



Year: 2019

Benchmarks in Pancreatic Surgery: A Novel Tool for Unbiased Outcome Comparisons

Sánchez-Velázquez, Patricia ; Muller, Xavier ; Malleo, Giuseppe ; Park, Joon-Seong ; et al ; Puhan, Milo A

Abstract: **OBJECTIVE** To use the concept of benchmarking to establish robust and standardized outcome references after pancreatoduodenectomy (PD). **BACKGROUND** Best achievable results after PD are unknown. Consequently, outcome comparisons among different cohorts, centers or with novel surgical techniques remain speculative. **METHODS** This multicenter study analyzes consecutive patients (2012-2015) undergoing PD in 23 international expert centers in pancreas surgery. Outcomes in patients without significant comorbidities and major vascular resection (benchmark cases) were analyzed to establish 20 outcome benchmarks for PD. These benchmarks were tested in a cohort with a poorer preoperative physical status (ASA class 3) and a cohort treated by minimally invasive approaches. **RESULTS** Two thousand three hundred seventy-five (38%) low-risk cases out of a total of 6186 PDs were analyzed, disclosing low in-hospital mortality (1.6%) but high morbidity, with a 73% benchmark morbidity rate cumulated within 6 months following surgery. Benchmark cutoffs for pancreatic fistulas (B-C), severe complications (grade 3), and failure-to-rescue rate were 19%, 30%, and 9%, respectively. The ASA 3 cohort showed comparable morbidity but a higher in-hospital mortality (3% vs 1.6%) and failure-to-rescue rate (16% vs 9%) than the benchmarks. The proportion of benchmark cases performed varied greatly across centers and continents for both open (9%-93%) and minimally invasive (11%-62%) PD. Centers operating mostly on complex PD cases disclosed better results than those with a majority of low-risk cases. **CONCLUSION** The proposed outcome benchmarks for PD, established in a large-scale international patient cohort and tested in 2 different cohorts, may allow for meaningful comparisons between different patient cohorts, centers, countries, and surgical techniques.

DOI: <https://doi.org/10.1097/SLA.0000000000003223>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-175255>

Journal Article

Published Version

Originally published at:

Sánchez-Velázquez, Patricia; Muller, Xavier; Malleo, Giuseppe; Park, Joon-Seong; et al; Puhan, Milo A (2019). Benchmarks in Pancreatic Surgery: A Novel Tool for Unbiased Outcome Comparisons. *Annals of Surgery*, 270(2):211-218.

DOI: <https://doi.org/10.1097/SLA.0000000000003223>

Benchmarks in Pancreatic Surgery

A Novel Tool for Unbiased Outcome Comparisons

Patricia Sánchez-Velázquez, MD, PhD,* Xavier Muller, MD,* Giuseppe Malleo, MD,†
 Joon-Seong Park, MD, PhD,‡ Ho-Kyoung Hwang, MD, PhD,‡ Niccolò Napoli, MD,§ Ammar A. Javed, MD,¶
 Yosuke Inoue, MD, PhD,|| Nassiba Beghdadi, MD,** Marit Kalisvaart, MD,†† Emanuel Vigia, MD,‡‡
 Carrie D. Walsh, BA,§§ Brendan Lovasik, MD,¶¶ Juli Busquets, MD, PhD,|||| Chiara Scandavini, MD,***
 Fabien Robin, MD,††† Hideyuki Yoshitomi, MD, PhD,‡‡‡ Tara M. Mackay, MD,§§§ Olivier R. Busch, MD,§§§
 Hermien Hartog, MD, PhD,¶¶¶ Stefan Heinrich, MD,||||| Ana Gleisner, MD, PhD,**** Julie Perinel, MD,††††
 Michael Passeri, MD,‡‡‡‡ Nuria Lluís, MD,§§§§ Dimitri A Raptis, MD, PhD,* Christoph Tschuor, MD, PhD,*
 Christian E. Oberkofler, MD,* Michelle L. DeOliveira, MD, PhD,* Henrik Petrowsky, MD,*
 John Martinie, MD,‡‡‡‡ Horacio Asbun, MD,§§§§ Mustapha Adham, MD,†††† Richard Schulick, MD,****
 Hauke Lang, MD,||||| Bas Groot Koerkamp, MD, PhD,¶¶¶¶ Marc G. Besselink, MD, PhD,§§§
 Ho-Seong Han, MD, PhD,¶¶¶¶ Masaru Miyazaki, MD,||||||| Cristina R. Ferrone, MD,*****
 Carlos Fernández-del Castillo, MD,***** Keith D. Lillemoe, MD,***** Laurent Sulpice, MD, PhD,†††
 Karim Boudjema, MD, PhD,††† Marco Del Chiaro, MD, PhD,*** Joan Fabregat, MD, PhD,||||
 David A. Kooby, MD,¶¶ Peter Allen, MD,††††† Harish Lavu, MD,§§ Charles J. Yeo, MD,§§
 Eduardo Barroso, MD,‡‡ Keith Roberts, MD, PhD,†† Paolo Muijsan, MD, PhD,†† Alain Sauvanet, MD, PhD,**
 Akio Saiura, MD, PhD,|| Christopher L. Wolfgang, MD, PhD,¶ John L. Cameron, MD, PhD,¶
 Ugo Boggi, MD,§ Dong-Sup Yoon, MD, PhD,‡ Claudio Bassi, MD,† Milo A. Puhan, MD, PhD,‡‡‡‡‡ and
 Pierre-Alain Clavien, MD, PhD*

Objective: To use the concept of benchmarking to establish robust and standardized outcome references after pancreaticoduodenectomy (PD).

Background: Best achievable results after PD are unknown. Consequently, outcome comparisons among different cohorts, centers or with novel surgical techniques remain speculative.

Methods: This multicenter study analyzes consecutive patients (2012–2015) undergoing PD in 23 international expert centers in pancreas surgery. Outcomes in patients without significant comorbidities and major vascular resection (benchmark cases) were analyzed to establish 20 outcome

benchmarks for PD. These benchmarks were tested in a cohort with a poorer preoperative physical status (ASA class ≥ 3) and a cohort treated by minimally invasive approaches.

Results: Two thousand three hundred seventy-five (38%) low-risk cases out of a total of 6186 PDs were analyzed, disclosing low in-hospital mortality ($\leq 1.6\%$) but high morbidity, with a 73% benchmark morbidity rate cumulated within 6 months following surgery. Benchmark cutoffs for pancreatic fistulas (B-C), severe complications (\geq grade 3), and failure-to-rescue rate were 19%, 30%, and 9%, respectively. The ASA ≥ 3 cohort showed comparable morbidity but a higher in hospital-mortality (3% vs 1.6%) and failure-to-rescue

From the *Department of Surgery and Transplantation, University Hospital Zurich, Zurich, Switzerland; †Department of Surgery, University Hospital of Verona, "Pancreas Institute," Verona, Italy; ‡Department of Surgery, Yonsei Severance Hospital, Seoul, South Korea; §Division of General and Transplant Surgery, University of Pisa, Pisa, Italy; ¶The John L. Cameron Division of Hepatobiliary and Pancreatic Surgery, Department of Surgery, Johns Hopkins University School of Medicine, Baltimore, MD; ||Department of Surgery, Cancer Institute Hospital, Tokyo, Japan; **Department of Surgery, Hôpital Beaujon, Clichy, France; ††Department of Surgery, Queen Elizabeth Hospital, Birmingham, UK; ‡‡Department of Surgery, Curry Cabral Hospital, CHLC, Lisboa, Portugal; §§Department of Surgery, Thomas Jefferson University, Philadelphia, PA; ¶¶Department of Surgery, Emory University School of Medicine, Atlanta, GA; ||||Department of Surgery, University Hospital of Bellvitge, Barcelona, Spain; ***Department of Surgery, Karolinska University Hospital, Stockholm, Sweden; †††Department of Digestive Surgery, E Herriot University Hospital, HCL, UCBL1, Lyon, France; ‡‡‡Department of General Surgery, Chiba University Hospital, Chiba Japan; §§§Department of Surgery, Cancer Center Amsterdam, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands; ¶¶¶Department of Surgery, Erasmus University Medical Center, Rotterdam, The Netherlands; |||||Department of Surgery, University Hospital of Mainz, Mainz, Germany; *****Department of Surgery, University of

Colorado, Denver, CO; ††††Department of Surgery, E Herriot University Hospital, HCL, Lyon, France; ‡‡‡‡Department of Surgery, Carolina's Health Care Hospital, NC; §§§§Department of Surgical Oncology Miami Cancer Institute, Miami, FL; ¶¶¶¶Department of Surgery, Seoul National University Bundang Hospital, Seoul, South Korea; |||||Department of Surgery, International University of Health and Welfare, Mita Hospital, Tokyo, Japan; *****Department of Surgery, Massachusetts General Hospital, Boston, MA; †††††Department of Surgery, Memorial Sloan-Kettering Cancer Center (MSKCC), New York, NY; and ‡‡‡‡‡Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland.

All authors approved the final manuscript.

P.S. and X.M. contributed equally as first authors.

The authors report no conflicts of interest.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.annalsofsurgery.com).

Reprints: Pierre-Alain Clavien, MD, PhD, Department of Surgery and Transplantation, University Hospital Zurich, Raemistrasse 100, CH-8091 Zurich, Switzerland. E-mail: clavien@access.uzh.ch.

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 0003-4932/19/27002-0211

DOI: 10.1097/SLA.0000000000003223

rate (16% vs 9%) than the benchmarks. The proportion of benchmark cases performed varied greatly across centers and continents for both open (9%–93%) and minimally invasive (11%–62%) PD. Centers operating mostly on complex PD cases disclosed better results than those with a majority of low-risk cases.

Conclusion: The proposed outcome benchmarks for PD, established in a large-scale international patient cohort and tested in 2 different cohorts, may allow for meaningful comparisons between different patient cohorts, centers, countries, and surgical techniques.

Keywords: benchmarks, minimally invasive surgery, outcomes, pancreatic surgery, pancreatoduodenectomy, quality of care, surgical complications

(Ann Surg 2019;270:211–218)

Pancreatico-duodenectomy (PD) is frequently performed for a number of diseases, particularly in patients with pancreatic adenocarcinoma.¹ This procedure is, however, still associated with significant perioperative mortality and high morbidity, which strongly impacts patient’s quality of life and costs.^{2–7} The major shortcoming of quality comparisons in complex surgery is the absence of standardized benchmarks for outcome indicators in a well-defined patient population, which may serve as reference values.⁸ For instance, novel minimally invasive laparoscopic or robotic approaches are promoted by many surgeons claiming superiority, although genuine reference data in the classical open procedure are still lacking.

The concept of benchmarking is a widely used tool for quality improvement in the manufacturing industry and economy, but its use in medicine has been more versatile and imprecise.⁹ The novel idea of applying the benchmark concept to surgery is to assess best achievable results in a well-defined low-risk patient cohort to establish meaningful reference values for comparisons, for example among centers or over time, or to evaluate the implementation of novel surgical techniques.¹⁰ Recent studies in the field of surgery have applied this methodology to set reference figures in liver resection, transplantation, and esophagectomy.^{11–13}

Outcome indicators should no longer be based exclusively on mortality, but rather cover other relevant endpoints for the patients and health care system. Consistent with the concept of value-based medicine, special emphasis should be put on cumulative morbidity extending well after index-hospital discharge.¹⁴

Consequently, this study aims at establishing a variety of clinically relevant outcome benchmark values for PD from a low-risk cohort treated at high volume centers covering 3 continents. As an example of utility, similar data were collected from a few expert centers performing laparoscopic and robotic PD, and compared with the established benchmark values for open PD. The proposed standardized benchmark methodology for outcome analysis and reporting might be an important step to improve quality and validate novel strategies for complex surgical procedures.

METHODS

Study Design

Establishing benchmarks in surgery followed a standardized methodology, previously applied and reported for major hepatectomy, liver transplantation, and esophagectomy.^{11–13} First, a large patient cohort from international expert centers was selected to define *best practice*. Only patients with a low preoperative risk profile were selected, using well-defined inclusion criteria (Table 1). Second, relevant outcome indicators for quality of surgical care were assessed in this low risk population. For each outcome indicator, a numeric benchmark cutoff was calculated indicating the

TABLE 1. Selection Criteria for Low-risk Patients Included in the Benchmark Analysis

Inclusion criteria	
Age	≥ 18 yrs
Open PD	for resectable malignant and benign pancreatic disease
Exclusion criteria-surgical	
Extended or total pancreatico-duodenectomy	
Resection of the mesenteric artery/cealic trunk	
Macroscopic positive margins (R2)	(16)
Extrapancreatic metastases (non-nodal)	(16)
Previous major abdominal surgery (ie, bariatric surgery, liver surgery, splenectomy)	
Minimally invasive procedures (laparoscopic or robotic)	
Exclusion criteria-comorbidities	
American Society of Anesthesiologists (ASA) classification	≥ 3 (16)
Body mass index	≥ 35 kg/m ² (15)
Cardiac disease defined as (17,18):	
Congestive heart failure (CHF) onset or exacerbation in 30 d prior to surgery	
History of angina pectoris within 1 month of surgery	
Myocardial infarct within 6 months prior to surgery	
History of percutaneous coronary intervention or cardiac surgery	
Atrial fibrillation	
Chronic renal failure MDRD	≥ Stage 3 (19)
GRF < 60 mL/min per 1.73 m ² or Serum creatinine > 1.8 mg/dL or 160 μmol/L	
Chronic obstructive pulmonary disease with FEV1	< 80% (22)
Exclusion criteria-medication	
Use of anticoagulants (20,21)	
Nonvitamin K antagonist oral anticoagulants (NOACs)	
Vitamin K antagonist	
Clopidogrel	
≥ 2 oral antidiabetic drugs or insulin	(16)

“best achievable” results. These benchmark cutoffs were applied to identify clinically relevant performance gaps in different patient cohorts, internal, or external to the expert centers, which may then be addressed by specific quality improvement measures. This process is commonly referred to as benchmarking.

Patient and Hospital Selection

Hospitals in Asia, Europe, and America were screened to identify centers performing high-volume pancreatic surgery (≥ 50 complex pancreas interventions per year).^{4,15,16} Only the largest programs having previously published their results and holding a prospective database were contacted to participate to the study. The final collaborative consortium included 23 centers: 13 from Europe (Amsterdam, Barcelona, Birmingham, Lisbon, Lyon, Mainz, Paris, Pisa, Rennes, Rotterdam, Stockholm, Verona, and Zurich), 6 from USA (Atlanta, Baltimore, Boston, Denver, New York, and Philadelphia), and 4 from Asia (Tokyo, Chiba, and 2 from Seoul).

From the 23 included centers, we selected only patients undergoing open PD for malignant or benign pancreatic diseases localized in the pancreatic head, without distant metastases or involvement of the arterial axis. To further narrow the selection to patients with a low-preoperative risk-profile, comorbidities proven to negatively impact the postoperative course, for example American Society of Anesthesiologists (ASA) classification ≥ 3, body mass index > 35 kg/m², and anticoagulation drug therapy, were considered as exclusion criteria.^{17–24} (Table 1)

Patient and outcome data were centralized in a secure and anonymized online data management system provided by the University Hospital of Zurich (<https://www.whipplebenchmarks.org>). The data were audited and checked for completeness by 3 local

investigators (P.S.V., X.M., and C.T.). Only patients with a minimal follow-up of 1 year were included in the study.

Ethical approval from the institutional ethical board of the University Hospital of Zurich (KEK 2017-01638) and from each respective center was obtained before analysis of the data.

Outcome Indicators

Twenty clinically relevant intra- and postoperative outcome indicators were selected to establish serial outcome benchmark cutoffs at discharge, 3, 6, and 12 months. In detail, perioperative parameters, which have been previously shown to be associated with surgical performance, namely duration of the intervention, percentage of intraoperative blood transfusions, and hospital stay, were included.^{25–28} Special emphasis was put on postoperative complications, as they are recognized to reflect surgical outcomes more comprehensively than mortality alone.^{29,30} To accurately reflect morbidity, each postoperative complication was assessed for every single patient during the postoperative course. Additionally, the severity of the complications were graded using the Clavien–Dindo classification, which defines the severity of the complication by the treatment intervention required.³¹ Of note, the index-complication of PD, postoperative pancreatic fistulas (POPF), was also graded using the updated International Study Group on Pancreatic Fistula classification.³² To investigate the proportion of fatalities following a postoperative complication, we calculated the failure-to-rescue rate according to a previously published methodology: the number of deaths in patients with a complications graded > II (numerator) to the total number of patients with a complication graded > II (denominator).³³

The comprehensive complication index (CCI[®]) was used to cover the total number of complications by severity for every single patient.^{34,35} The CCI[®] adds each complications graded by the Clavien–Dindo system into a continuous numeric score ranging from 0 (uneventful course) to 100 points (death). This index has been shown to enable a quantification of the cumulative morbidity burden for a single patient over a certain period of time, extending for example to 1 year.^{12,13}

For the subgroup of patients with pancreatic adenocarcinoma, validated oncological quality indicators including resection margins, number of lymph nodes resected, and disease-free survival were additionally assessed.^{18,36,37}

Definition and Applications of Benchmarks

The benchmark cutoff for a given outcome indicator is calculated as the 75th percentile of the median value for each center.¹⁰ To account for variability across centers, the medians as well as the ranges are also depicted.

To illustrate the ability of the established benchmarks to identify small differences in outcomes, we analyzed outcomes in a cohort of patients from the same centers and differing from the benchmark cohort by a poorer preoperative physical status, defined by an ASA class ≥ 3 .³⁸

To further illustrate the utility of the benchmark concept, we tested the benchmark values in cohorts of patients gathered from 3 recognized institutions performing minimal-invasive PD using either a robotic or laparoscopic approach.^{39–42} Of note, apart from the minimally invasive approach, this cohort fulfilled the same selection criteria as the benchmark cohort for open PD (Table 1).

Center Specific Risk Profile

To inquire how many high-risk cases a center is performing, we stratified the 23 centers according to their respective proportion of benchmark versus nonbenchmark cases and analyzed the relation of this center specific risk profile with outcomes in the benchmark cohort.

Statistical Analysis

Statistical analysis was performed using the SPSS statistical software package (SPSS, version 20, IBM, Armonk, NY), Stat 10.1 (version 19), and Graph Pad Prism7. The Kolmogorov–Smirnov test was used to determine the distribution of the variables. Categorical variables were compared using the Fischer exact or the Pearson chi-square tests, where appropriate. The Student *t* test was used to make pairwise comparisons of normally distributed parameters and the Mann–Whitney *U* test was used for nonparametric data. Pearson correlation was performed to compute correlation coefficients between measured variables. Tests were considered statistically significant at a 2-sided *P* value of < 0.05.

RESULTS

From the 23 high-volume international centers performing a total 6186 PD cases during the 3-year study period, 2375 (38%) low-risk patients constituted the cohort for the benchmark analysis. The median patient follow-up after surgery was 23 months (IQR: 13–34 mo). In nearly half (*n* = 1091, 46%) of the cases, the indication for PD was adenocarcinoma, followed by ampullary carcinoma (*n* = 327, 13.8%) and cholangiocarcinoma (*n* = 266, 11.2%). Of all the patients with an oncological indication for PD, the majority (55%) presented a stage T3 disease. For patients with adenocarcinoma (*n* = 1091), 14% received neo-adjuvant therapy and 65% adjuvant treatment. Benign diseases such as Intraductal Papillary Mucinous Neoplasia (*n* = 197, 8.3%), chronic pancreatitis (*n* = 47, 2%), and other cystic lesions (*n* = 38, 8.3%) were less frequent. The texture of the pancreatic gland was reported as soft in 42.1% (*n* = 999) and hard in 34.7% (*n* = 824) of the cases. No information on texture was available in 23% of the cases (*n* = 551). Of note, among the 1682 patients with documented pancreatic duct measurement, more than half (62.7%, *n* = 1490) presented a duct size > 3 mm. Detailed baseline characteristics can be found in Supplementary Digital Content Table 1, <http://links.lww.com/SLA/B583>.

Outcome Benchmarks

Twenty benchmark cutoffs, derived from the 75th percentile of the medians for each outcome indicator per center, were established from this low-risk patient cohort (Table 2). The 6-months time point was chosen as follow-up to assess these outcome indicators, since the overall median cumulative patient morbidity expressed by the CCI[®] increased $\times 2.4$ from index-hospital discharge (CCI[®]: 8.6) until 6 months after surgery (CCI[®]: 20.9) for the entire benchmark cohort and did not increase significantly thereafter. (Supplementary Digital Content Figure S1, <http://links.lww.com/SLA/B583>).

In-hospital mortality rates and overall 6-months morbidity were $\leq 1.6\%$ and $\leq 73\%$, respectively with a readmission rate of $\leq 21\%$. The benchmark cutoff for postoperative clinically relevant pancreatic fistula rate (grade B–C) was $\leq 19\%$. The benchmark rate for failure-to-rescue was $\leq 9\%$ with the most frequent complication-related death being postoperative bleeding (0.3%).

Outcome Comparisons

The applicability of the benchmark values for outcome comparisons was tested in 2 independent cohorts of patients (Table 3).

First, we identified 931 patients, who underwent PD in the same 23 centers only differing from benchmark cases by a poorer preoperative physical status (ASA class ≥ 3). In this cohort, morbidity was within the benchmark while in-hospital mortality and failure-to-rescue rate were outside the established benchmark values (16% vs 9%).

TABLE 2. Twenty Outcome Benchmarks After PD in 2375 Low-risk Cases From Twenty-three International High-volume Centers

	Median Across the 23 Centers (Range)	Benchmark Cutoff
Operating duration	6.1 h (3.4–8.6)	≤7.5 h
Blood transfusions	13% (2–36.4)	≤23%
Hospital stay	12 d (6–31)	≤15 d
Postoperative 6-mos morbidity [†]		
Patients with at least 1 complication (graded as Clavien–Dindo)	65.3% (43.5–89.6)	≤73%
Patients with a complication Grades I–II	53% (30.6–86.5)	≤62%
Patients with a complication Grade ≥ III	19% (4.4–52.3)	≤30%
Patients with a complication Grade IV	2% (0–14)	≤5%
CCI [®] *	20.9 (0–33.5)	≤20.9
Pancreatic fistula rate (ISGPF classification B–C)	10% (0–35.4)	≤19%
Grade B	8% (0–35.4)	≤15%
Grade C	2% (0–12)	≤5%
Biochemical leak	8% (1.3–22.7)	≤13%
Severe postoperative bleeding (≥ III)	4% (0–14)	≤7%
In-hospital mortality	0% (0–4)	≤1.6%
Failure-to-rescue rate	0% (0–25)	≤9%
Readmission rate	16% (1.6–29.1)	≤21%
Oncological outcomes in patients with PADC (n = 1091)		
R1 rate	27% (2.3–67)	≤39%
Number of lymph nodes resected	19 (14–43)	≥16
1-yr DFS survival (actual)	42% (22.6–100)	≥53%
3-yrs DFS survival (actuarial)	0% (0–15.4)	≥9%

*CCI[®] ranges from 0 (no complication) to 100 (death).

[†]The values at discharge, 3 mo and 12 mo are listed in the supplementary material (Table S4, <http://links.lww.com/SLA/B583>).

Second, in a cohort of 224 patients undergoing minimally invasive PD in 3 centers, 79 patients (35%) fulfilled the selection criteria for benchmark (low-risk) cases. During the study period, the proportion of benchmark cases was higher in one of the centers performing robotic PD (62%) than in the other two minimal-invasive

PD (11% and 16%). In benchmark patients undergoing minimally invasive PD, operation duration and readmission rates exceeded the benchmark cutoffs in 2 of the 3 centers, while other more specific morbidity indicators, for example POPF rates, showed a large variation (0%–23%).

TABLE 3. Examples of Benchmark Comparisons in 4 Different Cohorts: One Cohort With a Poorer Preoperative Physical Status (ASA Class ≥3) and Three Single-center MIS Cohorts

	ASA Class ≥3 (n = 931)	Center MIS 1 (n = 8)	Center MIS 2 (n = 8)	Center MIS 3 (n = 62)	Benchmark Cutoff (75th Percentile of Medians From 23 Centers)
% Benchmark cases		8/37 (22%)	8/51 (16%)	62/100 (62%)	–
Operating duration	6 h	7.6 h [†]	n.a.	8.3 h [†]	≤7.5 h
Blood transfusions	14.7%	12.5%	n.a.	4.8%	≤23%
Hospital stay	10 d	4.5 d	6.5 d	16 d	≤15 d
6 mo postoperative morbidity					
Patients with at least 1 complication (graded as Clavien–Dindo)	60%	38%	63%	68%	≤73%
Patients with a complication Grades I–II	50%	38%	12.5%	62%	≤62%
Patients with a complication Grade ≥ III	19%	13%	50% [†]	13%	≤30%
Patients with a complication Grade IV	4%	0%	0%	0	≤5%
CCI [®] *	20.9	4	17.5	20.9	≤20.9
Pancreatic fistula rate (ISGPF Classification B–C)	10%	13%	0%	23% [†]	≤19%
Grade B	7%	13%	0%	21% [†]	≤15%
Grade C	3%	0%	0%	2%	≤5%
Biochemical leak	5%	13%	0%	8%	≤13%
Severe postoperative Bleeding (≥ III)	3%	0%	0%	10% [†]	≤7%
In-hospital mortality	3% [†]	0%	0%	3% [†]	≤1.6%
Failure-to-rescue	16% [†]	0%	0%	25% [†]	≤9%
1-yr readmission rate	20%	25% [†]	50% [†]	7%	≤21%
Oncological outcomes in patients with PADC					
R1 rate	21%	0	n.a.	n.a.	≤39%
Total of harvested Lymph nodes (number)	22	32	13 [†]	35	≥16
1-yr DFS survival (actual)	41% [†]	n.a.	n.a.	n.a.	≥53%
3-yrs DFS survival (actuarial)	2.5% [†]	n.a.	n.a.	n.a.	≥9%

*CCI[®] ranges from 0 (no complication) to 100 (death).

[†]Values are outside of the benchmark.

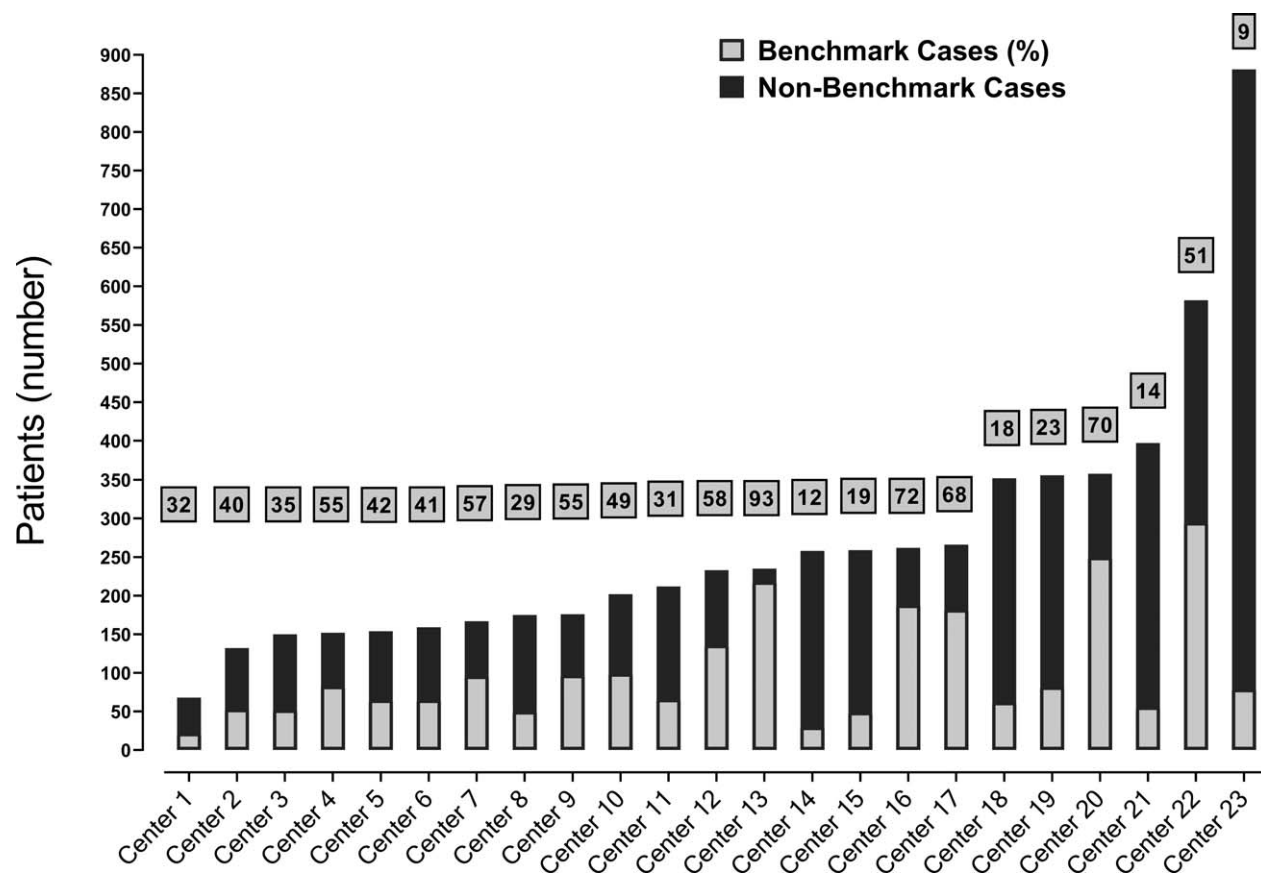


FIGURE 1. Large variations in the proportion of benchmark cases performed across the 23 centers.

Center-specific Risk Profile

The proportion of benchmark cases respective to the total PD volume greatly varied among centers, ranging from 9% to 93% (Fig. 1). This center-specific case-mix, however, did not correlate with the center-volume ($r = -0.14$, $P = 0.52$) (Fig. 2). We observed that higher proportion of benchmark cases correlated with higher fistula rates in the benchmark cases ($r = 0.53$; $P = 0.009$). For example, center #23 performed 9% benchmark cases ($n = 79$ out of total 881) with a postoperative pancreatic fistula rate (grades B–C) of 6% ($n = 5$). In contrast, center #13 performed mostly benchmark cases (93%, $n = 218$ out of a total of 235 patients) with a postoperative fistula rate (grades B–C) of 32% ($n = 70$) (Fig. 2). A similar correlation was observed for overall morbidity rates and proportion of benchmark cases ($r = 0.39$, $P = 0.06$, Figure S2, <http://links.lww.com/SLA/B583>). In addition, the centers performing <50% of benchmark cases disclosed a $\times 2$ lower failure-to-rescue rate (3.6% vs 7.0%, $P = 0.138$).

The same correlation between proportion of benchmark cases and outcomes was also observed in the higher risk cohort for example for POPF rates ($r = 0.45$, $P = 0.04$, Figure S3, <http://links.lww.com/SLA/B583>). When analyzing the 3 minimally invasive PD subgroups, the 2 centers performing the smallest proportion of benchmark cases displayed the lowest fistula rate (grades B–C) and in-hospital mortality (Table 3).

In addition, this center-specific risk profiles – based on the proportion of benchmark versus non benchmark cases – followed a

geographical distribution (Fig. 3). Centers in the US operated on a significantly higher proportion of cases outside of the benchmark criteria (82%), when compared with European (56%) and Asian (32%) centers ($P < 0.001$). In line with the correlation between the risk profile and outcomes, the overall postoperative fistula rate in the benchmark cases was lower in the US centers (6.5%), when compared with European and Asian centers (17% and 20%, $P < 0.001$) (Fig. 3). Of note, factors influencing the occurrence of POPF, for example soft pancreatic texture or the use of chemotherapy prior to surgery in patients with pancreas adenocarcinoma, had similar distributions across the 3 continents.

DISCUSSION

This study provides 20 benchmark values for surgical and oncological outcomes after PD showing low in-hospital mortality, but high morbidity rates increasing up to 6 months after surgery. The study also offers benchmarks for oncological indicators such a 1-year disease-free survival ($\geq 53\%$) or minimal number of harvested lymph nodes (≥ 16). The benchmark values were tested in 2 different cohorts of patients, 1 in patients from the same centers with a poorer preoperative physical status (ASA class ≥ 3) and another series of cases undergoing minimally invasive PD, which enabled us to detect and quantify specific outcome parameters beyond the benchmark values. This novel concept of benchmarks for complex surgical procedures may serve as a robust methodology to improve outcome comparisons among different patient groups, centers, and countries,

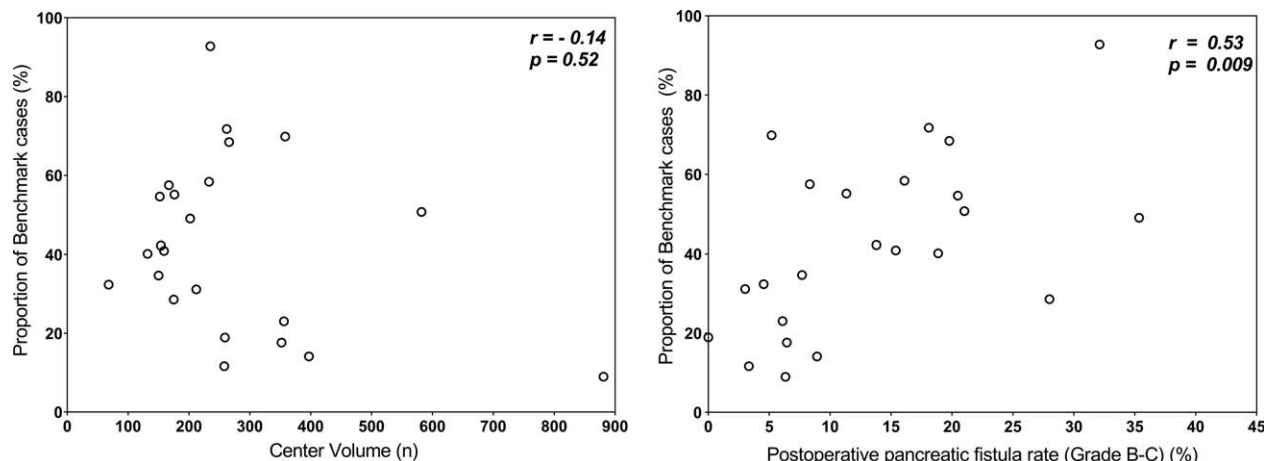


FIGURE 2. Relation of the proportion of benchmark cases with the total center volume and postoperative fistula rates (Grades B–C)*. *Each point in the graph represents 1 of the 23 centers included in the study.

or even to test whether single cases lie inside or outside benchmark values. Additionally, the results showed a correlation between the proportion of difficult cases performed in a center and center-specific outcomes, independently of the total case volume.

In an effort to apply the benchmark concept for pancreas surgery, this study followed a standardized and previously reported methodology by focusing on a low-risk patient population, called benchmark cases, from 23 international high-volume centers.¹⁰ From this large-scale patient cohort, reference values for 20 outcome parameters, that is, benchmark values or cutoffs, were established for indicators reflecting postoperative mortality, morbidity, and other clinically relevant endpoints like hospital stay and oncological indicators.

The results of the study show that even in benchmark cases undergoing PD, postoperative morbidity doubled between discharge and 6 months postoperatively as assessed by the overall cumulative morbidity index, CCI[®]. This finding is also well reflected by a benchmark readmission value of 21%. Based on this observation, reliable reports on outcomes in patients undergoing PD must cover at

least the first 6 postoperative months. Coupled to a recent study published in this journal showing a close correlation between CCI[®] and costs, we would expect that the financial burden for PD extends up to 6 months after surgery, which should be taken into account for proper reimbursement.⁴³

To illustrate the applicability of the established benchmark cutoffs for outcome comparisons with other cohorts undergoing PD, we first analyzed nearly thousand patients, only differing from the benchmark cases by a poorer preoperative physical status (ASA class ≥ 3). In this cohort, morbidity rates increased but stayed within the benchmark cutoffs. However, the failure-to-rescue rate showed a 2-fold increase compared with the benchmark, which lead to an in-hospital mortality outside the benchmark cutoffs.³³ Interestingly, similar outcome differences in patients differing by ASA class have been observed in previous studies.^{38,44} We believe that this example nicely shows that the established benchmarks enable us to disclose novel insights, by evaluating selective outcome indicators in patients differing only by baseline characteristics.

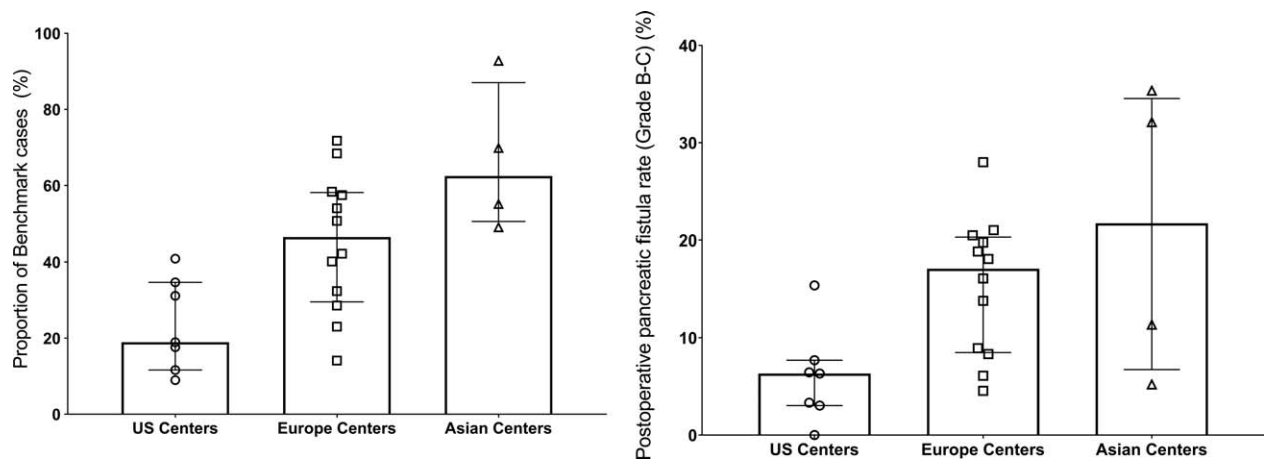


FIGURE 3. Geographical distribution of the single center proportion of benchmark cases and associated postoperative fistula rates (Expressed as median and interquartile range).

As a second illustration, we assessed outcomes after minimally invasive (laparoscopic or robotic) PD in benchmark patients treated at 3 centers with expertise in such approaches. Here, while the majority of outcome indicators were within the cutoffs, comparisons with the benchmarks allowed us to identify center-specific outcomes outside of the cutoffs, for example operation duration and readmission rates.

These examples show how benchmarks can be used for outcome comparisons within and between single center cohorts, and identify center-specific differences.

An unexpected finding of this study was the large variation in the proportion of benchmark cases across the 23 high-volume centers, all performing over 50 complex pancreatic procedures per year. We observed that centers operating on a high proportion of benchmark cases had inferior outcomes compared with centers performing more difficult cases, independently of the overall volume. The strongest correlation was observed for clinically relevant pancreatic fistulas (B and C). In other words, centers exposed to difficult cases had better results in their benchmark cases. A possible explanation for this observation is that a high proportion of complex cases enhance surgeon's skills, which may in turn improve management of complications through an experienced multidisciplinary team, and thereby preventing failure-to-rescue.⁴⁵ Indeed, in the current study the centers having less experience with difficult PD cases disclosed a $\times 2$ higher failure-to-rescue rates. We also noted a geographical distribution in the rate of complications and fistulas with superior results in the USA. This finding could be explained by the different risk profile among the regions, although of course other factors cannot be excluded.

A similar variation in the proportion of benchmark cases was observed among the 3 centers performing minimally invasive PD, which likewise may explain differences in outcomes. The center with the largest case volume disclosed the greatest proportion of benchmark cases, and also more indicators outside the benchmark values, than the other 2 centers with a smaller case load but more "high risk" cases.

A new finding of this study is therefore that center-risk profile may add to the current volume-based surgical assessment and centralization strategies.^{4,15,16,45-47} Case volume alone is an insufficient proxy for performance, if not coupled with information on the proportion of benchmark cases. The results of this study suggest that to gain a sufficient expertise in PD with better results, centers must accept and perform difficult cases irrespective of their caseload. The main message of this study, however, remains that assessment of centers or perhaps individual surgeons or new minimally invasive approaches should be based on the well-defined benchmark population to measure outcomes in a standardized patient population without the need of cumbersome risk adjustment. An application at the level of individual centers might be to use the benchmark cutoffs to identify cases to be presented at morbidity-mortality conferences. We have implemented this policy in Zurich, Switzerland to review all the specific steps within the care process in patients with a CCI[®] outside of the benchmark. Learning from these cases could reduce the failure-to-rescue rate, which is known to be a key factor for improved outcome in specialized centers.³³

This study has unavoidable limitations. First, the selection criteria for the low-risk benchmark patients did not include intraoperative features such as gland texture, pancreatic duct size, or other technical variations in performing the surgery. Although these factors do impact on outcomes, they are rarely available preoperatively, and thus lack pragmatic values. Therefore, we made the choice to exclusively focus on readily available and well-defined preoperative features. Second, while reporting of complications and pancreatic fistulas was standardized for all centers, oncological indicators such

as resection margins and number of resected lymph nodes were recorded according to the local protocols. Thus, some variations related to the difference in definitions and quality of the pathological examinations at different centers cannot be excluded. Since this study was not designed to assess outcomes after laparoscopic or robotic PD, no generalizable conclusions should be drawn from these comparisons, which serve as an illustration of the utility of benchmark values in outcome assessment.

The current study is a first attempt to apply benchmarking in pancreatic surgery. Further studies should now test these cutoffs in independent cohorts undergoing PD to validate and broaden this concept.

In conclusion, this international large-scale study proposes a novel approach for outcome analysis of PD by proposing benchmark values for clinically relevant outcome indicators. Individual outcome parameters from single patients, centers, geographical areas, or after minimally invasive procedures may be compared with the established benchmark values to enable an objective assessment of performance gaps and learning curves. In addition, center-risk profiles, based on the proportion of benchmark versus nonbenchmark cases may provide additional information on center practice and may reflect outcomes more accurately than center volume alone.

REFERENCES

- Hartwig W, Werner J, Jager D, et al. Improvement of surgical results for pancreatic cancer. *Lancet Oncol*. 2013;14:e476-e485.
- van Heek NT, Kuhlmann KF, Scholten RJ, et al. Hospital volume and mortality after pancreatic resection: a systematic review and an evaluation of intervention in the Netherlands. *Ann Surg*. 2005;242:781-788.
- Yoshioka R, Yasunaga H, Hasegawa K, et al. Impact of hospital volume on hospital mortality, length of stay and total costs after pancreaticoduodenectomy. *Br J Surg*. 2014;101:523-529.
- Hata T, Motoi F, Ishida M, et al. Effect of hospital volume on surgical outcomes after pancreaticoduodenectomy: a systematic review and meta-analysis. *Ann Surg*. 2016;263:664-672.
- Finks JF, Osborne NH, Birkmeyer JD. Trends in hospital volume and operative mortality for high-risk surgery. *N Engl J Med*. 2011;364:2128-2137.
- Wang J, Ma R, Churilov L, et al. The cost of perioperative complications following pancreaticoduodenectomy: a systematic review. *Pancreatol*. 2018;18:208-220.
- Wang J, Ma R, Eleftheriou P, et al. Health economic implications of complications associated with pancreaticoduodenectomy at a University Hospital: a retrospective cohort cost study. *HPB (Oxford)*. 2018;20:423-431.
- Porter ME, Larsson S, Lee TH. Standardizing patient outcomes measurement. *N Engl J Med*. 2016;374:504-506.
- von Eiff W. International benchmarking and best practice management: in search of health care and hospital excellence. *Adv Health Care Manag*. 2015;17:223-252.
- Staiger RD, Puhan M, Schwandt H, et al. Improving surgical outcomes through benchmarking. *Br J Surg*. 2019;106:59-64.
- Rosler F, Sapisochin G, Song G, et al. Defining benchmarks for major liver surgery: a multicenter analysis of 5202 living liver donors. *Ann Surg*. 2016;264:492-500.
- Schmidt HM, Gisbertz SS, Moons J, et al. Defining benchmarks for transthoracic esophagectomy: a multicenter analysis of total minimally invasive esophagectomy in low risk patients. *Ann Surg*. 2017;266:814-821.
- Muller X, Marcon F, Sapisochin G, et al. Defining benchmarks in liver transplantation: a multicenter outcome analysis determining best achievable results. *Ann Surg*. 2018;267:419-425.
- Porter ME. What is value in health care? *N Engl J Med*. 2010;363:2477-2481.
- van der Geest LG, van Rijssen LB, Molenaar IQ, et al. Volume-outcome relationships in pancreatoduodenectomy for cancer. *HPB (Oxford)*. 2016;18:317-324.
- Krautz C, Nimptsch U, Weber GF, et al. Effect of hospital volume on in-hospital morbidity and mortality following pancreatic surgery in Germany. *Ann Surg*. 2018;267:411-417.
- Williams TK, Rosato EL, Kennedy EP, et al. Impact of obesity on perioperative morbidity and mortality after pancreaticoduodenectomy. *J Am Coll Surg*. 2009;208:210-217.

18. Hartwig W, Hackert T, Hinz U, et al. Pancreatic cancer surgery in the new millennium: better prediction of outcome. *Ann Surg.* 2011;254:311–319.
19. Ronnekleiv-Kelly SM, Greenblatt DY, Lin CP, et al. Impact of cardiac comorbidity on early outcomes after pancreatic resection. *J Gastrointest Surg.* 2014;18:512–522.
20. Kazaure HS, Roman SA, Tyler D, et al. The significance of atrial fibrillation in patients aged \geq 55 years undergoing abdominal surgery. *World J Surg.* 2015;39:113–120.
21. Squires MH 3rd, Mehta VV, Fisher SB, et al. Effect of preoperative renal insufficiency on postoperative outcomes after pancreatic resection: a single institution experience of 1,061 consecutive patients. *J Am Coll Surg.* 2014;218:92–101.
22. Mita K, Ito H, Takahashi K, et al. Postpancreatectomy hemorrhage after pancreatic surgery in patients receiving anticoagulation or antiplatelet agents. *Surg Innov.* 2016;23:284–290.
23. Kristensen SD, Knuuti J, Saraste A, et al. 2014 ESC/ESA Guidelines on non-cardiac surgery: cardiovascular assessment and management: The Joint Task Force on non-cardiac surgery: cardiovascular assessment and management of the European Society of Cardiology (ESC) and the European Society of Anaesthesiology (ESA). *Eur J Anaesthesiol.* 2014;31:517–573.
24. Fields AC, Divino CM. Surgical outcomes in patients with chronic obstructive pulmonary disease undergoing abdominal operations: an analysis of 331,425 patients. *Surgery.* 2016;159:1210–1216.
25. Cornella N, Sancho J, Sitges-Serra A. Short and long-term outcomes after surgical procedures lasting for more than six hours. *Sci Rep.* 2017;7:9221.
26. Glance LG, Dick AW, Mukamel DB, et al. Association between intraoperative blood transfusion and mortality and morbidity in patients undergoing noncardiac surgery. *Anesthesiology.* 2011;114:283–292.
27. Hopewell S, Omar O, Hyde C, et al. A systematic review of the effect of red blood cell transfusion on mortality: evidence from large-scale observational studies published between 2006 and 2010. *BMJ Open.* 2013;3. e002154.
28. Thomas JW, Guire KE, Horvat GG. Is patient length of stay related to quality of care? *Hosp Health Serv Adm.* 1997;42:489–507.
29. Clavien PA, Puhan MA. Measuring and achieving the best possible outcomes in surgery. *Br J Surg.* 2017;104:1121–1122.
30. Tevis SE, Kennedy GD. Postoperative complications and implications on patient-centered outcomes. *J Surg Res.* 2013;181:106–113.
31. 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.
32. Bassi C, Marchegiani G, Dervenis C, et al. The 2016 update of the International Study Group (ISGPS) definition and grading of postoperative pancreatic fistula: 11 years after. *Surgery.* 2017;161:584–591.
33. Ghaferi AA, Birkmeyer JD, Dimick JB. Complications, failure to rescue, and mortality with major inpatient surgery in Medicare patients. *Ann Surg.* 2009;250:1029–1034.
34. Clavien PA, Vetter D, Staiger RD, et al. The comprehensive complication index (CCI (R)) added value and clinical perspectives 3 years “down the line”. *Ann Surg.* 2017;265:1045–1050.
35. Slankamenac K, Graf R, Barkun J, et al. The comprehensive complication index a novel continuous scale to measure surgical morbidity. *Ann Surg.* 2013;258:1–7.
36. Huebner M, Kendrick M, Reid-Lombardo KM, et al. Number of lymph nodes evaluated: prognostic value in pancreatic adenocarcinoma. *J Gastrointest Surg.* 2012;16:920–926.
37. de Geus SWL, Kasumova GG, Sachs TE, et al. Neoadjuvant therapy affects margins and margins affect all: perioperative and survival outcomes in resected pancreatic adenocarcinoma. *HPB (Oxford).* 2018;20:573–581.
38. Hackett NJ, De Oliveira GS, Jain UK, et al. ASA class is a reliable independent predictor of medical complications and mortality following surgery. *Int J Surg.* 2015;18:184–190.
39. Croome KP, Farnell MB, Que FG, et al. Pancreaticoduodenectomy with major vascular resection: a comparison of laparoscopic versus open approaches. *J Gastrointest Surg.* 2015;19:189–194.
40. Asbun HJ, Stauffer JA. Laparoscopic vs open pancreaticoduodenectomy: overall outcomes and severity of complications using the Accordion Severity Grading System. *J Am Coll Surg.* 2012;215:810–819.
41. Mesleh MG, Stauffer JA, Bowers SP, et al. Cost analysis of open and laparoscopic pancreaticoduodenectomy: a single institution comparison. *Surg Endosc.* 2013;27:4518–4523.
42. Stauffer JA, Coppola A, Villacreses D, et al. Laparoscopic versus open pancreaticoduodenectomy for pancreatic adenocarcinoma: long-term results at a single institution. *Surg Endosc.* 2017;31:2233–2241.
43. Staiger RD, Cimino M, Javed A, et al. The comprehensive complication index (CCI(R)) is a novel cost assessment tool for surgical procedures. *Ann Surg.* 2018;268:784–791.
44. Hightower CE, Riedel BJ, Feig BW, et al. A pilot study evaluating predictors of postoperative outcomes after major abdominal surgery: physiological capacity compared with the ASA physical status classification system. *Br J Anaesth.* 2010;104:465–471.
45. Vonlanthen R, Lodge P, Barkun JS, et al. Toward a consensus on centralization in surgery. *Ann Surg.* 2018;268:712–724.
46. Urbach DR. Pledging to eliminate low-volume surgery. *N Engl J Med.* 2015;373:1388–1390.
47. Epstein AM. Volume and outcome—it is time to move ahead. *N Engl J Med.* 2002;346:1161–1164.