



Published in final edited form as:

J Am Coll Surg. 2014 September ; 219(3): 407–415. doi:10.1016/j.jamcollsurg.2014.01.064.

Variations in Definition and Method of Retrieval of Complications Influence Outcomes Statistics after Pancreatoduodenectomy: Comparison of NSQIP with Non-NSQIP Methods

Dominic E Sanford, MD¹, Cheryl A Woolsey, PA-C¹, Bruce L Hall, MD, PhD, MBA, FACS², David C Linehan, MD, FACS¹, William G Hawkins, MD, FACS¹, Ryan C Fields, MD, FACS¹, and Steven M Strasberg, MD FACS FRCS(C) FRCS(Ed)¹

¹Section of Hepato-Pancreato-Biliary Surgery, Washington University in St Louis, and Barnes-Jewish Hospital St Louis MO

²Washington University in St Louis Department of Surgery, Olin Business School, and Center for Health Policy; St Louis VA Medical Center; BJC Healthcare Saint Louis

Abstract

Background—NSQIP and the Accordion Severity Grading System have recently been used to develop quantitative methods for measuring the burden of postoperative complications. However, other audit methods such as chart reviews and prospective institutional databases are commonly used to gather postoperative complications. The purpose of this study was to evaluate discordance between different audit methods in pancreatoduodenectomy - a common major surgical procedure. The chief aim was to determine how these different methods could affect quantitative evaluations of postoperative complications.

Study Design—Three common audit methods were compared to NSQIP in 84 patients who underwent pancreatoduodenectomy. The methods were: use of a prospective database, a chart review based on discharge summaries only, and a detailed retrospective chart review. The methods were evaluated for discordance with NSQIP and among themselves. Severity grading was performed using the Modified Accordion System.

Results—53 complications were listed by NSQIP, and 31 complications were identified that were not listed by NSQIP. There was poor agreement for NSQIP type complications between NSQIP and the other audit methods for mild and moderate complications (Kappa 0.381-0.744), but excellent agreement for severe complications (Kappa 0.953-1.00). Discordance was usually due to variations in definition of the complications in non-NSQIP methods. There was good

© 2014 American College of Surgeons. Published by Elsevier Inc. All rights reserved.

Correspondence address: Steven M. Strasberg, MD, Washington University in Saint Louis, 4990 Children's Place, Suite 1160, Box 8109, St Louis, MO 63110, Phone: 314-362-7147, Fax: 314-367-1943, strasbergs@wustl.edu.

Disclosure Information: Nothing to disclose.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

agreement among non-NSQIP methods for non-NSQIP complications for moderate and severe complications, but not for mild complications.

Conclusions—There are important differences in perceived surgical outcomes based on the method of complication retrieval. The non-NSQIP methods used in this study could not be substituted for NSQIP in a quantitative analysis unless that analysis was limited to severe complications.

Introduction

Complications are key short-term outcome measures of surgical procedures. For many years, there was no standardized reporting of complications. In 1992, a definition and a method of severity grading of postoperative complications were proposed.¹ This method has been expanded and modified by Clavien et al.² as well as our group, who presented the Accordion Severity Grading System in 2009.³ Recent efforts have been directed toward development of techniques for quantifying the burden of complications. The result has been the “Postoperative Morbidity Index” (PMI) and severity spectrograms, which display burden of complications by severity level.⁴ These methods are based on definitions of postoperative complications found in the American College of Surgeons’ National Surgical Quality Improvement Program (ACS NSQIP, or NSQIP) and the 6 level Modified Accordion Grading System.^{4,5}

Quantitative techniques require rigorous methodology. Variations either in the definition of what constitutes a complication or the method of gathering complications could significantly affect the results of a quantitative evaluation. Therefore, exact definitions and thresholds for complications as well as a reliable method of gathering complications are essential. Another essential element is a validated severity grading system with quantitatively weighted grades, such as the Modified Accordion system.³ The PMI was based on the NSQIP, because the NSQIP has a methodologically rigorous complication gathering system with high inter-rater reliability.^{4,6} It also defines a set of postoperative complications and provides exact criteria for these complications.⁷ However, the NSQIP repertoire of what is considered to be a complication is limited and complications are recorded only for the first 30 day postoperative period.

There are several common methods of gathering postoperative complications other than the NSQIP method of using trained raters who prospectively gather a specific group of highly defined complications. Among these alternatives are prospective clinical recording of complications in databases and retrospective chart reviews, which sometimes rely on detailed examination of charts or may merely extract complications from discharge summaries. The purpose of this study was to evaluate the effect of defining and gathering complications of varying severity by different methods for pancreatoduodenectomy- a common major abdominal surgical procedure. The chief aim was to determine how these different methods could affect quantitative evaluations of postoperative complications.

Methods

NSQIP Audit Methods

The NSQIP was adopted in our institution in 2001 and takes place under an Institutional Review Board (IRB) approved protocol.⁸ Collection of pancreas-specific data within NSQIP was recently instituted. This recent addition incorporates pancreas-specific complications, such as postoperative pancreatic fistula and delayed gastric emptying (Table 1). The variables and definitions used in basic NSQIP can be accessed at http://nsqip.healthsofionline.com/lib/Documents/Ch_4_Variables_Definitions_062810.pdf (accessed October 4, 2013). The specifics of how NSQIP audits are conducted have previously been described in detail.⁶ Of note, NSQIP includes a category referred to as “postoperative other occurrences” which are used at some institutions to capture non-NSQIP complications. However, our institution does not utilize this category and therefore was not taken into account in our study.

The NSQIP monitors intraoperative and postoperative transfusion. Essentially, a blood transfusion given at any time from the beginning of a procedure to 72 hours later is logged into the data collection. As this study was concerned solely with postoperative complications, it was desired to filter these data. NSQIP was developed for a broad set of procedures of different magnitudes. For some operations, such as hernia repair or cholecystectomy, blood transfusion during or within 3 days of surgery would likely be an unexpected event. However, defining any such blood transfusion as a complication in pancreatoduodenectomy might not be appropriate. Because this is an unsettled area and because this study focused on postoperative complications, transfusion was not a complication by any of the audit methods in this work.

Other Audit Methods

Three other methods of complication audit were compared to the NSQIP for the same patient group. In the first, complications were gathered prospectively by a physician assistant and discussed at weekly hepato-pancreato-biliary surgery (HPB) service case management conference. Complications were confirmed in the expert surgical group discussion, graded according to the Modified Accordion Severity Grading System,³ and entered into a prospective database. The second and third audits were performed by a physician research fellow with expertise in the field of HPB surgery. The second audit method was a retrospective review using only discharge summaries, previously dictated by surgical residents, nurse practitioners, or physician assistants. The third audit method was a more detailed retrospective review of the entire clinical record including clinical notes, investigations, consultations and procedures. Results of the second and third type of audits were also recorded in a database, and complications were also graded by the Modified Accordion Severity Grading System. The three non-NSQIP audits were not limited to gathering complications in the 25 defined NSQIP categories. Therefore, complications derived from the three non-NSQIP audits could either be of a type listed by NSQIP or not listed by NSQIP. For instance, Superficial Surgical Site Infection (Superficial SSI) is listed by NSQIP as a postoperative complication, but the complications of tachycardia or delirium are not. The diagnosis of a complication using these non-NSQIP methods was predicated on

having a gradable intervention. For example, a UTI was considered as such if the patient was treated with antibiotics, and did not take laboratory data into account, unlike NSQIP. Also, the three non-NSQIP methods were temporally aligned with NSQIP, and only included NSQIP complications which occurred within the first 30 postoperative days. All non-NSQIP audit methods were conducted under IRB approval.

To look for discordance among methods of audit, two types of analyses were conducted. The first examined discordance among audit methods for NSQIP-type complications, and the second examined non-NSQIP complications. Understandably, NSQIP results are available for comparison only for the former. In the first analysis, the NSQIP result was used as the reference or gold standard. Four types of outcomes were possible. A non-NSQIP audit could 1) be in agreement with NSQIP that a particular complication had occurred, or 2) that it had not occurred, 3) it could disagree that it had occurred when it was recorded as occurring by NSQIP, or finally 4) it could conclude that a complication had occurred even when NSQIP had not recorded it as having occurred. The last is possible because the diagnosis of a complication can be made on clinical criteria that do not meet NSQIP definitions. In the second type of analysis, which examined non-NSQIP complications, similar methodology was used except that the prospective data gathering method was used as the reference.

Severity Grading

As noted, complications were classified by severity using the Modified Accordion Severity Grading System.⁵ The Accordion System, described in 2009,³ was slightly changed subsequently based on results of a validation study⁵— hence the term “Modified” Accordion Severity Grading System. For this study, complications were graded into three categories - mild (Grade 1), moderate (Grade 2), and severe (Grades 3-5). Severe complications of grades 3-5 were considered as a group because only a relatively small number of these complications were available for analysis. There were no postoperative deaths (i.e. Grade 6 complications).

For the purposes of this study, the following additional clarifications regarding grade 1 and grade 2 complications were made:

Grade 1—Drugs in use for co-morbidities preoperatively may be continued into the postoperative period without considering that a postoperative complication exists. Discharge with a drain in place in the peritoneal cavity in an asymptomatic patient is considered a grade 1 complication. Discharge with a trans-anastomotic stent in place in an otherwise well patient is not considered to be a complication. Readmission alone without identification of a higher grade complication is considered to be grade 1.

Grade 2—Prophylactic antibiotics may be continued for 24 hours after surgical start. Drugs started intraoperatively as part of the anesthetic management (e.g. pressors) may be continued for 12 hours postoperatively. Delirium is considered to be a Grade 1 complication if more than one dose of any drug is used to treat it or if the state persists for longer than 24 hours. As noted, transfusion was not considered a complication.

Statistical methods—Kappa statistics were used to determine agreement between methods of data collection.⁹ Mean kappas were calculated for each grade of complication using the kappa values of the individual gathering methods (vs. NSQIP) for all complications within a given severity grade.¹⁰ The Kappa value is a surrogate for agreement; the inverse of discordance (i.e. perfect agreement would have a kappa of 1 whereas agreement no more likely than by chance alone would have a kappa of zero). The following scoring system was utilized: a score of 0-0.2 = slight agreement, 0.21-0.40 = fair agreement, 0.41-0.60 = moderate agreement, 0.61-0.80 = substantial agreement, 0.81-0.99 = almost perfect agreement, and a value of 1 = perfect agreement.¹¹ Sensitivity, specificity, and predictive accuracies were calculated as standard. All statistics were performed using SAS version 9.3 and Graph Pad Prism version 6.

Results

The study started in November 2011 when pancreas specific NSQIP was instituted in our institution and ended in February 2013. Eighty-four pancreatoduodenectomies were entered into NSQIP during the study period. The study population consisted of 36 females and 48 males with a mean age of 64.4 +/- 11.7 years (range 34-83 years). The mean postoperative length of stay was 9.8 +/- 4.1 days and ranged from 4 to 27 days. Thirty-one patients (36.9%) had 53 postoperative complications listed by NSQIP. Additionally, 25 patients (29.8%) had 31 complications that were not listed by NSQIP. In total, 66.7% of the entire study population had a complication of some type (i.e. NSQIP and/or non-NSQIP) recorded by any method.

Complications listed by NSQIP

The overall results for complications listed by NSQIP are displayed by severity grade in Table 2 (top, mild complications; middle, moderate complications; bottom, severe complications). In each section, the 4 audit methods are identified on the left, while rates of false positives and false negatives as well as Kappa values are presented on the right. Between these are a series of vertical columns, which compare the non-NSQIP audit results with results of the NSQIP audit for individual complications. From left to right are four types of results: 1) instances in which the NSQIP audit and the three non-NSQIP audits were in agreement that a particular complication had occurred (“all Pos.”); 2) instances in which one or more of the non-NSQIP audits were false positive in respect to NSQIP (“False Pos.”); 3) instances in which the NSQIP audit and the three non-NSQIP audits were in agreement no complication had occurred (“all Neg.”) and lastly 4) instances in which one or more of the non-NSQIP audits were falsely negative in respect to a complication recorded by NSQIP (“False Neg.”). There was a high false negative rate for grade 1 (mild) and grade 2 (moderate) complications that ranged from 24% to 48% (Table 2). The false positive rate for grade 1 and grade 2 complications ranged between 1% and 19% (Table 2, top and middle). However, unlike for mild and moderate complications, the category of severe complications had only one complication with a false positive result and no complications with a false negative result (Grades 3-5, Table 2 bottom). Similarly, when examined by kappa analysis, there was only moderate agreement between the 3 non-NSQIP methods and NSQIP for complications of mild and moderate severity (mean kappa = 0.629 for mild and 0.485 for

moderate). However, agreement was almost perfect for severe complications (mean kappa = 0.969). The sensitivity and specificity as well as positive and negative predictive values of the three non-NSQIP audit methods with respect to the NSQIP audit are given in Table 3.

The moderate complication group (Table 2) contained the largest number of complications. In this group, the prospective gathering method and the detailed chart review method had very similar results and differed from the discharge summary method in that they had fewer false negatives and more false positives. Detailed examination of moderate complications by individual complication types (Figure 1) revealed that delayed gastric emptying, which was the commonest complication, displayed both false positive and false negative results. Of note, “Grade 2 Organ Space SSI” had no false positive or false negative results. Some other complications were prone to false negative results while yet others were more likely to display false positive results. For instance, for the complications “Grade 2 DVT”, “Grade 2 Superficial SSI”, “Grade 2 Myocardial Infarction”, and “Grade 2 Sepsis”, there were 14 instances of false negative results but no false positive results (Figure 1). Put another way, these complications were more likely to be captured by NSQIP while not counted as complications by the non-NSQIP methods. False negative results were usually due to an occurrence meeting NSQIP criteria in circumstances not judged to be a complication by the non-NSQIP reviewers. For instance, in the case involving an “occurrence of a Deep Venous Thrombosis,” NSQIP criteria resulted in a NSQIP positive status when a small subclavian vein thrombosis was diagnosed postoperatively. Anticoagulation was recommended, but not started in this patient because of concern for postoperative bleeding. This meets NSQIP criteria of “the record indicates that treatment was warranted but there was no additional appropriate treatment option available”.⁷

However, non-NSQIP methods did not record this as a complication. In another case, postoperative hypotension and tachycardia successfully treated by fluid replacement was associated with a troponin level of 0.86 ng/ml. This set of circumstances fits within NSQIP criteria for myocardial infarction, but was considered to be demand ischemia and not listed as a complication by non-NSQIP methods (i.e. did not require specific treatment). Conversely, for the complications “Grade 2 Urinary Tract Infection”, “Grade 2 Pneumonia”, and “Grade 2 Pancreatic Fistula”, there were 14 false positive results and only one false negative result (Figure 1). In other words, there was a higher likelihood that the non-NSQIP methods captured these occurrences when they were not counted by NSQIP. The cause for the discrepancy for pneumonia and urinary tract infection was that clinicians made a clinical diagnosis when it was below the NSQIP threshold for these diagnoses. For instance, some patients were treated for a urinary tract infection based on a positive urine culture, but with a colony count below the NSQIP threshold.⁷ Importantly, the criteria for clinicians making these diagnoses were not only insufficient for NSQIP, but were also variable.

In the mild complication category, superficial surgical site infection was a common area of discordance with NSQIP. One cause of false negativity was post-discharge diagnosis of a superficial site infection made by telephone follow-up by NSQIP, which was not present in the clinical record. Another was “purulence” at a drain site, which was not considered to be a superficial surgical site infection by one or more of the non-NSQIP methods. The main cause of false positivity was the diagnosis of superficial surgical site infections by one or

more of the non-NSQIP methods based on criteria which did not meet the stringent threshold for the NSQIP diagnosis.

Complications not listed by NSQIP

Some postoperative complications occurring in patients having undergone pancreatoduodenectomy are not defined by NSQIP. Urinary retention and delirium were the most common mild (grade 1) complications of this type, while cardiac arrhythmias and *C. difficile* colitis (and other causes of postoperative diarrhea) were the most common moderate (grade 2) complications. A listing of these complications by severity grade is given in Table 4.

In analyzing discordance among non-NSQIP complication gathering methods for these types of complications, the prospective gathering method was used as the reference method (Table 5). The pattern for mild complications was similar to that seen in NSQIP-type complications. The discharge summary tended to have more false negatives than the chart review method. Notably, there was near perfect agreement among the non-NSQIP methods for non-NSQIP type complications of moderate and severe grades, with only 1 discordant complication. This was a patient with grade 2 *C. difficile* colitis not captured by the discharge summary method. The sensitivity and specificity as well as the positive and negative predictive values of the discharge summary and chart review audit methods in respect to the prospective audit are given in Table 6.

Discussion

For quantitative methods to be reliable, they must be based on rigorous methods, otherwise differences in results may be due to variations in methods rather than real differences in what is being observed. NSQIP provides rigorous definitions of complications and rigorous methods of complication gathering.⁶ In terms of the latter, the inter-rater reliability in the past has been over 98%.⁶ Because of these characteristics, NSQIP methods were linked to the Modified Accordion Severity Grading System when PMI was being developed.^{4, 5} However, NSQIP is not universally available in American hospitals and is currently available only on a limited basis outside the USA. This study was performed to determine the degree of discordance that exists between NSQIP and other audit methods in order to better understand how differences in audit methods might affect quantitative evaluations. It would also establish whether these methods could substitute for NSQIP in computing PMI.

This study shows that the level of discordance between the NSQIP method and the other methods is too great for any of these methods to substitute for NSQIP in computing the PMI (i.e. to substitute for NSQIP methods and to be comparable to PMI computed using NSQIP as a foundation). Discordance exists between these methods and NSQIP within the group of postoperative events considered to be complications by NSQIP. Also, discordance is compounded by the fact that non-NSQIP methods identify a number of complications that are not considered to be postoperative occurrences by NSQIP. Consequently, PMI calculated by any of the non-NSQIP methods would yield a very different result from that of an NSQIP-based formulation.

The fact that NSQIP fails to identify a number of complications means that the PMI, as presently calculated, is not a total summation of all postoperative morbidity. NSQIP is a system in evolution and its imperfections are recognized. While working to maintain its basic principles of clear definition and meticulous gathering of complications, it is adding events that were not included in the initial iterations of NSQIP. That is the reason for adding the pancreas specific complications to basic NSQIP. Since NSQIP does not identify all postoperative complications of pancreatoduodenectomy, it must be asked whether one of the other methods could be substituted for NSQIP in calculating an overall morbidity index for this operation. Clearly, discharge summary information could not be used as it misses many grade 1 and grade 2 complications, both non-NSQIP and NSQIP type. The other two non-NSQIP methods were largely in agreement, but this is really only a reflection of the ability of both to faithfully extract clinical diagnoses using information entered into a hospital chart. In actuality, the thresholds for diagnosis of complications by physicians such as delayed gastric emptying, pneumonia, and urinary tract infections varied from patient to patient. Therefore, basing a quantitative method on such variable data would be undesirable since even within a single institution what would be considered a complication could vary substantially using these complication gathering methods. While not examined in this study, it would be very surprising if there were not similar inter-institution variability in thresholds for considering an event to be a complication.

A surprising and unexpected result was the almost complete lack of discordance among methods when the complications were severe. This was true even for the discharge summary method. In retrospect, this is understandable, as severe complications are notable and unlikely to be missed by NSQIP or other audit methods. Even grading in the Modified Accordion System becomes easier at the higher severe levels and is reflected in the contrast between the complex instructions for rating grade 1 and 2 complications compared to those instructions for grade 3-6 complications.^{3, 5} It seems possible that an index based only on severe complications might be sufficiently stable to be based on data gathered by methods less rigorous than NSQIP. The development of a Severe Complication index (SCI) deserves further exploration.

While this study was performed to evaluate how different methods of audit might affect results of quantitative methods such as PMI, a corollary finding is that for non-NSQIP methods the lack of sharply defined thresholds for complications is a critical issue. As a result, even simple enumeration (i.e., a listing of complications) will be variable under these circumstances. In 1992, Clavien and Strasberg introduced severity grading based mainly on the treatment provoked by the complication.¹ While this approach and its modifications^{2, 3} have been very useful, they have the inherent limitation that the thresholds for treatments of the various postoperative complications are undefined. Consequently, this results in an instability which can be overcome only by defining those thresholds. This is largely what the combination of the severity grading system with NSQIP has achieved. NSQIP indicates whether a complication has occurred and the Accordion system then grades it.

In summary, variations in definition and methods of retrieval greatly influence what is rated as a complication. This is especially true of lower grades of complications (i.e. grade 1 and 2). These issues can completely invalidate attempts to assess the aggregated severity or

burden of complications. NSQIP methods have sufficient rigor in definition and methods for collecting complications to support quantitative evaluation of complications, features which other complication gathering methods seem to lack. While NSQIP does not identify all complications, this will hopefully be remedied as NSQIP continues to evolve.

References

1. Clavien PA, Sanabria JR, Strasberg SM. Proposed classification of complications of surgery with examples of utility in cholecystectomy. *Surgery*. 1992; 111:518–526. [PubMed: 1598671]
2. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg*. 2004; 240:205–213. [PubMed: 15273542]
3. Strasberg SM, Linehan DC, Hawkins WG. The accordion severity grading system of surgical complications. *Ann Surg*. 2009; 250:177–186. [PubMed: 19638919]
4. Strasberg SM, Hall BL. Postoperative morbidity index: a quantitative measure of severity of postoperative complications. *J Am Coll Surg*. 2011; 213:616–626. [PubMed: 21871822]
5. Porembka MR, Hall BL, Hirbe M, Strasberg SM. Quantitative weighting of postoperative complications based on the accordion severity grading system: demonstration of potential impact using the american college of surgeons national surgical quality improvement program. *J Am Coll Surg*. 2010; 210:286–298. [PubMed: 20193891]
6. Shiloach M, Frencher SK Jr, Steeger JE, et al. Toward robust information: data quality and inter-rater reliability in the American College of Surgeons National Surgical Quality Improvement Program. *J Am Coll Surg*. 2010; 210:6–16. [PubMed: 20123325]
7. ACS NSQIP Classic Variables & Definitions. [October 4, 2013] http://nsqip.healthsoftonline.com/lib/Documents/Ch_4_Variables_Definitions_062810.pdf. pp. data file
8. Khuri SF, Henderson WG, Daley J, et al. Successful implementation of the Department of Veterans Affairs' National Surgical Quality Improvement Program in the private sector: the Patient Safety in Surgery study. *Ann Surg*. 2008; 248:329–336. [PubMed: 18650645]
9. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 1960; 20:37–46.
10. Sprent, P.; Smeeton, NC. *Applied nonparametric statistical methods*. 3. Boca Raton: Chapman & Hall/CRC; 2001.
11. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med*. 2005; 37:360–363. [PubMed: 15883903]

Method	Grade 2 DVT	Method	Grade 2 Superficial SSI				Method	Grade 2 Myocardial Infarction			Method	Grade 2 Sepsis		
DC Summary	-	DC Summary	-	-	+	+	DC Summary	-	-	+	DC Summary	-	-	+
Chart Review	-	Chart Review	-	-	+	+	Chart Review	-	+	+	Chart Review	+	+	+
Prospective	-	Prospective	-	+	+	+	Prospective	-	+	+	Prospective	+	+	+
NSQIP	+	NSQIP	+	+	+	+	NSQIP	+	+	+	NSQIP	+	+	+

Method	Grade 2 Urinary Tract Infection				Method	Grade 2 Pneumonia		Method	Grade 2 Pancreatic Fistula				Method	Grade 2 Organ Space SSI	
DC Summary	-	+	+	+	DC Summary	+	+	DC Summary	+	+	+	+	DC Summary	+	+
Chart Review	+	+	+	+	Chart Review	+	+	Chart Review	+	+	+	+	Chart Review	+	+
Prospective	+	+	+	-	Prospective	+	+	Prospective	+	+	+	+	Prospective	+	+
NSQIP	+	-	-	-	NSQIP	+	-	NSQIP	+	+	+	-	NSQIP	+	+

Method	Grade 2 Delayed Gastric Emptying												
DC Summary	-	-	-	-	+	+	+	+	+	+	+	-	-
Chart Review	-	-	-	+	+	+	+	+	+	+	+	+	+
Prospective	-	-	-	+	+	+	+	+	+	+	+	+	+
NSQIP	+	+	+	+	+	+	+	-	-	-	-	-	-

Figure 1. Grade 2 complications by type of complication. For each complication, all false positive and false negative results are shown as well as all instances of agreement that the complication had occurred. Note the tendency for some complications to have false positive results and others to have false negative results.

Table 1

NSQIP Pancreatectomy-Specific Postoperative Occurrences and Definitions

Pancreatic fistula	Delayed gastric emptying	Percutaneous drainage
Persistent drainage (a drain output of any measurable volume of fluid on or after postoperative day 3) of amylase-rich fluid (amylase content > 3x serum amylase) OR Clinical diagnosis of pancreatic fistula by attending surgeon AND one of following: Drain continued > 7 d Percutaneous drain placed Reoperation performed	Gastrostomy tube to external drainage or Nasogastric tube present or reinserted after postoperative day 7 OR No oral intake by postoperative day 14	Placement of a percutaneous drain after completion of surgery, but within 30 d postoperatively.

[C.E. Mark as 3 column list; no rule lines between subhead row and bottom.]

Table 2
 Comparison of Results for NSQIP-Type Complications Gathered by NSQIP and the 3 Other Audit Methods*

Method	All positive*	False positive*	All negative*	False negative*	False-positive rate	False-negative rate	Kappa value	Mean kappa
All grade 1 NSQIP complications								
DC Summary	10 +	1+ 5 -	61 -	4 - 1 -	1/68 (1.4%)	5/16 (31.3%)	0.744	0.629
Chart Review	10 +	1+ 5 +	61 -	4 - 1 -	6/68 (8.8%)	5/16 (31.3%)	0.585	
Prospective	10 +	1+ 5 +	61 -	4 - 1 +	7/68 (10.3%)	5/16 (31.3%)	0.558	
NSQIP	10 +	1- 5 -	61 -	4 + 1 +	NA	NA	NA	NA
All grade 2 NSQIP complications								
DC Summary	13 +	8 + 1+ 2 -	48 -	6 - 1 -	9/59 (15.3%)	12/25 (48.0%)	0.381	0.485
Chart Review	13 +	8 + 1+ 2 +	48 -	6 - 1 -	11/59 (18.6%)	7/25 (28.0%)	0.510	
Prospective	13 +	8 + 1- 2 +	48 -	6 - 1 +	10/59 (16.9%)	6/25 (24.0%)	0.564	
NSQIP	13 +	8 - 1- 2 -	48 -	6 + 1 +	NA	NA	NA	
All grade 3 to 5 NSQIP complications								
DC Summary	12 +	1- 1+	71 -	NA	0/72 (0%)	0/12 (0%)	1.000	0.969
Chart Review	12 +	1+ 1+	71 -	NA	1/72 (1.4%)	0/12 (0%)	0.953	
Prospective	12 +	1+ 1+	71 -	NA	1/72 (1.4%)	0/12 (0%)	0.953	
NSQIP	12 +	1- 1-	71 -	NA	NA	NA	NA	

* Numbers indicate the number of patients with the particular pattern in those categories.

Table 3
Sensitivity, Specificity, and Predictive Accuracy of Audit Methods by Severity of NSQP Complications

Method	Sensitivity	Specificity	Positive predictive value	Negative predictive value
Grade 1				
Discharge summary	11/16 (68.8%)	67/68 (98.5%)	11/12 (91.7%)	62/72 (86.1%)
Retrospective chart review	11/16 (68.8%)	62/68 (91.2%)	11/17 (64.7%)	62/67 (92.5%)
Prospective	11/16 (68.8%)	61/68 (89.7%)	11/18 (61.1%)	61/66 (92.4%)
Grade 2				
Discharge summary	13/25 (52.0%)	50/59 (84.7%)	13/22 (59.1%)	50/62 (80.6%)
Retrospective chart review	18/25 (72.0%)	48/59 (81.4%)	18/29 (62.1%)	48/55 (87.2%)
Prospective	19/25 (76.0%)	49/59 (83.1%)	19/29 (65.5%)	49/55 (89.1%)
Grade 3 to 5				
Discharge summary	12/12 (100%)	72/72 (100%)	12/12 (100%)	72/72 (100%)
Retrospective chart review	12/12 (100%)	71/72 (98.6%)	12/13 (92.3%)	71/71 (100%)
Prospective	12/12 (100%)	71/72 (98.6%)	12/13 (92.3%)	71/71 (100%)

Table 4

Non-NSQIP Complications by Grade and Collection Method

	Prospective method	Retrospective chart review	Discharge summary
Mild, grade 1			
Urinary retention	6	5	2
Delirium	5	6	5
Failure to thrive*	2	2	2
Ketoacidosis	1	1	1
Orthostatic hypotension [†]	1	1	1
Ileus	1	1	1
Arterial line complication [‡]	0	1	0
Pathological vertebral fracture	1	1	1
Incisional hematoma	1	0	0
Moderate, grade 2			
Colitis/diarrhea/ileus	4	4	3
Cardiac arrhythmias	5	5	5
Chyle leak	1	1	1
Bile leak	1	1	1
Superior mesenteric vein occlusion	1	1	1
Severe, grade 3 to 5			
Cardiovascular failure [§]	1	1	1

* Readmission for weakness; no specific cause found; discharged.

[†] Readmission for orthostatic hypotension; treated with IV rehydration.

[‡] Hand ischemia from arterial line; treated with line removal and aspirin.

[§] Patient with postoperative hypotension refractory to IV fluids; required vasopressor doses meeting Accordion Classification for organ failure.³

Table 5

Comparison of Non-NSQIP Methods for Non-NSQIP Complications

Method	All positive*	False positive*	All negative*	False negative*	False-positive rate	False-negative rate	Kappa value	Mean kappa
All grade 1 NSQIP complications								
DC Summary	11 +	2 +	63 -	3 -	2/66 (3.0%)	7/18 (3.9%)	0.646	0.717
Chart Review	11 +	2 +	63 -	3 -	3/66 (4.5%)	3/18 (16.7%)	0.788	
Prospective	11 +	2 -	63 -	3 +	NA	NA	NA	NA
All grade 2 NSQIP complications								
DC Summary	11 +	NA	72 -	1 -	0/72 (0%)	1/12 (8.3%)	0.950	0.975
Chart Review	11 +	NA	72 -	1 +	0/72 (0%)	0/12 (0%)	1.000	
Prospective	11 +	NA	72 -	1 +	NA	NA	NA	NA
All grade 3 to 5 NSQIP complications								
DC Summary	1 +	NA	83 -	NA	0/83 (0%)	0/1 (0%)	1	1
Chart Review	1 +	NA	83 -	NA	0/83 (0%)	0/1 (0%)	1	
Prospective	1 +	NA	83 -	NA	NA	NA	NA	NA

* Numbers indicate the number of patients with the particular pattern in those categories.

Table 6
Sensitivity, Specificity, and Predictive Accuracy of Audit Methods by Severity of Non-NSQIP Complications

Method	Sensitivity	Specificity	Positive predictive value	Negative predictive value
Grade 1				
Discharge summary	11/18 (61.1%)	64/66 (97.0%)	11/13 (84.6%)	64/71 (90.1%)
Retrospective chart review	15/18 (83.3%)	63/66 (95.5%)	15/18 (83.3%)	63/66 (95.5%)
Grade 2				
Discharge summary	11/12 (91.7%)	72/72 (100%)	11/11 (100%)	72/73 (98.6%)
Retrospective chart review	12/12 (100%)	72/72 (100%)	12/12 (100%)	72/72 (100%)
Grade 3 to 5				
Discharge summary	1/1 (100%)	83/83 (100%)	1/1 (100%)	83/83 (100%)
Retrospective chart review	1/1 (100%)	83/83 (100%)	1/1 (100%)	83/83 (100%)