Annals of Surgery Publish Ahead of Print
DOI:10.1097/SLA.0000000000007014

Setting the Standard in Robotic Whipple Surgery: International Multicenter Benchmark Analysis

Matthias Pfister, MD[1,2], Zhihao Li, MD[2,3], Fariba Abbassi, MD[2,4], Raphael L C Araujo, MD, PhD[5], Vikram A Chaudhari, MD, PhD[6], Rim Cherif, MD[7,8], Laleh Foroutani, MD[9], Fabio Giannone, MD, PhD[10], Michael Ginesini, MD[11], Abdallah Iben-Khayat, MD[12], Boram Lee, MD[13], Ricardo Nunez, MD[14], Chie Takishita, MD[15], Patrick W Underwood, MD[16], Emanuel Vigia, MD, PhD[17], Elaina Vivian, MPH[18], Yusuke Watanabe, MD, PhD[19], Taiga Wakabayashi, MD, PhD[20], Adnan Alseidi, MD[9], Fabrizio Di Benedetto, MD, PhD[21], Ugo Boggi, MD[11], Raffaele Brustia, MD, PhD[7,8], Tan-To Cheung, MD[22], Mary E Dillhoff, MD[16], Brian K P Goh, MD[23], Ho-Seong Han, MD, PhD[13], Benedetto Ielpo, MD, PhD[24], Jae Hoon Lee, MD, PhD[25], Marcel A Machado, MD[26], John Martinie, MD[5], Hugo Pinto Marques, MD, PhD[17], Alejandro Mejia, MD[18], Yuichi Nagakawa, MD, PhD[15], Masafumi Nakamura, MD, PhD[19], Patrick Pessaux, MD, PhD[10], Patricio M Polanco, MD[14], Olivier Saint-Marc, MD[12], Shailesh V Shrikhande, MD[6], Go Wakabayashi, MD, PhD[20], Pierre-Alain Clavien, MD, PhD[1, 2, ✉]

[1]Hub for Translational Research and Liver/GI Health, Zurich, Switzerland.
[2]Department of Surgery and Transplantation, University of Zurich, Zurich, Switzerland.
[3]Division of Transplantation Surgery, Department of Surgery, Mayo Clinic, Rochester, Minnesota, USA.
[4]Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland.

[5]Division of HPB Surgery, Department of Surgery, Carolina's Health Care Hospital, Charlotte, NC, USA.
[6]GI and HPB Surgical Oncology, Tata Memorial Centre, Homi Bhabha National Institute, Mumbai, India.
[7]Université Paris-Est Créteil, UPEC, Créteil, France, INSERM, U955, Team "Virus Hépatologie Cancer", Créteil, France.
[8]Department of Digestive and Hepatobiliary Surgery, Assistance Publique-Hôpitaux de Paris, Henri Mondor University Hospital, Créteil, France, GRCT Optima.
[9]Department of Surgery, University of California - San Francisco, San Francisco, California, USA.
[10]Division of Hepato-Biliary and Pancreatic Surgery, Nouvel Hôpital Civil, Hôpitaux Universitaires de Strasbourg, Strasbourg, France.
[11]Division of General and Transplant Surgery, University of Pisa, Pisa, Italy.
[12]Department of Surgery, Centre Hospitalier Regional Orleans, Orleans, France.
[13]Department of Surgery, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Korea.
[14]Department of Surgery, University of Texas Southwestern Medical Center, Dallas, TX, USA.
[15]Department of Gastrointestinal and Pediatric Surgery, Tokyo Medical University, Tokyo, Japan.
[16]Division of Surgical Oncology, Department of Surgery, The Ohio State University, Columbus, Ohio, USA.
[17]Hepato-Biliary-Pancreatic and Transplantation Centre, Curry Cabral Hospital, Centro Hospitalar Universitário de Lisboa Central, Lisbon, Portugal.
[18]The Liver Institute, Methodist Dallas Medical Center, Dallas, Texas, USA.
[19]Department of Surgery and Oncology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan.
[20]Center for Advanced Treatment of Hepatobiliary and Pancreatic Diseases, Ageo Central General Hospital, Saitama, Japan.
[21]HPB Surgery and Liver Transplant Unit, University of Modena and Reggio Emilia, Modena, Italy.
[22]Department of Surgery, The University of Hong Kong, Hong Kong, China.

[23]Department of Hepatopancreatobiliary and Transplant Surgery, Singapore General Hospital and National Cancer Centre Singapore, Singapore, Singapore.
[24]Hepato-Biliary and Pancreatic Surgery Unit, Hospital del Mar, Barcelona, Spain.
[25]Division of Hepato-Biliary and Pancreatic Surgery, Department of Surgery, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea.
[26]Department of Surgery, University of São Paulo, São Paulo, Brazil.

✉**Correspondence and Reprints**
Pierre-Alain Clavien, MD, PhD
Toblerstrasse 51
8044 Zurich, Switzerland
Tel.: +41 43 268 28 80
E-mail: clavien@access.uzh.ch

Keywords: pancreatoduodenectomy, Whipple, robotic surgery, benchmarking

Running head: Benchmarks for Robotic Whipple

Word count (main text): ~~2643~~2776

Tables and figures: 2 tables and 2 figures

Online-only supplemental data content: 4 supplemental tables, 2 supplemental figures

Funding: The authors received no commercial funding for this research. This study is supported by the LGID (Liver and Gastrointestinal Disease) Foundation and FZ-Healthcare Foundation.

ABBREVIATIONS

ASA	American Society of Anesthesiologists
BMI	Body Mass Index
CCI®	Comprehensive Complication Index®
CI	Confidence Interval
HPB	Hepato-Pancreato-Biliary

ICU	Intensive Care Unit
IDEAL	Idea, Development, Exploration, Assessment, Long-term follow-up
IQR	Interquartile Range
MIS	Minimally Invasive Surgery
OR	Odds Ratio
OS	Overall Survival
PDAC	Pancreatic Ductal Adenocarcinoma
PD-ROBOSCORE	Difficulty Score for Robotic Pancreatoduodenectomy
POPF	Postoperative Pancreatic Fistula
PPH	Post-Pancreatectomy Hemorrhage
RCT	Randomized Controlled Trial
RFS	Recurrence-Free Survival

STRUCTURED ABSTRACT

Objective

To establish international benchmark values for relevant outcome parameters in robotic Whipple.

Summary Background Data

For safe adoption of surgical innovation, robust quality control is essential. Benchmarking is a validated tool for assessing surgical performance. Recent international consensus identified establishing benchmark values for robotic Whipple as top priority.

Methods

We analyzed consecutive patients undergoing robotic Whipple between 2020-2023 with a minimum one-year follow-up. Reference centers were required to perform ≥15 cases/year, be scientifically active in the field, and maintain a prospective database. Benchmark criteria included benign or resectable malignant disease without neoadjuvant therapy, arterial resection, major co-morbidities, or significant previous abdominal surgery. Benchmarks were established for 13 outcome parameters.

Result

The benchmark cohort comprised 418 patients from 12 centers across four continents. Benchmark values were: conversion rate ≤4.3%, transfusion rate ≤2.1%, 6-month mortality ≤2.2%, major complications ≤23.2%, and CCI® ≤20.9. Clinically relevant pancreatic fistula

(grade B/C) and hemorrhage (grade B/C) rates were ≤23.6% and ≤12.7%, respectively. For pancreatic ductal adenocarcinoma (n=123), the benchmark for lymph node yield was ≥20. Higher surgical difficulty was associated with increased overall postoperative morbidity (R^2=0.38, p=0.019), higher center caseload with reduced pancreas-specific complications (R^2=0.28, p=0.044). Independent POPF predictors included duct diameter ≤4mm (OR 1.37, 95% CI: 1.03, 1.82), anticoagulation (OR 2.45, 95% CI: 1.47, 3.99), and indication other than PDAC (OR 2.33, 95% CI: 1.68, 3.27).

Conclusions

This study establishes the first international benchmarks for robotic Whipple, demonstrating oncologic outcomes and morbidity comparable to open surgery with the benefits of minimally invasive surgery.

INTRODUCTION

The adoption of minimally invasive surgery (MIS) in pancreatic surgery has progressed unevenly through the IDEAL framework for the development of surgical innovation [1,2], with Whipple (i.e., pancreatoduodenectomy) procedures lagging behind distal pancreatectomy [3]. This discrepancy is primarily related to the technical limitations of laparoscopy in advanced procedures requiring complex reconstruction, which have resulted in steep learning curves [4] and significant safety concerns [5]. Laparoscopic Whipple remains restricted to a few high-volume centers with advanced laparoscopic skills, hindering widespread implementation and denying most patients the benefits of MIS known from other procedures, such as reduced blood loss, lower conversion rates, and faster functional recovery [6-8].

Robotic surgery, with its technical advantages over laparoscopy, holds the potential to overcome these limitations and facilitate broader adoption of MIS for Whipple procedures. While accumulating observational data supports its feasibility and safety in selected patients [9-11], recent randomized-controlled trials (RCTs) show conflicting results [12,13]. The EUROPA trial [13], comparing robotic and open Whipple surgery, reported higher rates of pancreas-specific complications without demonstrating the expected advantages of MIS. Albeit not significant, higher R1 resection rates additionally raised concerns about the oncologic outcomes of the robotic approach in malignant disease. Therefore, the recent Paris Consensus on robotic HPB surgery [3] emphasized the need for further robust data and identified benchmarking as a top research priority. Meanwhile–fueled by many surgeons' enthusiasm–robotic Whipple surgery is undoubtedly here to stay [14,15], making robust performance quality control more important for its safe adoption.

Benchmarking is a validated tool for improving quality of surgical performance and enables the comparison of outcomes across surgical techniques on an international scale [16,17]. It has been recognized as one of the critical features and milestones in the structured assessment of surgical innovation and new procedures [18]. Critical elements in the lifecycle of a new advancement include the creation of an international registry, a timely consensus conference, the generation of level-one data, and finally the iterative designation of benchmark values (Supplemental Figure 1, supplemental data content 1, http://links.lww.com/SLA/F711) [18]. Such best-achievable results have been established for numerous surgical procedures —including robotic distal pancreatectomy [19,20] and open Whipple surgery [21,22]—but are lacking for robotic Whipple.

To address this gap, this large-scale, international study aimed to establish benchmark cutoffs for key outcome parameters in robotic Whipple surgery. Using validated benchmarking methodology, these reference values provide a crucial framework for evaluating robotic Whipple outcomes and enable direct comparisons with the open approach.

METHODS

Study Design

We analyzed consecutive adult (≥ 18 years) patients who underwent robotic Whipple surgery (interchangeably used with pancreatoduodenectomy throughout this manuscript) for benign or malignant indications from January 2020 until December 2023 in international expert centers. We employed the validated 10-step approach to establish benchmark cutoffs [16]. Benchmark centers (i.e., international reference institutions from which benchmark values are derived) had to fulfil the following criteria: i) caseload of ≥ 15 robotic Whipple procedures/year, ii) active research in the field, and iii) maintenance of prospective data collection for pancreatic surgery. The first 40 patients provided by each center that did not embark on robotic Whipple surgery prior to the study period were omitted from the analysis to account for a potential learning curve (threshold defined by learning curve (CUSUM) analyses [23] and applied in the most recent RCT [12]). A minimum patient follow-up of one year was required. Patients were stratified into a benchmark (i.e., low risk) and a non-benchmark cohort. Benchmark patients were defined as those with benign disease or upfront resectable malignant disease, no neoadjuvant therapy, no arterial resection, American Society of Anesthesiologists (ASA) score ≤ 2, body mass index (BMI) ≥ 18.5 and < 35, no major comorbidities, no history of previous pancreatic or major abdominal surgery (Supplemental Table 1, supplemental data content 1, http://links.lww.com/SLA/F711). We applied the same selection criteria as for published benchmark studies on open Whipple surgery to allow for comparability of results [22]. Then, we defined cutoff values for relevant outcome parameters. Ethical approval was obtained from the Cantonal Ethics Committee of Zurich (BASEC 2024-01461) and by each participating center according to applicable regulations.

Outcome Parameters and Definitions

Benchmark cutoffs were defined for clinically relevant outcome parameters. Operative metrics: operative time, conversion to open (i.e., procedure started robotically but requiring laparotomy for any reason except specimen extraction), blood transfusion rate, intensive care unit (ICU) length of stay; pancreas-specific complications: any pancreas-specific complication (grade B/C), Post-Operative Pancreatic Fistula (POPF, grade B/C [24]), Post-Pancreatectomy Hemorrhage (PPH, grade B/C [25]); surgical complications at 6 months: any complication, major complication (Clavien Dindo grade ≥ 3a [26-28]), mortality, Comprehensive Complication Index® (CCI® [29]); oncologic outcome: number of lymph nodes harvested. The 6-month follow-up cutoff to assess morbidity was chosen according to established benchmark analyses for open Whipple surgery [22], where morbidity plateaued after this period.

Oncological resectability was defined as in the National Comprehensive Cancer Network guidelines [30]. Surgical difficulty was assessed using the PD-ROBOSCORE (1.6-23 points) [31] and patients were classified into low-, intermediate-, and high-difficulty subgroups based on tertiles of the score. This is in line with the previously published approach [31]. Overall survival (OS) was defined as the time from surgery to death from any cause or last follow-up; Recurrence-Free Survival (RFS) as the time from surgery to either disease recurrence, death from any cause, or last oncological follow-up. The use of anticoagulation was defined as any anticoagulation medication other than single platelet inhibition (e.g., acetylic salicylic acid) in therapeutic dose administered at least preoperatively.

Statistical Analysis

The 75th and 25th percentile across the benchmark centers' medians represented the benchmark cutoff values for negative and positive outcome parameters, respectively. Continuous variables were presented as median with its interquartile range (IQR), and categorical variables as count with percentage. Continuous variables were compared using Mann-Whitney U Test, categorical variables using Chi-square/ Fisher's exact Test and logistic regression. Time-to-event outcomes were analyzed using Kaplan-Meier and Cox proportional hazards regression model. The Pearson correlation coefficient (R^2) was calculated to evaluate the association between center characteristics and outcomes. Multivariable logistic regression using bidirectional stepwise variable selection was employed to assess predictors for POPF grade B/C. P-values are two-sided and reported up to

3 decimal places. A predefined significance level of ≤ 0.05 was used. Statistical analyses were performed using R Statistical Software (v4.3.2, R Core Team, Vienna, Austria).

RESULTS

Baseline Characteristics

21 centers contributed a total of 1526 cases. Of these, 12 qualified as benchmark centers spanning four continents (Asia n=6, Europe n=2, North America n=3, and South America n=1) performing a total of 1079 robotic Whipple procedures. The final benchmark cohort consisted of 418 patients (39 %). The proportion of benchmark cases varied considerably among centers (range: 11 – 80 %) (Supplemental Figure 2, supplemental data content 1, http://links.lww.com/SLA/F711). The median follow-up among benchmark patients was 24.9 months (IQR: 14.5-42.1). Baseline characteristics for the benchmark cohort and the entire study population are depicted in Supplemental Table 2, supplemental data content 1, http://links.lww.com/SLA/F711. Benchmark patients had a median age of 61 years (IQR: 53-68); the majority (54 %) were male. The median PD-ROBOSCORE was 4.8 (IQR: 4.8-8.2), reflecting low surgical difficulty. Thirty per cent (n=123) underwent robotic Whipple for pancreatic ductal adenocarcinoma (PDAC). Other indications included ampullary carcinoma (18 %), distal cholangiocarcinoma (15 %), pancreatic neuroendocrine neoplasms (12 %), and benign disease (23 %).

Benchmark Cutoffs and Long-Term Outcome in the Benchmark Cohort

Benchmark cutoffs were defined for 13 outcome parameters (Table 1). The cutoffs for operative metrics included operative time ≤ 514 min, conversion to open ≤ 4.3 %, blood transfusion ≤ 2.1 %, and ICU length of stay ≤ 1 day. The cutoffs for pancreatic-specific complications were any pancreas-specific complication B/C ≤ 35.2 %, POPF B/C ≤ 23.6 % and PPH B/C ≤ 12.7 %. The cutoffs for surgical complications after 6 months were ≤ 90.6 % for any complication, ≤ 23.2 % for major (Clavien-Dindo ≥ 3a) complications, ≤ 2.2 % for mortality, and ≤ 20.9 CCI® points. The cutoffs for lymph nodes harvested was ≥ 20. Compared to non-benchmark patients, the benchmark cohort exhibited significantly lower conversion rates, less blood transfusions, lower postoperative pancreas-specific and overall

morbidity (Supplemental Table 3, supplemental data content 1, http://links.lww.com/SLA/F711).

Figure 1 illustrates the Kaplan-Meier survival analyses in PDAC patients meeting benchmark criteria. Median OS (Figure 1A) was 33.7 months; median RFS (Figure 1B) 29.6 months. Actuarial OS and RFS were ≥ 92 % and ≥ 84 % at 1 year, and ≥ 45 % and ≥ 44 % at 3 years, respectively.

Correlation of Center Characteristics with Postoperative Outcome in the Entire Cohort

Figure 2A-D presents the association between center characteristics, i.e., case complexity and caseload per center, and overall and pancreas-specific postoperative morbidity, across all participating centers. Higher surgical complexity, measured by the PD-ROBOSCORE, positively correlated with an increased CCI® score (R^2=0.38, p=0.019). Subgroup analysis (Supplemental Table 4, supplemental data content 1, http://links.lww.com/SLA/F711) further confirmed that cases of higher difficulty show higher rates of conversion to open (p<0.001), higher rates of any (p<0.001) and major (p=0.004) complications, as well as an increased CCI® score (p<0.001). Surgical difficulty, however, was not associated with increased pancreas-specific morbidity (R^2=0.02, p=0.66). In contrast, higher technical expertise as in increased surgical caseload was associated with a lower incidence of pancreas-specific complications in both the entire cohort (R^2=0.28, p=0.044) and specifically in non-benchmark patients (R^2=0.28, p=0.042), but did not correlate with overall postoperative morbidity (R^2=0.11, p=0.26).

Risk Factors for POPF grade B/C in the Entire Cohort

Multivariable logistic regression identified three independent risk factors for developing POPF grade B/C (Table 2): anticoagulation (OR 2.45, 95% CI: 1.47, 3.99), duct diameter < 4mm (OR 1.37, 95% CI: 1.03, 1.82), and indication other than PDAC (OR 2.33, 95% CI: 1.68, 3.27). Surgical difficulty, increased BMI, previous major abdominal surgery, and cases requiring venous resection were not associated with increased POPF rates.

DISCUSSION

This international multicenter study establishes benchmark cutoffs for robotic Whipple surgery for key surgical outcome parameters. Applying validated benchmarking

methodology, we demonstrate that outcomes achieved in expert centers are broadly comparable to those previously reported for the open approach [21,22] while introducing the benefits of MIS to this cohort and thereby supporting its broader adoption. Besides substantiating the feasibility and safety of robotic Whipple, this study further confirms its oncological benefit in patients treated for PDAC with favorable patient survival.

By applying the same inclusion criteria as open Whipple benchmark studies [21,22], this analysis enables direct comparison across surgical modalities–a key benefit of benchmarking (Table 1). Still, potential differences in patient selection not captured by the inclusion criteria, as well as potential temporal bias cannot be excluded. While the duration of surgery for the robotic approach is one hour longer compared to the open benchmark (≤ 514min vs. ≤ 450min), the robot demonstrates lower blood transfusion rates (≤ 2.1% vs. ≤ 23%) with comparable rates of surgical morbidity and mortality. This is in line with the commonly discussed trade-off of minimally invasive surgery also observed in other procedures. Importantly, operative time–though dependent on technical complexity–is anticipated to approach standards in open surgery with increasing adoption and standardization. Notably, benchmarks for pancreas-specific complications including the index complication POPF are similar between the robotic and open approach, underscoring the safety of robotic reconstruction even in complex surgical settings. Multivariable analysis confirms that traditional risk factors for POPF—such as small duct diameter and indication other than PDAC—retain their predictive value irrespective of surgical technique, while the robotic approach may possibly even be an independent protective factor as evidenced by multicentric observational evidence [32]. Anticoagulation is most likely a surrogate for the presence of medical risk factors indicating a higher risk subgroup.

These important findings stand in contrast to the recent EUROPA trial comparing robotic to open Whipple [13], which reported a significantly higher incidence of grade B/C pancreas-specific complications following robotic resection (59% *vs.* 33%, p=0.046) without demonstrating MIS-related benefits. EUROPA concluded the safety of the robotic approach based on the non-inferiority in 90-day CCI® (34 *vs.* 36 points, p=0.713), but reported values beyond the newly established benchmark cutoffs in both groups. While not directly comparable to the current benchmark cohort because of differences in selection criteria, this may support prior observations that adverse outcomes in randomized trials may be subject to an institutional learning curve or procedural immaturity, rather than intrinsic limitations of

the robotic technique. In fact, recent evidence indicates that achieving maturity in terms of mastery with robotic Whipple requires far more procedures than those needed to complete the competency learning curve [33]. Mastery is associated not only with improved outcomes but also with the capacity to manage more complex cases. This highlights the critical role of benchmarking as an indispensable milestone in the systematic adoption of surgical innovation [18]. Importantly, another recent RCT from China comparing robotic to open Whipple after the initial learning curve [12], found postoperative morbidity within both open and robotic benchmark cutoffs, thus further reinforcing the reproducibility of robotic Whipple in mature programs.

A central concern in robotic Whipple surgery pertains to the noninferiority of oncologic outcomes. The EUROPA trial reported increased R1-resection rates of 18% to 0% observed in the robotic vs the open group [13]. This difference, however, did not reach statistical significance. The current benchmark study establishes a benchmark of ≥ 20 lymph nodes harvested, which exceeds established benchmarks for open Whipple (≥ 16). This finding likely reflects the combination of two critical aspects: i) the high surgical quality achieved in low-risk patients treated at high-volume expert centers after the learning curve, and ii) the enhanced visualization and dexterity afforded by robotic systems, particularly during deep dissection and vascular margin clearance [3]. One limitation of this study was the inconsistent reporting of surgical margins without uniform definition ~~(i.e., > 1 mm)~~ for R0 across centers. Therefore, no robust benchmark value for R0 resection rate could be determined. Future benchmarking initiatives on robotic Whipple surgery should focus on deriving robust R0 resection rate cutoffs based on standardized reporting practice across centers (e.g., > 1 mm). Corresponding survival data—although limited through the lack of adjustment for ~~not stratified by~~ margin status, the receipt of adjuvant therapy and tumor stage—however, ~~strongly~~ support favorable oncologic outcome of the robotic approach in upfront resectable PDAC [34].

Conversion remains a critical quality indicator in minimally invasive surgery. In line with available observational evidence [35], our data show conversion rates (≤ 4.3%) consistently lower than laparoscopy, thus underlining the technical advantages inherent with robotic systems. This technical advantage is particularly relevant during the reconstruction phase, where robotic articulation enables precise suturing in constrained anatomical environments. In line with recent international consensus [3], this supports the superior

reconstructive capabilities of the robot and strengthens its case in high-complexity procedures such as Whipple surgery. Still, surgical difficulty remains an important factor in clinical decision making, as it increases the risk of conversion and surgical morbidity [31]. When interpreting conversion rates, however, inconsistent distinction between urgent vs nonurgent conversion across studies needs to be considered. Similarly, because of the retrospective nature of the current study and limited information on the intraoperative course we could not reliably determine the reason for conversion. Importantly, urgent robotic conversions may represent a logistical challenge that impair surgical performance and outcome [36]. Therefore, for emergent scenarios–particularly in robotic surgery–standardized team-based conversion protocols are indispensable to ensure patient safety [3].

This study has some limitations. First, its retrospective design inherently introduces potential biases related to data collection and reporting, mitigated by prospective data capture and meticulous control of reporting quality. Second, the robotic approach remains in an intermediate phase of adoption, and although early institutional learning curve effects were excluded, residual center-specific and inter-surgeon variability cannot be entirely ruled out. Third, caseload varied considerably across participating centers with large centers contributing most cases. The robust benchmark methodology, however, counteracts such potential imbalances by weighing each center outcome parameter equally. Still, residual heterogeneity across centers cannot be excluded and may have influenced the results. Finally, no data on pancreatic texture was available. This well-established predictor of POPF in open surgery [37] remains more difficult to assess using the robotic technique as it is exclusively based on visual cues and needs further standardization.

In conclusion, this global analysis provides robust benchmark cutoffs for robotic Whipple surgery, supporting its evidence-based adoption. It affirms the approach's safety and oncologic adequacy in expert hands and highlights the key advantages of the robot compared to other modalities. Careful patient selection remains critical also in the robotic approach to ensure a favorable patient outcome. By defining objective reference values, this study provides guidance for adopting and established robotic HPB units and represents a critical tool for surgical quality improvement. The establishment of an international registry will be paramount to guide the robotic Whipple's further adoption, particularly in more technically challenging scenarios like borderline tumors or the need for vascular resection.

REFERENCES

1. McCulloch M, Altman DG, Campbell WB, et al. No surgical innovation without evaluation: the IDEAL recommendations. Review. *The Lancet*. 2009;374(9695):1105-1112.
2. Barkun JS, Dimick JB, Clavien PA. Surgical Research in Patients: Ideal Time for an IDEAL Checklist. *Ann Surg*. Feb 2019;269(2):208-210.
3. Hobeika C, Pfister M, Geller D, et al. Recommendations on Robotic Hepato-Pancreato-Biliary Surgery. The Paris Jury-Based Consensus Conference. *Ann Surg*. Jan 1 2025; 281(1):136-153.
4. Müller PC, Kuemmerli C, Cizmic A, et al. Learning Curves in Open, Laparoscopic, and Robotic Pancreatic Surgery: A Systematic Review and Proposal of a Standardization. *Ann Surg Open*. Mar 2022;3(1):e111.
5. van Hilst J, de Rooij T, Bosscha K, et al. Laparoscopic versus open pancreatoduodenectomy for pancreatic or periampullary tumours (LEOPARD-2): a multicentre, patient-blinded, randomised controlled phase 2/3 trial. *Lancet Gastroenterol Hepatol*. 03 2019;4(3):199-207.
6. Mariette C, Markar SR, Dabakuyo-Yonli TS, et al. Hybrid Minimally Invasive Esophagectomy for Esophageal Cancer. *N Engl J Med*. 01 10 2019;380(2):152-162.
7. de Rooij T, van Hilst J, van Santvoort H, et al. Minimally Invasive Versus Open Distal Pancreatectomy (LEOPARD): A Multicenter Patient-blinded Randomized Controlled Trial. *Ann Surg*. 01 2019;269(1):2-9.
8. Bonjer HJ, Deijen CL, Abis GA, et al. A randomized trial of laparoscopic versus open surgery for rectal cancer. *N Engl J Med*. Apr 02 2015;372(14):1324-1332.
9. Nickel F, Wise P, Müller PC, et al. Short-term Outcomes of Robotic Versus Open Pancreatoduodenectomy - Propensity Score-matched Analysis. *Ann Surg*. April 1 2024;279(4):665-670.

10. de Graaf N, Zwart MJW, van Hilst J, et al. Early experience with robotic pancreatoduodenectomy versus open pancreatoduodenectomy: nationwide propensity-score-matched analysis. *Br J Surg*. Jan 31 2024;111(2)

11. Zureikat AH, Postlewait LM, Liu Y, et al. A Multi-institutional Comparison of Perioperative Outcomes of Robotic and Open Pancreaticoduodenectomy. *Ann Surg*. Oct 2016;264(4):640-649.

12. Liu Q, Li M, Gao Y, et al. Effect of robotic versus open pancreaticoduodenectomy on postoperative length of hospital stay and complications for pancreatic head or periampullary tumours: a multicentre, open-label randomised controlled trial. *Lancet Gastroenterol Hepatol*. May 2024;9(5):428-437.

13. Klotz R, Mihaljevic AL, Kulu Y, et al. Robotic versus open partial pancreatoduodenectomy (EUROPA): a randomised controlled stage 2b trial. *Lancet Reg Health Eur*. Apr 2024;39:100864.

14. Khachfe HH, Nassour I, Hammad AY, et al. Robotic Pancreaticoduodenectomy: Increased Adoption and Improved Outcomes: Is Laparoscopy Still Justified? *Ann Surg*. Sep 01 2023;278(3):e563-e569.

15. Emmen AMLH, de Graaf N, Khatkov IE, et al. Implementation and outcome of minimally invasive pancreatoduodenectomy in Europe: a registry-based retrospective study - a critical appraisal of the first 3 years of the E-MIPS registry. *Int J Surg*. Apr 01 2024;110(4):2226-2233.

16. Gero D, Muller X, Staiger RD, et al. How to Establish Benchmarks for Surgical Outcomes?: A Checklist Based on an International Expert Delphi Consensus. *Ann Surg*. Jan 01 2022;275(1):115-120.

17. Domenghino A, Walbert C, Birrer DL, et al. Consensus recommendations on how to assess the quality of surgical interventions. *Nat Med*. Apr 2023;29(4):811-822.

18. Linecker M, Pfister M, Kambakamba P, et al. Assessing Surgical Innovation. ALPPS: An IDEAL Example of Disruptive Innovation. *Ann Surg*. Aug 07 2025. Online ahead of print.

19. Müller PC, Breuer E, Nickel F, et al. Robotic Distal Pancreatectomy: A Novel Standard of Care? Benchmark Values for Surgical Outcomes From 16 International Expert Centers. *Ann Surg*. Aug 01 2023;278(2):253-259.

20. van Ramshorst TME, Giani A, Mazzola M, et al. Benchmarking of robotic and laparoscopic spleen-preserving distal pancreatectomy by using two different methods. *Br J Surg*. Dec 13 2022;110(1):76-83.

21. Raptis DA, Sánchez-Velázquez P, Machairas N, et al. Defining Benchmark Outcomes for Pancreatoduodenectomy With Portomesenteric Venous Resection. *Ann Surg*. Nov 2020;272(5):731-737.

22. Sánchez-Velázquez P, Muller X, Malleo G, et al. Benchmarks in Pancreatic Surgery: A Novel Tool for Unbiased Outcome Comparisons. *Ann Surg*. 08 2019;270(2):211-218.

23. Zhang T, Zhao ZM, Gao YX, et al. The learning curve for a surgeon in robot-assisted laparoscopic pancreaticoduodenectomy: a retrospective study in a high-volume pancreatic center. Article. *Surgical Endoscopy*. 2019;33(9):2927-2933.

24. 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*. 03 2017;161(3):584-591.

25. Wente MN, Veit JA, Bassi C, et al. Postpancreatectomy hemorrhage (PPH): an International Study Group of Pancreatic Surgery (ISGPS) definition. *Surgery*. Jul 2007;142(1):20-25.

26. 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*. Aug 2004;240(2):205-213.

27. Clavien PA, Barkun J, de Oliveira ML, et al. The Clavien-Dindo classification of surgical complications: five-year experience. *Ann Surg*. Aug 2009;250(2):187-196.

28. Abbassi F, Pfister M, Lucas KL, et al. Milestones in Surgical Complication Reporting. Clavien-Dindo Classification 20 Years & Comprehensive Complication Index (CCI®) 10 Years. *Ann Surg*. Nov 1 2024;280(5):763-771.

29. Slankamenac K, Graf R, Barkun J, et al. The comprehensive complication index: a novel continuous scale to measure surgical morbidity. *Ann Surg*. Jul 2013;258(1):1-7.

30. Tempero MA. NCCN Guidelines Updates: Pancreatic Cancer. *J Natl Compr Canc Netw*. May 01 2019;17(5.5):603-605.

31. Napoli N, Cacace C, Kauffmann EF, et al. The PD-ROBOSCORE: A difficulty score for robotic pancreatoduodenectomy. *Surgery*. Jun 2023;173(6):1438-1446.

32. Vining CC, Kuchta K, Berger Y, et al. Robotic pancreaticoduodenectomy decreases the risk of clinically relevant post-operative pancreatic fistula: a propensity score matched NSQIP analysis. *HPB (Oxford)*. Mar 2021;23(3):367-378.

33. Napoli N, Ginesini M, Kauffmann EF, et al. Navigating the learning curve of robotic pancreatoduodenectomy: Competency, proficiency, and mastery in a first-generation robotic surgeon with established open pancreatic expertise. *Surgery*. Aug 2025;184:109347.

34. Conroy T, Hammel P, Hebbar M, et al. FOLFIRINOX or Gemcitabine as Adjuvant Therapy for Pancreatic Cancer. *N Engl J Med*. Dec 20 2018;379(25):2395-2406.

35. Lof S, Vissers FL, Klompmaker S, et al. Risk of conversion to open surgery during robotic and laparoscopic pancreatoduodenectomy and effect on outcomes:

international propensity score-matched comparison study. Article. *British Journal of Surgery*. 2021;108(1):80-87.

36. Slavin M, Ross SB, Sucandy I, et al. Unplanned conversions of robotic pancreaticoduodenectomy: short-term outcomes and suggested stepwise approach for a safe conversion. *Surg Endosc*. Feb 2024;38(2):964-974.

37. Schuh F, Mihaljevic AL, Probst P, et al. A Simple Classification of Pancreatic Duct Size and Texture Predicts Postoperative Pancreatic Fistula: A classification of the International Study Group of Pancreatic Surgery. *Ann Surg*. Mar 01 2023;277(3):e597-e608.

Figure 1. Overall and Recurrence-Free Survival in Benchmark Patients with Pancreatic Ductal Adenocarcinoma

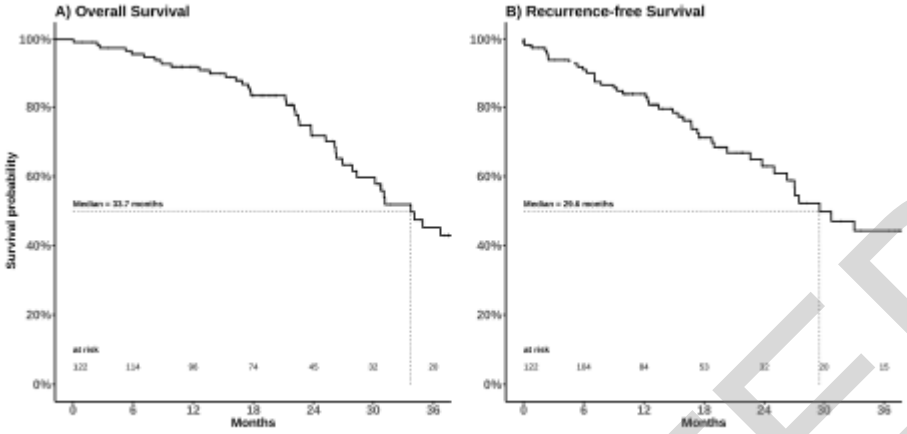

Figure 1.

Figure 2. Correlation of Center Characteristics with Postoperative Outcome in all Patient

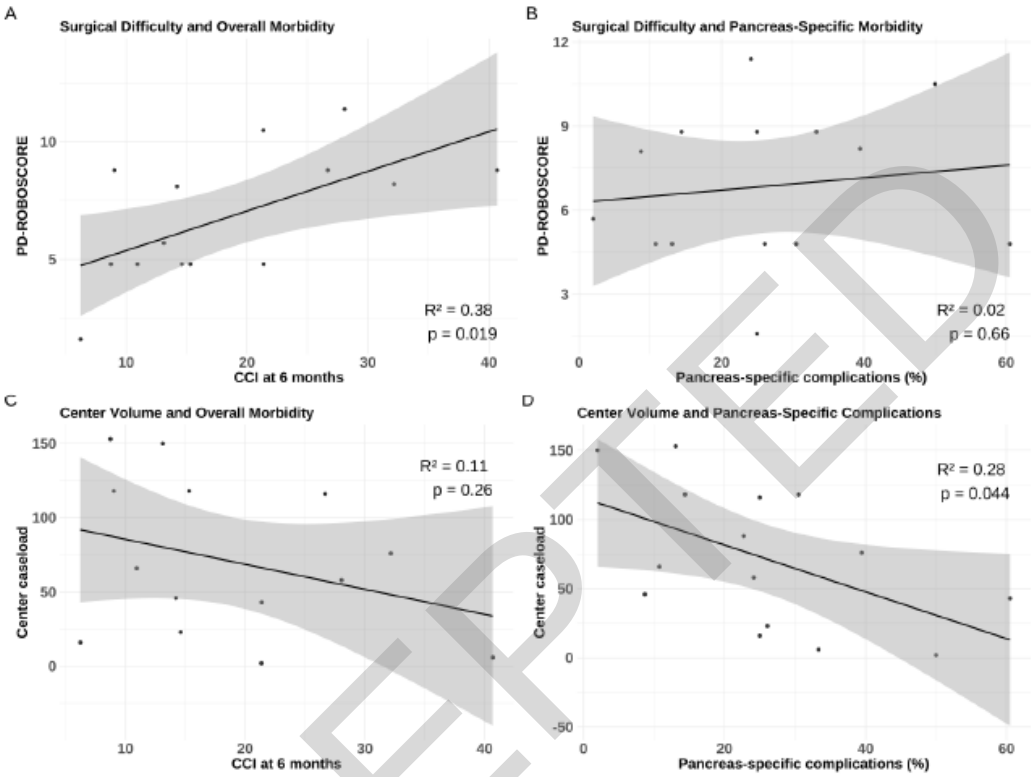

Figure 2.

Table 1: Benchmark Cutoffs for Robotic Whipple Compared to Open Whipple

	Robotic Whipple		**Open Whipple**[1]	
	(12 centers, 418 patients)		(23 centers, 2375 patients)	
	Median	**Benchmark***	**Median**	**Benchmark***
Operative metrics				
Duration of surgery (min)	≤402	**≤514**	≤366	≤450
Conversion to open (%)	≤0.0	**≤4.3**	-	-
Blood transfusion (%)	≤0.0	**≤2.1**	≤13	≤23
ICU length of stay (days)	≤1	**≤1**	-	-
Pancreas-specific complications				
Any complication grade B/C (%)	≤18.8	**≤35.2**	-	-
POPF grade B/C (%)	≤13.4	**≤23.6**	≤10	≤19
PPH grade B/C (%)	≤3.2	**≤12.7**	≤4	≤13
Surgical complications (6 months)				
Any complication (%)	≤62.1	**≤90.6**	≤65.3	≤73
Clavien-Dindo grade ≥3a (%)	≤17.5	**≤23.2**	≤21.0	≤35
Mortality (%)	≤0.0	**≤2.2**	-	-
CCI®	≤0.0	**≤20.9**	≤20.9	≤20.9
PDAC (n = 123)				
Lymph nodes harvested (n)	≥22	**≥20**	≥19	≥16

*75th/25th-percentile of centers medians.

[1]Sánchez-Velázquez et al., [2]Mean across centers, [3]Disease-free survival

Abbreviations: Min (minutes), ICU (Intensive Care Unit), POPF (Post-Operative Pancreatic Fistula), PPH (Post-Pancreatectomy Hemorrhage), DGE (Delayed Gastric Emptying), CCI® (Comprehensive Complication Index®), OS (Overall Survival), RFS (Recurrence-free survival)

Table 2: Independent Predictors of Post-Operative Pancreatic Fistula Grade B/C Development

	Univariate analysis[1]		**Multivariable logistic regression**[1]	
	OR (95% CI)	p-value	OR (95% CI)	p-value
PD-ROBOSCORE[2]	0.99 (0.94, 1.04)	0.7		
BMI (kg/m^2)	0.99 (0.95, 1.04)	0.7		
ASA ≥3	0.82 (0.54, 1.24)	0.4		
Borderline-resectable	0.36 (0.09, 0.99)	0.088		
Duct diameter <4mm	1.46 (0.98, 2.19)	0.062	**1.37 (1.03, 1.82)**	0.031
Prior major abdominal surgery	0.29 (0.02, 1.38)	0.2		
Anticoagulation	2.77 (1.40, 5.18)	0.002	**2.45 (1.47, 3.99)**	<0.001
Venous resection	0.43 (0.13, 1.07)	0.11	0.56 (0.26, 1.09)	0.11
Conversion	0.35 (0.06, 1.18)	0.2		
Indication other than PDAC	2.14 (1.39, 3.39)	<0.001	**2.33 (1.68, 3.27)**	<0.001

[1]Analysis of entire cohort (N=1079), [2]Napoli et al.

Abbreviations: OR (Odds Ratio), CI (Confidence Interval), BMI (Body Mass Index), ASA (American Society of Anesthesiologists), PDAC (Pancreatic Ductal Adenocarcinoma).