Meta-analysis of drain amylase content on postoperative day 1 as a predictor of pancreatic fistula following pancreatic resection

M. C. Giglio, D. R. C. Spalding, A. Giakoustitdis, A. Zarzavadjian Le Bian, L. R. Jiao, N. A. Habib and M. Pai

HPB Surgical Unit, Department of Surgery and Cancer, Imperial College, Hammersmith Hospital Campus, Du Cane Road, London W12 0HS, UK

Correspondence to: Mr M. Pai (e-mail: madhava.pai@imperial.ac.uk)

Background: Drain amylase content in the days immediately after major pancreatic resection has been investigated previously as a predictor of postoperative pancreatic fistula (POPF). Its accuracy, however, has not been determined conclusively. The purpose of this study was to evaluate the accuracy of drain amylase content on the first day after major pancreatic resection in predicting the occurrence of POPF.

Methods: A literature search of the MEDLINE, Embase and Scopus® databases to 13 May 2015 was performed to identify studies evaluating the accuracy of drain amylase values on day 1 after surgery in predicting the occurrence of POPF. The area under the hierarchical summary receiver operating characteristic (ROC) curve (AUC$_{hSROC}$) was calculated as an index of accuracy, and pooled estimates of accuracy indices (sensitivity and specificity) were calculated at different cut-off levels. Subgroup and meta-regression analyses were performed to test the robustness of the results.

Results: Thirteen studies involving 4416 patients were included. The AUC$_{hSROC}$ was 0.89 (95 per cent c.i. 0.86 to 0.92) for clinically significant POPF and 0.88 (0.85 to 0.90) for POPF of any grade. Pooled estimates of sensitivity and specificity were calculated for the different cut-offs: 90–100 units/l (0.96 and 0.54 respectively), 350 units/l (0.91 and 0.84) and 5000 units/l (0.59 and 0.91). Accuracy was independent of the type of operation, type of anastomosis performed and octreotide administration.

Conclusion: Evaluation of drain amylase content on first day after surgery is highly accurate in predicting POPF following major pancreatic resection. It may allow early drain removal and institution of an enhanced recovery pathway.
**A: Introduction**

In recent years, as a result of advances in surgical techniques and centralization of services\(^1\), the perioperative mortality rate following major pancreatic resections has decreased to less than 5 per cent\(^2,3\). Despite this achievement, morbidity rates remain high, even in large-volume centres\(^4\). Postoperative pancreatic fistula (POPF) remains a major cause of morbidity; it has a reported incidence of 35 per cent\(^4,5\), and can lead to intra-abdominal collections, subsequent infection with sepsis and even death\(^6,7\).

In 2005, the International Study Group on Pancreatic Fistula (ISGPF) provided a clear definition of POPF that has since been accepted worldwide, defining POPF as an amylase level in the drain fluid three times higher than the upper normal serum value on or after postoperative day (POD) \(^3,8\). Based on these criteria and clinical impact, pancreatic fistulas have since been classified as subclinical (POPF grade A), clinical (grade B) or severe (grade C)\(^8\). Grades B and C have now been merged into the group of clinically relevant fistula, requiring a change in patient management\(^9\).

Early prediction of the likely non-occurrence of a POPF would be clinically useful, allowing a patient to be placed safely on a fast-track postoperative protocol\(^10\). Several preoperative, intraoperative and postoperative risk factors have been analysed previously and found to be associated with POPF\(^3,11–14\). A number of recent studies\(^1,12,15–17\) have emphasized the importance of early postoperative drain amylase values, by evaluating the accuracy of this simple test in predicting POPF.

The aim of the present study was to analyse current evidence of the importance of drain amylase levels, explore potential sources of heterogeneity, and provide a qualitative and quantitative analysis of the accuracy of drain amylase values in the prediction of POPF.

**B: Methods**

**Search strategy**

MEDLINE, Embase and Scopus\(^\circ\) databases were searched, with no language or publication status restrictions, using the following search terms: “amylase”, “drain\(^*\)”, “whipple\(^*\)”, “pancreat\(^*\)”, “fistula” (Appendix S1, supporting information). The last search was run on 13 May 2015. The reference list of retrieved articles was used to identify additional eligible studies.

**Study selection**

To be eligible, a study had to include adult patients undergoing elective pancreatic resection for any disease, evaluate the accuracy of drain amylase levels on POD 1 in predicting the development of a POPF, use an appropriate reference standard for diagnosis of POPF (ISGPF criteria or dedicated imaging), and report absolute numbers of true-positive, false-positive, true-negative and false-
negative observations or other information allowing these data to be calculated. Conference abstracts and case–control studies were excluded, as they may have led to a spectrum bias resulting in overestimation of the accuracy of a test. Potentially eligible studies were identified by two reviewers who assessed the title and abstracts of retrieved references. Divergences, if any, were resolved by a third reviewer. The full text of these studies was then retrieved for further analysis.

+B: Data extraction and quality assessment

Two authors independently extracted or calculated relevant data from each included study, and entered the following information into an electronic database: first author, publication year, country of origin, study design, number of patients, reference test used as the standard for diagnosis of POPF, rate and type of POPF, cut-off value(s) tested, the POD on which the drain amylase content was evaluated, absolute number of true-positive, true-negative, false-positive and false-negative observations, type and number of drains used, prophylactic use of octreotide in the perioperative period, type of resection and type of pancreatic–enteric anastomosis, as well as the percentage of each type in the study. Corresponding authors were contacted to retrieve information regarding the follow-up, if not reported in the study. In addition, one author was contacted to clarify the cut-off value used. The quality of each study was assessed using the quality assessment of diagnostic accuracy studies instrument (QUADAS-2). In particular, this tool assesses the risk of bias and concerns about the applicability of the test in each study by exploring four domains: patient selection, index test, reference standard, and time and flow (including follow-up).

+B: Data synthesis

Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR) and negative LR were calculated for each cut-off. Youden’s J index (sensitivity + specificity − 1), the diagnostic odds ratio together with its standard error were also calculated as summaries of the diagnostic performance.

Two different analyses were conducted using a bivariable random-effects model. In the first analysis, as the aim was to define the accuracy of drain amylase values in predicting POPF, a hierarchical summary receiver operating characteristic (hSROC) curve, which plots the individual estimates of sensitivities and specificities, was generated. When a study reported multiple estimates of different cut-offs, the one with the highest Youden’s J index was used to avoid duplicate reporting of the same population. The area under the hSROC curve (AUC$_{hSROC}$) was calculated as an indicator of diagnostic accuracy; an AUC$_{hSROC}$ of between 0.9 and 1.0 indicated that the test in question was highly accurate. In the second analysis, the aim was to evaluate pooled accuracy indices (sensitivity, specificity, positive and negative LR) for the cut-offs tested in the studies, as this information would be useful in clinical practice. PPV and NPV were not pooled,
as they depend strongly on the prevalence of POPF in the population and therefore have no external validity. The pooled analysis was performed for each cut-off evaluated at least in three studies. It was planned a priori to group and analyse the accuracy of different but close cut-offs, falling in an interval wide 100 units/l. The presence of a threshold effect was searched for by the analysis of the correlation (Spearman’s ρ) between sensitivity and specificity.

The Q value and the inconsistency index ($I^2$) test were used to estimate the heterogeneity between each study$^{26}$. $I^2$ values of 25, 50 and 75 per cent were considered to be indicative of low, moderate and high statistical heterogeneity respectively$^{26}$. When heterogeneity was present, its possible sources were investigated at different levels. The proportion of heterogeneity due to threshold effects ($P^{2}_{THR}$) was estimated as the squared correlation coefficient estimated from the between-study covariance parameter. Residual heterogeneity not explained by the threshold effect was calculated as $P_{RES}^2 = (1 - I^{2}_{THR}) P^2$.

The robustness of the results was tested further. Subgroup analysis was performed of studies using the same reference standard and including only pancreaticoduodenectomies. To identify whether the drain amylase accuracy (estimated as log diagnostic odds ratio and its s.e.) was influenced by study co-variables, a random-effects meta-regression analysis$^{27}$ was performed. The variables, POPF prevalence, cut-off, prophylactic use of octreotide, type of resection and pancreatic-enteric anastomosis, as well as the percentage of each type, and results of QUADAS-2 assessment for each domain, were tested in this analysis.

The Fagan nomogram was elaborated for each cut-off analysed; it is based on the pooled positive and negative LR values, and shows the influence of a test’s result on the probability of a patient having a disease$^{28}$. The mean prevalence of the disease in the included studies was used as a pretest probability.

+B: Statistical analysis

The Review Manager Calculator tool (RevMan, version 5.3; The Cochrane Collaboration, The Nordic Cochrane Centre, Copenhagen, Denmark) was used to determine accuracy values (sensitivity, specificity, PPV, NPV, positive and negative LR) from the available data in the selected studies. Other statistical analyses were performed using STATA® 12 statistical software (StataCorp, College Station, Texas, USA). Although the present study focused on the predictive accuracy of a test, recommendations for systematic reviews of diagnostic test accuracy$^{29,30}$ were followed; these are the only available guidelines that oversee the application of the statistical methods used. Results were reported according to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement$^{31}$. 
+A: Results

+B: Study selection

The search of MEDLINE, Embase and Scopus® databases revealed a total of 1011 citations; 497 of these were found to be duplicated and were therefore removed. Of the remaining 514 studies, 465 were discarded because they did not meet the inclusion criteria after review of the title or abstract. The full text of each of the remaining 49 articles was examined in more detail. This led to the exclusion of a further 36 studies, leaving 13 studies that were subsequently included in the meta-analysis.\(^1,^{10,12,15–17,19,32–38}\) (Fig. 1).

+B: Study characteristics

The included studies involved 4416 patients undergoing pancreatic resection. Of these, 4079 (92.4 per cent) had a pancreatoduodenectomy, 321 (7.3 per cent) underwent distal pancreatectomy and 16 had an enucleation (0.4 per cent). The ISGPF criteria for the definition and diagnosis of POPF were used as a reference standard by all but two of the studies, in which the Cambridge definition of POPF\(^38\) and CT findings\(^17\) were used as the reference standard.

All studies graded POPF using the ISGPF criteria. Drain amylase levels were evaluated as a predictor of POPF of any grade in seven studies\(^1,^{15,16,32–34,37,38}\), as a predictor of clinically relevant POPF in four studies\(^^{10,12,17,36}\) and as a predictor of both in two studies\(^^{19,37}\). The prevalence of POPF of any grade ranged from 0.10 to 0.43, and the prevalence of clinically relevant POPF ranged from 0.10 to 0.39. Table S1 (supporting information) shows the characteristics of the included studies.

+B: Quality assessment

Four studies\(^{16,19,34,36}\) were retrospective. The QUADAS-2 instrument highlighted concerns of bias in the index test category in seven studies as their threshold was not prespecified (Fig. S1, supporting information). In the patient selection domain, concerns were raised over a multicentre study by Lee and colleagues\(^{33}\), in which drains were not used routinely in all patients. The authors did not explain the reason behind this choice, which may reflect variability in preference of individual surgeons; however, the drain might have been positioned selectively in patients at greater risk of developing POPF, thus representing a selection bias. A quality score was not generated owing to problems associated with the use of such scores\(^{39,40}\); the results of QUADAS-2 assessment, however, were considered in the meta-regression analysis.

+B: Drain amylase accuracy

The accuracy of drain amylase level on POD 1 as a predictor of POPF of any grade was tested in eight studies\(^1,^{15,16,32–34,37,38}\): the AUC\(_{\text{hSROC}}\) was 0.88 (95 per cent c.i. 0.85 to 0.90) (Fig. 2a), and there was residual heterogeneity (\(Q = 60, P = 0.001\); \(I^2 = 97\) per cent; \(I^2\text{THR} = 0.36; I^2\text{RES} = 62\) per cent). In five studies\(^{10,12,17,19,36}\) drain amylase content on POD 1 was tested as predictor of clinically
relevant POPF: the hSROC showed high predictive accuracy (AUC_{hSROC} 0.89, 95 per cent c.i. 0.86 to 0.92) (Fig. 2b). Heterogeneity was completed explained by the presence of a threshold effect \( (Q = 79, P = 0.013; I^2 = 97\) per cent; \( I^2_{\text{THR}} = 1; I^2_{\text{RES}} = 0 \) per cent).

+B: Cut-off analysis

Eleven different cut-off values were analysed on POD 1 in the included studies (Table S2, supporting information). The existence of a threshold effect was confirmed by the correlation between estimated values of sensitivity and specificity \( (r_S = -0.595, P = 0.033) \). The lower the cut-off value, the higher the sensitivity and the lower the specificity (Fig. S2, supporting information).

For the prediction of POPF of any grade, a cut-off value of 90–100 units/l (group analysis)\(^{16,32,33}\), 350 units/l\(^{1,10,32,33}\) and 5000 units/l\(^{15,16,32,34}\) had pooled sensitivities of 0.96 (95 per cent c.i. 0.88 to 0.98), 0.91 (0.76 to 0.97) and 0.59 (0.40 to 0.76), and specificities of 0.54 (0.37 to 0.70), 0.84 (0.59 to 0.95) and 0.91 (0.84 to 0.95) respectively (Table 1). Other estimates of accuracy indices of drain amylase values using these cut-offs are shown in Table S2 (supporting information).

Fagan diagrams were used to show the clinical impact of the test using the first and last cut-off values. There was a 27 per cent pretest probability of developing a POPF of any grade (mean prevalence in the included studies). The pretest probability of developing a POPF of any grade after a negative drain amylase test (cut-off 100 units/l) was 3 per cent (Fig. 3a), and that after a positive test (cut-off 5000 units/l) was 70 per cent (Fig. 3b).

+B: Subgroup and meta-regression analyses

Results were confirmed when the analysis was restricted to the nine studies\(^{1,10,12,16,17,19,36–38}\) that only included patients undergoing pancreaticoduodenectomies (AUC_{hSROC} 0.87, 95 per cent c.i. 0.84 to 0.92). Two studies\(^{17,38}\) used reference standards other than the ISGPF definition, but this did not affect the results, which were confirmed after excluding these two studies (AUC_{hSROC} 0.88, 0.85 to 0.90).

Quality assessment using the QUADAS-2 tool showed a risk of bias in patient selection in the study of Lee and co-workers\(^{33}\). After exclusion of this study from the analysis, the accuracy of drain amylase values in predicting POPF still remained high (AUC_{hSROC} 0.91, 88 to 93), and, interestingly, there was no residual heterogeneity \( (Q = 58, P = 0.001; I^2 = 97\) per cent; \( I^2_{\text{THR}} = 1; I^2_{\text{RES}} = 0 \) per cent). The meta-regression analysis showed the accuracy of drain amylase content to be independent of all investigated co-variables, including type of operation, type of anastomosis, use of external pancreatic stents and octreotide prophylaxis (Table S3, supporting information). The risk of bias as assessed by the QUADAS-2 tool did not influence the results.
A: Discussion

This meta-analysis demonstrates that estimation of the drain amylase content on POD 1 is highly accurate in predicting the occurrence of POPF. This confirms the usefulness of this simple and economic test in playing a significant role in the postoperative management of these patients. Indeed, evaluation of drain amylase levels on POD 1 allows the identification of a group of patients at low risk of POPF who can proceed safely to early drain removal and an enhanced recovery pathway (ERP) after surgery.

In recent years there has been a lively debate regarding drain versus no drainage after major pancreatic resections. The idea that drains might be more harmful than helpful has been supported by some evidence; however, a recent randomized trial found significantly increased morbidity and mortality in patients who did not undergo drainage. Against this background, early drain removal is presented as an excellent compromise. Molinari, Bassi and colleagues first showed a correlation between drain amylase content on POD 1 and the incidence of POPF, and this Italian group has subsequently also shown the benefits of early drain removal driven by drain amylase values on POD 1. These authors randomized patients at low risk of POPF (amylase level less than 5000 units/l) to an early (POD 3) versus late (POD 5 or later) removal of the drain. They reported a decreased incidence of POPF (1.8 versus 26.3 per cent), abdominal and pulmonary complications, length of hospital stay, and cost in the early drain removal group. Kawai and co-workers also found a lower rate of intra-abdominal infection in patients who had drain removal on POD 4, compared with a historical cohort in which the drain was removed on POD 8 (8 versus 38 per cent respectively). Based on the drain amylase level on POD 1, Sutcliffe et al. implemented an ERP in patients undergoing pancreaticoduodenectomy, with early oral intake (oral fluids on POD 1, diet in the morning of POD 2) and early drain removal (POD 3) in low-risk patients (POD 1 drain amylase level below 350 units/l). These patients had a shorter hospital stay (7 versus 9 days) and a lower 30-day readmission rate (2 versus 17 per cent respectively).

In the present meta-analysis, the sensitivity and specificity of the drain amylase test were found to be strictly dependent on the chosen cut-off, as a threshold effect exists. The lower the cut-off value used, the higher the sensitivity and the lower the specificity. Therefore, the use of a single cut-off value could render the test neither sensitive nor specific enough to be useful in a clinical setting. Adopting two cut-off values, the lower to improve sensitivity (rule-out value) and the higher to improve the specificity (rule-in value), would therefore be more useful.

To select patients at low-risk of POPF, a ruling-out cut-off value with high-sensitivity is needed to reduce the number of false-negative results. Moreover, this cut-off needs to have high
applicability, encompassing a reasonable proportion of patients who would benefit from early drain removal or an ERP\textsuperscript{37}. In the pooled analysis, cut-off values of 100 units/l and 350 units/l had sensitivities of 0.96 and 0.91 respectively, and encompassed 34 and 65 per cent of the population respectively. Thus, at present a drain amylase cut-off value of 350 units/l might represent the best compromise between safety and applicability.

Patients at high risk of POPF can be identified by using a higher, ruling-in, drain amylase cut-off with higher specificity (few false-positive results). In pooled analysis, a cut-off value of 5000 units/l had the highest specificity (0.91) and would therefore be ideal for this aim. These patients might benefit from the prophylactic administration of somatostatin analogues\textsuperscript{49} and close monitoring, in order to detect and treat POPF-related complications early\textsuperscript{48}. Nonetheless, a proportion of these patients will develop only a subclinical POPF, necessitating no change in clinical management\textsuperscript{9}. For this reason, further prophylactic actions in this group of patients (nil by mouth, intravenous parenteral nutrition) could be of no value and potentially harmful.

A major concern regarding the use of drain amylase content is the external validity of a single-institution experience\textsuperscript{37}. Indeed, a number of technical choices vary across different institutions, such as the type of anastomosis performed, use of external pancreatic stents, number and type of drains placed, and use of perioperative administration of octreotide (Table S1, supporting information). Each of these factors might theoretically affect the predictive accuracy of drain amylase levels on POD 1. This hypothesis was disproved by the meta-regression analysis, which showed the accuracy of the test to be independent of all these factors (Table S3, supporting information). The absence of statistical heterogeneity in the accuracy analysis, despite the varying techniques, further corroborates this fact. This suggests that, regardless of the surgeon’s choice, whenever a POPF develops the drain amylase test on POD 1 is a good tool to predict this event.

A further impediment to external validity and applicability of the test might be the presence of variation bias – a test performance significantly affected by the pretest probability (the risk of POPF in the population). The presence of variation bias was excluded, as the accuracy of the drain amylase content on POD1 in predicting POPF was shown to be independent of the prevalence of POPF (Table S3, supporting information).

NPV and PPV, in contrast to sensitivity, specificity, and positive and negative LR, are strongly dependent on pretest probability and therefore have no external validity. For this reason NPV and PPV were not pooled. However, the Fagan nomograms (Fig. 3) demonstrate the impact of testing drain amylase values on POD 1 in their own setting: the post-test probability of developing
POPF can easily be calculated when the pretest probability (the incidence of POPF in the institution or in a particular subgroup of patients) is known.

This study has some limitations. A number of cut-off values have been tested by different institutions, but few centres validated in their own population values against those proposed by others. Therefore, the pooled analysis was limited to the three most commonly used cut-offs. Further studies evaluating the sensitivity and specificity of several cut-off values (Table S2, supporting information), even on data collected retrospectively, would be useful in providing further external validation of these values and facilitating future meta-analyses. No heterogeneity was present in the pooled analysis of lower, high sensitivity, cut-offs (100 and 350 units/l). Conversely, analysis of the 5000 units/l cut-off value found significant heterogeneity ($I^2 = 85$ per cent), which remained unexplained. However, this heterogeneity is due to variation in sensitivity across the studies (from 0.39 to 0.71) (Table S2, supporting information), rather than in specificity (from 0.84 to 0.94), which is the characteristic for which this cut-off should be used in clinical settings. Although the findings of this study support the external validity and applicability of this test, a further prospective validation of the test for different patient populations with a well defined risk of POPF and set of surgeons is necessary. Finally, the safety and benefits of postoperative management tailored on drain amylase content on POD 1 need to be shown in the field, by adding further high-quality evidence to the literature15,48.

+A: Disclosure
The authors declare no conflict of interest.

+A: References


Operative drainage following pancreatic resection: analysis of 1122 patients resected over Correa
pancreaticoduodenectomy?

Wang Q, Jiang Y


Supporting information

Additional supporting information may be found in the online version of this article:

**Fig. S1** Results of the study quality assessment by the QUADAS-2 tool (EPS file)

**Fig. S2** Relationship between the cut-off used in the study and estimated values of sensitivity and specificity (EPS file)

**Fig. S3** Forest plot of the pooled sensitivity and specificity of drain amylase values on postoperative day 1 in the prediction of postoperative pancreatic fistula of any grade at different cut-off levels (EPS file)

**Appendix S1** Electronic search strategy for MEDLINE (Word document)

**Table S1** Characteristics of included studies (Word document)

**Table S2** Summary of reported values for accuracy of amylase drain content on postoperative day 1 in predicting postoperative pancreatic fistula (Word document)

**Table S3** Results of univariable random-effects meta-regression analysis to test the potential effect of co-variables on drain amylase accuracy (Word document)

<TYPESETTER: PLEASE FOLLOW MARK-UPS OF FIGS 1-3>

**Fig. 1** PRISMA diagram showing selection of articles for review. POD, postoperative day

**Fig. 2** Hierarchical summary receiver operating characteristic curves evaluating the accuracy of drain amylase values on postoperative day 1 in predicting postoperative pancreatic fistula (POPF) of **a** any grade and **b** clinically significant POPF. Each symbol represents a study in the meta-analysis, and the area under the curve (AUC) represents the diagnostic accuracy. **a** AUC 0.88 (95 per cent c.i. 0.85 to 0.90), **b** AUC 0.89 (0.85 to 0.92)

**Fig. 3** Fagan's nomograms. A pretest probability of 24 per cent for postoperative pancreatic fistula (POPF) of any grade was fixed, as estimated by the mean prevalence of POPF in the included studies. **a** For a drain amylase level of 100 units/l or less on postoperative day (POD) 1, the post-test probability of developing a POPF is 3 per cent, with a negative likelihood ratio. **b** For a drain amylase level above 5000 units/l on POD 1, the post-test probability of POPF is 70 per cent, with a positive likelihood ratio
Table 1 Summary of pooled accuracy indices of drain amylase on postoperative day 1 at different cut-off values to predict postoperative pancreatic fistula of any grade

<table>
<thead>
<tr>
<th>Cut-off (units/l)</th>
<th>No. of studies</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive LR</th>
<th>Negative LR</th>
<th>Heterogeneity (Q test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>90–100</td>
<td>3,10,32,33</td>
<td>0.96 (0.88, 0.98)</td>
<td>0.54 (0.3, 0.70)</td>
<td>2.1 (1.4, 3.0)</td>
<td>0.08 (0.03, 0.22)</td>
<td>0.161 0</td>
</tr>
<tr>
<td>350</td>
<td>4,10,32,33</td>
<td>0.91 (0.76, 0.97)</td>
<td>0.84 (0.59, 0.95)</td>
<td>5.8 (1.80, 18.9)</td>
<td>0.10 (0.03, 0.36)</td>
<td>0.341 0</td>
</tr>
<tr>
<td>5000</td>
<td>6,16,32–34</td>
<td>0.59 (0.40, 0.76)</td>
<td>0.91 (0.84, 0.95)</td>
<td>7.6 (3.5, 16.4)</td>
<td>0.43 (0.25, 0.74)</td>
<td>&lt; 0.010 85</td>
</tr>
</tbody>
</table>

Values in parentheses are 95 per cent c.i. LR, likelihood ratio.