Prognostic Survival Predictors for R0-resected pancreatic ductal adenocarcinoma patients

April 12th, 2023

This is an R project sourced from an open-source genomics dataset. This is a second edition.

¤

Abstract

Pancreatic ductal adenocarcinoma, representing over 80% of pancreatic cancers, is a notoriously lethal malignancy with a dismal five-year overall survival rate of around 10% as of 2021. Late stage at diagnosis, attributed to both the lack of reliable screening protocols and obvious symptoms at earlier stages, is almost always the first reason stated regarding why this particular cancer is so uniquely lethal. However, it must be noted that pancreatic ductal adenocarcinoma is a deadly malignancy *in overall*, given the limited survival and high recurrence rates of patients that have underwent a successful surgery with clean margins; the five-year overall survival rate increases to only around 20%, still far below the five-year overall survival of all cancer patients regardless of primary site or stage at diagnosis.

## Methods

A 2016 dataset from *The Cancer Genome Atlas* listing the tumor characteristics and patient status for 185 patients is accessed. Only the 47 patients known to have passed away from pancreatic ductal adenocarcinoma are included in this analysis. Linear regressions in patient log-survival (logarithm of the overall survival in months) over patient and tumor characteristics are performed to measure the rate of patients dying over time.

## Results

Both univariate and multivariate regression showed that large tumor size, nodal involvement, poor differentiation, advanced age, and male gender are significant negative predictors on patient overall survival. Patients are not expected to survive five years after successful resection regardless of small tumor size, node negativity, or young age, hence justifying the tumor’s uniquely lethal reputation.

Keywords: Pancreatic Ductal Adenocarcinoma, Overall Survival, Disease Free Survival, Resection Status

Prognostic Survival Predictors for R0-resected pancreatic ductal adenocarcinoma patients

Pancreatic cancer is commonly cited to be the deadliest of the cancers. The “silent” nature of pancreatic cancer is the most cited reason for the uniquely poor prognosis. There are no effective screening programs for pancreatic cancer, meaning that most patients are diagnosed only when the cancer has already reached distant organs. Among the patients that have underwent a clean (R0) surgical resection, only about 20% of them will survive for another five years.

Note that I only consider *pancreatic ductal adenocarcinoma* (PDAC) cases. PDAC represents around 85% of all pancreatic cancer cases, and is the histological type associated with aggressive behavior and the signature poor prognosis. The other major type of pancreatic cancer, pancreatic neuroendocrine tumor (PNET), is associated with a much more favorable prognosis, with a five-year overall survival of 51.3% according to one population-based study (Yadav, Sharma, & Zakalik, 2018).

## Objective

The aim of this report is to determine the average overall and disease-free survival (OS and DFS respectively in this article) of *pancreatic ductal adenocarcinoma* patients that underwent a successful (R0) surgical resection. Current reports on resectable pancreatic cancer patient survival include all PDAC patients that underwent resection, including those which underwent R1 resection. I will apply statistical analysis to determine if the predictors are similarly able to predict the survival metrics of the small proportion of PDAC patients that underwent R0 resection. The target variable is the patient’s overall survival after surgery, in months.

I also sought to determine whether the conditional survival after a certain time varies little within the first five years, by observing whether linear regression is applicable on the patient’s log-survival. Please see the appendix for literature review.

# Methods

## Dataset

I analyzed a 2016 dataset accessed by <www.cbioportal.org>. This dataset is part of *The Cancer Genome Atlas*, a cancer genomics program and a joint effort between the National Cancer Institute and the National Human Genome Research Institute that collects patient cancer samples internationally. A total of 185 patients are present in this dataset, and resection margin status is stated for all but twelve of these patients. I excluded 62 patients because of failed or unknown resection status (53 R1, 5 R2, 4 RX, 12 missing), leaving 108 patients known to have an R0 resection for consideration. 87 of these patients have tumors that are histologically pancreatic ductal adenocarcinoma (representing 80% of the dataset); there is one undifferentiated pancreatic carcinoma, two colloid carcinomas, eleven other exocrine cancers, and six PNETs. 50 of the PDAC patients are deceased. Two out of these 50 passed away early due to surgical complications, and one patient passed away from another malignancy; these three are excluded from analysis. I will include the patients whose patient death reason is not reported as part of the analysis along with patients explicitly stated to have demised because of PDAC.

## Variable Selection

This dataset contains 104 attributes. Many of the patients have missing attributes, meaning that they either declined to answer or that the physician(s) did not report this metric. The relevant variables are:

* Longest dimension of tumor
* Lymph node ratio
* Tumor grade (categorical)
* Age at diagnosis
* Patient gender (1 for male)
* Year of diagnosis

## Training and Test Data

I divided the remaining 47 patients into two groups:

1. Training: Patients explicitly known to have passed away from pancreatic cancer
2. Test: Patient death reason is omitted

## Linear Regression

I will assume that the conditional survival of the patient after a certain time period remains constant and is predetermined by the six candidate predictors plus a normally distributed and homoscedastic error term. Thus, for a given patient of the same age and gender, diagnosed at the same time, and having the same tumor staging and grading metrics, the patient’s survival should roughly follow a log-normal distribution with constant parameters. This is the homoscedasticity assumption of the regression. Since each patient is independent of other patients, I assume that the independence assumption also holds when we perform linear regression. I will measure the logarithm of the patient’s survival to determine whether linear regression is applicable, and whether the negative trend is present as expected. The natural logarithm is used throughout this report.

# Results

## Univariate Analysis

After performing univariate analysis, the strongest predictors on patient overall log survival. Large tumor size and poor tumor differentiation compared to moderate differentiation both demonstrated to be strong predictors on the log survival of the patient. The dataset did not show that patients with well differentiated tumors have a survival advantage compared to patients with moderately differentiated tumor. Males are shown to have a significantly poorer survival compared to females. Age and lymph node ratio did not strongly correlate to patient log survival with p-values in excess of 0.500. Interestingly, the year at diagnosis demonstrated a negative correlation to survival with a surprisingly high p-value of 0.020.

| Predictor | Slope | Intercept | Std. Error | t-value | RSE | Multiple R2 | F-statistic[[1]](#footnote-2) | p-value |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | -0.22275 | 3.63982 | 0.08655 | -2.574 | 0.5315 | 0.2398 | 6.624 | 0.0177 |
|  | -0.3988 | 2.8644 | 0.5911 | -0.675 | 0.6031 | 0.02121 | 0.4552 | 0.5072 |
|  | 0.00147 | 2.89445 | 0.32372 | 0.005 | 0.6002 | 0.0000015 | 0.0002063 | 0.9964 |
|  | -0.3438 | 2.8945 | 0.2785 | -1.235 | 0.5759 | 0.08697 | 1.524 | 0.2348 |
|  | -0.004816 | 3.110275 | 0.012013 | -0.401 | 0.6073 | 0.007595 | 0.1607 | 0.6926 |
|  | -0.5774 | 3.0908 | 0.2211 | -2.612 | 0.5296 | 0.2452 | 6.821 | 0.01629 |
|  | -0.17510 | 3.51276 | 0.06986 | -2.507 | 0.5348 | 0.2303 | 6.283 | 0.02048 |

Table 1: Univariate Linear Regression by Predictor

## Multivariate Analysis

Multivariate linear regression on the survival on the variables (except for grade) shown to display the expected results confirmed that they are independent predictors of reduced patient survival. Notably, age becomes a much stronger independent survival factor in terms of p-value. Unlike the univariate case, high tumor grade did not strongly indicate a poorer patient prognosis given similar patients with the same tumor size, node ratio, age at diagnosis, and gender.

| Predictor | Coefficient | Std. Error | t-value | p-value |
| --- | --- | --- | --- | --- |
| (Intercept) | 4.78599 | 0.91782 | 5.215 | 0.000051 |
|  | -0.21314 | 0.10157 | -2.098 | 0.0502 |
|  | -0.22262 | 0.55126 | -0.404 | 0.6911 |
|  | -0.01380 | 0.01095 | -1.261 | 0.2236 |
|  | -0.42542 | 0.22845 | -1.862 | 0.0790 |

Table 2: Multivariate Linear Regression excluding grade

| Predictor | Coefficient | Std. Error | t-value | p-value |
| --- | --- | --- | --- | --- |
| (Intercept) | 4.76901 | 0.95839 | 4.976 | 0.000115 |
|  | -0.20979 | 0.10942 | -1.917 | 0.072176 |
|  | -0.22502 | 0.56754 | -0.396 | 0.696684 |
|  | -0.02568 | 0.24904 | -0.103 | 0.919081 |
|  | -0.01364 | 0.01137 | -1.200 | 0.246711 |
|  | -0.42208 | 0.23723 | -1.779 | 0.093085 |

Table 3: Multivariate Regression including grade

# Discussion

Univariate and multivariate analysis both demonstrated that the relationship between patient overall log-survival is negatively correlated with tumor size, node ratio, and advanced age, though the high p-values still imply that patient survival is highly variable. High grade and male gender are also confirmed as a negative prognostic indicator.

### Interpreting the Log Survival

Using as the base, the log-survival of the patient roughly corresponds to the logarithm of the time interval where patient survival drops to a factor of . A patient with zero log-survival survived only one month after surgery. For every unit increase in log-survival, the patient’s survival increases by a factor of . The correlation between the value of the predictors and the patient log-survival is surprisingly linear, and as expected, exhibits a negative slope. That means, higher readings indicate that the chance of the patient surviving is reduced.

### Lymph Node Ratio

The lymph node ratio itself is a sample statistic, because it only factors in harvested lymph nodes rather than all the neighboring lymph nodes present throughout the body. It is impossible to harvest and count all of the lymph nodes involved in a patient. Thus, a node-negative patient may actually turn out to be node-positive in case if the surgeons elected to extend their lymph node resection, reflecting occult lymph node infiltration of the tumor. A patient could have its lymph node ratios overestimated to 100% if the surgeons under-reported the number of lymph nodes, and the remaining lymph node(s) turn out to be positive for malignancy. Hence, survival is highly variable when factor in lymph node ratio.

### Grade

Tumor grade is expected to be an independent prognosticator for reduced patient survival. However, multivariate analysis showed that grade is not a strong independent prognosticator given the positive correlation between grade and the size and nodal involvement of the tumor.

### Year of Diagnosis

The latest patients present on the dataset are diagnosed in 2012. There are more patients with short survival periods than long survival periods, and that this dataset looks only at patients diagnosed between 2007 to 2012.

### Overall Result

The prognosis of pancreatic cancer remains dismal even among patients that underwent surgical resection. Based on following cited articles, older age at diagnosis, resection margin involvement, tumor size (T stage), lymph node involvement (N stage), and tumor differentiation (G grade) are significant negative prognostic predictors for survival.

⁂

# Bibliography

Canadian Cancer Society. (2022). *Cancer-specific stats 2021.* Retrieved from https://cdn.cancer.ca/: https://cdn.cancer.ca/-/media/files/research/cancer-statistics/2022-statistics/2022\_cancer-specific-stats.pdf

Canadian Cancer Society. (n.d.). *Staging and Grading*. Retrieved from Canadian Cancer Society: https://cancer.ca/en/cancer-information/what-is-cancer/stage-and-grade

Cazacu, I. M., Singh, B. S., & McAllister, F. (2022). Pancreatic Cancer Screening. In M. S. Bhutani, M. H. Katz, A. Maitra, J. M. Herman, & R. A. Wolff, *Pancreatic Cancer: A Multidisciplinary Approach* (1 ed.). Switzerland: Springer International Publishing AG. doi:10.1007/978-3-031-05724-3\_4

Chandrasegaram, M. D., Goldstein, D., Simes, J., Gebski, V., Kench, J. G., Gill, A. J., . . . Barbour, A. P. (2015, November). Meta-analysis of radical resection rates and margin assessment in pancreatic cancer. *British Journal of Surgery, 102*(12), 1459–1472. doi:https://doi.org/10.1002/bjs.9892

Ghaneh, P., Kleeff, J., Halloran, C., Raraty, M., Jackson, R., Melling, J., . . . Padbury, R. (2019). The Impact of Positive Resection Margins on Survival and Recurrence Following Resection and Adjuvant Chemotherapy for Pancreatic Ductal Adenocarcinoma. *Annals of Surgery, 269*(3), 520-529. doi:10.1097/SLA.0000000000002557

Javadi, S., & Vincenzo Wong. (2022). Diagnosis and Staging of Pancreatic Cancer: Imaging Evaluations—Pancreatic Protocol CT and MRI, PET-CT. In M. S. Bhutani, M. H. Katz, A. Maitra, J. M. Herman, & R. A. Wolff, *Pancreatic Cancer: A Multidisciplinary Approach* (1 ed.). Switzerland: Springer International Publishing AG. doi:10.1007/978-3-031-05724-3\_4

La Torre, M., Cavallini, M., Ramacciato, G., Cosenza, G., del Monte, S., Nigri, G., . . . Mercantini, P. a. (2011, 11 01). Role of the Lymph node ratio in pancreatic ductal adenocarcinoma. Impact on patient stratification and prognosis. *Journal of Surgical Oncology, 104*(6), 629-633. doi:10.1002/jso.22013

Lau, M. K., Davila, J. A., & Shaib, Y. H. (2010). Incidence and Survival of Pancreatic Head and Body and Tail Cancers. *Pancreas, 39*(4), 458-462. doi:10.1097/MPA.0b013e3181bd6489

Mukherjee, S. (2011). *The emperor of all maladies : a biography of cancer .* New York: Scribner.

Tzeng, C.-W. D. (2022). Pancreatic Cancer Screening. In M. S. Bhutani, M. H. Katz, A. Maitra, J. M. Herman, & R. A. Wolff, *Pancreatic Cancer: A Multidisciplinary Approach* (1 ed.). Switzerland: Springer International Publishing AG. doi:10.1007/978-3-031-05724-3\_4

Yadav, S., Sharma, P., & Zakalik, D. (2018). Comparison of Demographics, Tumor Characteristics and Survival Between Pancreatic Adenocarcinomas and Pancreatic Neuroendocrine Tumors. *American Journal of Clinical Oncology, 41*(5), 485-491. doi:10.1097/COC.0000000000000305

# Appendix

## Five-year overall survival of cancers by type

Data is from Projected estimates of cancer in Canada in 2022 (Tables 1 and 2) available through cancer.ca/statistics and trend and survival estimates are from Canadian Cancer Statistics 2021 (Tables 1.7, 2.7 and 3.1). (Canadian Cancer Society, 2022)

|  | Incidence | | | Mortality | | | OS |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Type | Rank | Count | Trend | Rank | Count | Trend | 5-year (%) |
| Pancreas | 11 | 6,900 | → | 3 | 5,700 | → | 10 |
| Esophagus | 19 | 2,500 | ↓ | 10 | 2,400 | → | 16 |
| Lung and bronchus | 1 | 30,000 | ↓ | 1 | 20,700 | ↓ | 22 |
| Liver | 16 | 3,500 | → | 16 | 1,650 | → | 22 |
| Brain/CNS | 17 | 3,200 | ↓ | 9 | 2,500 | ↓ | 22 |
| Stomach | 14 | 4,100 | ↓ | 12 | 2,000 | ↓ | 29 |
| Ovary | 18 | 3,000 | ↓ | 14 | 1,950 | ↓ | 44 |
| Multiple myeloma | 15 | 4,000 | ↑ | 16 | 1,650 | ↓ | 50 |
| Leukemia | 13 | 6,700 | ↓ | 6 | 3,100 | ↓ | 61 |
| *All cancers* |  | *233,900* | *↓* | *—* | *85,100* | *↓* | *64* |
| Head and neck | 10 | 7,500 | ↑ | 11 | 2,100 | → | 64 |
| Colorectal | 4 | 24,300 | ↓ | 2 | 9,400 | ↓ | 67 |
| NHL | 6 | 11,400 | → | 7 | 3,000 | ↓ | 69 |
| Kidney and renal pelvis | 9 | 8,100 | → | 14 | 1,950 | ↓ | 73 |
| Cervix | 20 | 1,450 | ↓ | 19 | 380 | → | 74 |
| Bladder | 5 | 13,300 | → | 9 | 2,500 | → | 77 |
| Uterus (body, NOS) | 9 | 8,100 | → | 17 | 1,500 | ↑ | 82 |
| Hodgkin lymphoma | 22 | 1,050 | ↓ | 21 | 110 | ↓ | 85 |
| Breast | 2 | 28,900 | ↓ | 4 | 5,500 | ↓ | 89 |
| Melanoma | 7 | 9,000 | ↑ | 18 | 1,200 | ↓ | 89 |
| Prostate | 3 | 24,600 | ↓ | 5 | 4,600 | ↓ | 91 |
| Thyroid | 13 | 6,700 | ↓ | 20 | 250 | → | 97 |
| Testis | 21 | 1,200 | ↑ | 22 | 35 | ↓ | 97 |
| All other cancers |  | 24,500 | ↑ |  | 10,900 | ↓ |  |

Table 4: Five-year overall survival

## Log Survival and Survival

The five-year survival mark corresponds to a log survival just greater than four.

| Overall Survival (mo) | Log survival |
| --- | --- |
| 1 | 0 |
| 6 | 1.791759 |
| 12 | 2.484907 |
| 24 | 3.178054 |
| 36 | 3.583519 |
| 60 | 4.094345 |

## Literature Review

Cancer, also called malignancy, cements its place as among the deadliest and most intractable of the maladies across the globe and time. Patients diagnosed with malignancy suffer from debilitating and progressive physical symptoms, while feeling a sense of existential dread and uncertainty regarding their fate. Current treatment options for cancer are physically disfiguring and rife with side effects. A significant proportion of curatively treated patients will recur after a time period from months to decades, eventually passing away from recurrent disease. Siddhartha Mukherjee referred cancer as The Emperor of All Maladies, highlighting the storied attempts and failures to treat malignancy for centuries (Mukherjee, 2011).

Cancer is actually a multitude of significantly differing diseases, but all sharing a commonality where cellular mutation has been involved in such a way that it prevented the organized destruction (apoptosis) of affected cells. As these cells accumulate further mutations while resisting apoptosis, they may gain malignant behaviors, such as the invasion and destruction of adjacent cells, uncontrolled proliferation, and eventually, metastasis. Physicians specify cancers by the organ where its cells started becoming malignant; lung cancer is cancer that starts in the lungs, and leukemia originates from mutated bone marrow cells resulting in the overproduction of ineffective white blood cells. They further specify the malignancy by their histology; cancers originating from the same organ may originate from different specialized cells. The prognosis of the patient depends widely on the origin and histologic type of their malignancy.

Physicians categorize the extent of the cancer by staging, usually on a four-stage ordinal scale ranging from I to IV. They then further specify the tumor extent by measuring tumor size (T), lymph node involvement (N), and distant metastasis (M), called the AJCC classification system. Distant metastasis always results in a Stage IV categorization. Prognosis of cancer is also dependent on biological behavior; cancers of the same histological type and AJCC classification might be well-differentiated and exhibit indolent behavior (low-grade), or they might be undifferentiated and exhibit aggressive behavior that defy surgical or chemotherapeutic treatment (high-grade). The treatment options and the prognosis of cancer are highly dependent on the staging, since cancers are considered incurable once it has metastasized or progressed such that complete resection can no longer be done safely (Canadian Cancer Society, n.d.).

Current methods of curative treatment display a morbid imagery. They involve body-altering procedures, are rife with side effects, and moreover patients pass away from recurrence, usually only a few years after treatment. Instead, people quantify cancer survival as the proportion of patients still alive or disease-free given a time period after treatment (5 years is the benchmark).

## Pancreatic Ductal Adenocarcinoma

Of all the types of cancers (specified by organ), one cancer stands apart from the rest: pancreatic cancer. With a five-year overall survival (since diagnosis) of 10% as of 2021, and furthermore with little improvement in survival after decades of research, pancreatic cancer is regarded as the deadliest type of cancer by a large margin (Canadian Cancer Society, 2022). Surgery with clear margins is the only possibly curative therapy for pancreatic cancer. More concerning, the remaining patients diagnosed at a lower stage are deemed unresectable due to vascular involvement of the tumor rendering effective surgery impossible. Only around 15-20% of patients diagnosed with pancreatic cancer are eligible to undergo surgery (Cazacu, Singh, & McAllister, 2022, pp. 39-41).

The prognosis of resected patients still remains well below the overall survival across all cancer patients (64% five-year) due to the complicated nature of the surgery (the Whipple procedure), the high risk of treatment failure (R1 or R2 resection, characterized by microscopic or gross tumor involvement at the surgical margins), and the significant risk of recurrence even after successful surgery.

The following section explains the rationale behind data selection.

### Resection Status

Pancreatic cancer resection is one of the most complicated major surgical procedures (Tzeng, 2022, pp. 127-137). Surgeons must perform a pancreaticoduodenectomy (Whipple procedure) to resect tumors occurring on the head of the tumor that represents around 77.5% of the PDAC cases (Lau, Davila, & Shaib, 2010). This procedure is associated with a considerable failure rate, as one meta-analysis showed that only 70-80% of the resections have R0 margins. The 8th edition of the AJCC now considers close resections with clearance margins less than 1 mm as R1 as well; only 15-24% of the resections are R0 under the new definition (Chandrasegaram, et al., 2015).

According to, *Annals of Surgery*, “R1-direct resections were associated with significantly reduced overall and recurrence-free survival following pancreatic cancer resection” (Ghaneh, et al., 2019). I will exclude patients who underwent R1 or R2 resection from our analysis.

### Tumor Size

Surgeons use the maximum dimension of the tumor to classify the T stage. The length of the maximum dimension determines the T stage of the tumor if the tumor does not involve vascular structures so as to make it unresectable. The T stage ranges from 0 to 4, with T0 indicating absence of primary, and T4 reserved for tumors that became locally unresectable (Javadi & Vincenzo Wong, 2022). Higher T stage corresponds to higher Roman numeral staging and is thus an independent prognosticator of patient demise.

### Nodal Involvement

The hallmark of cancer is the ability to metastasize. Cancers usually metastasize first to the lymph nodes, bloodstream, or the neural system before evidence of distant metastases becomes present. A lymph node that has been occupied by cancer en route to distant metastases is referred to as a *positive lymph node*. Lymph node involvement is usually quantified in terms of the number of harvested lymph nodes, the number of lymph nodes known to be positive, and the lymph node ratio (known positive nodes divided by harvested nodes). Nodal involvement contributes to the N stage of the tumor, with N0 indicate node-negative disease (absence of any detected positive lymph nodes), and higher classifications indicating various levels of lymph node involvement by location and distance from primary site.

Positive nodal involvement is a significant negative predictor on the patient prognosis, as nodal involvement indicate that underlying and undetected metastatic processes are present, which might be behind the patient’s likely recurrence (Jamiyan, et al., 2020).

According to a 2011 Journal of Surgical Oncology study, the lymph node ratio is a more powerful predictor for survival than the categorical lymph node status in patients who underwent the Whipple procedure. A node-positive disease with high lymph node ratio has been demonstrated to be a significant predictor of both early recurrence and poor overall survival at both the univariate and multivariate analysis (La Torre, et al., 2011, p. 632).

### Grade

The survival of patients of a given AJCC staging system is highly variable, suggesting that other factors might be at play. Tumor grade is often omitted when categorizing the Roman numeric staging of the tumor, but has been shown, in a single-center study that this metric is one of the strongest independent predictors of prognosis (Rochefort, et al., 2013). The lack of cell differentiation in a tumor suggests aggressive biological behavior that negatively impacts patient survival given similar tumor sizes and lymph node ratios. Like nodal involvement, high tumor grade is a candidate for an independent predictor for underlying metastatic process, recurrence, and the consequent demise of the patient (Jamiyan, et al., 2020).

### Age

Evidence shows, from an Iranian multi-center study, that advanced age (namely over sixty years old at first diagnosis) is also an independent negative predictor to patient survival (Bahardoust, et al., 2022). According to the study’s multivariate Cox regression, patients at least sixty years old confer a hazard ratio of 1.308 (95% CI of 1.026–1.66) and 1.25 (95% CI of 1.03–1.49) compared to patients diagnosed before the age of sixty.

### Year of Diagnosis

With advances in pancreatic cancer screening and treatment options, patients diagnosed in later years expect to have a better prognosis than cases recorded in earlier years. The five-year overall survival of patients has increased slowly, but steadily within the past decade from 6% to around 12% due to improved surgical practices, better early detection methods, and more effective chemotherapeutic regimens such as folfirinox (Scholten, 2023).

⁂

1. The number of degrees of freedom is 21 for all predictors, except for high grade (df = 16) and low grade (df = 14). [↑](#footnote-ref-2)