Scores for Prediction of Fistula after Pancreatoduodenectomy: A Systematic Review

Marta Sandini\textsuperscript{a} Giuseppe Malleo\textsuperscript{b} Luca Gianotti\textsuperscript{a}

\textsuperscript{a}School of Medicine and Surgery, Milano-Bicocca University, San Gerardo Hospital, Monza, and
\textsuperscript{b}Unit of General Surgery B, The Pancreas Institute, University of Verona Hospital Trust, Verona, Italy

Key Words
Score · Pancreatic fistula · Pancreatoduodenectomy · Review

Abstract
Background/Aim: Different scoring systems to predict the occurrence of postoperative pancreatic fistula (POPF) after pancreatoduodenectomy have been described, but the considered risk factors often suffer subjective scaling. The aim of this review is to evaluate and compare all published risk metrics predictive of POPF. Methods: All existing scores were retrieved by literature web search. Inclusion criteria were ISGPF classification of POPF and the development of a risk score metric. Results: From a total of 286 publications, 10 studies were selected. Most of them were retrospective and single center. The models considered a median number of 3 items (range from 2 to 5); in 5 of 10 trials only pre or intraoperative variables were included. The median number of patients/study was 186 (IQR 111.1–229.0). External validation was performed in 6 of 10 studies. The most recurrent items were abdominal fat (4/10), main pancreatic duct diameter (in 4/10), and pancreatic texture (3/10). Conclusion: POPF risk estimation should be easy, accurate, and objective. It should consider preoperative patient-related and gland-related features, and intraoperative events. None of the published systems completely adhere to these principles. Large heterogeneous multicentric validations should be endorsed, to account for the case-mix and evaluate the reproducibility of each scoring system.

Introduction
Postoperative pancreatic fistula (POPF) is the commonest complication after pancreatic resection, with a rate of appearance up to 30% even in high-volume centers [1–3]. Clinically relevant POPF (grade B or C) [4] can be treated with conservative therapy, such as antibiotics or prolonged drain and in selective cases of infected collections and disruption of pancreatic anastomosis, radiologic, endoscopic, or surgical procedures may be required for complete healing. The latter scenario may delay adjuvant treatment and affect oncologic outcome [5].

Since there are limited tools to minimize the occurrence of POPF [6, 7], the attention has been focused on the assessment of the risk factors. After the publication of the ISGPF definition [4], several studies addressed the role of single factors, such as age [8–10], fat distribution...
[11–14], operative time [15], blood loss [16–18], pathologic diagnosis [16–19], diameter of the main pancreatic duct (MPD) and texture of pancreatic parenchyma [20–24] on the occurrence of POPF. Other authors proposed comprehensive risk scores based on multivariable modeling. In general, risk assessment may help caregivers to set up protocols for a strict and early detection of warning clinical signs, to tailor the clinical management of different risk classes, or to select high-risk patients who might be excluded from surgical resection. However, current international guidelines and recommendations for perioperative care after pancreatic resections do not endorse any of the proposed fistula risk score [25]. Moreover, a recent web-based survey, distributed to almost 900 surgeons to evaluate the intra-abdominal fat thickness at CT scan and texture of pancreatic parenchyma was evaluated in 3 of 10 studies. Two scoring systems [16, 30] used a subjective intraoperative evaluation, while Gaujoux et al. [28] appraised the degree of pancreatic fibrosis and fatty infiltration by histology.

The quality of the selected study is summarized in table 3. Scores ranged from 7 to 10. In 2 of 10 studies, the definition of exposure and outcome were blinded to assessors and 4 of 10 described potential confounding factors.

Table 1 summarizes the characteristics of the selected studies. All studies but one [30] were retrospective. Nine studies involved PD and only one study addressed a mixed resection type, with a proportion of 70.8% of PD [32]. Reconstruction was standardized in 3 of 10 studies [28, 30, 34]. The median number of subjects per study was 186 (IQR 112–229) and the median fistula rate was 27.7% (IQR 23.2–31.8). Scores were associated with the prediction of any POPF grade in 3 of 10 studies and in 7 of 10 with clinically relevant POPF (B/C grade).

As shown in table 2, the variables were both subjective and objective. In 3 of 10 studies [21, 29, 35], only preoperative items were considered, while in 7 studies, intra and postoperative items were also included in the score. The median number of considered risk factors was 3 (range from 2 to 5). For statistical analyses and assessment of score accuracy, receiver characteristics-curve was performed in 6 of 10 studies [16, 28, 29, 32, 34, 35] with a range of the area under the curve from 0.780 to 0.950. Wellner et al. [21] considered the Spearman rank correlation coefficient, Ansorge et al. [30] analyzed the OR of 2 risk factors and Assifi et al. [31] used the Cochrane Armitage trend test. The median number of citations was 21 (range from 1 to 74). Validation was performed in 6 of 10 reports [16, 21, 29, 30, 32, 35] and external validation in 2 of 10 [16, 27].

The accuracy of validation set was remarkable, as depicted in table 2.

Figure 2 depicts the number of studies using specific risk factors in the score. Abdominal fat composition and MPD diameter were used in 40% of the scores. Fat mass was preoperatively assessed by calculating the body mass index (BMI) [28, 33, 35] or by evaluating the intra-abdominal fat thickness at CT scan [29]. The diameter of the MPD was measured preoperatively at CT scan [29, 35] or intraoperatively [16, 30].

The texture of the pancreatic parenchyma was evaluated in 3 of 10 studies. Two scoring systems [16, 30] used a subjective intraoperative evaluation, while Gaujoux et al. [28] appraised the degree of pancreatic fibrosis and fatty infiltration by histology.

Table 1 summarizes the characteristics of the selected studies. All studies but one [30] were retrospective. Nine studies involved PD and only one study addressed a mixed resection type, with a proportion of 70.8% of PD [32]. Reconstruction was standardized in 3 of 10 studies [28, 30, 34]. The median number of subjects per study was 186 (IQR 112–229) and the median fistula rate was 27.7% (IQR 23.2–31.8). Scores were associated with the prediction of any POPF grade in 3 of 10 studies and in 7 of 10 with clinically relevant POPF (B/C grade).

As shown in table 2, the variables were both subjective and objective. In 3 of 10 studies [21, 29, 35], only preoperative items were considered, while in 7 studies, intra and postoperative items were also included in the score. The median number of considered risk factors was 3 (range from 2 to 5). For statistical analyses and assessment of score accuracy, receiver characteristics-curve was performed in 6 of 10 studies [16, 28, 29, 32, 34, 35] with a range of the area under the curve from 0.780 to 0.950. Wellner et al. [21] considered the Spearman rank correlation coefficient, Ansorge et al. [30] analyzed the OR of 2 risk factors and Assifi et al. [31] used the Cochrane Armitage trend test. The median number of citations was 21 (range from 1 to 74). Validation was performed in 6 of 10 reports [16, 21, 29, 30, 32, 35] and external validation in 2 of 10 [16, 27].

The accuracy of validation set was remarkable, as depicted in table 2.

Figure 2 depicts the number of studies using specific risk factors in the score. Abdominal fat composition and MPD diameter were used in 40% of the scores. Fat mass was preoperatively assessed by calculating the body mass index (BMI) [28, 33, 35] or by evaluating the intra-abdominal fat thickness at CT scan [29]. The diameter of the MPD was measured preoperatively at CT scan [29, 35] or intraoperatively [16, 30].

The texture of the pancreatic parenchyma was evaluated in 3 of 10 studies. Two scoring systems [16, 30] used a subjective intraoperative evaluation, while Gaujoux et al. [28] appraised the degree of pancreatic fibrosis and fatty infiltration by histology.

The quality of the selected study is summarized in table 3. Scores ranged from 7 to 10. In 2 of 10 studies, the definition of exposure and outcome were blinded to assessors and 4 of 10 described potential confounding factors.

Table 1 summarizes the characteristics of the selected studies. All studies but one [30] were retrospective. Nine studies involved PD and only one study addressed a mixed resection type, with a proportion of 70.8% of PD [32]. Reconstruction was standardized in 3 of 10 studies [28, 30, 34]. The median number of subjects per study was 186 (IQR 112–229) and the median fistula rate was 27.7% (IQR 23.2–31.8). Scores were associated with the prediction of any POPF grade in 3 of 10 studies and in 7 of 10 with clinically relevant POPF (B/C grade).

As shown in table 2, the variables were both subjective and objective. In 3 of 10 studies [21, 29, 35], only preoperative items were considered, while in 7 studies, intra and postoperative items were also included in the score. The median number of considered risk factors was 3 (range from 2 to 5). For statistical analyses and assessment of score accuracy, receiver characteristics-curve was performed in 6 of 10 studies [16, 28, 29, 32, 34, 35] with a range of the area under the curve from 0.780 to 0.950. Wellner et al. [21] considered the Spearman rank correlation coefficient, Ansorge et al. [30] analyzed the OR of 2 risk factors and Assifi et al. [31] used the Cochrane Armitage trend test. The median number of citations was 21 (range from 1 to 74). Validation was performed in 6 of 10 reports [16, 21, 29, 30, 32, 35] and external validation in 2 of 10 [16, 27].

The accuracy of validation set was remarkable, as depicted in table 2.

Figure 2 depicts the number of studies using specific risk factors in the score. Abdominal fat composition and MPD diameter were used in 40% of the scores. Fat mass was preoperatively assessed by calculating the body mass index (BMI) [28, 33, 35] or by evaluating the intra-abdominal fat thickness at CT scan [29]. The diameter of the MPD was measured preoperatively at CT scan [29, 35] or intraoperatively [16, 30].

The texture of the pancreatic parenchyma was evaluated in 3 of 10 studies. Two scoring systems [16, 30] used a subjective intraoperative evaluation, while Gaujoux et al. [28] appraised the degree of pancreatic fibrosis and fatty infiltration by histology.

The quality of the selected study is summarized in table 3. Scores ranged from 7 to 10. In 2 of 10 studies, the definition of exposure and outcome were blinded to assessors and 4 of 10 described potential confounding factors.
Discussion

A variety of strategies may potentially modify the risk of POPF, that is, the optimization of the patient preoperative status, the surgical technique (placement of external MPD stenting) [38], the type of pancreatic anastomosis [6, 39] and the use of somatostatin analogues [7]. Despite the remarkable effort to prevent, predict, mitigate, and treat, the burden of postoperative morbidity related to pancreatic fistula has not substantially diminished [40]. In fact, the appearance of POPF in patients undergoing PD remains the most relevant clinical hazard for its severe consequences. Thus, any attempt to predict POPF occurrence through scoring systems, identifying high-risk patients and modifying the indication to surgery or the clinical management should be fully supported.

POPF recognizes a multifactorial pathogenesis related to patient characteristics, gland features, and intraoperative events. With regard to intrinsic patient features, obesity seems to be a relevant factor. In fact, 3 of 10 studies [28, 33, 35] identified an elevated BMI as a significant predictor of POPF. We recently reported that, rather than BMI, the distribution of excessive abdominal fat, measured at preoperative CT scan, strongly predicted the onset of clinically relevant POPF [41]. These data are in line with the findings of Yamamoto et al. [29] and several other authors [11, 12, 14] who reported a significant association between increased adipose abdominal compartment and postoperative complications after pancreatic surgery. Yet, we excluded the above studies from the present review since radiological features were not used to create a risk metric system to predict the likelihood of POPF onset. It may be speculated that the use of BMI and other adiposity measures are just surrogates of pancreatic fat content or gland softness. Nonetheless, BMI and pancreatic texture have been identified as 2 independent risk factors for POPF [42]. This might be partially explained by the arising concept of visceral fat as an endocrine or-

Fig. 1. Study selection according to the PRISMA statement.
gan, capable of modulating inflammatory pathways [43] and consequently predisposing to POPF occurrence. Since the distribution rather than the absolute quantity of body fat seems relevant, means of objective measure of fat compartments, such as CT scan or magnetic resonance, may be additional helpful tools.

The intrinsic characteristics of the pancreatic gland appear to be the second strong determinant of POPF risk. It has been repeatedly shown that soft pancreatic texture and small MPD diameter are highly predictive of fistula onset. Four studies [16, 29, 30, 35] considered MPD diameter to be a parameter in the risk stratification and 3 trials [16, 28, 30] evaluated pancreatic texture. In the studies by Ansorge et al. [30] and Callery et al. [16], the surgeon subjective evaluation of pancreatic consistency was considered one of the items to calculate risk scores. Despite the subjectivity of the manual perception of the pancreatic stiffness may limit the reproducibility of the score, the surgeon evaluation remains the gold standard for pancreatic texture assessment [40]. Gaujoux et al. [28] proposed an objective measurement of the degree of fatty/fibrosis infiltration of the pancreatic specimen by pathological examination. Yet, the advantages of objectivity and reproducibility are blunted by the lack of practicality due to the time of assessment. In fact, the delayed information may limit the possibility of tailoring perioperative strategies for high-risk patients. The histologic score proposed by Belyaev et al. [44] suffers comparable limitations.

Useful information may be achieved by CT scan [13, 41], magnetic resonance [45] or instruments such as the durometer [46], with the aim of reducing judgment bias and maintaining the opportunity of evaluating pancreatic texture in a pre- or intraoperative setting.

Intraoperative events are the third relevant factor identified as predictors of POPF. Even before the ISGPF definition, an excessive intraoperative bleeding was considered as a risk variable [47]. Estimated blood loss was found to be a relevant parameter in 2 studies and the amount was used to calculate the risk score. Ross et al. [48] suggested

---

Table 1. Characteristics of the included studies

<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>Design</th>
<th>Population</th>
<th>POPF grade</th>
<th>POPF rate, %</th>
<th>Reconstruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wellner [21]</td>
<td>2010</td>
<td>Retrospective monocentric</td>
<td>62</td>
<td>A–C</td>
<td>30.6</td>
<td>PJ/PG</td>
</tr>
<tr>
<td>Gaujoux [28]</td>
<td>2010</td>
<td>Retrospective monocentric</td>
<td>100</td>
<td>A–C</td>
<td>31</td>
<td>PG</td>
</tr>
<tr>
<td>Yamamoto [29]</td>
<td>2011</td>
<td>Retrospective monocentric</td>
<td>279</td>
<td>B/C</td>
<td>53</td>
<td>PJ</td>
</tr>
<tr>
<td>Ansorge [30]</td>
<td>2012</td>
<td>Prospective observational single center</td>
<td>164</td>
<td>A–C</td>
<td>21.8 (A) 15.5 (B/C)</td>
<td>PJ</td>
</tr>
<tr>
<td>Assifi [31]</td>
<td>2012</td>
<td>Retrospective monocentric</td>
<td>553</td>
<td>A–C</td>
<td>11</td>
<td>PJ</td>
</tr>
<tr>
<td>Callery [16]</td>
<td>2013</td>
<td>Retrospective monocentric</td>
<td>233</td>
<td>A–C</td>
<td>24.7</td>
<td>PJ/PG</td>
</tr>
<tr>
<td>Fujiwara [32]</td>
<td>2013</td>
<td>Retrospective monocentric</td>
<td>208</td>
<td>B/C</td>
<td>20.2</td>
<td>NA</td>
</tr>
<tr>
<td>Graham [33]</td>
<td>2013</td>
<td>Retrospective monocentric</td>
<td>146</td>
<td>A–C</td>
<td>34</td>
<td>PJ</td>
</tr>
<tr>
<td>Kosaka [34]</td>
<td>2014</td>
<td>Retrospective monocentric</td>
<td>100</td>
<td>B/C</td>
<td>15 (A) 32 (B)</td>
<td>PJ</td>
</tr>
<tr>
<td>Roberts [35]</td>
<td>2014</td>
<td>Retrospective monocentric</td>
<td>217</td>
<td>A–C</td>
<td>23.7</td>
<td>PJ/PG</td>
</tr>
</tbody>
</table>

PJ = Pancreateojejunostomy; PG = pancreatogastrostomy; NA = not available.
<table>
<thead>
<tr>
<th>First author</th>
<th>Number of items</th>
<th>Timing of assessment</th>
<th>Subjective/objective</th>
<th>Statistical analysis</th>
<th>Accuracy</th>
<th>Scale</th>
<th>Citations</th>
<th>Validation</th>
<th>Accuracy in validation set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wellner [21] 5</td>
<td>Pre</td>
<td>O</td>
<td>Spearman rank correlation</td>
<td>$r = 0.47$</td>
<td>$p &lt; 0.001$</td>
<td>(-3)–2</td>
<td>30</td>
<td>Yes, consecutive cohort</td>
<td>$r = 0.35$</td>
</tr>
<tr>
<td>Gaujoux [28] 3</td>
<td>Pre/post</td>
<td>S/O</td>
<td>OR</td>
<td>AUC $= 0.78$</td>
<td>0–3</td>
<td>74</td>
<td>No</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Yamamoto [29] 5</td>
<td>Pre</td>
<td>O</td>
<td>AUC</td>
<td>AUC $= 0.808$</td>
<td>0–7</td>
<td>13</td>
<td>Yes, consecutive cohort</td>
<td>AUC $= 0.834$</td>
<td></td>
</tr>
<tr>
<td>Ansorge [30] 2</td>
<td>Intra</td>
<td>S/O</td>
<td>OR</td>
<td>OR 12.18, $95%$ CI 4.20–35.34; $p &lt; 0.001$</td>
<td>0–2</td>
<td>1</td>
<td>Yes, other centers</td>
<td>SE = 100%</td>
<td>SP = 100%</td>
</tr>
<tr>
<td>Assifi [31] 3</td>
<td>Intra</td>
<td>O</td>
<td>Cochran-Armitage trend test</td>
<td>$p = 0.043$</td>
<td>0–10</td>
<td>22</td>
<td>No</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Callery [16] 4</td>
<td>Intra/post</td>
<td>S/O</td>
<td>AUC</td>
<td>OR (95% CI) for: Soft gland: 5.02 (1.97–12.81) Pathology: 2.98 (1.36–6.54) MPD diameter: 9.66 (2.32–40.26) Blood loss: 3.99 (1.20–13.21)</td>
<td>0–10</td>
<td>60</td>
<td>Yes, consecutive cohort and other centers</td>
<td>AUC $= 0.942$</td>
<td></td>
</tr>
<tr>
<td>Fujiwara [32] 2</td>
<td>Post</td>
<td>O</td>
<td>AUC</td>
<td>AUC albumin: 0.621; $p = 0.015$ AUC CRP: 0.644; $p = 0.004$</td>
<td>0–2</td>
<td>1</td>
<td>Yes, consecutive cohort</td>
<td>OR 18.45, 95% CI 1.69–201.1 $p &lt; 0.001$, (Joncheere-Terpstra test)</td>
<td></td>
</tr>
<tr>
<td>Graham [33] 4</td>
<td>Pre/post</td>
<td>O</td>
<td>OR</td>
<td>SE = 0.72, SP = 0.81; $p &lt; 0.001$</td>
<td>0–1</td>
<td>1</td>
<td>No</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Kosaka [34] 3</td>
<td>Post</td>
<td>O</td>
<td>AUC</td>
<td>AUC $= 0.95$; $p &lt; 0.01$</td>
<td>0–1</td>
<td>4</td>
<td>No</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Roberts [35] 2</td>
<td>Pre</td>
<td>O</td>
<td>ROC</td>
<td>AUC $= 0.832$; $p &lt; 0.001$</td>
<td>0–1</td>
<td>11</td>
<td>Yes, consecutive cohort</td>
<td>$p &lt; 0.001$, (Joncheere-Terpstra test)</td>
<td></td>
</tr>
</tbody>
</table>

AUC = Area under curves; ROC = receiver characteristics curves; NA = not available; SE = sensitivity; SP = specificity; PPV = positive predictive value; NPV = negative predictive value.
that elevated blood loss, with subsequent need of fluid replacement, should be considered an indirect measure of technical difficulties. In the last decades, the debate on fluid therapy in surgical patients yielded to growing evidences that a positive fluid balance and subsequent overload cause tissue edema, activation of inflammatory pathways and poor wound healing [49–51]. All these elements may contribute to increase the risk of pancreatic fistula.

We recognized several critical elements in the evaluation of the overall quality and value of the scoring systems. One is the timing of the item assessment. The level of amylase in the drain, increased serum C-reactive protein (CRP) and leukocyte count on postoperative day 4 [34] are more likely to be early clinical signs of POPF rather than predictive factors. Ideally risk evaluation should guide preoperative counseling and help tailoring periop-
operative strategies. Thus, early signs of POPF prediction might be of little use in the clinical setting. For this reason, an ideal score should include only pre- and intraoperative items.

Another puzzling element in the analyzed metrics is the use of generic parameters, such as variations of mean arterial pressure, heart rate, CRP values and white blood cell count [31, 32]. These factors may be of value in predicting generic postoperative complications but are less specific for POPF, which is peculiar of pancreatic resection.

We also observed conflicting results on the role of some risk factors. Wellner et al. [21] found that active smoking habit and a poor nutritional status were protective on POPF development, probably because they promote a fibrotic transformation of the pancreatic parenchyma. Other authors reported that malnutrition was associated with an increased rate of pancreatic fistula [52, 53] and active smoking correlated with a higher rate of overall complications and mortality after pancreatic resections [54].

From a methodological standpoint several limitations should be highlighted. First, the scores proposed by Assifi et al. [31], Graham et al. [33] and Roberts et al. [35] did not stratify the results on POPF grading. It has been clearly established that grade A fistula is clinically irrelevant, so that its prediction seems of marginal value in daily practice. Moreover, Graham et al. [33] score has been described in a short communication not allowing an accurate and critical validation of the methodology. Second, all but one of the analyzed studies [30] were retrospective, blind outcome assessment was described in 2 trials [28, 35] and the different statistical methods used to evaluate the metric accuracy did not allow a direct comparison among all the trials. Third, surgical and perioperative procedures were not standardized increasing the possibility of interference of technical and management features on the POPF onset.

Fourth, we tried to assess the study quality by using a modified checklist [27] even though gold standards and references to estimate the validity of observational research are lacking. Indeed, it has been suggested that the application of numerical scales in systematic reviews are of limited value, since the global evaluation limits the ability to judge the degree and weight of bias [55].

The ultimate POPF risk metrics should be easy, feasible, accurate, objective, reproducible and transferable. The model should take into account preoperative patient-related and gland-related features and intraoperative technical aspects. None of the published systems completely adhere to these principles, although the score proposed by Ansorge et al. [30] and Callery et al. [16] has been validated by other centers, suggesting reproducible results in different settings [56–58].

We endorse the necessity of a large heterogeneous multicentric validation set, regardless of the surgical technique or habits related to the local practice, to account for case-mix and to evaluate the accuracy and reproducibility of each scoring system.

**Funding**

There was no funding for this work.

---

**References**


