Optimizing strategies in pancreatic and hepato-biliary surgery

Mungroop, T.H.

Citation for published version (APA):
CHAPTER 6

Alternative Fistula Risk Score for Pancreatoduodenectomy (a-FRS): Design and International External Validation


* shared senior authorship

Annals of Surgery. 2019 May; 269(5).
ABSTRACT

Objective: To develop an alternative fistula risk score (a-FRS) for postoperative pancreatic fistula (POPF) after pancreateoduodenectomy, without blood loss as a predictor.

Background: Blood loss, one of the predictors of the original-FRS, was not a significant factor during two recent external validations.

Methods: The a-FRS was developed in two databases: the Dutch Pancreatic Cancer Audit (18 centers) and the University Hospital Southampton NHS. Primary outcome was grade B/C POPF according to the 2005 ISGPS definition. The score was externally validated in two independent databases (University Hospital of Verona and University Hospital of Pennsylvania), using both 2005 and 2016 ISGPS definitions. The a-FRS was also compared to the original-FRS.

Results: For model design, 1,924 patients were included of whom 12% developed POPF. Three predictors were strongly associated with POPF: soft pancreatic texture (odds ratio [OR] 2.58, 95% Confidence Interval [CI] 1.80-3.69), small pancreatic duct diameter (per mm increase, OR: 0.68 [95% CI: 0.61-0.76]), and high body mass index (BMI) (per kg/m² increase, OR: 1.07 [95% CI: 1.04-1.11]). Discrimination was adequate with an Area Under Curve (AUC) of 0.75 (95% CI: 0.71-0.78) after internal validation, and 0.78 (0.74-0.82) after external validation. The predictive capacity of a-FRS was comparable with the original-FRS, both for the 2005 definition (AUC 0.78 vs. 0.75, P = 0.03), and 2016 definition (AUC 0.72 vs. 0.70, P = 0.05).

Conclusions: The a-FRS predicts POPF after pancreateoduodenectomy based on three easily available variables (pancreatic texture, duct diameter, BMI) without blood loss and pathology, and was successfully validated for both the 2005 and 2016 POPF definition.
INTRODUCTION

Postoperative pancreatic fistula (POPF) is one of the most threatening complications after pancreatoduodenectomy (PD). POPF occur in up to 20% of patients and are typically associated with increased hospital stay, costs, re-intervention rates, and mortality.1-3 Prediction of the risk of POPF can optimize individual treatment decisions (e.g., drain placement, use of somatostatin analogues) and therefore several fistula risk models have been proposed.2, 4-13 These models were typically built with single-center data, often lacked adequate (external) validation, and most have not been widely implemented in daily practice. The validated Fistula Risk Score (FRS) by Callery and colleagues is the most cited and best used POPF prediction model.2 The FRS predicts POPF based on gland texture, pancreatic duct diameter, intraoperative blood loss, and definitive pathology.

Some have argued that a FRS without blood loss could be preferred.14 First, blood loss was not a significant factor at two recent external validations.14, 15 Second, it is currently not registered in several audits, for example, the National Surgical Quality Improvement Program (US-NSQIP) but also the Dutch Pancreatic Cancer Audit (DPCA). Third, blood loss depends on surgical quality, and is therefore not a suitable prognostic factor for adjusting POPF-risk for benchmarking. Fourth, several studies have argued that estimation of blood loss during surgery is unreliable and inaccurate, and this metric therefore should not be used to judge physician performance or patient outcomes.16-18 Finally, in future patients this factor may be even less predictive for POPF, because for example minimally-invasive PD seems to lead to less blood loss but similar rates of POPF.14, 19

In 2016, a new definition of POPF was presented by the International Study Group on Pancreatic Surgery (ISGPS) and up to our knowledge no POPF prediction model has been validated for prediction of POPF according to this new ISGPS classification.

The aim of this study was therefore to develop and externally validate a POPF prediction model for both definitions, without blood loss.

METHODS

The TRIPOD guidelines20 for multivariable prediction models were followed for the design, internal and external validation, and reporting of this clinical prediction model. The Medical Ethics Review Committee of the Academic Medical Center Amsterdam (the Netherlands) waived the need for consent.


**Patients and methods**

For model design, data from two combined databases were used: the Dutch Pancreatic Cancer Audit (DPCA, 18 centers; January 2014 to March 2016) and the University Hospital Southampton (United Kingdom, National Health Service; January 2007 to December 2016). The DPCA is a nationwide, mandatory, prospective audit of pancreatic surgery. All 18 centers in the Netherlands performing PD participate in this audit, each performing a minimum of 20 PDs annually. The PD database of the University hospital Southampton NHS Trust Foundation, UK, is a prospectively maintained database for pancreatic surgery.

For external validation, patient outcomes data from two high-volume pancreatic surgery centers were used: the University of Verona Hospital Trust (Verona, Italy; 2014 - 2016) and the University Hospital of Pennsylvania (Philadelphia (PA), United States; 2004 - 2015).

**Definitions, outcome and predictors**

Primary outcome was POPF defined according to the 2005 ISGPS definition of grade B/C fistula. An exploratory univariate analysis (by comparison of means) was performed in the DPCA data. Significant predictors were added in the multivariable analysis and combined with Southampton data. Preoperative variables assessed included: age, sex, body mass index (BMI), American Society of Anesthesiologist (ASA) physical status, pancreatic duct diameter (on preoperative imaging, measured at location where the pancreas is to be transected), presence of (medically-treated) diabetes mellitus, preoperative biliary drainage, year of surgery, any comorbidity, cardiac comorbidity, vascular comorbidity, pulmonary comorbidity, neurologic/psychiatric comorbidity, gastrointestinal comorbidity, urogenital comorbidity, thrombotic comorbidity, comorbidity related to disease of muscles and joints, endocrine comorbidity, infectious diseases, history of malignancy, and neo-adjuvant treatment. Intraoperative variables included were pancreatic texture (soft or not soft), vascular resection, and multi-visceral resection. Duct size was assessed in the design cohort on the most recent preoperative CT-scan. The median interval between imaging and surgery was 5 weeks (IQR: 3-7). The a-FRS and original-FRS were compared in the external validation cohort since these centers (Verona and Pennsylvania) were not used for the design of either score. In the Verona database both the ISGPS 2005 and 2016 definition were scored, and in the Pennsylvania database the ISGPS 2016 definition was scored. In the external validation cohort duct size was assessed intraoperatively.

**Statistical analysis**

Statistical analysis was performed using SPSS version 22.0 (IBM corp., Armonk (NY), United States) and STATA version 14.2 (StataCorp., College Station (TX), United States).
Normal distribution of data was assessed using visual inspection of histograms. Normally distributed data were presented as mean with standard deviation (SD) and non-normally distributed data were presented as median with interquartile range (IQR). Categorical variables are presented as counts and proportions. The level for statistical significance was set at $P < 0.05$. Non-linearity was tested in continuous variables, but this did not lead to transformations. Duct size was truncated at 5 mm, and ASA class 3 and 4 were merged since there were only 2 patients with ASA class 4. Missing data were imputed using multiple imputation (five permutations). This imputation procedure assumes that the missing data patterns can be modeled based on covariates and the outcome as observed in the data set. Pancreatic texture was missing in 5% of cases, BMI in 4%, and duct size in 27%.

A prediction model was developed using multivariable logistic regression modeling. Variables included were selected based on a univariate screen in the DPCA data set, a literature search, and clinical relevance. The Akaike information criterion (AIC) was used for variable selection, resulting in a parsimonious model with minor loss of predictive ability. This implies the model with the smallest difference in predictive capacity is selected in case this only leads to marginal loss of predictive value. The performance was assessed by of Area Under the receiver operating Curve (AUC) and calibration plots. Internal validation was undertaken using bootstrap resampling to obtain shrunken regression coefficients and an optimism corrected AUC. Three risk groups were proposed to allow clinically-relevant risk stratification. The original-FRS and a-FRS were calculated in the external validation cohort (Verona and Pennsylvania). The DeLong test was used to compare the difference between the AUCs of both scores.

RESULTS

Cohort characteristics
In total, data of 2,850 patients undergoing PD from 21 institutions in 4 countries were included in the model design and external validation cohorts. In the design cohort, 1,924 consecutive patients were included with 232 (12%) patients developing POPF (2005 definition), see Table 1. The median age in this cohort was 67 (IQR: 60-74) years, median BMI was 25 (IQR 22-28) kg/m², and 57% of patients were male. A surgical drain was placed in 99% of patients, octreotide was administered postoperatively in 63% of patients, and pancreatojejunostomy was the most frequently used pancreatic anastomosis (in 75% of patients).

The external validation cohort consisted of 926 patients with 131 (24%) patients developing POPF according to the 2005 definition (Verona only) and 154 (17%) in the
2016 definition (Verona and Pennsylvania). In this external validation cohort (Verona and Pennsylvania), median age was 65 (IQR: 56-72) years, median BMI was 25 (IQR: 22-28), and 53% of patients was male, see Table 1.

**Table 1.** Baseline Characteristics of the Study Cohorts

<table>
<thead>
<tr>
<th></th>
<th>Model design cohort (n=1,924)</th>
<th>External validation cohort (n=926)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, median (IQR), mm</td>
<td>67 (60-74)</td>
<td>65 (56-72)</td>
</tr>
<tr>
<td>Sex (male), no. (%)</td>
<td>778 (57%)</td>
<td>491 (53%)</td>
</tr>
<tr>
<td>BMI, median (IQR), mm</td>
<td>25 (22-28)</td>
<td>25 (22-28)</td>
</tr>
<tr>
<td>ASA Class, no. (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>283 (15%)</td>
<td>45 (5%)</td>
</tr>
<tr>
<td>2</td>
<td>1273 (67%)</td>
<td>516 (56%)</td>
</tr>
<tr>
<td>3</td>
<td>339 (18%)</td>
<td>359 (39%)</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Diabetes mellitus, no. (%)</td>
<td>403 (26%)</td>
<td>174 (19%)</td>
</tr>
<tr>
<td>Soft pancreatic texture, no. (%)</td>
<td>970 (54%)</td>
<td>511 (55%)</td>
</tr>
<tr>
<td>Duct size, median (IQR), mm</td>
<td>4 (2-5)</td>
<td>3 (2-4)</td>
</tr>
<tr>
<td>Pathology, no. (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Pancreatic ductal adenocarcinoma</td>
<td>816 (43%)</td>
<td>402 (43%)</td>
</tr>
<tr>
<td>- Chronic pancreatitis</td>
<td>50 (3%)</td>
<td>33 (4%)</td>
</tr>
<tr>
<td>- Neuroendocrine</td>
<td>104 (5%)</td>
<td>80 (9%)</td>
</tr>
<tr>
<td>- Cholangio-carcinoma</td>
<td>265 (14%)</td>
<td>61 (7%)</td>
</tr>
<tr>
<td>- Ampullary carcinoma</td>
<td>265 (14%)</td>
<td>95 (10%)</td>
</tr>
<tr>
<td>- Duodenal carcinoma</td>
<td>110 (6%)</td>
<td>46 (5%)</td>
</tr>
<tr>
<td>- Cystic neoplasms (IPMN, MCN and serous cysts)</td>
<td>128 (7%)</td>
<td>114 (12%)</td>
</tr>
<tr>
<td>- Other</td>
<td>221 (12%)</td>
<td>95 (10%)</td>
</tr>
</tbody>
</table>

**Model design**

The full model (based on male sex, age, soft pancreatic texture, duct size, ASA physical status, BMI, DM and tumor location) had adequate discriminative ability (AUC 0.75 [95% CI: 0.71-0.78]; Table 2). The final model consisted of three strong predictors: soft pancreatic texture (odds ratio (OR): 2.58 [95% CI: 1.80-3.69]), decreasing pancreatic duct diameter (continuous, truncated at 5mm, OR per mm increase: 0.68 [95% CI: 0.61-0.76], and increasing BMI per index point (OR: 1.07 [95% CI: 1.04-1.11]) (see table 2). Discrimination of the model was adequate with an AUC of 0.75 (95% CI: 0.71-0.78) after
Alternative Fistula Risk Score for Pancreatoduodenectomy

internal validation, see figure 1a. For calibration after internal validation see figure 2. An online calculator was made available at pancreascalculator.com based on the following calculation:

\[
P = \frac{\exp(-3.136 + 0.947[\text{texture}] + 0.0679[\text{BMI}] - 0.385[\text{PD size}])}{1 + \exp(-3.136 + 0.947[\text{texture}] + 0.0679[\text{BMI}] - 0.385[\text{PD size}])}
\]

with \(P\) = probability, texture 1 = soft, and 0 if not soft, PD size = pancreatic duct size in mm (truncated at 5).

Table 2. Model Design

<table>
<thead>
<tr>
<th>Odds ratio</th>
<th>95% CI</th>
<th>(P)</th>
<th>Odds ratio</th>
<th>95% CI</th>
<th>(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area Under Curve (AUC)</td>
<td>0.75</td>
<td>0.71 - 0.78</td>
<td>0.75</td>
<td>0.71 - 0.78</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.46</td>
<td>1.08 - 2.00</td>
<td>0.015</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Age, year</td>
<td>0.99</td>
<td>0.98 - 1.01</td>
<td>0.485</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Soft pancreatic texture</td>
<td>2.29</td>
<td>1.59 - 3.29</td>
<td>&lt;0.001</td>
<td>2.78</td>
<td>1.97 - 3.91</td>
</tr>
<tr>
<td>Duct size, per mm increase</td>
<td>0.69</td>
<td>0.57 - 0.83</td>
<td>&lt;0.001</td>
<td>0.68</td>
<td>0.61 - 0.76</td>
</tr>
<tr>
<td>ASA Class</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1 Reference</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>0.69</td>
<td>0.45 - 1.07</td>
<td>0.125</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3 / 4</td>
<td>1.17</td>
<td>0.71 - 1.94</td>
<td>0.652</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
<td>1.07</td>
<td>1.03 - 1.11</td>
<td>&lt;0.001</td>
<td>1.07</td>
<td>1.04 - 1.11</td>
</tr>
<tr>
<td>DM</td>
<td>0.94</td>
<td>0.64 - 1.38</td>
<td>0.957</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tumor location (Pancreatic vs other)</td>
<td>0.64</td>
<td>0.46 – 0.89</td>
<td>0.009</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

External validation – 2005 and 2016 ISGPS definitions

External validation was performed in the external cohort using the 2005 POPF definition. Discrimination was adequate (figure 1a, AUC of 0.78 [95% CI: 0.74-0.82]). Discrimination in the external cohort was similar using the 2016 POPF definition (figure 1b, AUC of 0.72 [95% CI: 0.68-0.76]) and calibration was adequate, see figure 2.
Figure 1. Comparison of alternative-FRS and original-FRS for prediction of POPF according to the 2005 ISGPS definition (1a) and 2016 ISGPS definition (1b)
Figure 2. Calibration plot of the a-FRS in the design cohort and external validation cohort (ISGPS 2016). The triangles indicate the observed frequencies by deciles of predicted probabilities. This LOESS curve is a locally weighted polynomial regression.

Risk groups

Three risk groups were identified based on the risk distribution; low (0 - 5%), intermediate (> 5 – 20%) and high (20 ≥ %) of POPF using both the design and external validation cohort, see figure 3. Using the 2005 definition, the observed mean risk was 2.7% (95% CI: 1.4-4.0%) in the low-risk group, 14% (95% CI: 12-15%) the intermediate-risk group, and 35% (95% CI: 30-39%) in the high-risk group (Fishers exact test: \( P < 0.001 \)). Using the 2016 definition, the observed mean risk was 2.8% (95% CI: 0.8-4.9%) in the low-risk group, 18% (95% CI: 14-21%) in the intermediate-risk group, and 31% (95% CI: 24-37%) in the high-risk group (Fishers exact test: \( P < 0.001 \)).

Comparison with the original-FRS

In the external validation cohort, the a-FRS displayed a marginal improvement compared to the original-FRS for prediction POPF according to the 2005 definition (AUC 0.78 [95% CI: 0.74-0.82] vs. 0.75 [0.71-0.80], DeLong test: \( P = 0.029 \)). Furthermore, for prediction of POPF according to the 2016 definition, the a-FRS performed equally well compared to the original-FRS (AUC 0.72 [95% CI: 0.68-0.76] vs. 0.70 [0.66-0.74], DeLong test: \( P = 0.050 \)).
Chapter 6

DISCUSSION

The a-FRS is an externally and internationally validated score to predict the risk of POPF after PD without the need for intra-operative blood loss or pathology as variables. The score is based on three intraoperatively available variables: duct size, pancreatic texture and BMI, and was designed using (nationwide) data from over 1,900 patients. The a-FRS is the first POPF prediction model, which was successfully validated for both the ISGPS 2005 and 2016 definitions, and performed at least as good as the original-FRS but without blood loss as a variable.

At least eleven prediction models for POPF after PD have been proposed. Most of these models have not been (adequately) externally validated, which limits applicability in clinical practice. At least eleven prediction models for POPF after PD have been proposed. Most of these models have not been (adequately) externally validated, which limits applicability in clinical practice.^

Figure 3. Risk groups of the a-FRS

DISCUSSION

The a-FRS is an externally and internationally validated score to predict the risk of POPF after PD without the need for intra-operative blood loss or pathology as variables. The score is based on three intraoperatively available variables: duct size, pancreatic texture and BMI, and was designed using (nationwide) data from over 1,900 patients. The a-FRS is the first POPF prediction model, which was successfully validated for both the ISGPS 2005 and 2016 definitions, and performed at least as good as the original-FRS but without blood loss as a variable.

At least eleven prediction models for POPF after PD have been proposed. Most of these models have not been (adequately) externally validated, which limits applicability in clinical practice. Besides, some models also included clinically irrelevant 2005 ‘grade A’ POPF. In the 2016 update of the ISGPS definition, grade A POPF has been reclassified to ‘biochemical leak’. Finally, a number of the existing risk models include postoperative factors, such as drain amylase, which limit the applicability to guide early treatment decision. The predictors included in the a-FRS are strong predictors in the cohorts used in this study (P < 0.001) and have been recognized as universal predictors.

A recent POPF prediction model without blood loss and based on NSQIP data showed poor performance at external validation (AUC of 0.62). The original-FRS however, showed a high AUC (> 0.90) in design. Typically, prediction models show impressive accuracy in the (single-center) derivation cohort, because the model is optimized to fit the data. Performance is usually lower at external validation.
The (a-)FRS can be applied to select patients for somatostatin analogues or placement of surgical drains. A recent worldwide survey showed that large variations exist between surgeons in the management after PD. Large variations in practice may be indicative for room for improvement. The (a-)FRS could help surgeons to objectify the risk of POPF and design, study and implement decision rules. Low-risk patients could be selected for a no-drain strategy. This is especially relevant since several studies have recently shown that complications may arise from these surgical drains. Furthermore, the (a-)FRS can lead to improved cost-effectiveness. For example with somatostatin analogues such as Pasireotide, where surgeons might want to use different thresholds for ‘high-risk’, because of heterogeneity in pharmacy costs and financial consequences of treatment and complications. In low-risk patients these (expensive) analogues can be omitted, since the absolute risk reduction in these patients may not justify the costs. As a result, medication can be allowed in a specific setting. If this medication was not used otherwise, or less optimally, this could lead to reduction in POPF. At last, for research purposes models such as FRS, and especially a-FRS can be used for evaluation of surgeons technical performance, risk-adjusted comparison of surgical outcome, stratification in RCTs as well as high-risk subgroup analysis. We have recently used this score in a pilot study wherein we applied a somatostatin analogue intra-operatively only in high-risk patients (>15% risk). Future prospective studies can stratify treatment based on the outcome of this score (e.g. regarding use of drains, somatostatin analogues and other measures) and could provide treatment algorithms to be tested in randomized trials.

We were able to successfully simplify POPF prediction since the a-FRS consists of three (easily-available) variables without loss of predictive capacity as compared to the original-FRS. The variables blood loss and pathology were successfully omitted. Blood loss is partially related to surgical quality and may therefore rather not be used for risk-adjusted comparisons. Although experienced pancreatic surgeons might predict final pathology outcome correctly in 90% of patients, it may differ from the preoperative assessment.

Ideally, we would have designed a score which defines the risk of POPF already prior to surgery. This could be useful in starting preventive treatment prior to surgery, or referral of high-risk patients to undergo surgery in expert centers. Although we attempted to do so, this led to a substantially lower predictive ability. This was caused mostly because of the strong predictive value of pancreatic texture, which independently of pancreatic duct size predicts POPF (Table 2, Appendix 2). Future studies should therefore attempt to determine or predict pancreatic texture preoperatively.
This study has some limitations. First, in the design cohort duct size was measured on preoperative imaging and in the external validation database this was done during surgery. However, duct size on preoperative imaging is an accurate measurement of the actual duct size. Second, the a-FRS was developed to predict the risk of POPF using the 2005 definition, since the 2016 ISGAPS definition was not available during the build-up of this cohort. However, since the a-FRS was successfully validated for both the 2005 and 2016 ISGAPS definitions, we feel this confirms its validity and broad applicability. Third, the observational nature of the study makes it impossible to rule out residual confounding.

Strengths of this study include the large dataset, including variation between the international cohorts, with many surgeons and hospitals working according to their local practices, and a design cohort with nationwide audit data of PDs performed including the full spectrum of indications, and variations in operative techniques and patient management. A second strength is the use of sound statistical methods for the development of this prediction model. The optimal method consists of using a large multi-institutional dataset for development, using multiple imputation for missing data, and using continuous predictors as such instead of categorizing without assuming linearity. We decided to make an online calculator, since this allows for an exact risk calculation, as cut-offs for high-intermediate-low risk groups may differ between institutions. Furthermore, the calculator provides a 95%-confidence interval to reflect uncertainty in the risk estimate. We made POPF risk groups of 0-5%, 5-20% and ≥20%. This could be useful in stratifying fistula management (e.g., decision to leave drains or administer Pasireotide).

Future studies should investigate the benefits of using a-FRS on individual health outcomes and/or cost-effectiveness of care, preferably within a randomized trial using the a-FRS to guide decision-making in high-risk patients. Besides, the addition of extra variables to the a-FRS should be studied but always considering the effort-reward balance, especially with non-readily available variables are considered. Such studies should involve large multicenter datasets with geographical external validation.

CONCLUSION

POPF can be predicted accurately during surgery with the a-FRS using pancreatic texture, duct size, and BMI as variables. The model, which does not require blood loss or pathology, may guide intraoperative decision-making, improve stratification and analysis in RCTs, and allow risk-adjusted comparisons.
Appendix
The supplementary appendix is available on the journal website.

Acknowledgements
The authors gratefully thank Rick Pleijhuis and Rob Mentink from Evidencio.com for their assistance with the online risk calculator.

Conflicts of interest and source of funding
No conflicts of interest. This research was funded in part by a grant from the Dutch Cancer Society (grant number UVA2013-5842)
REFERENCES


