ontwikkelen. In **hoofdstuk 6** werden meerdere kwesties besproken met betrekking tot de implementatie van AI, van de ontwikkeling tot de daadwerkelijke implementatie. Het gaf suggesties voor het gebruik van AI bij het voorspellen van postoperatieve complicaties na grote buikoperaties.

Een nauwkeurige voorspelling van POPF kan de chirurg helpen bij het aanbieden van op maat gemaakte behandelbeslissingen. Radiomic features, computer-afgeleide wiskundig geëxtraheerde kenmerken, zijn geïntroduceerd als een nieuw instrument om een voorspellingsmodel voor POPF te ontwikkelen. In **hoofdstuk 7** werd een systematische review uitgevoerd om de voorspellende prestaties van radiomic features bij het voorspellen van POPF te evalueren en de methodologische kwaliteit, rapportagetransparantie en het risico op vertekening systematisch te beoordelen. Het toonde aan dat radiomic features het potentieel hebben om POPF nauwkeurig te voorspellen. Hoewel deze resultaten veelbelovend waren, was het bewijs ter ondersteuning van klinische toepassing zwak.

Ook de operatiekamer is een data-rijke omgeving waarin AI een belangrijke rol kan spelen. **Hoofdstuk 8** beoordeelde de huidige toepassingen van AI in de perioperatieve zorg. Het toonde aan dat toepassingen gericht waren op het ondersteunen van perioperatieve besluitvorming en het verbeteren van chirurgische vaardigheden en veiligheid. Het benadrukte het belang van het standaardiseren van gegevens, klinische validatie van AI-systemen, zorgvuldige evaluatie van het implementatieproces en het vaststellen van ethische en juridische kaders.

Deel III: Ontwikkeling van voorspellingsmodellen

Er zijn veel soorten machine learning, een subcategorie van AI, modellen die kunnen worden gebruikt om voorspellingsmodellen te ontwikkelen. Tree-based modellen worden steeds vaker aanbevolen voor het ontwikkelen van voorspellingsmodellen. Het is echter onduidelijk of deze modellen superieur zijn aan logistische regressie bij het gebruik van gestructureerde data. **Hoofdstuk 9** gebruikte landelijke gegevens van 16 centra in het Nederlandse Pancreas Kanker Audit van 799 patiënten en toonde aan dat een tree-based model niet beter presteerde dan logistische regressie bij het voorspellen van postoperatieve complicaties na pancreaschirurgie.

In **hoofdstuk 10** wordt de Radiomics preoperative-Fistula Risk score (RAD-FRS), een nieuw voorspellingsmodel voor POPF, geïntroduceerd. Radiomic features werden geëxtraheerd uit preoperatieve CT-scans van patiënten die een pancreatoduodenectomie zouden ondergaan. Het model werd ontwikkeld en intern gevalideerd in Nederland en extern gevalideerd in Italië. Het model presteerde goed met een area under the curve van 0,81 in de externe validatiecohort. Dit model heeft de potentie om geïntegreerd te worden in het patiëntendossier om zo de patiëntenvoorlichting vóór de operatie te verbeteren.

Hoofdstuk 11 toont een ander model voor de voorspelling van POPF, de Artificial Intelligence-based Fistula Risk Score (AI-FRS). Dit model voorspelde het risico op POPF na pancreatoduodenectomie op basis van radiomic features van automatische pancreassegmentaties. Na verdere validatie kan dit model automatische preoperatieve voorspelling van POP mogelijk maken.

In **hoofdstuk 12** wordt het studieprotocol van de A1Check-studie gepresenteerd. Deze prospectieve studie valideert externe AI-algoritmen die CAL voorspellen. Deze modellen maken gebruik van pre- en intra-operatieve gegevens van patiënten die een colorectale resectie ondergaan met een primaire anastomose.

Appendix

List of publications
Dankwoord
Curriculum Vitae

List of publications and contributions

Ingwersen E.W.*, B.T. Bootsma*, F. Daams, D.L. van der Peet. Artificiële intelligentie: toepassingen voor de operatiekamer. *NTVG (Nederlands Tijdschrift voor Geneeskunde)*. 2023 Aug 23:167:D7603. *Equal contribution to this work

Ingwersen, E.W., P.M.W. Rijssenbeek, H.A. Marquering, G. Kazemier, F. Daams. Radiomics for the prediction of a postoperative pancreatic fistula following a pancreatoduodenectomy: a systematic review and radiomic score quality assessment. *Accepted for publication at Pancreatology.*

Ingwersen, E.W.*, J.I Bereska*, A. Balduzzi, B.V. Janssen, M.G. Besselink, G. Kazemier, G. Marchegiani, G. Malleo, H.A. Marquering, C.Y. Nio, R. De Robertis, R. Salvia, E.W. Steyerberg, J. Stoker, F. Struik, I.M. Verpalen, F. Daams. Radiomics preoperative-Fistula Risk Score (RAD-FRS) for pancreatoduodenectomy: development and external validation. *BJS open*. 2023 Sep 5;7(5):zrad100.

*Equal contribution to this work

Ingwersen, E.W.*, J.I Bereska*, A. Balduzzi, B.V. Janssen, M.G. Besselink, G. Kazemier, G. Marchegiani, G. Malleo, H.A. Marquering, C.Y. Nio, R. De Robertis, R. Salvia, E.W. Steyerberg, J. Stoker, F. Struik, I.M. Verpalen, F. Daams. Artificial Intelligence-based fully automatic prediction of postoperative pancreatic fistula from pre-operative CT scans of patients undergoing a pancreatoduodenectomy. *Submitted 2023 (Radiology: Artificial Intelligence)*.

*Equal contribution to this work

Ingwersen E.W., Stam W.T., Meijs B.J.V., Roor J., Besselink M.G., Groot Koerkamp B., de Hingh I.H.J.T., van Santvoort H.C., Stommel M.W.J., Daams F. Machine learning versus logistic regression for the prediction of complications after pancreatoduodenectomy. *Surgery*. 2023 May 5:S00396060(23)00158-7.

Huisman D.E., Bootsma B.T., **Ingwersen E.W.**, Reudink M., Slooter G.D., Stens J., Daams F.; LekCheck Study group. Fluid management and vasopressor use during colorectal surgery: the search for the optimal balance. *Surg Endosc*. 2023 Aug;37(8):6062-6070.

Ingwersen E.W., Stam W.T., van Kesteren L.J., Wissink I.J.A., van Berge Henegouwen M.I., Besselink M.G., Busch O.R., Erdmann J.I., Eshuis W.J., Gisbertz S.S., Kazemier G., van der Peet D.L., Swijnenburg R.J., Zonderhuis B,. Daams F. Impact of merging two university

hospitals on surgical outcome after esophagogastric and hepato-pancreato-biliary surgery: Results from a retrospective study. *Surg Open Sci*. 2023 Jan 24;12:1-8.

Stam W.T., **Ingwersen E.W**., Ali M., Spijkerman J.T., Kazemier G., Bruns E.R.J., Daams F. Machine learning models in clinical practice for the prediction of postoperative complications after major abdominal surgery. *Surg Today*. 2023 Oct Oct;53(10):1209-1215.

Ingwersen E.W., van der Beek P.J.K., Dekker J.W.T., van Dieren S., Daams F. One Decade of Declining Use of Defunctioning Stomas After Rectal Cancer Surgery in the Netherlands: Are We on the Right Track? *Dis Colon Rectum*. 2023 July 1;66(7):1003-1011.

Stam W.T., Goedknegt L.K., **Ingwersen E.W**., Schoonmade L.J., Bruns E.R.J., Daams F. The prediction of surgical complications using artificial intelligence in patients undergoing major abdominal surgery: A systematic review. *Surgery*. 2022 Apr;171(4):1014-1021.

Huisman D.E., **Ingwersen E.W**., Reudink M., Bootsma B.T., Slooter G.D., Dekker J.W.T., Daams F. The Selective Use of a Diverting Stoma in Rectal Surgery. *J Gastrointest Surg*. 2022 Jul;26(7):1509-1512.

Dankwoord 1/2

Dankwoord 2/2

Curriculum Vitae

Ewout Ingwersen was born on July 27, 1995, in Haarlem, the Netherlands. He grew up in Bloemendaal, where he attended primary school. He completed his secondary education at "Het Eerste Christelijk Lyceum" in Haarlem, obtaining his gymnasium diploma in 2013. During this time, he enjoyed playing the piano, playing soccer, and spending time with friends.

His interest in science, particularly biology, developed during his high school years. In 2013, Ewout began studying Medicine at the Vrije Universiteit Amsterdam. This year also marked his move to Amsterdam and his subsequent joining of a fraternity in 2014. His master's studies introduced him to the field of surgical oncology research, which led to his PhD candidacy in 2021. During this period, his passion for radiology was solidified.

In 2023, Ewout started working as a resident not in training at the Spaarne Gasthuis in Haarlem and Hoofddorp. In 2024, he began his residency in the Department of Radiology at the Amsterdam UMC.

In his free time, Ewout continues to enjoy playing the piano, running, and spending time with his girlfriend and friends.