External validation of risk prediction platforms for pancreatic fistula after pancreatoduodenectomy using nomograms and artificial intelligence

So Jeong Yoon1,*, Wooil Kwon2,*, Ok Joo Lee1, Ji Hye Jung1, Yong Chan Shin3, Chang-Sup Lim4, Hongbeom Kim2, Jin-Young Jang2, Sang Hyun Shin1, Jin Seok Heo1, In Woong Han1

1Division of Hepatobiliary-Pancreatic Surgery, Department of Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea
2Department of Surgery, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea
3Department of Surgery, Ilsan Paik Hospital, Inje University College of Medicine, Goyang, Korea
4Department of Surgery, Seoul Metropolitan Government - Seoul National University Boramae Medical Center, Seoul National University College of Medicine, Seoul, Korea

**INTRODUCTION**

Postoperative pancreatic fistula (POPF) is one of the most fatal complications after pancreatoduodenectomy (PD) and is associated with postoperative hemorrhage, intraabdominal infections, and increased mortality [1,2]. The incidence of POPF remains high despite advances in surgical techniques and perioperative management [2]. Particularly for clinically relevant POPF (CR-POPF), defined by the grading system of the International Study Group of Pancreatic Fistula (ISGPF) [3], the

*So Jeong Yoon and Wooil Kwon contributed equally to this work as co-first authors.

**Corresponding Author:** In Woong Han
Division of Hepatobiliary-Pancreatic Surgery, Department of Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Korea
Tel: +82-2-3410-0772, Fax: +82-2-3410-6980
E-mail: cardioman76@gmail.com
ORCID: https://orcid.org/0000-0001-7093-2469

**Key Words:** Artificial intelligence, Nomograms, Pancreatic fistula, Pancreatoduodenectomy, Postoperative complications
reported incidence was 14.5% in a previous meta-analysis [4].

Many studies have analyzed the risk factors for POPF. Based on traditional risk factors, there were attempts to develop platforms for predicting POPF [5,6]. However, due to the limited predictive values of previous platforms [7,8], we developed new risk prediction platforms using nomograms [9] and artificial intelligence (AI) technology [10]. The nomogram consisting of 6 preoperatively available data provides a better insight into the risk factors and their contributions. The AI model was based on an in-depth analysis of risk factors using machine learning algorithms. The model finally included 16 preoperative and intraoperative variables. Both models are readily available in the form of a calculator at http://popf.smchbp.org and http://popfrisk.smchbp.org.

In this study, we performed external validation of the aforementioned platforms with multicenter datasets to verify the reproducibility and generalizability of the models and to determine their clinical utilities.

METHODS

This study was approved by the Institutional Review Boards of Samsung Medical Center (Seoul, Korea: No. 2020-09-181), Seoul National University Hospital (Seoul, Korea: No. SNUH 2010-147-110), Ilsan Paik Hospital (Goyang, Korea: No. 2021-06-009), and Boramae Medical Center (Seoul, Korea: No. 30-2021-72). The study was performed in accordance with the Declaration of Helsinki and written informed consent was waived due to its retrospective nature.

Patient database

The cohort for external validation included 1,576 patients who underwent PD between January 2007 and December 2016 at 3 different centers: Seoul National University Hospital, Ilsan Paik Hospital, and Borame Medical Center. The patients’ demographic data, preoperative laboratory results, imaging findings and surgical outcomes were retrospectively reviewed.

Perioperative data and risk calculation

The individual risks of CR-POPF were calculated using previously developed nomogram- and AI-based web calculators, which are available at the websites (Fig. 1). The nomogram was based on the following 6 preoperative variables: sex, body mass index (BMI), the American Society of Anesthesiology physical status (ASA PS) classification, serum albumin, tumor location, and the diameter of the main pancreatic duct (MPD) measured via CT or magnetic resonance cholangiopancreatography. The AI calculator was developed using the following preoperative and intraoperative variables: age, sex, BMI, underlying heart disease, ASA PS classification, preoperative platelet count, serum albumin, serum lipase, preoperative endoscopic biliary drainage, neoadjuvant radiotherapy, amount of intraoperative fluid infusion, pancreatic texture, the diameter of MPD, portal vein resection, coexisting pancreatitis detected preoperatively or intraoperatively, and tumor location.

There were no missing values in the categorical variables.

Fig. 1. (A) The web-based nomogram calculator (http://popf.smchbp.org). (B) The web-based artificial intelligence (AI) calculator (http://popfrisk.smchbp.org).
of the validation cohort. Median imputation was used for the missing continuous variables to replace the missing data with medians.

**Postoperative outcomes**

POPF was diagnosed and graded according to the 2016 ISGPF definition and grading. POPF grades B and C (CR-POPF) were included as the outcomes in the analysis, and biochemical leak was not counted.

**Statistical analysis and external validation**

Receiver operating characteristic (ROC) curve analyses were performed using IBM SPSS Statistics ver. 26 (IBM Corp., Armonk, NY, USA). The area under the curve (AUC) for the logistic regression model was reported with 95% confidence intervals (CIs). The AUC values with P-values of less than 0.05 were regarded as statistically significant.

Backward elimination was performed to obtain the optimal AUC value of the AI model by selecting features that had no significant prognostic value. The AUC for each model was calculated based on the stepwise selection in accordance with the development process.

**RESULTS**

The clinical demographics and surgical outcomes of 1,576 patients in the validation cohort are presented in Table 1. The patients' mean age was 63.6 years, and 697 patients (44.2%) had underlying heart disease including hypertension. Preoperative endoscopic biliary drainage was performed in 825 patients (52.3%). Intraoperatively, 1,056 patients (67.0%) had soft pancreas and the mean diameter of MPD was 3.5 mm. CR-POPF was developed in 270 patients (17.1%).

In the external validation of the nomogram with 1,576 patients, the ROC curve was drawn with the AUC of 0.679 (95% CI, 0.645–0.713; P < 0.001) (Fig. 2A). Fig. 3 shows the values of the AUC after backward elimination. All 16 independent variables were entered into the analysis first and each variable was eliminated one by one. The maximal AUC was 0.672 (95% CI, 0.637–0.706; P < 0.001) (Fig. 2B), including the following 13 variables: the diameter of MPD, BMI, serum albumin, amount of intraoperative fluid infusion, age, preoperative platelet count, tumor location, portal vein resection, coexisting pancreatitis, serum lipase, neoadjuvant radiotherapy, ASA PS classification, and sex.

**DISCUSSION**

In the absence of a standardized management protocol for POPF until now, early prediction and recognition are crucial to identify patients at high risk of POPF requiring careful

| Variable                                      | Data     |
|----------------------------------------------|----------|
| No. of patients                              | 1,576    |
| Age (yr)                                     | 63.6 ± 10.3 |
| Sex                                          |          |
| Male                                         | 957 (60.7) |
| Female                                       | 619 (39.3) |
| Body mass index (kg/m²)                      | 23.2 ± 3.2 |
| Underlying heart disease                     |          |
| No                                           | 879 (55.8) |
| Yes                                          | 697 (44.2) |
| ASA PS classification                         |          |
| I                                            | 441 (28.0) |
| II                                           | 1,020 (64.7) |
| III                                          | 112 (7.1)  |
| IV                                           | 3 (0.2)    |
| Preoperative platelet (X10³/μL)              | 267.8 ± 95.7 |
| Preoperative albumin (g/dL)                  | 3.9 ± 0.6   |
| Preoperative lipase (IU)                     | 118.6 ± 257.6 |
| Preoperative endoscopic biliary drainage      |          |
| No                                           | 751 (47.7) |
| Yes                                          | 825 (52.3) |
| Neoadjuvant therapy                          |          |
| No                                           | 1,519 (96.4) |
| Yes                                          | 57 (3.6)   |
| Radiotherapy                                 | 1          |
| Chemotherapy                                 | 32         |
| CCRT                                         | 24         |
| Intraoperative fluid infusion (mL)            | 3,231.1 ± 257.6 |
| Pancreatic texture                           |          |
| Soft                                         | 1,056 (67.0) |
| Moderate/hard                                | 520 (33.0)  |
| Pancreatic duct diameter (mm)                | 3.5 ± 2.3   |
| Portal vein resection                        |          |
| No                                           | 1,535 (97.4) |
| Yes                                          | 41 (2.6)    |
| Intraoperative pancreatitis                  |          |
| None                                         | 1,370 (86.9) |
| Yes                                          | 206 (13.1)  |
| Pathology                                    |          |
| Pancreatic tumor                             | 732 (46.4)  |
| Others                                       | 844 (53.6)  |
| POPF                                         |          |
| No                                           | 574 (36.4)  |
| Yes                                          | 1,002 (63.6) |
| BCL                                          | 732 (46.5)  |
| Grade B                                      | 267 (16.9)  |
| Grade C                                      | 3 (0.2)     |

Values are presented as number only, mean ± standard deviation, or number (%).

ASA PS, American Society of Anesthesiologist physical status; CCRT, concurrent chemoradiotherapy; POPF, postoperative pancreatic fistula; BCL, biochemical leakage.

*Data were not available in 1,688, and 2 patients in order.*
observation [2]. A series of traditional risk scoring systems, such as the original fistula risk score (o-FRS) [5] and the alternative fistula risk score (a-FRS) [6], have been used. However, several studies performed external validation and the predictability varied with the study population [7,8,11,12]. Therefore, our institution suggested new predictive models using nomograms and AI [9,10] and this study validated the predictability of the new platforms with an external cohort.

The nomogram including 6 simple variables had an AUC value of 0.709 in the development process [9], and 0.679 in the external validation. A few other nomograms were recently proposed. Huang et al. [13] suggested a new nomogram with the following 3 variables: BMI, the diameter of MPD, and drain fluid amylase level (DFA) on postoperative day (POD) 1. The AUC value was 0.744 in the external validation. Another nomogram by Suzuki et al. [14] also included drain fluid lipase level on POD 1 and decreased rate of DFA, which was defined as a change in levels from POD 1 to 3. The accuracy of the nomogram was 0.810, as stated in the study. Compared to these recently developed nomograms, our nomogram showed limited predictive value in the external validation. The AUC value might have decreased in the process of external validation with highly heterogeneous data collected from different centers. But most of all, the predictive power of the variables in the nomogram could also be limited. Previous studies suggested that DFA is a strong predictive factor of POPF [15,16], and DFA on POD 1 is currently widely accepted as an indicator for early drain removal after PD [17]. Considering that our nomogram is composed of only preoperative and intraoperative factors, updating the platform with DFA may improve the accuracy and predictability of the model.

Machine learning is a branch of AI technology designed to enable rapid analytical model building. It has been used in various medical fields including surgery. As far as we know, we invented the first AI-based prediction model for CR-POPF. The most remarkable advantage of machine learning is that it can identify complex structures in high-dimensional data [18] and detect latent variables, which are not directly measured using conventional analytical methods [19]. The new AI model yielded the maximal AUC value of 0.74 with 16 variables, and in the external validation with backward elimination, the AUC was 0.672 with 13 variables. This value is acceptable but leaves considerable room for improvement. First, there were quite a few missing values in both development and validation cohorts.

Fig. 2. (A) The receiver operating characteristic (ROC) of the nomogram. Area under the curve (AUC) = 0.679, P < 0.001. (B) The ROC of the artificial intelligence predictor. AUC = 0.672, P < 0.001.

Fig. 3. The area under the curve (AUC) values with backward elimination.
Despite attempts to handle missing data in deep learning [20], it remains the main obstacle to model development. Also, the AI model includes a few variables, which cannot be objectively measured. For example, coexisting pancreatitis or pancreatic texture detected by surgeons intraoperatively can be highly subjective. In order to resolve this issue, several studies are underway to objectively measure those variables using preoperative images and machine learning [21,22]. These efforts are expected to establish a foothold for improving predictability of the platforms in the future.

The current study has several limitations. Since this study is based on retrospectively reviewed multicenter datasets, the results might have been affected by selection and information bias. Some potential factors that could influence the development of POPF, such as anastomosis technique and postoperative drain management, varied considerably among the surgeons and the institutions. Also, as previously stated, the platforms were developed and validated with datasets including missing values, which reduces the sample representativeness and complicates the analysis. Further studies with prospectively collected high-quality data are needed to upgrade the platforms with improved predictability.

In conclusion, this study was performed to externally validate the previously developed prediction platforms for POPF. The results suggest the need for improvement and future studies to build better prediction models with higher accuracy.

ACKNOWLEDGEMENTS

The authors would like to thank Hyemin Kim (data manager, Department of Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine) for help with data collection.

Fund/Grant Support
This study was supported by Samsung Medical Center grant (# SMO1210421).

Conflict of Interest
No potential conflict of interest relevant to this article was reported.

ORCID iD
So Jeong Yoon: https://orcid.org/0000-0001-5227-4490
Wooil Kwon: https://orcid.org/0000-0002-4827-7805
Ok Joo Lee: https://orcid.org/0000-0003-1077-2592
Ji Hye Jung: https://orcid.org/0000-0002-9456-6472
Yong Chan Shin: https://orcid.org/0000-0001-9808-1213
Chang-Sup Lim: https://orcid.org/0000-0002-2349-9647
Hongbeom Kim: https://orcid.org/0000-0002-1595-0135
Jin-Young Jang: https://orcid.org/0000-0003-3312-0503
Sang Hyun Shin: https://orcid.org/0000-0002-2533-4491
In Woong Han: https://orcid.org/0000-0001-7093-2469

Author Contribution
Conceptualization, Data Curation: SJY, WK, YCS, CSL, HK, JYJ, SHS, JSH, IWH
Formal Analysis: All authors
Writing – Original Draft: SJY, WK, IWH
Writing – Review & Editing: All authors

REFERENCES

1. McMillan MT, Vollmer CM Jr, Asbun HJ, Ball CG, Bassi C, Beane JD, et al. The characterization and prediction of ISGPF grade c fistulas following pancreateo-duodenectomy. J Gastrointest Surg 2016;20:262-76.
2. 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.
3. Bassi C, Marchegiani G, Dervenis C, Sarr M, Abu Hilal M, Adham M, et al. The 2016 update of the International Study Group (ISGPF) definition and grading of postoperative pancreatic fistula: 11 years after. Surgery 2017;161:584-91.
4. Eshmuninov D, Schneider MA, Tschuor C, Raptis DA, Kambakamba P, Muller X, et al. Systematic review and meta-analysis of postoperative pancreatic fistula rates using the updated 2016 International Study Group Pancreatic Fistula definition in patients undergoing pancreatic resection with soft and hard pancreatic texture. HPB (Oxford) 2018;20:992-1003.
5. Callery MP, Pratt WB, Kent TS, Chaikof EL, Vollmer CM Jr. A prospectively validated clinical risk score accurately predicts pancreatic fistula after pancreateo-duodenectomy. J Am Coll Surg 2013;216:1-14.
6. 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;269:937-43.
7. Ryu Y, Shin SH, Park DJ, Kim N, Heo JS, Choi DW, et al. Validation of original and alternative fistula risk scores in postoperative pancreatic fistula. J Hepatobiliary Pancreat Sci 2019;26:354-9.
8. Kang JS, Park T, Han Y, Lee S, Kim JR, Kim H, et al. Clinical validation of scoring systems of postoperative pancreatic
13. Huang XT, Huang CS, Liu C, Chen W, Cai JP, Cheng H, et al. Development and validation of a new nomogram for predicting clinically relevant postoperative pancreatic fistula after pancreatoduodenectomy. World J Surg 2021;45:261-9.

14. Suzuki S, Shimoda M, Shimazaki J, Oshiro Y, Nishida K, Shiithara M, et al. Drain lipase levels and decreased rate of drain amylase levels as independent predictors of pancreatic fistula with nomogram after pancreatocoduodenectomy. World J Surg 2021;45:1921-8.

15. Molinari E, Bassi C, Salvia R, Butturini G, Crippa S, Talamini G, et al. Amylase value in drains after pancreatic resection as predictive factor of postoperative pancreatic fistula: results of a prospective study in 137 patients. Ann Surg 2007; 246:281-7.

16. Bertens KA, Crown A, Clanton J, Alemi F, Alseidi AA, Biehl T, et al. What is a better predictor of clinically relevant postoperative pancreatic fistula (CR-POPF) following pancreatocoduodenectomy (PD): postoperative day one drain amylase (POD1DA) or the fistula risk score (FRS)? HPB (Oxford) 2017;19:75-81.

17. Melloul E, Lassen K, Roulin D, Grass F, Perinel J, Adham M, et al. Guidelines for perioperative care for pancreatoduodenectomy: Enhanced Recovery After Surgery (ERAS) Recommendations 2019. World J Surg 2020;44:2056-84.

18. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436-44.

19. Borsboom D, Mellenbergh GJ, van Heerden J. The theoretical status of latent variables. Psychol Rev 2003;110:203-19.

20. Song M, Greenbaum J, Luttrell J 4th. Zhou W, Wu C, Shen H, et al. A review of integrative imputation for multi-omics datasets. Front Genet 2020;11:570255.

21. Kambakamba P, Mannil M, Herrera PE, Müller FC, Kueemmerli 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;167:448-54.

22. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H. Artificial intelligence in radiology. Nat Rev Cancer 2018; 18:500-10.