ARTIFICIAL INTELLIGENCE

original reports

Toward an Optimized Staging System for Pancreatic Ductal Adenocarcinoma: A Clinically Interpretable, Artificial Intelligence–Based Model

Dimitris Bertsimas, PhD[1]; Georgios Antonios Margonis, MD, PhD[2,3]; Yifei Huang, MSc[1]; Nikolaos Andreatos, MD[4]; Holly Wiberg, MSc[1]; Yu Ma, MSc[1]; Caitlin Mcintyre, MD[2]; Alessandra Pulvirenti, MD[5]; Doris Wagner, MD[6]; J. L. van Dam, MD[7]; Francesca Gavazzi, MD[5]; Stefan Buettner, MD, PhD[7]; Katsunori Imai, MD, PhD[8]; Georgios Stasinos, BSc[9]; Jin He, MD, PhD[3]; Carsten Kamphues, MD[10]; Katharina Beyer, MD[10]; Hendrik Seeliger, MD[10]; Matthew J. Weiss, MD[11]; Martin Kreis, MD[10]; John L. Cameron, MD[3]; Alice C. Wei, MD, MSc[2]; Peter Kornprat, MD[6]; Hideo Baba, MD, PhD[8]; Bas Groot Koerkamp, MD, PhD[7]; Alessandro Zerbi, MD[5]; Michael D'Angelica, MD[2]; and Christopher L. Wolfgang, MD, PhD[3]

abstract

PURPOSE The American Joint Committee on Cancer (AJCC) eighth edition schema for pancreatic ductal adenocarcinoma treats T and N stage as independent factors and uses positive lymph nodes (PLNs) to define N stage, despite data favoring lymph node ratio (LNR). We used artificial intelligence–based techniques to compare PLN with LNR and investigate interactions between tumor size and nodal status.

METHODS Patients who underwent pancreatic ductal adenocarcinoma resection between 2000 and 2017 at six institutions were identified. LNR and PLN were compared through shapley additive explanations (SHAP) analysis, with the best predictor used to define nodal status. We trained optimal classification trees (OCTs) to predict 1-year and 3-year risk of death, incorporating only tumor size and nodal status as variables. The OCTs were compared with the AJCC schema and similarly trained XGBoost models. Variable interactions were explored via SHAP.

RESULTS Two thousand eight hundred seventy-four patients comprised the derivation and 1,231 the validation cohort. SHAP identified LNR as a superior predictor. The OCTs outperformed the AJCC schema in the derivation and validation cohorts (1-year area under the curve: 0.681 *v* 0.603; 0.638 *v* 0.586, 3-year area under the curve: 0.682 *v* 0.639; 0.675 *v* 0.647, respectively) and performed comparably with the XGBoost models. We identified interactions between LNR and tumor size, suggesting that a negative prognostic factor partially overrides the effect of a concurrent favorable factor.

CONCLUSION Our findings highlight the superiority of LNR and the importance of interactions between tumor size and nodal status. These results and the potential of the OCT methodology to combine them into a powerful, visually interpretable model can help inform future staging systems.

JCO Clin Cancer Inform 5:1220-1231. © 2021 by American Society of Clinical Oncology

INTRODUCTION

The American Joint Committee on Cancer (AJCC) TNM staging system remains the gold standard for prognostication in pancreatic ductal adenocarcinoma (PDAC). Three main criteria have been proposed to evaluate how well the TNM staging system fulfills its mission[1]: reproducibility, simplicity, and discriminatory ability. Striving for simplicity, however, involves an implicit trade-off with discriminatory ability, as the addition of more prognostic variables has been shown to lead to modest gains in the latter, but at the cost of increased model complexity. The compromise approach followed in AJCC staging systems and clinical practice to date tends to favor simplicity; as a result, the area under the curve (AUC) of both the AJCC seventh and eighth edition staging systems for PDAC remains quite low (approximately 0.6), reflecting suboptimal discriminatory ability.[1,2]

Increasing the discriminatory ability of the AJCC eighth edition without introducing new factors or unduly complicating the model is evidently a daunting challenge. One promising approach is to define the N component on the basis of the ratio of positive lymph nodes divided by the total number of harvested lymph nodes (LNR) rather than the number of positive lymph nodes (PLNs) alone, which is the current practice.[3-5] A second and possibly more widely applicable approach would be to stop considering the different components of the AJCC system as independent, additive prognostic factors, but to assess the possibility of interactions between them that may render the prognostic effect of T stage dependent on the respective N stage and vice versa (eg, it is possible that low T stage does not confer the same beneficial prognostic impact in patients with a high burden of nodal metastases compared with those without nodal involvement). In turn, if such

ASSOCIATED CONTENT

Appendix

Data Supplement

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on November 8, 2021 and published at ascopubs.org/journal/cci on December 22, 2021: DOI https://doi.org/10.1200/CCI.21.00001

CONTEXT

Key Objective

To our knowledge, this work marks the first successful use of artificial intelligence and game theory to modify an American Joint Committee on Cancer classification system not only among patients with pancreatic ductal adenocarcinoma but also in oncology as a whole. This novel approach to constructing staging systems may have significant future applications across different malignancies.

Knowledge Generated

This study contradicts the assumption that T and N components of the TNM staging system are independent of each other, by showing that they exhibit clinically important interactions. These interactions proved to be prognostically important as they leverage increased prognostic power without hindering clinical interpretation.

Relevance

Our findings may help resolve the long-standing debate regarding the optimal proxy of nodal status in pancreatic ductal adenocarcinoma, whereas they suggest that for a given patient, the beneficial impact on prognosis of low T or N stage may be overridden by that of coexisting high N or T stage to a greater extent than would be expected by simple addition.

interactions exist and are significant, their incorporation into the model can increase discriminatory power without the need to include additional prognostic factors.

We first assessed the relative prognostic power of LNR and PLN via the novel shapley additive explanations (SHAP) methodology, which uses mathematical techniques used in game theory to estimate a variable's contribution to a prognostic model.[6] After selecting the more promising of these two predictors to reflect nodal status and incorporating tumor size as an additional variable, we developed a prognostic model with the aid of optimal classification trees (OCTs), an artificial intelligence–based methodology that can readily capture interactions between variables and present them in a visually interpretable manner.[7] The OCT model was subsequently compared with the AJCC eighth edition staging system and XGBoost, a state-of-the-art, but less interpretable, machine learning approach.

METHODS

Patient Selection

All patients who underwent a partial or total pancreatectomy for PDAC as part of routine clinical care between January 1, 2000, and December 31, 2017, at Johns Hopkins Hospital, Memorial Sloan Kettering Cancer Center, Erasmus Medical Center, Fondazione IRCCS Istituto Nazionale Dei Tumori, Graz University, and Kumamoto University were retrospectively identified from institutional databases. The study was conducted in accordance with the ethical standards of the participating institutions and was approved by their institutional review boards. A detailed ethics statement is provided in the Data Supplement.

Patients with metastatic disease were excluded. We also excluded patients with grossly positive resection margins (as determined by review of the operative and pathology reports and clinical documentation authored by the attending surgeon) because the presence of macroscopic residual disease precludes accurate assessment of tumor size on pathology.[8] Similarly, patients treated with neoadjuvant therapy were excluded because of the lack of consensus definitions on how to assess gross tumor size after response to treatment.[9] Patients who died within 30 days of resection and those with missing data on survival outcomes, tumor size, or nodal status (as determined by pathologic examination of the resected specimen) were also excluded.

Data Extraction and Study Outcomes

We collected data on clinicopathologic variables with the aid of a centrally designed data report form that was used by all participating institutions (Table 1). All pathologic data were derived from the available pathology reports. All clinical and pathologic variables necessary to estimate AJCC stage for each patient were included in the respective data report form. Staging was subsequently determined centrally by the Johns Hopkins Hospital investigators. We used state-of-the-art nonlinear imputation methods to help replace missing entries with artificially created substitute values that aim to approximate the variable distribution that would be observed in a complete data set.[10] Resections were considered margin positive when tumor cells were found at the margin or within 1 mm of the margin, as defined by the Royal College of Pathologists in 2017.[8] T, N, and M stages were defined on the basis of the eighth edition of the respective AJCC staging manual. Overall survival was calculated from the date of surgery to the date of death or last follow-up. The process followed by each institution to determine survival outcomes is presented in the Data Supplement.

XGBoost and SHAP Analysis

To discern whether LNR or PLN is the stronger predictor of survival outcomes and, consequently, the best basis for defining nodal status, we trained two separate XGBoost

TABLE 1. Patient and Tumor Characteristics of the Derivation and Validation Cohorts

Characteristic	Derivation (n = 2,874)	Validation (n = 1,231)
Age, median (IQR)	68 (60.3-75.0)	69 (61.0-75.5)
Sex, No. (%)		
Male	1,477 (51.4)	637 (51.7)
Female	1,397 (48.6)	594 (48.3)
Diabetes, No. (%)		
Yes	671 (23.3)	299 (24.3)
No	2,203 (76.7)	932 (75.7)
Other malignancy, No. (%)		
Yes	272 (9.5)	128 (10.4)
No	2,602 (90.5)	1,103 (89.6)
CA19-9, median (IQR)	296 (87.0-663.5)	374 (104.5-689.1)
Surgery type, No. (%)		
Pancreaticoduodenectomy	2,205 (76.7)	948 (77)
Total pancreatectomy	126 (4.4)	49 (4)
Distal pancreatectomy	543 (18.9)	234 (19)
Splenectomy, No. (%)		
Yes	622 (21.6)	260 (21.1)
No	2,252 (78.4)	971 (78.9)
Vascular resection, No. (%)		
Yes	290 (10.1)	152 (12.3)
No	2,584 (89.9)	1,079 (87.7)
Length of operation, median (IQR), minutes	325 (244-400)	325 (251.0-393.1)
Resection margin status, No. (%)		
R0	1867 (65)	808 (65.6)
R1	748 (26)	321 (26.1)
Within 1 mm	167 (5.8)	68 (5.5)
Positive resected to negative	79 (2.7)	27 (2.2)
Others	13 (0.5)	7 (0.6)
Tumor size, median (IQR)	3 (2.2-3.9)	3 (2.2-4.0)
No. of PLNs, median (IQR)	2 (0-4.0)	2 (0-4.0)
No. of harvested lymph nodes, median (IQR)	18 (13-24)	18 (13.0-25.0)
Grade of tumor differentiation, No. (%)		
Well	130 (4.5)	56 (4.5)
Moderate	1,595 (55.5)	708 (57.5)
Poor	1,125 (39.1)	457 (37.1)
Others	24 (0.8)	10 (0.8)
Lymphovascular invasion, No. (%)		
Yes	1795 (62.5)	830 (67.4)
No	1,079 (37.5)	410 (32.6)
Perineural invasion, No. (%)		
Yes	2,473 (86)	1,051 (85.4)
No	401 (14)	180 (14.6)

Abbreviations: CA, cancer antigen; IQR, interquartile range; PLNs, positive lymph nodes.

models (probability of death was the outcome of interest) at the 1-year and 3-year time points with the aid of multiple prognostic factors (including both PLN and LNR) and subsequently used SHAP analysis to determine the relative importance of these predictors. Further details regarding XGBoost and SHAP analysis can be found in the Data Supplement. The magnitude of a given SHAP value reflects the importance of the predictor in question.

To illustrate the results of the SHAP analysis, we constructed variable importance plots that list the predictors incorporated in each XGBoost model on the basis of the respective SHAP values.

Estimating Interactions Between Predictors With the SHAP Model

To better assess the interplay of tumor size and nodal status, we used SHAP on the XGBoost models derived above and plotted the respective interaction values for a range of different predictor values for the 1-year and 3-year time points. Further details regarding how the SHAP method can help assess interactions between predictors can be found in the Data Supplement.

Developing the OCT Models

The selection of the OCT method to develop the final prognostic models is explained in the Data Supplement. Separate OCTs were trained to predict 1-year and 3-year death rates; outcomes at each time point were generated using the following principles: (1) patients who died before the time point of interest were coded as dead, (2) patients who died after the time point of interest were coded as alive, (3) patients who were censored before the time point of interest because they were lost to follow-up were excluded, and (4) patients who were censored after the time point of interest were coded as alive.

To develop the OCTs, the cohort was randomly divided into derivation (70%) and validation (30%) data sets. To limit overfitting, the OCTs were tuned with the use of the following two hyperparameters: maximum depth of tree and minimum number of samples per leaf. Grid search and cross-validation were used to help us select hyperparameters.

Model Evaluation

We compared the AUC of the 1-year and 3-year OCTs with that of the AJCC eighth edition staging system in both the derivation and validation data sets. To further assess the performance of the OCT methodology, we trained additional 1-year and 3-year XGBoost models on the basis of the same factors as the OCTs for the purpose of comparison. Details regarding computing can be found in the Data Supplement.

RESULTS

Patient Characteristics and Survival Analysis

The final study population included 4,105 patients (Appendix Fig A1). Of those, 2,874 patients (70%) comprised the derivation cohort and 1,231 patients (30%) the

validation cohort. Detailed clinicopathologic characteristics of the two cohorts are presented in Table 1; participating institutions were represented in proportion to their contribution to the overall study population (Appendix Table A1). The distribution of disease stages in the derivation and validation cohorts according to the AJCC eighth edition staging system is presented in Appendix Table A2. The proportion of cases with missing values in each variable is provided in Appendix Table A3. A total of 508 patients were censored before the 3-year time point; their characteristics in comparison with the remaining cohort are presented in Appendix Table A4. As can be seen, the two groups were well-balanced overall and any observed differences were generally of small magnitude.

In the derivation cohort, the 1-year and 3-year death rates were 25.7% and 71.4%, respectively. In the validation cohort, the 1-year and 3-year death rates were 24.8% and 71%, respectively.

SHAP Variable Importance Plot

According to the SHAP analysis of the XGBoost models (Appendix Fig A2), the mean absolute SHAP value of LNR was much higher than that of PLN at both time points, indicating that LNR is a stronger predictor. As such, we incorporated LNR rather than PLN as a proxy of nodal status in the 1-year and 3-year OCTs.

1-Year and 3-Year OCTs

The OCTs for predicting death within 1 and 3 years of surgery are illustrated in Figures 1 and 2 and in Ref. 11. The 1-year and 3-year overall survival estimates for a specific patient can be derived with the aid of the online calculator.[11]

OCT Performance and Comparison With the AJCC Eighth Edition Staging System and the XGBoost Model

In both the derivation and validation cohorts, the AUC of the OCT compared favorably with that of the AJCC eighth edition staging system at the 1-year and 3-year time points (Table 2). The OCTs slightly underperformed the XGBoost model in both the derivation and validation cohorts at the 1-year and 3-year time points (Table 2). Subgroup analyses restricted to patients with a low (< 15 lymph nodes) or a high (≥ 15 lymph nodes) number of harvested lymph nodes yielded similar results. Specifically, in the validation cohort, the AUC of the 1-year OCT increased to 0.647 (*v* 0.638), whereas the AUC of the 3-year OCT increased to 0.698 (*v* 0.675).

SHAP Interaction Values

With the aid of SHAP, we analyzed the feature interactions in the full XGBoost model (used to determine whether LNR or PLN is the stronger predictor as discussed above) to provide a data-driven validation of the OCT cutoffs. The results of this analysis for the 1-year and 3-year time points are depicted in Figures 3 and 4, respectively. Further details regarding the explanation of Figures 3 and 4 can be found in the Data Supplement.

DISCUSSION

The relative prognostic value of LNR and PLN has been debated for a long time; although most studies appear to support LNR as a more potent predictor, consensus on this point is lacking.[12-15] This might partly explain why LNR has not been incorporated into the AJCC staging system. Our analysis, performed with the aid of SHAP in a very large data set, clearly demonstrated that LNR was not only superior to PLN but also the most important of all commonly used clinicopathologic predictors at both time points. This is consistent with a recent study that reported that the presence of nodal disease was the most important among several predictors in a large PDAC cohort.[16] The prognostic value of LNR can likely be attributed to its ability to limit nodal disease misclassification and more accurately reflect tumor burden. It is also theoretically possible that LNR more closely reflects the adequacy of nodal resection, which, in

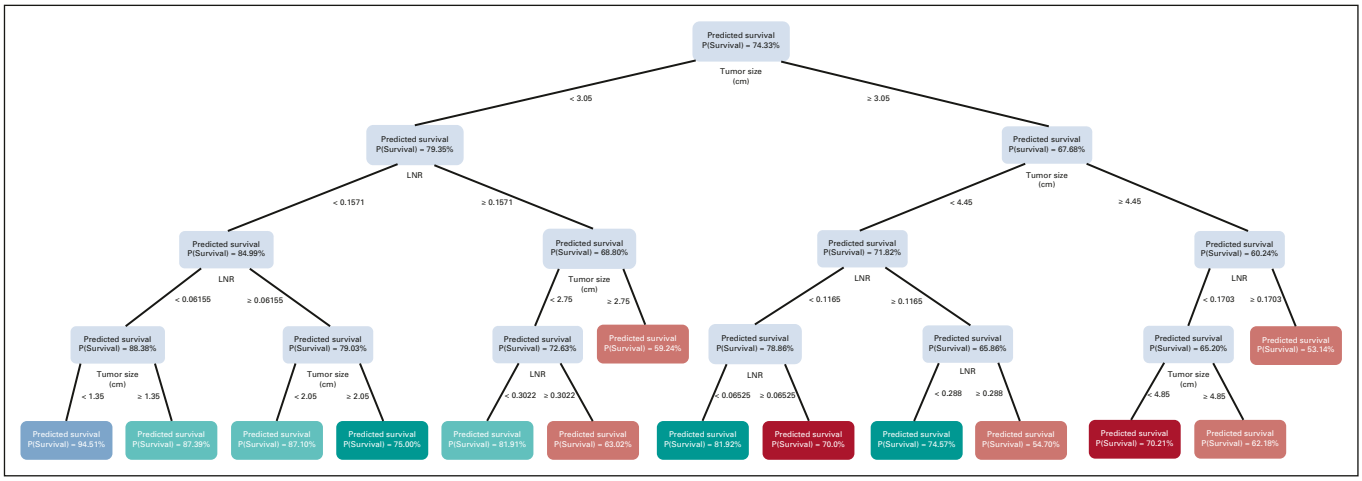

FIG 1. Prediction of 1-year overall survival with the aid of optimal classification trees. LNR, lymph node ratio.

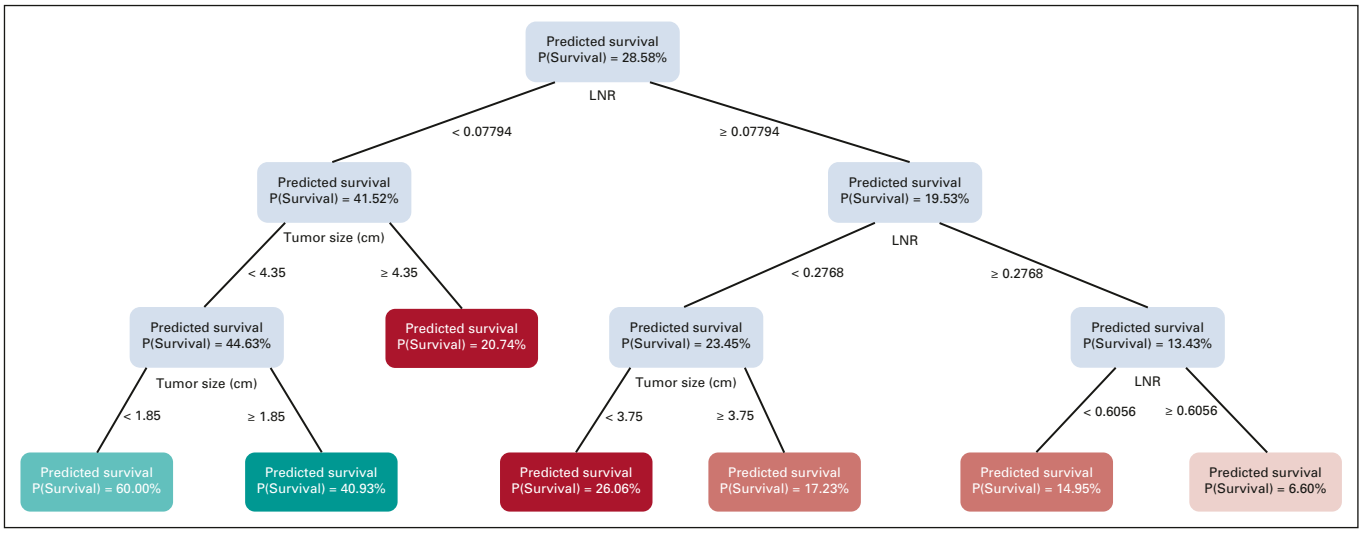

FIG 2. Prediction of 3-year overall survival with the aid of optimal classification trees. LNR, lymph node ratio.

turn, may be tied to outcomes. These findings may both help resolve the ongoing debate and pave the way for the use of LNR in future AJCC editions. We used Optimal Imputation to account for missing values in secondary prognostic variables. Although this did not affect the main analyses, which relied on variables that were not imputed, the comparison of PLN and LNR was based on a multivariable model and might have been indirectly affected by this approach. However, as the available data were reasonably complete for the majority of important predictors, any corresponding impact was likely limited.

After deciding in favor of LNR as a proxy of nodal status, we proceeded to train 1-year and 3-year OCTs and assess their discriminatory ability. In both the derivation and validation cohorts, the AUC of the present model compared favorably with that of the AJCC eighth edition staging system at the 1-year and 3-year time points and surpassed that of previous validations of the AJCC eighth edition reported in the literature.[2,8] The prognostic advantage of our model over the AJCC staging system was retained even among patients with accurate nodal staging (defined as $\geq$ 15 harvested nodes) in whom the benefit of using LNR rather than PLN may be less pronounced. In fact, the AUC of the 3-year OCT in this patient group was nearly 0.7, which is among the highest ever reported in the literature.

The OCT method has another important advantage over the AJCC schema, as it considers tumor size and nodal status concurrently, thus capturing underlying interactions between them; on the contrary, prognostic cutoffs in the AJCC staging system are developed independently for each variable. Interestingly, the 3-year OCT used essentially the same tumor size cutoffs as the AJCC eighth edition: 1.85 cm (AJCC T1: tumor size $<$ 2 cm), 1.85-4.35 cm (AJCC T2: 2 cm $<$ tumor size $\leq$ 4 cm), and $\geq$ 4.35 cm (AJCC T3: tumor size $>$ 4 cm). As these cutoffs were selected in an unbiased way after considering tumor size as a continuous variable, the resulting level of consistency is remarkable.

To further assess the internal validity of the cutoffs identified by the OCTs, we performed an analysis of interactions with the aid of the SHAP method in independently devised XGBoost prognostic models. Importantly, the same cutoffs were identified with the use of this alternative methodology. In the case of 1-year risk of death, a tumor size of 3.05 cm was not only the first point of divergence for the respective OCT but also the point where the interactions between LNR and tumor size switched direction, thus indirectly demonstrating the poor prognostic consequences of exceeding that tumor size, as discussed above. Similar results were noted for LNR at the 3-year time point. Taken together, the identification of the same cutoffs by two distinct mathematical methods not only illuminates the significance of the underlying variable interactions but also supports the overall reliability of the findings.

Beyond simply validating prognostic cutoffs, the SHAP method allowed us to explore the interactions of tumor size and LNR as continuous variables for the first time.

TABLE 2. OCT Performance and Comparison With the AJCC Eighth Edition Staging System and the XGBoost Model in the Derivation and Validation Cohorts

Year	OCT Derivation	AJCC Derivation	XGBoost Derivation	OCT Validation	AJCC Validation	XGBoost Validation
1	0.681	0.603	0.684	0.638	0.586	0.654
3	0.682	0.639	0.693	0.675	0.647	0.690

Abbreviations: AJCC, American Joint Committee on Cancer; OCT, optimal classification tree.

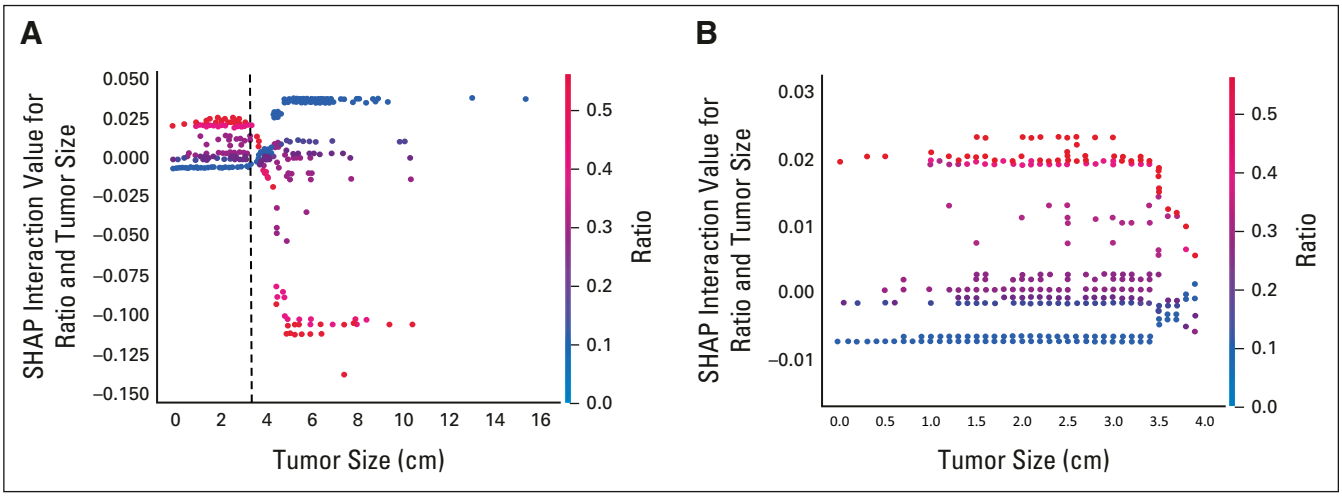

FIG 3. SHAP interaction values for different tumor sizes and lymph node ratios at the 1-year time point. (A) Full size; (B) zoom in of figure part (A). SHAP, shapley additive explanations.

Importantly, the results suggest that the combination of large tumor size (as determined by the above cutoff) with minimal nodal involvement leads to worse prognosis than would be expected by aggregating the impact of the two individual factors; a similar pattern is observed for patients with a LNR $\geq$ 0.08 and small tumors. This implies that the adverse prognostic effect of large tumor size or high LNR partly overrides the beneficial effect of low LNR or small tumor size, respectively. These results underline the importance of considering tumor size and nodal status concurrently.

The study has several limitations. The inclusion of multi-institutional data has introduced some heterogeneity; for example, there may be differences in surgical procedures and pathologic examination across centers that could theoretically result in variable lymph node yield or other discrepancies in reported pathologic variables. Moreover, outcome data are derived from patients who were successfully followed for a significant period of time, whereas patients who could not maintain follow-up until the time point of interest were censored and thus effectively excluded from the analysis. Although the two groups appear comparable on the basis of the available data, the possibility of underlying differences in unmeasured characteristics that could have affected the results cannot be entirely excluded. Finally, although available data on adjuvant chemotherapy administration are concordant with previous reports, detailed information regarding this aspect of care was not available in some institutional databases. Although this limits our ability to accurately estimate the percentage of patients treated with adjuvant chemotherapy, prescription practices at these two institutions are unlikely to have diverged substantially from the respective standards of care for the period. In turn, the positive impact of adjuvant chemotherapy on survival is likely reflected in our results, a fact that should be considered when using the model for prognostication.

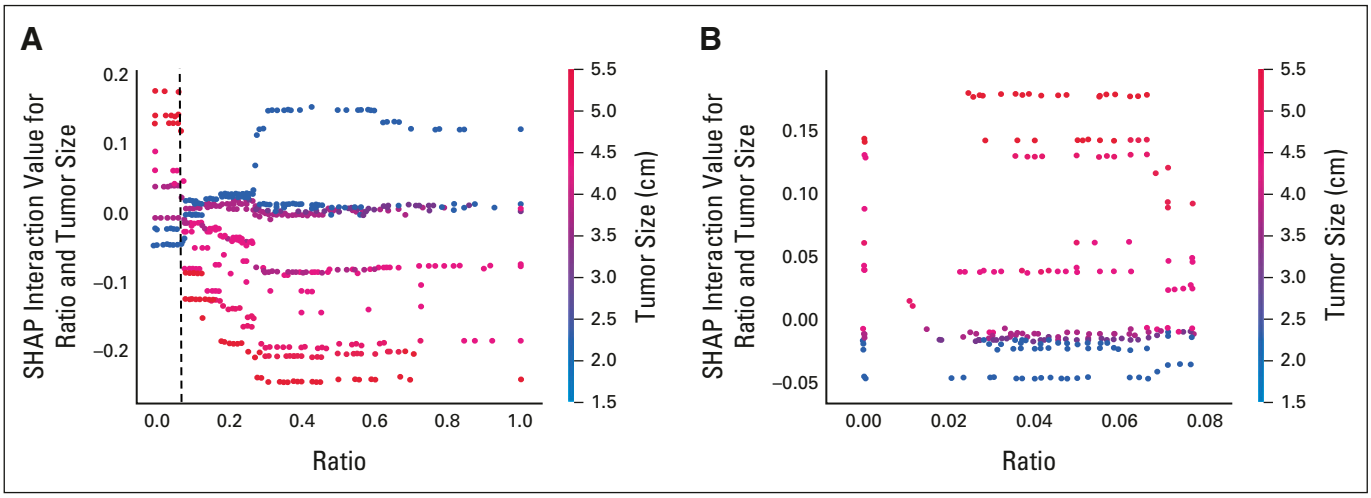

FIG 4. SHAP interaction values for different tumor sizes and lymph node ratios at the 3-year time point. (A) Full size; (B) zoom in of figure part (A). SHAP, shapley additive explanations.

In conclusion, we demonstrated the superiority of LNR over PLN as a prognostic factor and, subsequently, generated and validated a new prognostic model for patients with PDAC that incorporates interactions between tumor size and nodal status with the aid of the OCT methodology. Importantly, the resulting model comfortably outperformed the AJCC eighth edition staging schema and performed comparably with the state-of-the-art XGBoost model without sacrificing interpretability. Although the inclusion of additional predictors would have improved the model's AUC, it would have also precluded direct comparison with the AJCC schema and complicated the interpretation of interactions, thus frustrating the principal aims of the study. Additional performance optimization on the basis of a multivariable OCT model is feasible and should be evaluated in future reports. The resulting prognostic cutoffs identified by the model are highly original as they were derived by simultaneously considering tumor size in the context of nodal status and vice versa. Although external validation in patients from different geographic areas will have to await future studies, a supplementary analysis performed with the aid of the SHAP method confirmed the internal validity of our findings. Finally, the observed variable interactions are more than an abstract mathematical concept aimed at enhancing AUC, but can be readily interpreted at a clinical level; our findings suggest that unfavorable tumor size has a tendency to override favorable nodal status and vice versa, leading to more guarded prognosis than would be expected by adding the prognostic impact of each individual factor. If these observations are confirmed, a strong case can be made in favor of incorporating LNR and considering variable interactions in future AJCC editions. Perhaps more importantly, the ability of artificial intelligence–based interaction analysis to provide a sophisticated but interpretable view of the interplay between clinicopathologic variables may in due course lead to widespread prognostic, predictive, and, ultimately, prescriptive applications across different tumor types.

AFFILIATIONS

[1]Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA
[2]Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY
[3]Department of Surgery, Johns Hopkins University School of Medicine, Baltimore, MD
[4]Department of Internal Medicine and Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH
[5]Section of Pancreatic Surgery, Humanitas Clinical and Research Center-IRCCS, Milan, Italy
[6]Department of General Surgery, Medical University of Graz, Graz, Austria
[7]Department of Surgery, Erasmus MC, University Medical Center, Rotterdam, the Netherlands
[8]Department of Gastroenterological Surgery, Graduate School of Life Sciences, Kumamoto University, Kumamoto, Japan
[9]Technical Chamber of Greece (TEE-TCG), Athens, Greece
[10]Department of General, Visceral and Vascular Surgery, Charité Campus Benjamin Franklin, Berlin, Germany
[11]Department of Surgery, Northwell Health, Manhasset, NY

CORRESPONDING AUTHOR

Georgios Antonios Margonis, MD, PhD, Department of Surgery, Memorial Sloan Kettering Cancer Center, 1275 York Ave, New York, NY 10065; e-mail: antonis.margonis@gmail.com.

DISCLAIMER

Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation.

EQUAL CONTRIBUTION

D.B. and G.A.M. contributed equally to this work.

SUPPORT

Supported in part through NIH/NCI Cancer Center Support Grant No. P30 CA008748. H.W. was supported by the National Science Foundation Graduate Research Fellowship under Grant No. 174530.

AUTHOR CONTRIBUTIONS

Conception and design: Dimitris Bertsimas, Georgios Antonios Margonis, Nikolaos Andreatos, Christopher L. Wolfgang
Financial support: Georgios Antonios Margonis
Provision of study materials or patients: All authors
Collection and assembly of data: All authors
Data analysis and interpretation: Dimitris Bertsimas, Georgios Antonios Margonis, Yifei Huang, Holly Wiberg, Yu Ma
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Dimitris Bertsimas
Stock and Other Ownership Interests: Reclaim, Alexandria Health

Yifei Huang
Employment: Benefits Science Technologies, China Reform Health Management and Services Group Co, Ltd
Stock and Other Ownership Interests: UnitedHealthcare

Alice C. Wei
Consulting or Advisory Role: AstraZeneca, HistoSonics, Medtronic
Travel, Accommodations, Expenses: Intuitive Surgical

Hideo Baba
Honoraria: Taiho Pharmaceutical
Research Funding: Yakult Honsha (Inst), Chugai Pharma (Inst), Taiho Pharmaceutical (Inst), MSD K.K (Inst), Shin Nippon Biomedical (Inst), Lilly Japan, Johnson & Johnson K.K., Asahi Kasei

Bas Groot Koerkamp
Research Funding: Tricumed (Inst)

Christopher L. Wolfgang
Leadership: CATALIO HealthCor
Stock and Other Ownership Interests: CATALIO HealthCor

No other potential conflicts of interest were reported.

REFERENCES

1. Allen PJ, Kuk D, Castillo CF, et al: Multi-institutional validation study of the American Joint Commission on Cancer (8th edition) changes for T and N staging in patients with pancreatic adenocarcinoma. Ann Surg 265:185-191, 2017
2. Kamarajah SK, Burns WR, Frankel TL, et al: Validation of the American Joint Commission on Cancer (AJCC) 8th edition staging system for patients with pancreatic adenocarcinoma: A Surveillance, Epidemiology and End Results (SEER) analysis. Ann Surg Oncol 24:2023-2030, 2017
3. Elshaer M, Gravante G, Kosmin M, et al: A systematic review of the prognostic value of lymph node ratio, number of positive nodes and total nodes examined in pancreatic ductal adenocarcinoma. Ann R Coll Surg Engl 99:101-106, 2017
4. Riediger H, Keck T, Wellner U, et al: The lymph node ratio is the strongest prognostic factor after resection of pancreatic cancer. J Gastrointest Surg 13:1337-1344, 2009
5. Robinson SM, Rahman A, Haugk B, et al: Metastatic lymph node ratio as an important prognostic factor in pancreatic ductal adenocarcinoma. Eur J Surg Oncol 38:333-339, 2012
6. Lundberg SM, Nair B, Vavilala MS, et al: Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng 2:749-760, 2018
7. Bertsimas D, Dunn J, Velmahos GC, et al: Surgical risk is not linear: Derivation and validation of a novel, user-friendly, and machine-learning-based predictive OpTimal trees in emergency surgery risk (POTTER) calculator. Ann Surg 268:574-583, 2018
8. van Roessel S, Kasumova GG, Verheij J, et al: International validation of the eighth edition of the American Joint Committee on Cancer (AJCC) TNM staging system in patients with resected pancreatic cancer. JAMA Surg 153:e183617, 2018
9. Chatterjee D, Katz MH, Foo WC, et al: Prognostic significance of new AJCC tumor stage in patients with pancreatic ductal adenocarcinoma treated with neoadjuvant therapy. Am J Surg Pathol 41:1097-1104, 2017
10. Bertsimas D, Pawlowski C, Zhuo YD: From predictive methods to missing data imputation: An optimization approach. J Machine Learn Res 18:7133-7171, 2017
11. https://alexandriahealth.gitlab.io/pdac/trees.html
12. Valsangkar NP, Bush DM, Michaelson JS, et al: N0/N1, PNL, or LNR? The effect of lymph node number on accurate survival prediction in pancreatic ductal adenocarcinoma. J Gastrointest Surg 17:257-266, 2013
13. Yamada S, Fujii T, Hirakawa A, et al: Lymph node ratio as parameter of regional lymph node involvement in pancreatic cancer. Langenbecks Arch Surg 401:1143-1152, 2016
14. Mirkin KA, Hollenbeak CS, Wong J: Greater lymph node retrieval and lymph node ratio impacts survival in resected pancreatic cancer. J Surg Res 220:12-24, 2017
15. Huebner M, Kendrick M, Reid-Lombardo KM, et al: Number of lymph nodes evaluated: Prognostic value in pancreatic adenocarcinoma. J Gastrointest Surg 16:920-926, 2012
16. Paniccia A, Hosokawa P, Henderson W, et al: Characteristics of 10-year survivors of pancreatic ductal adenocarcinoma. JAMA Surg 150:701-710, 2015

APPENDIX

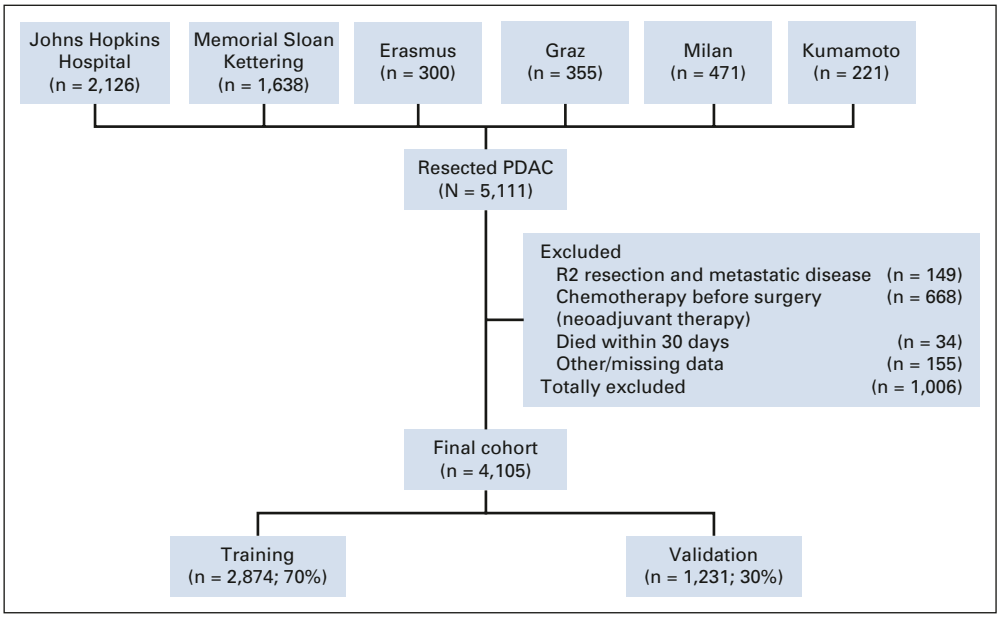

FIG A1. Flowchart of the study cohort. PDAC, pancreatic ductal adenocarcinoma.

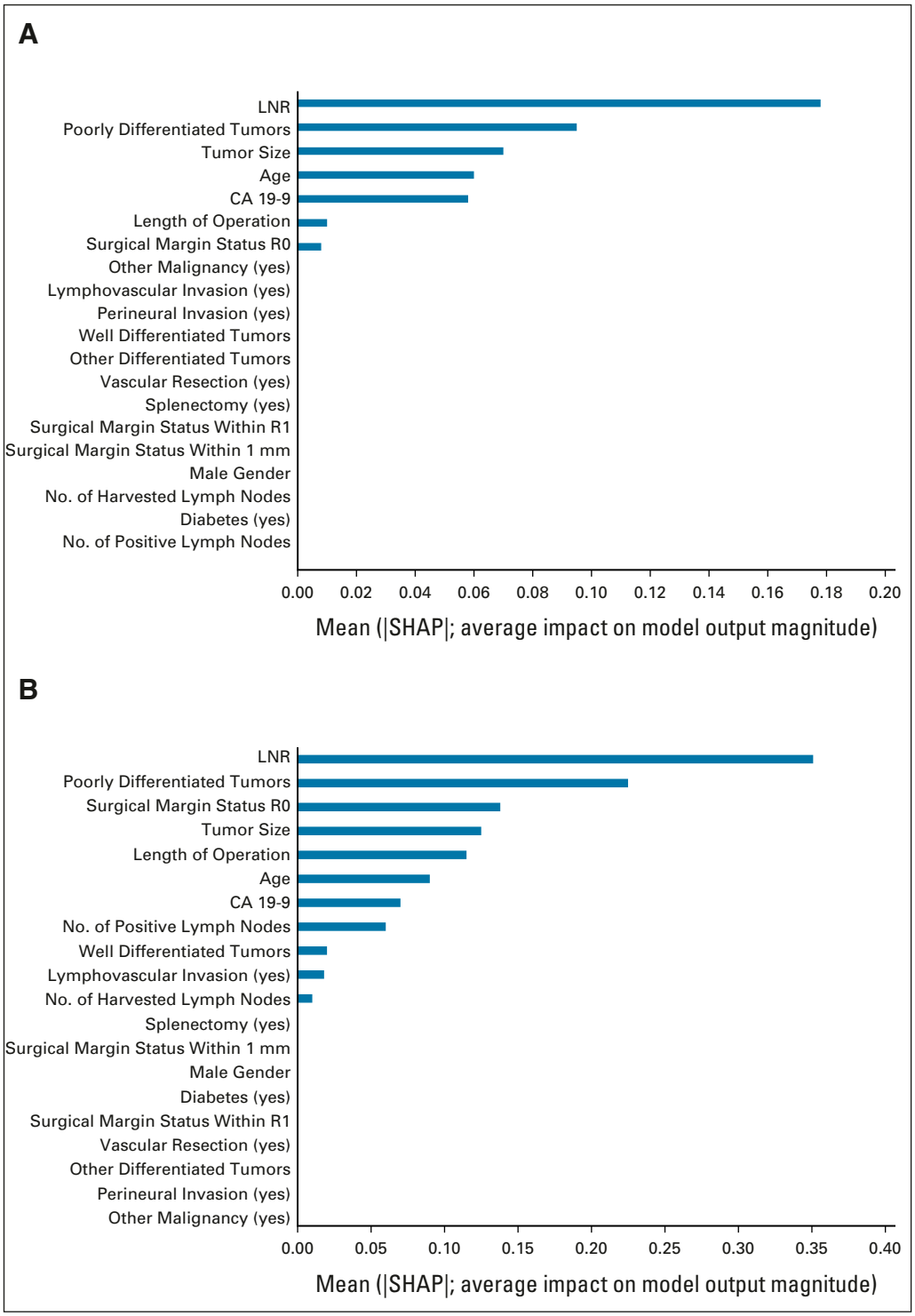

FIG A2. SHAP variable importance plot: (A) year 1 and (B) year 3. CA, cancer antigen; LNR, lymph node ratio; SHAP, shapley additive explanations.

TABLE A1. The Contribution of Participating Institutions to the Derivation and Validation Cohorts

Data Source	Derivation, No. (%)	Validation, No. (%)
Johns Hopkins Hospital	1,280 (44.5)	504 (40.9)
Memorial Sloan Kettering Cancer Center	898 (31.2)	437 (35.5)
Fondazione IRCCS Istituto Nazionale Dei Tumori	215 (7.5)	95 (7.7)
Erasmus Medical Center	191 (6.6)	62 (5)
Kumamoto University	85 (2.9)	35 (2.8)
Graz University	205 (7.1)	98 (8)
Total number	2,874	1,231

TABLE A2. Distribution of the American Joint Committee on Cancer Eighth Edition Staging System in the Derivation and Validation Cohorts

Staging	Derivation, No. (%)	Validation, No. (%)
IA	256 (8.9)	100 (8.1)
IB	458 (15.9)	181 (14.7)
IIA	97 (3.4)	42 (3.4)
IIB	1,152 (40.1)	518 (42.1)
III	911 (31.7)	390 (31.7)

TABLE A3. No. and Proportion of Cases With Missing Values in Each Variable

Characteristic	Derivation (n = 2,874), No. (%)	Validation (n = 1,231), No. (%)
Age	1 (0.03)	0 (0)
Sex	10 (0.35)	4 (0.32)
Diabetes	38 (1.32)	17 (1.38)
Other malignancy	47 (1.64)	19 (1.54)
CA19-9	1,278 (44.47)	556 (45.17)
Surgery type	54 (1.88)	24 (1.95)
Splenectomy	229 (7.97)	85 (6.90)
Vascular resection	717 (24.95)	248 (20.15)
Length of operations	412 (14.34)	191 (15.52)
Resection margin status	16 (0.56)	6 (0.49)
Tumor size	0 (0)	0 (0)
No. of PLNs	0 (0)	0 (0)
No. of harvested lymph nodes	0 (0)	0 (0)
Grade of tumor differentiation	31 (1.08)	8 (0.65)
Lymphovascular invasion	218 (7.59)	88 (7.15)
Perineural invasion	90 (3.13)	18 (1.46)

Abbreviations: CA, cancer antigen; PLNs, positive lymph nodes.

TABLE A4. Patient and Tumor Characteristics of Uncensored and Censored Cohorts for Year 3

Characteristic	Uncensored (n = 3,597)	Censored (n = 508)	*P*
Age, median (IQR[a])	68 (60.8-75.0)	68 (60.6-74.1)	.817
Sex, No. (%)			.628
Male	1,858 (51.7)	256 (50.4)	
Female	1,739 (48.3)	252 (49.6)	
Diabetes, No. (%)			.613
Yes	855 (23.8)	115 (22.6)	
No	2,742 (76.2)	393 (77.4)	
Other malignancy, No. (%)			< .001 (.064)
Yes	325 (9.0)	75 (14.8)	
No	3,272 (91.0)	433 (85.2)	
CA19-9, median (IQR)	338 (97.0-691.0)	225 (48.0-496.9)	< .001 (.095)
Surgery type, No. (%)			< .001 (.094)
Pancreaticoduodenectomy	2,816 (78.3)	337 (66.3)	
Total pancreatectomy	146 (4.0)	29 (5.7)	
Distal pancreatectomy	635 (17.7)	142 (28.0)	
Splenectomy, No. (%)			< .001 (.110)
Yes	712 (19.8)	170 (33.5)	
No	2,885 (80.2)	338 (66.5)	
Vascular resection, No. (%)			.976
Yes	388 (10.8)	54 (10.6)	
No	3,209 (89.2)	454 (89.4)	
Length of operation, median (IQR), minutes	329 (251.0-400.0)	291 (215.1-371.2)	< .001 (.086)
Resection margin status, No. (%)			.002 (.065)
R0	2,303 (64.0)	372 (73.2)	
R1	970 (27.0)	99 (19.5)	
Within 1 mm	211 (5.9)	24 (4.7)	
Positive resected to negative	94 (2.6)	12 (2.4)	
Others	19 (0.5)	1 (0.2)	
Tumor size, median (IQR)	3 (2.3-4.0)	2.5 (1.9-3.5)	.002 (.094)
No. of PLNs, median (IQR)	2 (0-4.0)	1 (0-3.0)	< .001 (.078)
No. of harvested lymph nodes, median (IQR)	18 (12.0-24.0)	20 (14.0-27.0)	< .001 (.082)
Grade of tumor differentiation, No. (%)			< .001 (.104)
Well	163 (4.5)	23 (4.5)	
Moderate	1,999 (55.6)	304 (59.8)	
Poor	1,417 (39.4)	165 (32.5)	
Others	18 (0.5)	16 (3.1)	
Lymphovascular invasion, No. (%)			< .001 (.075)
Yes	2,349 (65.3)	276 (54.3)	
No	1,248 (34.7)	232 (45.7)	
Perineural invasion, No. (%)			< .001 (.111)
Yes	3,140 (87.3)	384 (75.6)	
No	457 (12.7)	124 (24.4)	

Abbreviations: CA, cancer antigen; IQR, interquartile range; PLNs, positive lymph nodes.

[a]Effect sizes, or phi coefficients as appropriate depending on the variable are presented in brackets; effect size statistic for the Mann-Whitney *U* test is calculated as follows: r = z/square root of N, where N is the total number of cases, whereas the phi coefficient for the chi-square test is automatically calculated by SPSS.