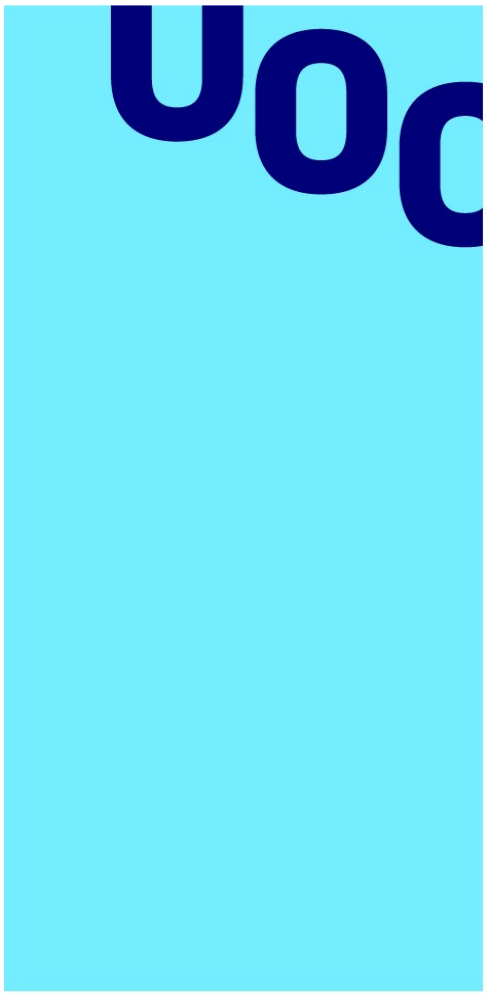
Cáncer de páncreas:

Análisis clínico y genómico, y modelización a partir de los datos del TGCA



Universitat Oberta de Catalunya

Nombre Estudiante
Carlos Gatón Rodriguez

Machine learning in biomarkers and clinical data analytics

Tutor/a de TF
Laura Jiménez Gracia
Profesor/a responsable de la asignatura
Laia Subirats Maté

28/12/2025

Ficha del Trabajo Final

Título del trabajo:	Cáncer de páncreas: Análisis clínico y genómico, y modelización a partir de los datos del TGCA.
Nombre del autor/a:	Carlos Gatón Rodriguez
Nombre del Tutor/a de TF:	Laura Jiménez Gracia
Nombre del/de la PRA:	Laia Subirats Maté
Fecha de entrega:	12/2025
Titulación o programa:	Máster Ciencia de Datos
Área del Trabajo Final:	Machine learning in biomarkers and clinical data analytics
Idioma del trabajo:	Castellano
Palabras clave	páncreas, supervivencia, análisis

Resumen del Trabajo

The Cancer Genome Atlas es un programa desarrollado por el National Cancer Institute de Estados Unidos entre los años 2006 y 2018. Se caracterizaron molecularmente más de 20000 muestras de tumores primarios abarcando 33 tipos diferentes de cáncer. Gracias a este proyecto se generaron una gran cantidad de datos genómicos, epigenéticos, transcriptómicos y proteómicos.

El trabajo se centra en volver a analizar estos datos, específicamente los de cáncer de páncreas. Se analizan los datos clínicos y genómicos, y se comparan los resultados con los presenten en la literatura.

Primero de todo se realiza una descripción de los datos demográficos, diagnósticos, de tratamientos y genómicos. Posteriormente, se realiza un extenso análisis de la supervivencia. Finalmente, se implementa un modelo de regresión basado en redes neuronales que permite estimar la curva de supervivencia de un hipotético paciente.

La mayoría de los resultados obtenidos durante la realización del trabajo concuerdan con los descritos en la literatura. Aquellos resultados discordantes han sido señalados y se ha intentado dar una posible explicación. El modelo desarrollado, aunque limitado por la poca cantidad de datos disponibles, ha conseguido obtener unos resultados aceptables.

Abstract

The Cancer Genome Atlas is a program developed by the National Cancer Institute of the United States between 2006 and 2018. More than 20,000 primary tumor samples, encompassing 33 different types of cancer, were molecularly characterized. This project generated a vast amount of genomic, epigenetic, transcriptomic, and proteomic data.

This work focuses on reanalyzing this data, specifically data related to pancreatic cancer. Clinical and genomic data are analyzed, and the results are compared with those presented in the literature.

First, a description of demographic, diagnostic, treatment, and genomic data is provided. Subsequently, an extensive survival analysis is performed. Finally, a regression model based on neural networks is implemented to estimate the survival curve of a hypothetical patient.

Most of the results obtained during this work are consistent with those described in the literature. These discrepancies have been noted, and an attempt has been made to provide a possible explanation. The model developed, although limited by the small amount of available data, has managed to obtain acceptable results.

Índice general

Índice de Figuras

Índice de tablas

1. Introducción

1.1. Contexto y justificación del Trabajo

El cáncer son un conjunto de enfermedades aún muy presentes en la sociedad y el desarrollo de nuevos tratamientos para combatirlos sigue siendo un campo de investigación en constante evolución. Por otro lado, la biología humana sigue siendo una ciencia que le queda mucho camino por recorrer para llegar a comprender todos los procesos biológicos y continuamente, sigue evolucionando la forma en la que entendemos el ciclo biológico del cáncer.

El trabajo se centra en el cáncer de páncreas. Se trata de un tipo de cáncer con una incidencia moderada y un pronóstico muy pobre. Las tasas de supervivencia a 5 años siguen siendo muy bajas, se sitúan en torno al 13% (1).

El ámbito en el que se centra el trabajo sigue siendo de gran interés tanto desde un punto de vista general, para la sociedad, como desde el punto de vista científico.

Existen gran cantidad de estudios alrededor de la biología del cáncer de páncreas y también se han realizado innumerables ensayos clínicos probando multitud de terapias en busca de mejorar las perspectivas que tienen los pacientes con cáncer de páncreas. Aun así, las mejoras que se han conseguido han sido limitadas y queda un largo camino que recorrer para mejorar la supervivencia de estos pacientes.

El trabajo busca reanalizar los datos producidos por el proyecto TCGA-PAAD y comparar los resultados obtenidos con aquellos presentes en la literatura relacionada. También, se pretende intentar extraer algún conocimiento nuevo. Como punto final del trabajo se busca generar un modelo que permite determinar la probabilidad de supervivencia de un paciente a lo largo del tiempo dada unas características clínicas y genómicas.

1.2. Explicación de la motivación personal

He escogido este trabajo debido al gran interés que me despierta este campo. Sigo la actualidad de este mundo y siempre estoy al tanto de nuevos avances en el tratamiento del cáncer. Actualmente, vivimos en una época revolucionaria para el tratamiento del cáncer con la aparición de multitud de modalidades terapéuticas que han mejorado las perspectivas de muchos pacientes, como pueden ser la inmunoterapia o las terapias dirigidas.

Actualmente, no tengo vinculación profesional con este ámbito, pero estoy abierto a que un día nuestros caminos se crucen y pueda aportar de grano de arena a la investigación del cáncer

1.3. Objetivos del Trabajo

Los objetivos del trabajo son variados. Por un lado, se buscan realizar tareas técnicas relacionadas con la ciencia de datos como pueden ser la limpieza de datos o la visualización de datos. Posteriormente, con los datos procesados se realizarán procesos de investigación científica para explorar diferentes hipótesis y obtener múltiples resultados. La lista detallada de tareas técnicas es:

- Extraer y procesar los datos clínicos y genómicos del TCGA-PAAD para que tengan un formato adecuado para su posterior análisis.
- Realizar un análisis exploratorio de datos (EDA) de los datos demográficos, diagnósticos y relacionados con los tratamientos.
- Actualizar la estadificación de los pacientes a la última versión y observar si mejora su capacidad pronóstica.

La lista detallada de objetivos científicos es:

- Analizar y describir los tratamientos de RHE comprobando la relación entre la dosis total de los tratamientos y las respuestas de los pacientes.
- Analizar y describir los tratamientos de quimioterapia, viendo la evolución que estos sufrieron a lo largo de los años.
- Analizar las mutaciones genómicas más comunes.
- Realizar un análisis completo de supervivencia, viendo que factores son los que tienen mayor capacidad pronóstica y serán seleccionados para el posterior modelo.
- Implementar un modelo de regresión de supervivencia a partir de un MLP para estimar la curva de supervivencia de un determinado paciente.

1.4. Impacto en sostenibilidad, ético-social y de diversidad

El impacto en sostenibilidad de este trabajo es mínimo. Para realizar el trabajo hay un cierto consumo energético relacionado con la creación y ejecución de los códigos de programación necesarios para el procesamiento y análisis de los datos. También tiene un pequeño consumo el entrenamiento del modelo MLP de regresión. Estas operaciones son ejecutadas en servidores con lo que es difícil medir el consumo de manera exacta.

El impacto en la dimensión de comportamiento ético y de responsabilidad social es nulo. Los datos utilizados están completamente anonimizados y no hay ninguna forma de que se vulnere la privacidad y confidencialidad de ninguna persona. El resultado del proyecto difícilmente se puede utilizar con fines fraudulentos en ningún ámbito. También sigue el código deontológico y no pone en riesgo ningún puesto de trabajo.

El impacto en la dimensión de diversidad, género y derechos humanos puede ser inherente a los propios datos. Durante el análisis de los datos demográficos se comprobarán si todas las partes están correctamente representadas o si existen minorías étnicas, raciales o de género que estén infrarrepresentadas en los datos. El trabajo que se realice con esos datos no tendrá ningún impacto es aspectos de género, diversidad o derechos humanos. Tampoco tendrá ningún impacto sobre ninguna legislación.

1.5. Enfoque y método seguido

La estrategia elegida empieza a partir de un producto existente e intenta acabar desarrollando un nuevo producto. Inicialmente, se extraen y procesan datos publicados de estudios anteriores. Con los datos procesados se realiza un EDA y se obtiene una descripción inicial. Esta descripción inicial está acompañada de tablas y figuras para ayudar a la comprensión. A continuación, se comparan los resultados obtenidos con aquellos presentes en la literatura. Se destacan las similitudes y diferentes entre el análisis realizado y los publicados.

Posteriormente, se intenta desarrollar un nuevo producto a partir de los resultados y conclusiones alcanzados en los apartados anteriores. Finalmente, se analizan todos los resultados obtenidos y se valora el grado de cumplimiento de los objetivos propuestos. Se presentan posibles mejores y futuras líneas de investigación.

1.6. Planificación del Trabajo

El trabajo se divide en 4 bloques: inicio, desarrollo, resultados y conclusiones.

En el primer bloque se pretende familiarizarse con el tema a desarrollar, el estado del arte y los posibles métodos y herramientas disponibles para el desarrollo del trabajo.

En el bloque principal, desarrollo, se llevan a cabo gran parte de las tareas necesarias para completar el trabajo. En este bloque se implementan todas las herramientas y recursos necesarios para completar el trabajo y se obtienen la mayor parte de productos.

En el bloque de resultados se analizan los resultados y productos obtenidos durante el desarrollo del proyecto y se empiezan a plasmar en la memoria asociada al trabajo.

En el bloque final, conclusión, se acaba de redactar la memoria incluyendo las conclusiones obtenidas de la realización del proyecto incluyendo posibles mejoras y líneas de investigación futuras.

La duración de cada bloque puede ser variable dependiendo de la disponibilidad, dedicación y dificultades encontradas en cada momento. Pero, cabe destacar que el primer y último bloque deben tener una dedicación bastante menor que los dos bloquees centrales.

Todo el proceso puede llevarse a cabo de una forma iterativa, donde los resultados que se vayan obteniendo lleven al planteamiento de nuevas hipótesis que hagan evolucionar los objetivos del trabajo según la información con la que se vaya contando.

1.7. Breve sumario de productos obtenidos

Los productos obtenidos durante la realización del trabajo son los siguientes:

- Conjuntos de datos sin procesar en formato tsv y cbt. Contienen los datos de las fuentes originales. Son el punto de partida del proyecto a partir de los cuales se empieza a desarrollar todo el trabajo.
- Conjunto de datos procesados en formato csv. Contienen los datos procesados a partir de los conjuntos originales. Ponen fin al primer paso del proyecto y son el punto de partida del posterior análisis.
- Conjunto de figuras que acompañan los resultados y conclusiones. Permiten mostrar información y patrones de una forma mucho más efectiva.
- Conjunto de tablas que ayudan a la descripción de los datos. Permiten condensar la información de múltiples variables de una forma efectiva y visual.
- Modelo de regresión de análisis de supervivencia que permite obtener las probabilidades de que el paciente este vivo en un determinado momento. Es el producto totalmente nuevo que pone fin al proyecto.

1.8. Breve descripción de los otros capítulos de la memoria

Los demás capítulos que conforman esta memoria son los siguientes:

- Estado del arte: Se trata de un capítulo introductorio al proyecto. En él se explican las diferentes investigaciones que se han desarrollado en el ámbito del proyecto o que estén relacionadas con este.
- Materiales y métodos: En este capítulo se detallan los aspectos más destacados del diseño, implementación y desarrollo del trabajo. Se describe la metodología utilizada y posibles alternativas o mejoras. También se detallan los productos obtenidos. Forma parte del bloque vertebrador del proyecto, pero con menor peso que el capítulo de los resultados.

- Resultados: Se trata de un capítulo en el que se detallan todos los resultados obtenidos durante la realización del proyecto. Este contiene figuras y tablas que facilitan la comprensión de aquello que se está exponiendo. Se trata de un capítulo central en el proyecto dado que es en el dónde se concentran la mayoría de los esfuerzos.
- Conclusiones y trabajos futuros: Se trata de un capítulo que recoge las conclusiones del trabajo y una reflexión sobre el cumplimiento de los objetivos y la planificación establecidos. También, se revisa el impacto en los criterios ASG definidos en la introducción. Finalmente, se proponen líneas de investigación para trabajos futuros que han surgido a partir del trabajo o que no han podido ser cubiertas. Por tanto, es el capítulo que cierra el trabajo realizado.
- Glosario: Se trata de un compendio que recoge la definición de los términos y acrónimos más relevantes utilizados en la confección de la memoria. Se trata de un capítulo transversal en el proyecto ya que está relacionado con todos los demás capítulos.
- Bibliografía: Se trata de una lista numerada de las referencias bibliográficas utilizadas en la memoria. También, se trata de un capítulo transversal en el proyecto.

2. Estado del arte

El cáncer son un conjunto de enfermedades que tienen una incidencia creciente (2). Cada vez más personas se ven afectadas por algún tipo de cáncer y se siguen teniendo unas tasas de supervivencia relativamente bajas en muchos casos. Por eso, aún queda un largo camino por recorrer en cuanto a la mejora de las expectativas de supervivencia en la mayoría de los casos.

Además, existe una gran diferencia de pronóstico entre los diferentes tipos de cáncer. En algunos tipos, como puede ser el de mama en el caso de las mujeres o el de próstata en caso de los hombres, la investigación ha avanzado mucho y las tasas de supervivencia a 5 años han mejorado mucho, en ambos casos se sitúan por encima del 90% (3,4). Esto se debe gracias a la aparición de tratamientos más efectivos y a campañas de sensibilización para mejorar la detección precoz.

En otros casos, como el cáncer de páncreas que se trata del objeto de estudio de este trabajo, no se han visto reflejados estos avances y las tasas de supervivencia se han mantenido muy bajas. A nivel mundial, el cáncer de páncreas se sitúa en el puesto 12 a nivel de incidencia, pero en el puesto 6 a nivel de mortalidad (5), reflejando el pobre pronóstico que tiene este tipo de cáncer. Por esto, las muertes asociadas a este cáncer siguen creciendo a medida que su incidencia sigue aumentando (6).

Actualmente, la tasa de supervivencia a 5 años del cáncer de páncreas se sitúa alrededor del 11%-13% y solo ha sufrido una ligera mejora respecto el 7% que tenía a principios de siglo. Además, los tipos con peor pronóstico, como pueden ser aquellos que son diagnosticados en estado metastático su tasa se ve reducida a solo el 3%-4% (1,7). Por eso, es fundamental progresar en el desarrollo de terapias dirigidas y nuevas modalidades terapéuticas que mejoren estos valores.

El TCGA fue un programa desarrollado entre el 2006 y el 2018 entre el NCI y el NHGRI. En él se caracterizaron molecularmente más de 20000 muestras de tumores primarios abarcando 33 tipos diferentes de cáncer. Gracias a este proyecto se generaron una gran cantidad de datos genómicos, epigenéticos, transcriptómicos y proteómicos. En el caso del cáncer de páncreas se caracterizaron 185 muestras y dio lugar a la publicación en 2017 de un estudio sobre la caracterización integral genómica de este cáncer (8,9).

El objetivo principal de este trabajo es la reanálisis de los datos óhmicos de secuenciación disponibles en el TGCA. Por lo tanto, el punto de partida debe de ser el trabajo que se realizó dentro de este programa y, a partir de las conclusiones que se extrajeron, plantear nuevas hipótesis y desarrollar el trabajo a partir de este punto.

A partir de lo expuesto y centrándose en la temática concreta del trabajo, el cáncer de páncreas, el TGCA público un estudio específico sobre este tipo de cáncer. En él se presentó el análisis molecular multiplataforma de 150 especímenes de PDAC (9).

Se confirmaron mutaciones en múltiples genes previamente identificadas relacionadas esta enfermedad. La principal mutación, presente en más del 90% de los casos, se trata del gen *KRAS*. Además, existe gran heterogeneidad en el perfil mutacional de la proteína *KRAS*, echo que dificulta identificar el papel que juega esta proteína en la progresión del cáncer. En algunos casos se identificaron múltiples mutaciones *KRAS* en una misma muestra. Se destaca el comportamiento diferencial de la variante *KRAS G12R*, dejando para futuras investigaciones la confirmación de este hecho.

También se destaca la presencia de otras mutaciones en la misma vía *RAS* o en otras vías proteicas. Se destaca la importancia de realizar perfiles moleculares en pacientes sin mutaciones *KRAS* para poder identificar otras proteínas con relevancia terapéutica.

Examinando la expresión proteica se identificaron diferentes subtipos proteómicos con diferencias en cuanto a pronóstico e implicaciones terapéuticas. Análisis de RNA también sugieren la existencia de múltiples subtipos con características diferenciadas como sugieren los otros análisis. Se deja para futuras investigaciones la caracterización más exhausta de estos varios subtipos.

La integración de los múltiples análisis que se realizaron reveló un perfil molecular complejo, pero pudieron proveer de una guía para poder desarrollar medicina de precisión.

Por lo tanto, el trabajo deberá empezar por volver a analizar estas muestras para caracterizarlas y ver que se obtienen algunos de los resultados expuestos en este estudio y a partir de ahí intentar ver si se puede generar algún conocimiento adicional o investigar sobre alguno de los aspectos que quedaron más abiertos en este estudio.

3. Materiales y métodos

El trabajo ha sido desarrollado íntegramente mediante Jupyter Notebooks utilizando el lenguaje de programación Python en la plataforma Google Colaborate. También se creó un repositorio en Git Hub como sistema de control de versiones, el enlace al cual se muestra a continuación:

https://github.com/carlosgaton/TFM-TCGA-PAAD

El trabajo se divide en cinco etapas que se describirán individualmente indicando el diseño y desarrollo de cada etapa y los productos obtenidos. A continuación, se muestra un esquema del proceso seguido.

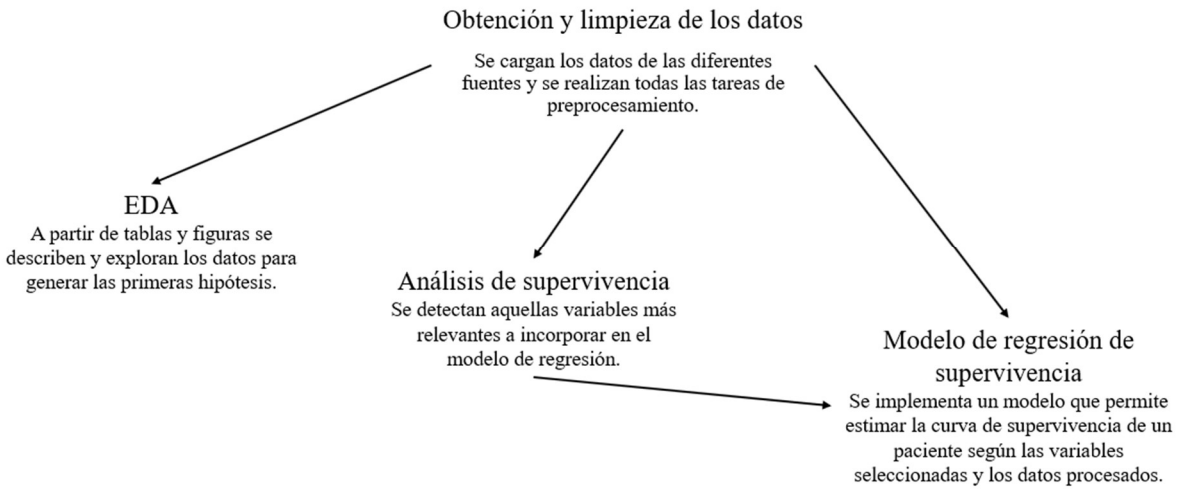


Figura 1: Esquema del flujo de trabajo seguido durante el proyecto

3.1. Obtención y carga de los datos

Se trata de la etapa inicial del proyecto. Su finalidad es la obtención de los diferentes conjuntos de datos y su carga para permitir su posterior preprocesado. Los productos obtenidos de esta etapa son los conjuntos de datos listos para ser preprocesados, concretamente se obtuvieron 5 archivos en formato tsv con los datos clínicos y un archivo cbt con los datos de las mutaciones genéticas.

Este proceso se llevó a cabo de dos formas diferentes en función del origen de los datos. Para los archivos clínicos se descargaron manualmente de la página web del GDC Data Portal y corresponden a los datos de la sección "clinical" del proyecto TCGA-PAAD (10). Posteriormente, se subieron a la plataforma Google Drive y se cargaron en un Notebook como un DataFrame para su posterior preprocesado.

Para el archivo de mutaciones genéticas se obtuvo a través de descarga directa realizando una petición HTTP a la página LinkedOmics (11). Se procesaron los datos, se guardaron y se volvieron a cargar como DataFrame para su posterior preprocesado.

3.2. Preprocesado de los datos

Se trata de una etapa muy importante dentro del trabajo. Un buen preprocesamiento de los datos permite dejarlos en un formato que facilite extraer la mayor información posible de ellos y permita realizar buenos análisis posteriores. En esta etapa se incluyen tareas de integración, selección, reducción, conversión y limpieza de datos. Todos los conjuntos de datos resultantes contienen el mismo identificador que permite relacionar la información de cualquiera de ellos entre sí.

Los datos relativos a las mutaciones genéticas ya estaban en un formato bastante limpio, solo se adecuo el nombre de las columnas para que coincidiese con el utilizado en los otros conjuntos de datos y se añadieron dos columnas resumen para facilitar el posterior análisis.

Los datos clínicos estaban mucho menos limpios y se requirió un preprocesamiento mucho más extenso. Estos se encontraban divididos en cinco archivos que se fueron procesando secuencialmente. El archivo clinical es el principal y contiene los datos básicos de los pacientes, sus diagnósticos y sus tratamientos. El archivo exposure contiene información sobre exposición a sustancias. El archivo family contiene información sobre el historial familiar de cáncer. El archivo follow_up contiene información sobre el seguimiento de los pacientes. El archivo path_details contiene datos histopatológicos adicionales.

La limpieza inicial de cada conjunto de datos es común en todos ellos. Tras substituirse diferentes formatos para indicar los valores nulos por un formato nulo reconocido se eliminan todas las columnas con solo valores nulos. Esto se debe a que todos los proyectos del TCGA comparten las mismas columnas, pero solo algunas de ellas se usan en cada proyecto.

Tras esta limpieza inicial se realiza una exploración de las diferentes columnas de cada uno de los conjuntos de datos para seleccionar aquellas que contengan información relevante. A continuación, varios de los conjuntos se dividen para capturar las particularidades de los diferentes datos que contienen. Por ejemplo, clinical se divide en demográfico, diagnóstico y tratamiento.

El siguiente paso común a todos ellos es el formateo del nombre de las columnas para que tengan un nombre descriptivo, eliminación de valores duplicados, eliminación de columnas vacías o innecesarias y la eliminación de filas vacías. También en muchas columnas de tipo categórico se unifican las categorías repetidas o que se solapen.

Toda la información relevante de exposure, family y follow_up se condensa con los datos de demográfico. El conjunto de datos diagnostico se vuelve a dividir en función del tipo de diagnóstico y también se crea un nuevo conjunto de datos con un resumen de los diagnósticos recibidos por cada paciente. Se puede destacar la creación de nuevas columnas en el conjunto de datos de diagnóstico primario donde se actualiza la estadificación de los pacientes a través de la información de los estadios actuales y la información histopatológica proveniente de path_details. El resultado es la obtención de cuatro conjuntos de datos relacionados con los tratamientos.

El conjunto de datos tratamiento también se vuelve a dividir en función del tipo de tratamiento. Se crea un conjunto de datos con los tratamientos de RHE. Se destaca la conversión de unidades, ya que, las dosis recibidas por los pacientes estaban expresadas en diferentes unidades.

Se crea un conjunto de datos con los tratamientos de quimioterapia. Este fue el conjunto de datos que más transformaciones sufrió hasta su forma final. Los datos estaban muy sucios, ya que, cada fármaco ocupaba una fila, aunque se hubiesen dado en simultaneo. Por lo tanto, hubo que aplicar múltiples suposiciones para ir agrupando las diferentes filas hasta llegar al resultado final. A la vista de los resultados obtenidos en los análisis posteriores, se puede afirmar que las suposiciones aplicadas son válidas y adecuadas, dado que no se observan incoherencias ni inconsistencias al comparar los resultados obtenidos con las referencias bibliográficas.

Finalmente, se vuelve a crear una tabla resumen con los diferentes tratamientos que recibió cada paciente para su posterior análisis.

Terminado el preprocesamiento de los datos, los diferentes conjuntos de datos generados se guardan en archivos csv. En resultado de esta etapa son ocho conjuntos de datos representando diferentes tipologías de datos.

3.3. Análisis exploratorio de los datos

En esta etapa se obtienen los primeros resultados del trabajo. Todo el proceso va acompañado de figuras y tablas que ayudan a la descripción de las variables y a la comprensión de la información.

Primero de todo se analizan los datos demográficos. Se describen las diferentes variables y analiza su representatividad. A continuación, se analizan los diferentes factores de riesgo presentes en los datos.

En segundo lugar, se analizan los diferentes datos relacionados con los diagnósticos. Se empieza analizando la distribución de los diagnósticos, para posteriormente entrar a explorar en detalle la información relacionada con cada tipo de diagnóstico. A destacar el análisis de la reclasificación de los estadios patológicos.

En tercer lugar, se analizan los diferentes datos relacionados con los tratamientos. Se empieza analizando la distribución de los tratamientos, para posteriormente entrar a explorar en detalle la información relacionada con cada tipo de tratamiento. A destacar el análisis en detalle de los tratamientos de RHE, utilizando el test de Wilcoxon, y de los tratamientos de quimioterapia donde se intenta observar la evolución que han sufrido estos tratamientos a partir del año de primer uso de un fármaco o combinación.

Finalmente, se analiza la información relacionada con las mutaciones genéticas. Se describen las mutaciones relevantes más frecuentes y se analizan los patrones de co-mutaciones entre ellas.

El resultado de esta etapa son un conjunto de figuras y tablas que acompañan la descripción de la información y los resultados y que facilitan la comprensión y la identificación de patrones.

3.4. Análisis de supervivencia

Primero de todo, se realiza una estimación de la supervivencia según el tipo de diagnóstico primario mediante diagramas de cajas. Observando las diferencias entre los diferentes tipos de cáncer de páncreas se seleccionan solo aquellos pacientes con adenocarcinoma ductal, el tipo mayoritario. Además, se filtran solo aquellos pacientes con datos genómicos disponibles. Tras esta selección se obtiene una muestra de 115 pacientes. Se muestra la curva de supervivencia global.

Se continua el análisis de supervivencia analizando el efecto de las diferentes variables demográficas. Se analiza la correlación entre la edad y la supervivencia mediante el coeficiente de correlación de Spearman debido a la no normalidad de los datos. Se analizan el resto de las variables demográficas, pero no se puede llegar a análisis más profundos por falta de representación de algunas categorías. Los valores nulos son imputados con la clase mayoritaria de cada variable. En todas las variables representan como máximo el 1%-2% teniendo un impacto muy limitado. Dado la escasez de datos se optó por esta opción preferentemente a eliminarlos.

A continuación, se analizan las variables diagnósticas y genómicas. Se utilizan curvas de supervivencia de Kaplan-Meier para estimar la supervivencia a lo largo del tiempo. Se implementan regresiones de Cox para evaluar como las diferentes variables afectan la tasa de riesgo de muerte. Primero de todo, se comparan las curvas de supervivencia de las dos versiones de clasificación de los estadios patológicos. Seguidamente, se van analizando el resto de las variables según la división que se aplicará en el posterior modelo de regresión. Simultáneamente, se va generando el conjunto de datos que servirá para generar el modelo de regresión

El resultado de esta etapa son el conjunto de figuras que permiten analizar la supervivencia y el conjunto de datos que se utilizará para entrenar y validar el modelo de regresión. Este conjunto de datos contiene un seguido de variables binarias y la edad que se escalará posteriormente.

3.5. Modelo de regresión de supervivencia

Se implementa un modelo de regresión de supervivencia mediante una red neuronal de Cox. Esta permite obtener una función de riesgo para cada paciente. El modelo devuelve un índice de concordancia que indica la capacidad de predicción de la red. Valores de 0.5 indican que el modelo realiza predicciones aleatorias y valores cercanos a 1 indican predicciones perfectas.

La red neuronal implementada es muy ligera, una sola capa de 4 neuronas, debido a la baja relación de instancias y atributos, se disponen de 115 instancias y 10 atributos. Se aplican técnicas de validación cruzada para optimizar el número de instancias.

El modelo se entrenó 5 veces reservando cada vez un 20% de los datos como test. Los resultados del modelo son un promedio de cada una de las evaluaciones. Así, se evita que pequeños errores cambien drásticamente el resultado del modelo. El mejor modelo de los cinco es guardado y cargado al final del entrenamiento para su uso. También, se aplica “early stopping” para optimizar el entrenamiento.

Una vez entrenado el modelo se realiza un análisis de importancia mediante el método SHAP para analizar que variables tienen una mayor contribución en el modelo y comparar si los resultados son coherentes con los análisis previos. Para finalizar se muestran las curvas de supervivencia individual para el conjunto de validación.

El resultado de esta etapa es una red neuronal entrenada para predecir la curva de supervivencia de un paciente dadas unas determinadas variables.

4. Resultados

4.1. Análisis de los datos demográficos

Se analizan las diferentes variables demográficas y se determinaran si son representativas de la población general y si pueden ser consideradas un factor de riesgo. La muestra cuenta con un total de 185 pacientes.

4.1.1. Edad

Se comienza por la edad. La mediana de edad de la muestra estudiada es de 65 años con un rango de valores entre los 35 años y los 88 años. La figura 1 muestra su distribución. Se observa que sigue una distribución cercana a la normal pero sesgada hacia los valores altos.

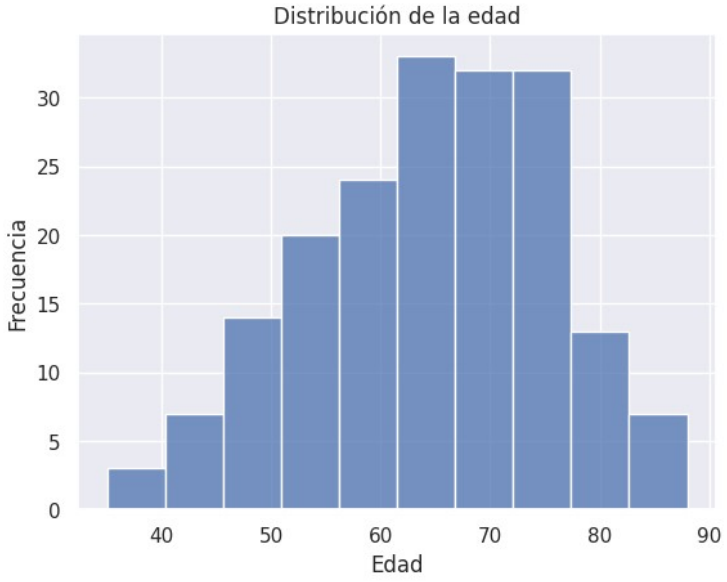


Figura 2: Distribución de la edad de los pacientes

Si se comparan con los datos del registro SEER que muestran una mediana de edad de 68 años vemos que esta variable sí que es representativa de la población general (12). La edad se considera un factor de riesgo, ya que, la probabilidad de desarrollar este cáncer aumenta con la edad. Es poco frecuente en personas menores de 40 años y alcanza un pico entre los 60 y 80 años (6). En el apartado de análisis de supervivencia se comprobará si tiene relación con la probabilidad de supervivencia.

4.1.2. Genero

En la muestra de datos 102 pacientes, el 55.1%, eran hombres. La incidencia de cáncer de páncreas es superior en hombres que en mujeres en una proporción 11/8 a nivel global (6). Por tanto, la muestra también recoge esa mayor proporción de hombres que de mujeres afectados por esta enfermedad. Se cree que esto es debido a la diferencia de estilos de vida o factores de riesgo ambientales. Se analizará si existen diferencias en la supervivencia.

4.1.3. Etnia y raza

En la muestra la etnia mayoritaria era la no hispano o latino con 137 pacientes, el 74.1%, y la raza mayoritaria era la blanca con 162 pacientes, el 87.6%. Se pueden ver la frecuencia de las otras etnias y razas en la tabla 1.

Etnia, n (%)	No reportado	43 (23.2)
	Hispano o latino	5 (2.7)
	No hispano o latino	137 (74.1)
Raza, n (%)	No reportado	5 (2.7)
	Asiática	11 (5.9)
	Negra o afroamericana	7 (3.8)
	Blanca	162 (87.6)

Tabla 1: Frecuencia de la diferentes etnias y razas

Se ha comprobado que la incidencia del cáncer de páncreas varia significativamente entre razas. Siendo los afroamericanos aquellos con mayor incidencia y los asiáticos los de menor incidencia. Parte de estas diferencias se pueden atribuir a factores de riesgo modificables, pero también se ha comprobado que existen diferencias a nivel molecular y genético entre las razas que pueden explicar las diferencias en las tasas de incidencia y supervivencia (6,13). Por eso es importante que todas las razas estén representadas en los ensayos clínicos y estudios, sobre todo las minoritarias para que puedan recibir tratamientos adecuados a sus características.

En cambio, está plenamente documentado que las minorías étnicas y raciales están infrarrepresentadas en la mayoría de los ensayos clínicos. En este caso también es así. Los pacientes de raza negra representan el 12.4% de los casos muy por encima de su representación en este estudio. También, los hispanos representan el 8.5% de los casos por encima de los datos de esta muestra (14).

4.1.4. Lugar de residencia

En la muestra la mayoría de los pacientes, un 87.5%, residen en Norteamérica. En la tabla 2 se muestra la distribución de los pacientes según el lugar de residencia. En la muestra también hay pacientes de Europa, Sudamérica, Asia y Oceanía. Pueden existir diferencias en las tasas de supervivencia entre países debido a múltiples factores como pueden ser el desarrollo económico del país o el nivel del sistema sanitario.

País de residencia, n (%)	Australia	5 (2.7)
	Brasil	2 (1.1)
	Canadá	40 (21.6)
	Alemania	6 (3.2)
	Rusia	1 (0.5)
	Corea del Sur	8 (4.3)
	Estados Unidos	122 (65.9)
	Vietnam	1 (0.5)

Tabla 2: Distribución de los pacientes según el país de residencia

4.1.5. Estado vital

El estado vital indica si un paciente ha fallecido o aún sigue con vida en el último seguimiento realizado. En la muestra 113 pacientes, el 61.1% aún siguen con vida. Estos datos se pueden considerar que aún son inmaduros, dado que aún no han ocurrido al menos la mitad de los eventos. Este hecho puede hacer que la forma de la curva de supervivencia y las estimaciones de supervivencia varíen cuando se vuelvan a analizar los mismos datos con un nivel de maduración superior.

4.2. Análisis de los factores de riesgo

Los factores de riesgo son aquellos elementos que afectan a las probabilidades de contraer un cáncer, en este caso cáncer de páncreas. En función de su naturaleza pueden modificables o no modificables. Como factores no modificables se estudiarán las comorbilidades y el historial familiar y como factores modificables la exposición a sustancias.

4.2.1. Comorbilidades

En la tabla 3 se muestra la prevalencia de diabetes y pancreatitis crónica en la población estudiada.

Diabetes, n (%)	No	147 (79.5)
	Yes	38 (20.5)
Pancreatitis crónica, n (%)	No	172 (93.0)
	Yes	13 (7.0)

Tabla 3: Distribución de comorbilidades de la muestra de estudio

Se ha descrito una asociación positiva entre ambos tipos de diabetes y el riesgo de cáncer de páncreas en múltiples estudios (6). Se estima que en Estados Unidos entre 2017 y 2020 la prevalencia de diabetes era del 14.5% en adultos entre 45 y 64 años y del 24.4% en adultos mayores de 65 años (15). Se puede estimar que la prevalencia normal en una muestra como la que disponemos, dada la distribución de edad, seria de alrededor del 19.5%. Por tanto, se puede estimar que la proporción de personas con diabetes es ligeramente superior en la muestra, pero no de forma significativa. Hay que tener en cuenta que la muestra es también ligeramente más antigua.

Múltiples estudios han determinado que la pancreatitis crónica es un fuerte factor de riesgo para el cáncer de páncreas (6). Se estima que la prevalencia de pancreatitis crónico está en torno a 100 casos por cada 100000 personas, es decir, entorno al 0.1% (16). En la muestra este porcentaje se sitúa en el 7%, un valor significativamente muy superior a la prevalencia en la población general. Por tanto, podemos determinar que la pancreatitis crónica es un factor de riesgo muy potente a la hora de contraer cáncer de páncreas.

4.2.2. Antecedentes familiares

En la muestra hay 66 pacientes, un 35.6%, que tienen antecedentes familiares de cáncer. En la tabla 4 se muestra la distribución de los diferentes tipos de cáncer de los familiares de los pacientes. En la mayoría de los casos no se especifica el tipo concreto de cáncer. Solo hay tres excepciones. El cáncer de mama y melanoma son dos tipos de cáncer bastante comunes y el cáncer de páncreas es el objeto de estudio.

Antecedentes de cáncer familiar, n (%)	Cáncer de mama	6 (3.2)
	Cáncer	44 (23.8)
	Melanoma	2 (1.1)
	Ninguno	121 (65.4)
	Cáncer de páncreas	12 (6.5)

Tabla 4: Distribución de los antecedentes familiares de cáncer

Se estima que alrededor de un 39% de las personas sufrirán un cáncer a lo largo de sus vidas (17). Por lo tanto, la proporción de pacientes con antecedentes familiares de cáncer es coherente con la población general. También se estima que alrededor de un 5%-10% de los pacientes de cáncer de páncreas tienen un historial familiar de cáncer de páncreas (18). La proporción en la muestra es coherente con esta estimación.

4.2.3. Exposición a sustancias

En la muestra se especifican dos tipos de sustancias, el alcohol y el tabaco. Se analizan cada una de ellas de forma individual.

En la muestra se especifica la intensidad de consumo de alcohol por parte de los pacientes. Se tiene información sobre los hábitos de alcoholismo del 54.5% de los pacientes. Se destaca que hay un 15.7% que no consume alcohol. En la tabla 5 se detalla la distribución de los pacientes en las diferentes categorías de consumo de alcohol.

Los estudios publicados no han llegado a determinar completamente la relación que existe entre el consumo de alcohol y el cáncer de páncreas. Existe una evidencia fuerte de que el consumo moderado o intenso de alcohol aumenta el riesgo de padecer cáncer de páncreas, pero en muchos casos este riesgo está también asociado al tabaquismo y es complicado aislar el efecto de cada una de las sustancias (6,19).

Por otro lado, un excesivo consumo de alcohol aumenta el riesgo de padecer pancreatitis y este sí que tiene una clara relación con el aumento del riesgo de padecer cáncer de páncreas. Así que, como mínimo, de forma indirecta el alcohol aumenta el riesgo de padecer cáncer de páncreas.

Consumo de alcohol, n (%)	Abstemio	29 (15.7)
	Consumidor ocasional	18 (9.7)
	Consumidor social	15 (8.1)
	Consumidor	15 (8.1)
	Consumidor intensivo	22 (11.9)
	No reportado	86 (46.5)

Tabla 5: Distribución del consumo de alcohol de los pacientes

En la muestra se especifica el historial de tabaquismo de los pacientes. Se tiene información de 149 pacientes, el 80.5%. Entre ellos, el 37.3% nunca ha fumado y el 42.7% es fumador o exfumador. En la tabla 6 se detalla la distribución de los pacientes en las diferentes categorías de tabaquismo. Se analizará si afecta a la supervivencia.

Tabaquismo, n (%)	Exfumador por 15 años o menos	23 (12.4)
	Exfumador por más de 15 años	29 (15.7)
	Exfumador, duración indeterminada	8 (4.3)
	Fumador	20 (10.8)
	No fumador	69 (37.3)
	No reportado	36 (19.5)

Tabla 6: Distribución del estado de tabaquismo de los pacientes

La relación entre tabaquismo e incremento del riesgo de padecer cáncer de páncreas está ampliamente documentada. A más años fumando o mayor consumo diario mayor es el riesgo (6). Aunque no están completamente establecidos los motivos de esta correlación. Se cree que los compuestos carcinogénicos que contiene el tabaco inducen la inflamación y fibrosis generando un entorno que promueve la aparición de este cáncer (20).

4.3. Análisis de los datos diagnósticos

Primero de todo se hace una descripción general de los diferentes diagnósticos para, a continuación, realizar un análisis individual de cada uno de ellos. En la figura 2 se muestra la distribución del número pacientes según el diagnóstico. Si han recibido múltiples veces el mismo diagnóstico solo se contabiliza una vez. Se observa que gran parte de los pacientes acaban desarrollando metástasis o recurrencia. También se observa una pequeña fracción de los pacientes desarrollaron otro tumor previo o posterior sin relación directa con este.

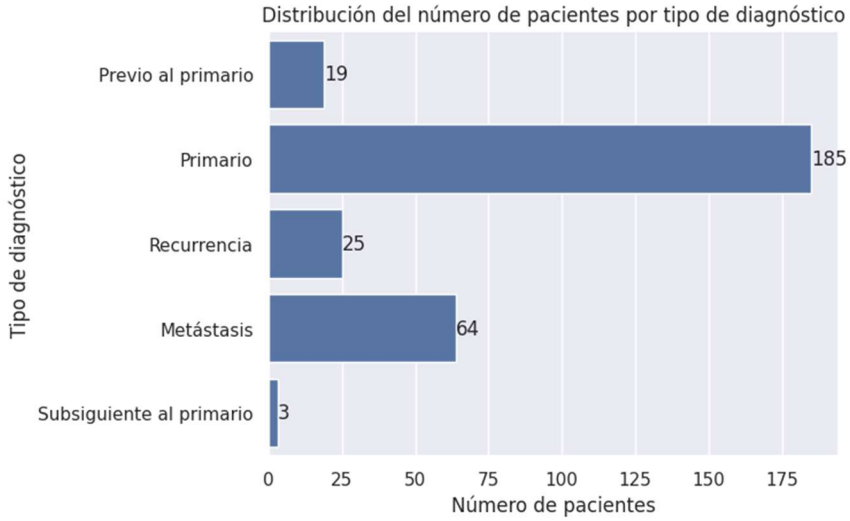


Figura 3: Distribución del número de pacientes según diagnostico

En la tabla 7 se observa la distribución del número de tratamientos que recibió cada paciente. Al analizar la tabla se puede deducir que hay pacientes que recibieron múltiples diagnósticos del mismo tipo. Por ejemplo, un paciente puede desarrollar múltiples metástasis.

Total de diagnósticos, n (%)	1	88 (47.6)
	2	77 (41.6)
	3	17 (9.2)
	4	2 (1.1)
	5	1 (0.5)

Tabla 7: Distribución del número total de diagnósticos de los pacientes

4.3.1. Diagnósticos previos

En la tabla 8 se observa la distribución de los canceres previos según el lugar donde se desarrolló. En la muestra no se observa ningún patrón concreto. Aquellos tipos más presentes también son aquellos que tienen mayor incidencia a nivel general. Se ha observado que personas diagnosticadas con determinados tipos de cáncer tienen mayor riesgo de padecer posteriormente cáncer de páncreas (21). Se ha determinado que puede deberse a predisposiciones genéticas o a factores de riesgo compartidos. Por otro lado, se ha observado que este hecho no comporta un peor pronóstico para estos pacientes (22). Debido a la mayor vigilancia y seguimiento que tienen estos pacientes suelen ser diagnosticados en estadios más tempranos contribuyendo a no tener un peor pronóstico.

Lugar cáncer previo, n (%)	Vejiga	1 (4.3)
	Mama	4 (17.4)
	Cuello uterino	1 (4.3)
	Cabeza, cara o cuello	2 (8.7)
	Riñón	1 (4.3)
	Extremidades inferiores	1 (4.3)
	Hipófisis	1 (4.3)
	Próstata	3 (13.0)
	Piel	4 (17.4)
	Tiroides	1 (4.3)
	Extremidades superiores	3 (13.0)
	Vagina	1 (4.3)

Tabla 8: Distribución del lugar del cáncer previo

4.3.2. Diagnósticos primarios

En la tabla 8 se puede observar la distribución de los diagnósticos primarios según el sistema de estadificación de la AJCC en la séptima y octava versión siguiendo el sistema TNM. Los principales cambios entre las dos versiones fueron en las categorías T y N. La categoría T paso a dividirse estrictamente en función del tamaño del tumor, menos para T4 que se asigna a tumores irresecables. En la categoría N se añadió una subdivisión adicional. En múltiples estudios se ha comprobado que esta nueva clasificación era más reproducible y permitía una mejor estratificación del pronóstico (23).

Efectivamente, se puede observar en la tabla que los principales cambios fueron entre las subcategorías T1 y T2, y entre las categorías N1 y N2. El primero se debe al cambio de criterio en la categoría T y el segundo se debe al incremento de subdivisiones de la categoría N. En el apartado de análisis de la supervivencia se explorará si esta nueva clasificación tiene mayor capacidad pronostica en la muestra de pacientes estudiada.

Si se analiza la distribución de los pacientes en los diferentes estadios se observa que la población de la muestra fue diagnosticada en un estadio mucho menos avanzado que en la población general. Se estima que alrededor del 50% de los casos son diagnosticados en estadio IV y otro 25% en estadio III (24). En la muestra es todo lo contrario, casi 3 de cada 4 pacientes son diagnosticados en estadio I o II. En la séptima versión, la original, esta proporción aún era mayor.

		7ª versión	8ª versión
Estadio patológico, n (%)	Estadio I	21 (11.4)	34 (18.4)
	Estadio II	152 (82.2)	89 (48.1)
	Estadio III	4 (2.2)	57 (30.8)
	Estadio IV	5 (2.7)	5 (2.7)
	No reportado	3 (1.6)	0 (0)
T, n (%)	T1	7 (3.8)	12 (6.5)
	T2	24 (13.0)	101 (54.6)
	T3	148 (80.0)	68 (36.8)
	T4	4 (2.2)	4 (2.2)
	TX	2 (1)	0 (0)
N, n (%)	N0	50 (27.0)	50 (27.0)
	N1	126 (68.1)	77 (41.6)
	N1b	4 (2.2)	0 (0)
	N2	0 (0)	55 (29.7)
	NX	5 (2.7)	3 (1.6)
M, n (%)	M0	85 (45.9)	85 (45.9)
	M1	5 (2.7)	5 (2.7)
	MX	95 (51.4)	95 (51.4)

Tabla 9: Distribución de la estadificación de los pacientes

En la figura 3, el diagrama de Sankey, se pueden observar los flujos de pacientes entre los diferentes estadios de las 2 versiones del sistema de clasificación patológico. En la séptima versión la gran mayoría de pacientes pertenecen al estadio II. Una pequeña fracción pertenecen al estadio I y casi no hay representación de los estadios III y IV. Además, hay 3 pacientes sin estadio asignado.

Tras la reclasificación se observa un mejor balance entre los diferentes estadios. El estadio II ya no es tan mayoritario y los estadios I y III tienen una mayor representación. La inclusión de los datos histopatológicos permitió asignar un estadio a los 3 diagnósticos sin clasificar. El estadio IV no vario, ya que, las dos versiones mantienen los mismos criterios para este estadio. Se analizará como estos cambios afectan a las curvas de supervivencia.

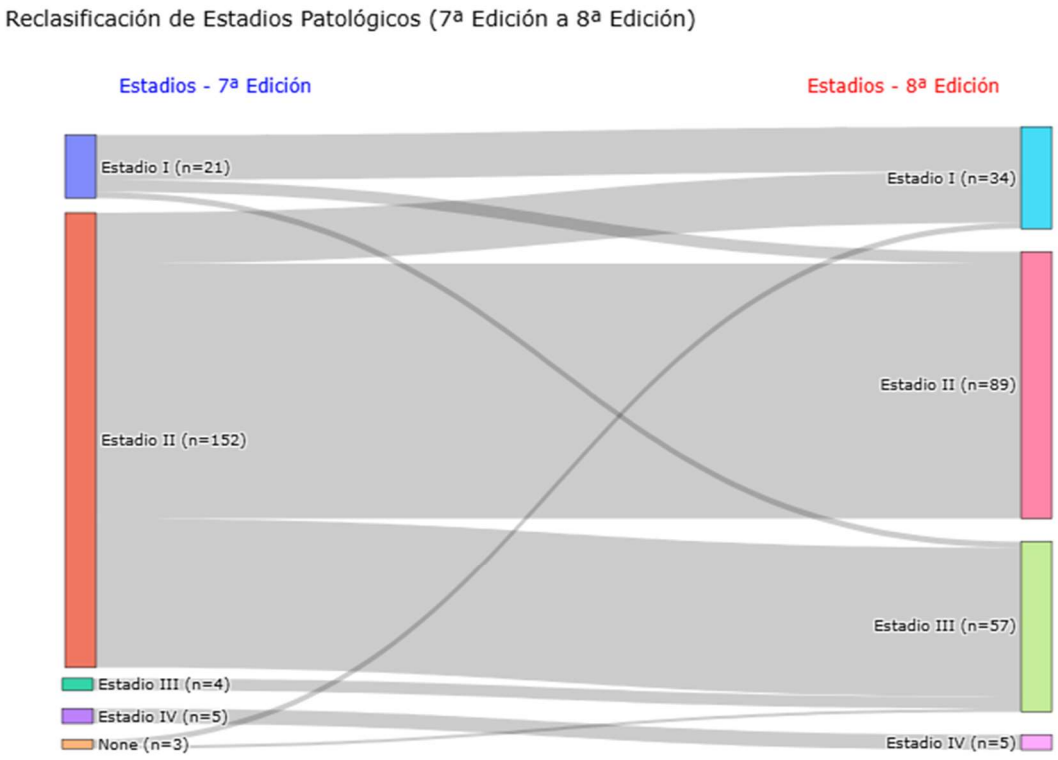


Figura 4: Diagrama de Sankey de la reclasificación de los estadios patológicos

En la tabla 10 se observa la distribución del margen de resección y del grado del tumor de los pacientes. Los márgenes de resección hacen referencia a los bordes del tejido extirpado tras la cirugía, donde R0 indica que no se encuentran células cancerosas. Se ha comprobado que los pacientes con márgenes positivos, R1 o R2, tienen peor pronóstico (25).

El grado del tumor describe la apariencia del tumor analizado bajo un microscopio. Se trata de otra clasificación alternativa al sistema TNM. El grado 1 indica que las células tumorales tienen una apariencia similar a las normales y el grado 3 o 4 que tienen una apariencia muy anormal. Suele estar relacionado con la velocidad de crecimiento del tumor. Se ha visto que está relacionado con un peor pronóstico independientemente del estadio (26).

Margen de resección, n (%)	R0	111 (60.0)
	R1	53 (28.6)
	R2	5 (2.7)
	RX	16 (8.7)
Grado del tumor, n (%)	G1	32 (17.3)
	G2	97 (52.4)
	G3	51 (27.6)
	G4	2 (1.1)
	GX	3 (1.6)

Tabla 10: Distribución del margen de resección y del grado del tumor de los pacientes

En la figura 4 se muestra la distribución de los diferentes tipos de cáncer de páncreas. El cáncer de páncreas se divide en dos grupos. Los canceres exocrinos afectan a las células que producen las enzimas digestivas y representan el 95% de los casos. Los canceres endocrinos afectan a las células que producen las hormonas y representan el 5% (27).

Dentro de los canceres exocrinos destaca el adenocarcinoma ductal que representa sobre el 90% del total. Este se desarrolla en los conductos del páncreas y se le considera agresivo debido a su tardío diagnóstico, típicamente en fase metastática (28). El adenocarcinoma mucinoso es un subtipo de tumor que se caracteriza por la producción de mucina. Representa el 2.6% de casos y suele tener mejor pronóstico con unas tasas de supervivencia a 5 años de alrededor del 65% (29). El carcinoma indiferenciado es un tipo muy raro de tumor donde las células cancerosas no se parecen a las células del tejido normal circundante, dificultando su diagnóstico. Representa el 0.1% de los casos y tiene una tasa de supervivencia a 5 años del 12% (29)

Dentro de los canceres endocrinos destaca el carcinoma neuroendocrino que afecta a las células neuroendocrinas del páncreas y representa menos del 2%. Suelen tener un mejor pronóstico con una tasa de supervivencia a 5 años del 48% (30).

El adenocarcinoma con subtipos mixtos es un tipo de tumor que afecta a diversos tipos de células y es una combinación de otros tipos. Puede combinar células endocrinas y exocrinas y se trata de un tipo muy raro (29).

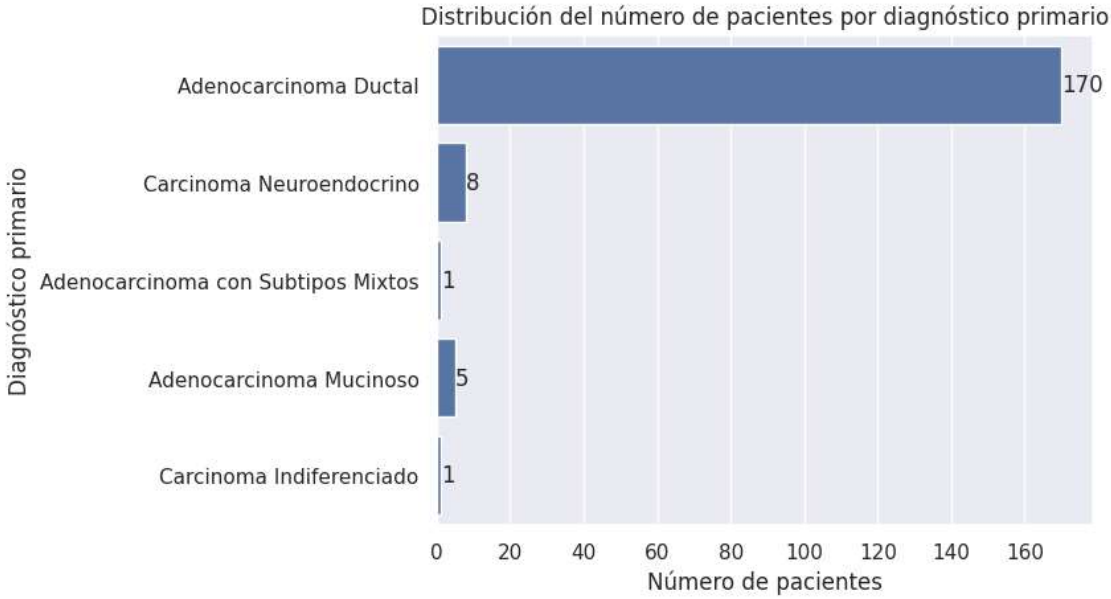


Figura 5: Distribución de los pacientes según el tipo de diagnóstico primario

El páncreas se divide en tres partes. La cabeza es la parte más ancha y próxima al intestino delgado, el cuerpo es la parte central del páncreas y la cola es la parte final y más estrecha del páncreas (31). En la figura 5 se puede ver la distribución de los pacientes según el lugar de participación del tumor. Se han establecido 4 categorías. La cabeza es la región predominante con un 80% de los casos. El cuerpo y la cola representan el 20% de los casos y tienen peor pronóstico debido a que presentan menos síntomas retrasando su detección (32). La categoría “solapado o difuso” representa aquellos tumores que debido a su tamaño abarcan varias regiones del páncreas. La categoría “no especificado” se reserva para aquellos tumores donde no se ha informado de su localización.

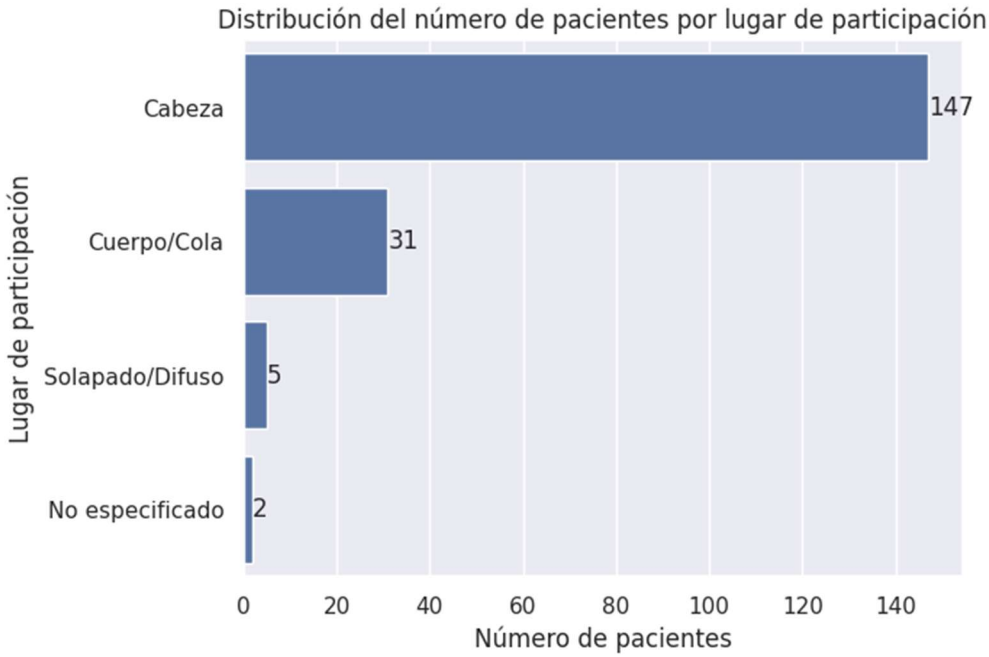


Figura 6: Distribución de los pacientes según el lugar de participación del tumor

4.3.3. Diagnósticos posteriores

		Metástasis	Recurrencia
Lugar de cáncer posterior, n (%)	Abdomen	1 (1.4)	2 (8.0)
	Glándula suprarrenal	1 (1.4)	0 (0.0)
	Hueso	1 (1.4)	0 (0.0)
	Tejidos blandos	1 (1.4)	0 (0.0)
	Hígado	41 (58.6)	3 (12.0)
	Pulmón	15 (21.4)	0 (0.0)
	Nódulo linfático	3 (4.3)	1 (4.0)
	Páncreas	0 (0.0)	1 (4.0)
	Peritoneo	5 (7.1)	8 (32.0)
	Retroperitoneo	1 (1.4)	1 (4.0)
	No reportado	1 (1.4)	9 (36.0)

Tabla 11: Distribución de las metástasis o recurrencias según localización

Se estima que más del 50% de los casos presentan metástasis en el momento del diagnóstico (33). En la muestra alrededor del 50% de los pacientes sufren metástasis o recurrencia local. Los lugares más típicos de metástasis en cáncer de páncreas son el hígado, el pulmón y el peritoneo (34). Los datos de la muestra concuerdan con las observaciones previas.

4.4. Análisis de los datos de tratamientos

Primero de todo se hace una descripción general de los diferentes tratamientos para, a continuación, realizar un análisis individual de algunos de ellos. En la figura 6 se observa el número total de tratamientos por paciente. Se observa un pico en pacientes entre 2 y 3 tratamientos indicando pacientes con diagnóstico más reciente y otro pico entre 5 y 9 tratamientos indicando pacientes con una mayor trayectoria. Se observa algunos pacientes con un largo historial de tratamientos, llegando hasta los 16.

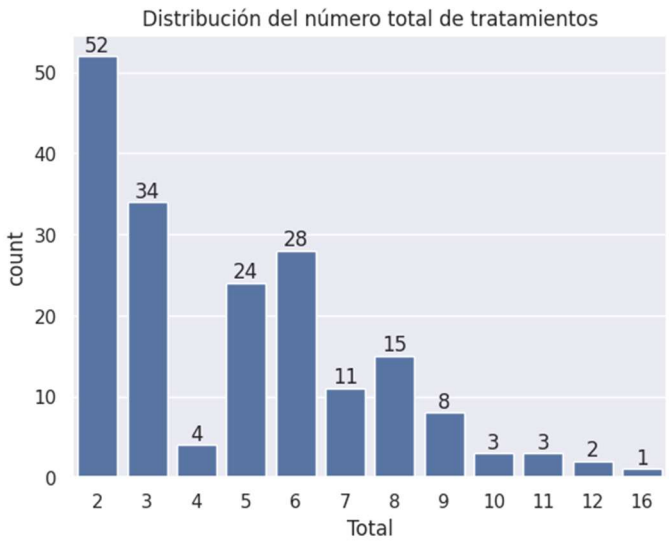


Figura 7: Distribución del número total de tratamientos por paciente

En la figura 7 se observa la distribución del número de pacientes que han recibido al menos una vez ese tratamiento. Se destaca que casi todos los pacientes han recibido el procedimiento de Whipple, cirugía que extirpa la cabeza del páncreas y algunos órganos adyacentes. También, gran parte de los pacientes han recibido otras cirugías. Se observa el mismo número de tratamientos farmacológicos y de radioterapia probablemente utilizados conjuntamente tras una cirugía.

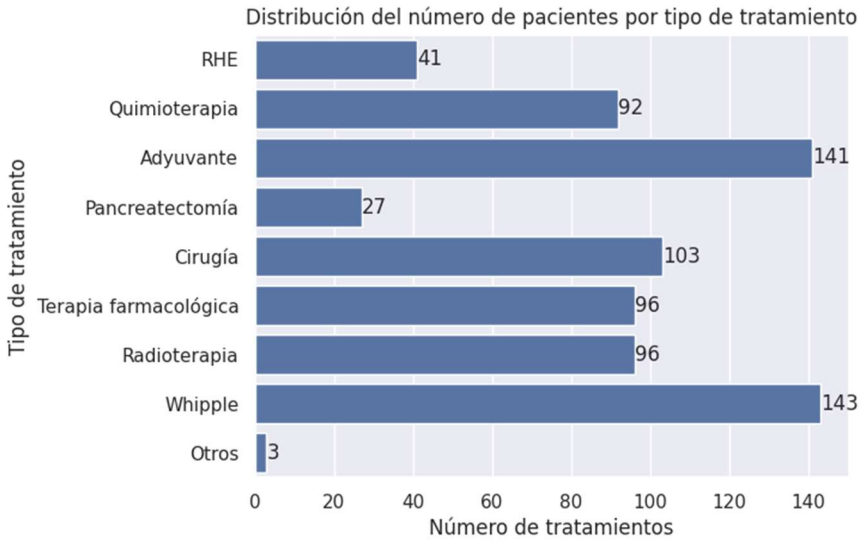


Figura 8: Distribución del número de pacientes por tratamiento

En la tabla 12 se observa la distribución de los tratamientos de algunas categorías. La opción de pancreatectomía preferente es la distal, en la que se extirpa el cuerpo y la cola, y se reserva la extirpación total del páncreas para los casos más severos donde el cáncer está más difuso.

La opción mayoritaria de tratamiento adyuvante es la radioterapia, seguida de la combinación de esta con tratamiento farmacológico, quimioterapia. Actualmente, el beneficio de la quimioterapia está mucho más demostrado que el de la radioterapia como tratamientos adyuvantes (35).

Finalmente, se reportan tres casos individuales de tratamiento hormonal, probablemente como tratamiento paliativo, biopsia excisional, donde se extrae la totalidad del tumor, y vacuna antineoplásica que se trata de un tratamiento experimental.

Adyuvante, n (%)	Farmacológica	1 (0.7)
	Farmacológica y radioterapia	63 (44.7)
	Radioterapia	77 (54.6)
Pancreatectomía, n (%)	Distal	24 (88.9)
	Total	3 (11.1)
Otros, n (%)	Vacuna antineoplásica - Inmunoterapia	1 (33.3)
	Biopsia, Excisional	1 (33.3)
	Megestrol Acetate & Dexamethasone - Hormone Therapy	1 (33.3)

Tabla 12: Distribución de los tratamientos de las categorías adyuvante, pancreatectomía y otros

4.4.1. Tratamientos de RHE

En la figura 8 se muestra la distribución de los resultados de los tratamientos de RHE en función de la dosis total del tratamiento. Se observa un patrón claro, para obtener respuestas completas es necesario que la dosis total supere los 5000 cGy, y dosis inferiores siempre producen progresión. Se plantea la hipótesis nula donde la dosis total es igual en las dos categorías y la hipótesis alternativa donde la dosis total de las respuestas completas es superior. Dada la no normalidad de los datos se aplica la prueba de Wilcoxon-Mann-Whitney y se obtiene un valor de p igual a 0.00336, inferior al valor de significancia de 0.05. Por tanto, se puede rechazar la hipótesis nula y concluir que la de la dosis total de los tratamientos que obtienen respuestas completas es superior a la dosis total de los tratamientos que obtienen progresiones.

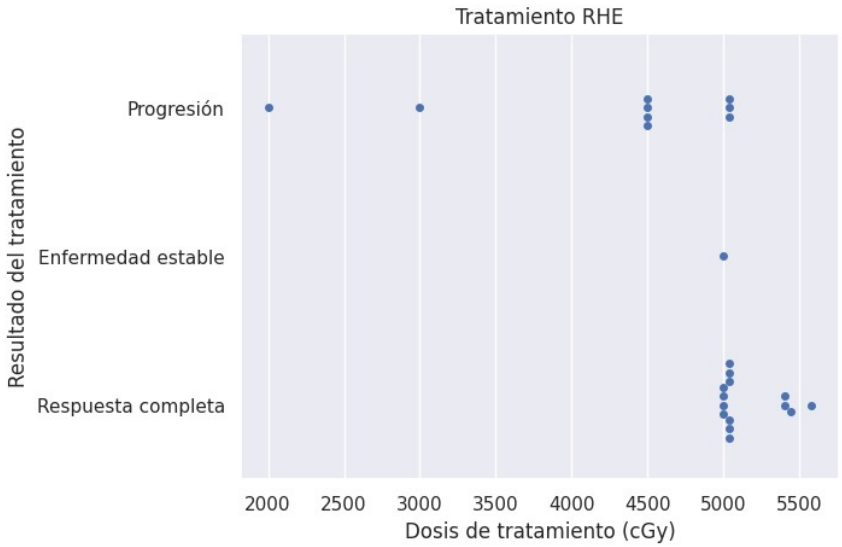


Figura 9: Grafico de enjambre de los resultados del tratamiento en función de la dosis

4.4.2. Tratamientos de quimioterapia

La figura 9 muestra la duración del tratamiento en función del tipo de tratamiento. Esta medida se puede relacionar indirectamente con el beneficio que aporta la terapia, ya que, se suelen administrar hasta que haya progresión. Se observa que los tratamientos combinados tienen mayor duración del tratamiento que los de monoterapia. Se comprueba si esta diferencia es estadísticamente significativa y se obtiene un valor de p de 0.093. Por tanto, no se puede descartar la hipótesis nula. Si observamos la respuesta de ambos tipos de tratamiento se obtienen valores similares, RC del 32.8% y del 32.1% y progresiones del 55.2% y del 42.9%, respectivamente para monoterapia y combinación.

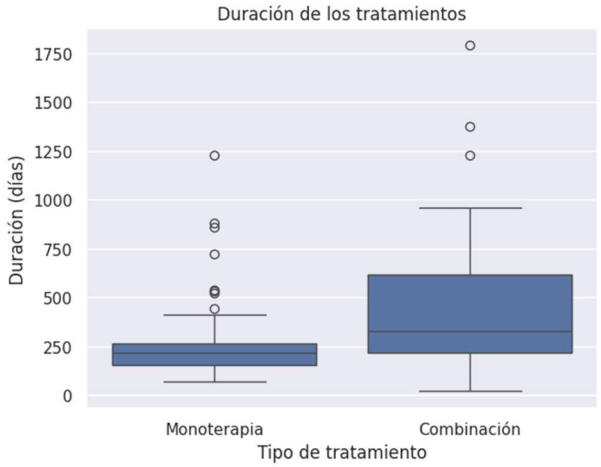


Figura 10: Duración del tratamiento en función del tipo de tratamiento

A partir de la información mostrada en la figura 10 se puede describir el panorama de tratamientos de quimioterapia entre los años 2007-2013 y analizar su evolución. Primero de todo se observa que Gemcitabina es el tratamiento principal. Este fue aprobado en 1997 y hasta 2007 fue la única opción disponible y que, posteriormente, siguió siendo muy utilizada para pacientes frágiles (36). En 2011 se aprobó un nuevo tratamiento, FOLFIRINOX, que mejoraba la supervivencia comparada con Gemcitabina (37). En 2013 se aprobó Gemcitabina con Nab-paclitaxel que volvía a mejorar la supervivencia y ha estado vigente casi hasta la actualidad cuando ha sido reemplazado por NALIRIFOX (38,39). También se observa la aparición de otros tratamientos como intento de mejorar los resultados que se obtenían y como segunda línea de tratamiento tras progresión con Gemcitabina u otros tratamientos (40,41).

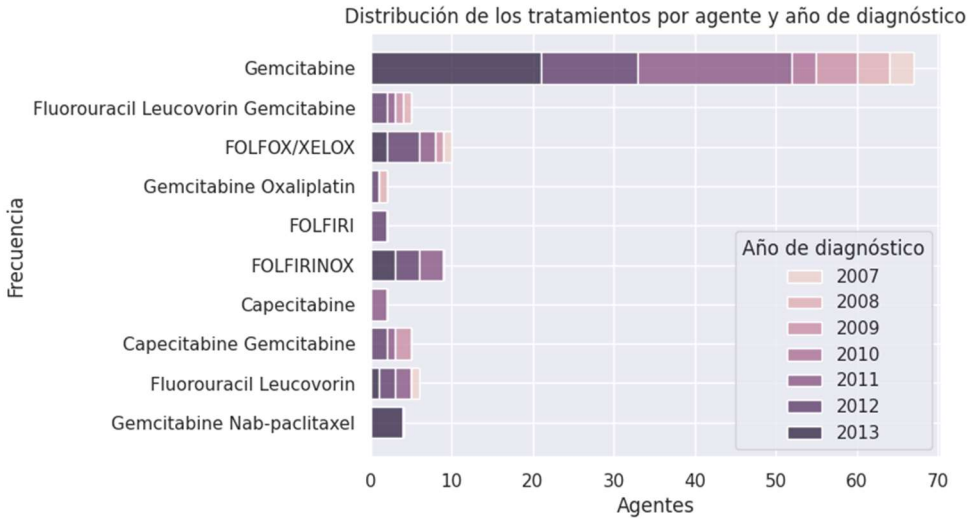


Figura 11: Distribución de los tratamientos por agentes y año de diagnóstico

4.5. Análisis mutaciones genéticas

En la figura 11 se observa la prevalencia de los 20 genes más mutados de la muestra. Se destacan los 4 principales oncogenes o genes supresores del tumor relacionados con el cáncer de páncreas. El gen *KRAS* se encuentra mutado en el 80% de las muestras, el *TP53* en el 65%, el *CDKN2A* por encima del 20% y el *SMAD4* en alrededor del 20%. Estas mutaciones están presentes con una menor prevalencia de lo descrito en la literatura (42).

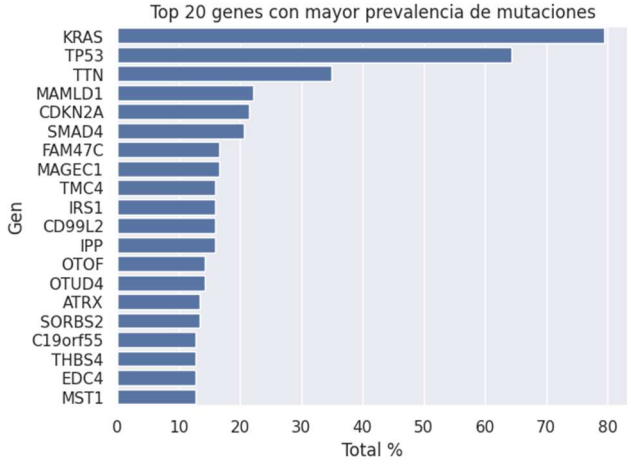


Figura 12: Principales genes con mayor prevalencia de mutaciones

En la figura 12 se muestran la prevalencia de co-mutaciones de los 4 principales genes implicados en el cáncer de páncreas. Las barras azules representan el porcentaje de co-mutaciones, mientras las barras rojas representan el porcentaje de mutaciones en total, permitiendo detectar fácilmente las principales discrepancias. En la mayoría de los casos hay mayor prevalencia de co-mutaciones en los genes *KRAS* y *TP53*. En los casos de co-mutaciones de *CDKN2A* y *SMAD4* hay mayor prevalencia de mutaciones en el gen WRN implicado en los procesos de reparación del ADN. En ambos casos, también, se muestra un perfil genómico distinto al de la muestra general, habiendo mayor prevalencia de mutaciones en genes que no se observan en la población general.

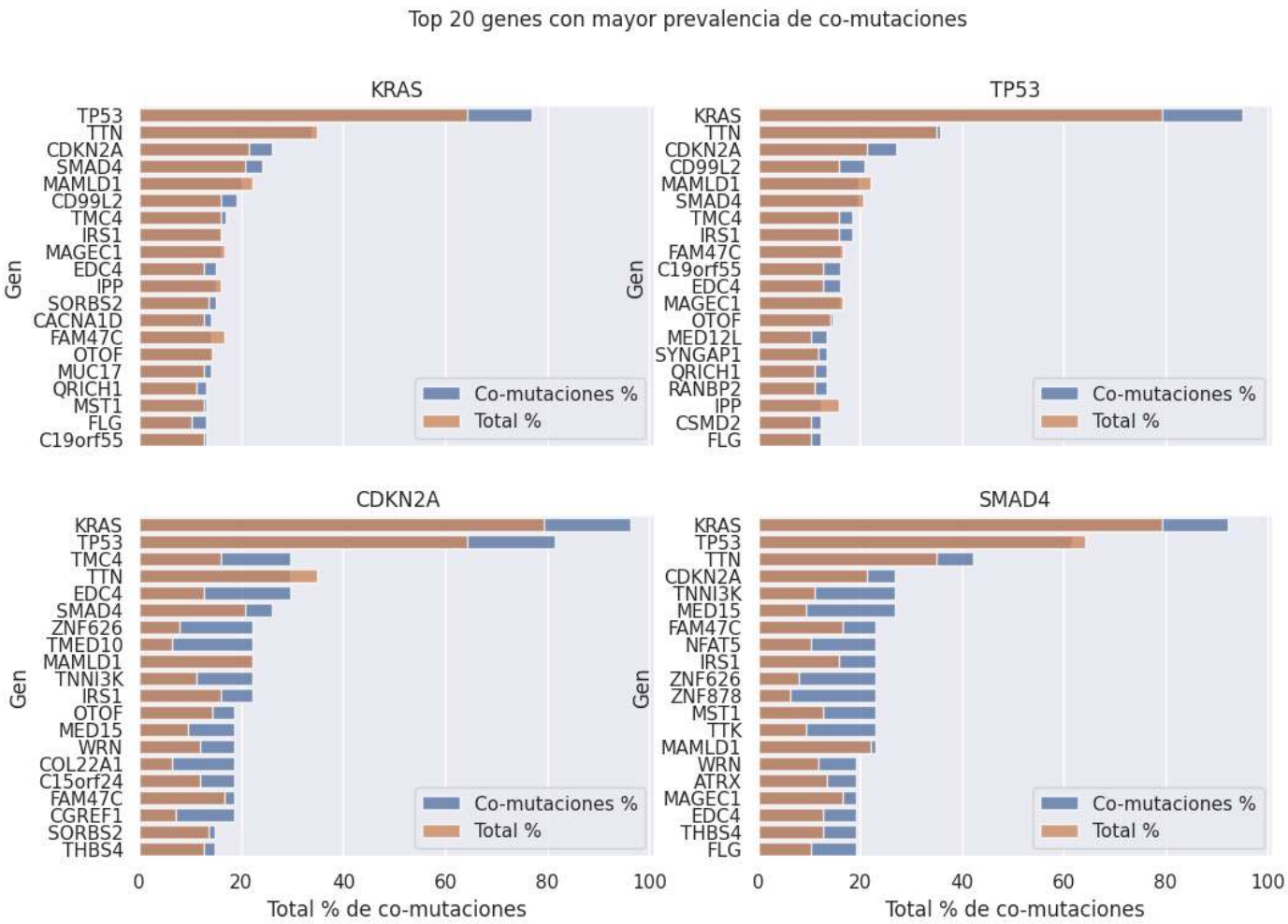


Figura 13: Principales co-mutaciones entre los genes con mayor relevancia

4.6. Análisis de supervivencia

Se empieza analizando la supervivencia según el diagnostico primario. En la figura 13 se muestra una estimación de la supervivencia según el diagnostico primario mediante un diagrama de cajas. Se observan diferencias significativas según el diagnostico, como previamente se había descrito. El carcinoma neuroendocrino tiene una supervivencia muy superior al resto y los tipos minoritarios tienen una supervivencia similar o ligeramente inferior al adenocarcinoma ductal. Debido a estos resultados, en los próximos análisis se realizarán filtrando solo los casos de adenocarcinoma ductal.

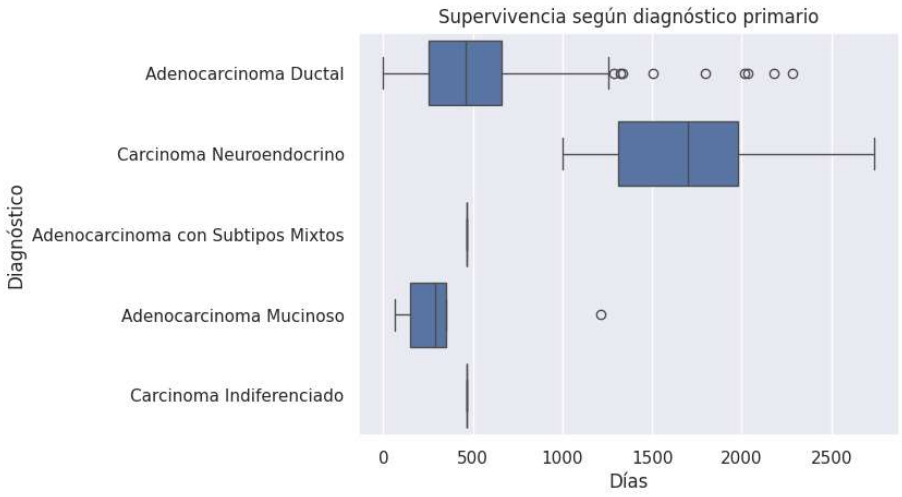


Figura 14: Supervivencia según el diagnóstico primario

En la figura 14 se muestra la curva de supervivencia global para los pacientes con adenocarcinoma ductal y mutaciones reportadas. Se observa que la supervivencia media es de 568 días, muy encima de lo común, pero en consonancia con la mayor proporción de pacientes en estadios tempranos.

A continuación, se analizan las diferentes variables demográficas. En la figura 15 se representa la supervivencia en función de la edad. Se observa una cierta correlación negativa entre las dos variables. Si se calcula el coeficiente de correlación de Spearman se obtiene un valor de -0.13, indicando una leve correlación negativa. Se utilizará la edad en el modelo de predicción.

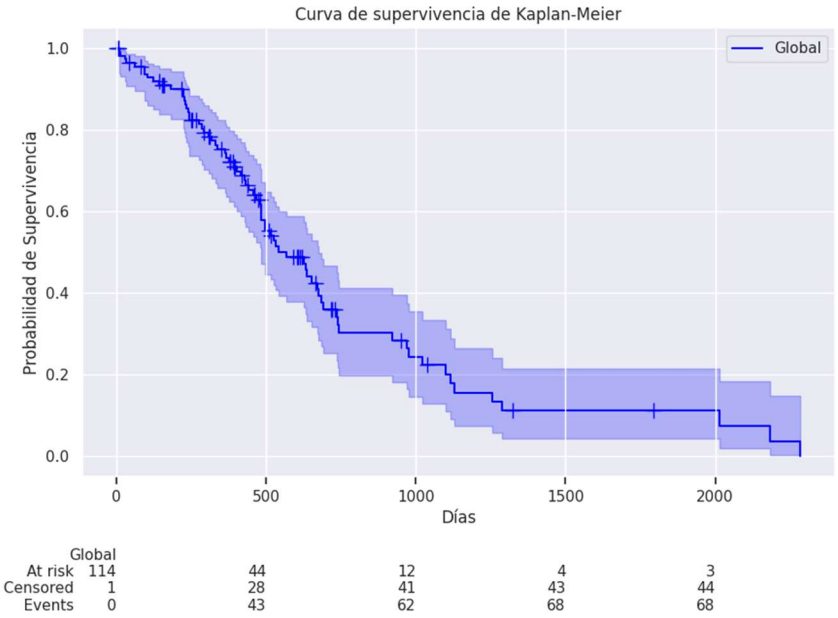


Figura 15: Curva de supervivencia global

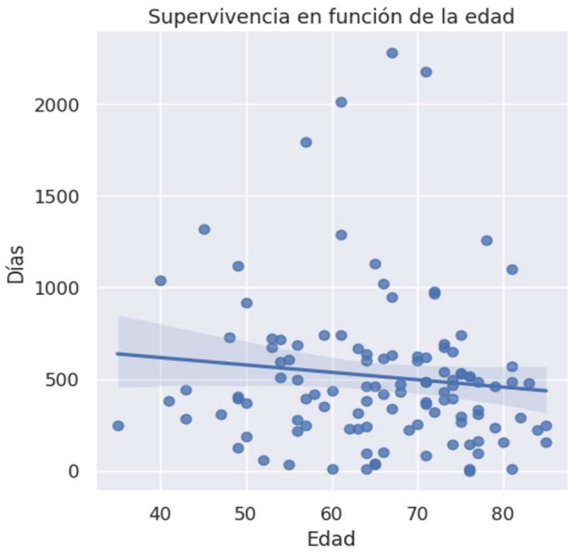


Figura 16: Supervivencia en función de la edad

En la figura 16 se muestra la supervivencia para múltiples variables demográficas. No se puede establecer grandes conclusiones, ya que, algunas categorías tienen una muy baja representación, pero se pueden destacar algunas observaciones. En general, no se aprecian grandes diferencias entre las diferentes categorías. Se observan una menor mortalidad para los hispanos y asiáticos, y una mayor mortalidad para los negros o afroamericanos. Estos patrones son similares a los descritos en la literatura (43).

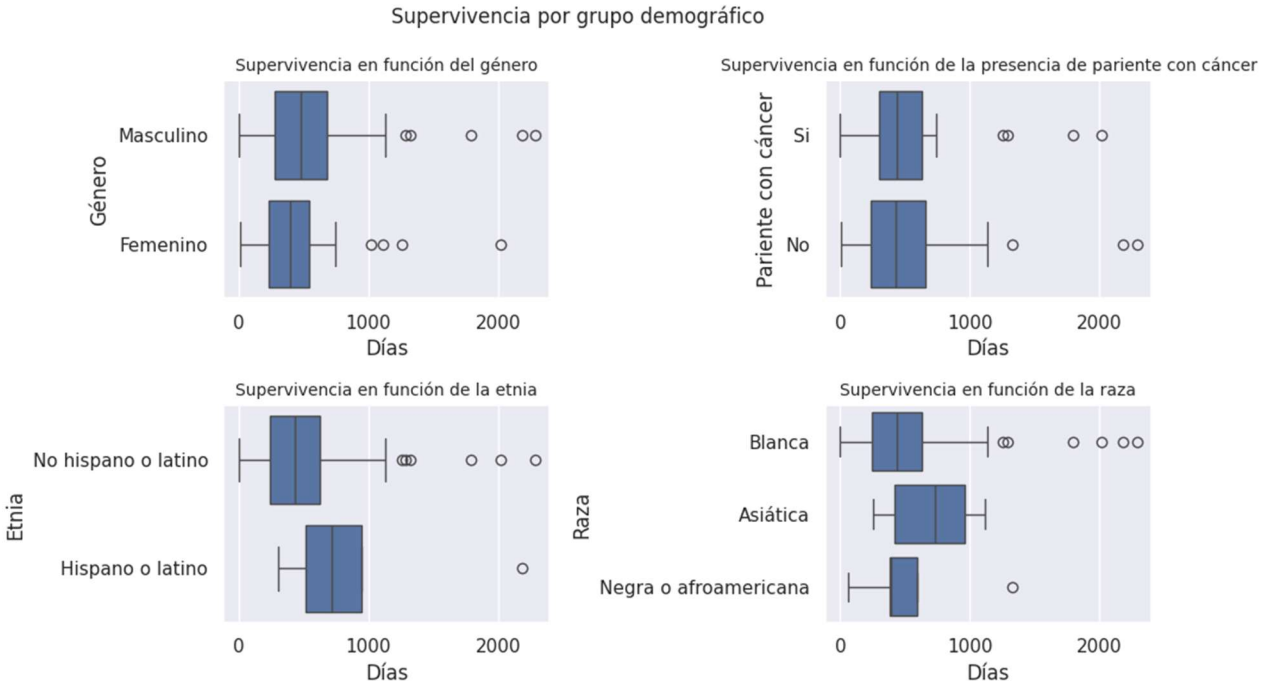


Figura 17: Análisis de supervivencia según variables demográficas

En la figura 17 se muestra la supervivencia de los pacientes en función de la exposición a sustancias. No se aprecian diferencias en ninguno de los dos casos.

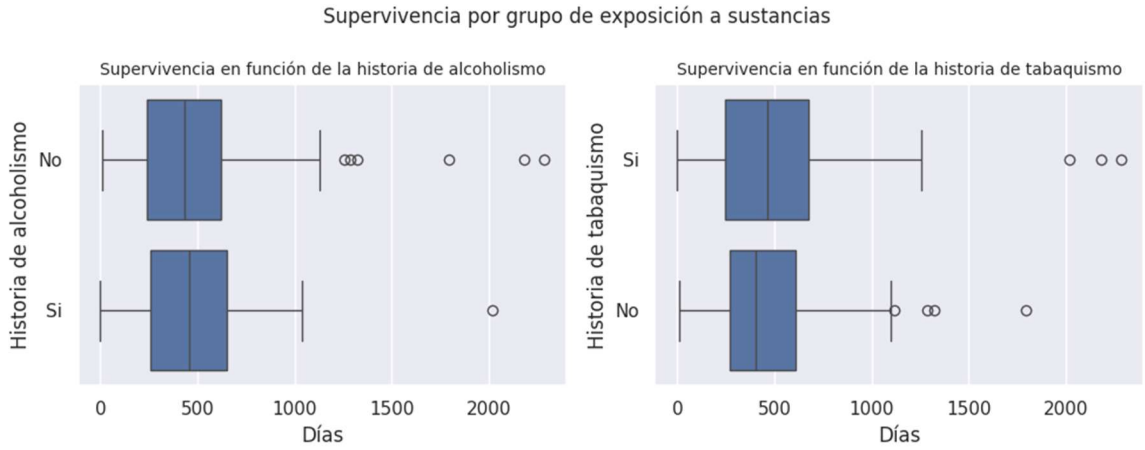


Figura 18: Supervivencia en función de la exposición a sustancias

En la figura 18 se muestran las curvas de supervivencia en función del estadio patológico para las dos versiones de clasificación. En la 8ª versión se observa una mejor capacidad predictiva que en la 7ª. Los pacientes de estadio I tienen una supervivencia muy superior a los de estadios más avanzados, unos 1000 días frente a 500. Se utilizará este criterio en el modelo de regresión.

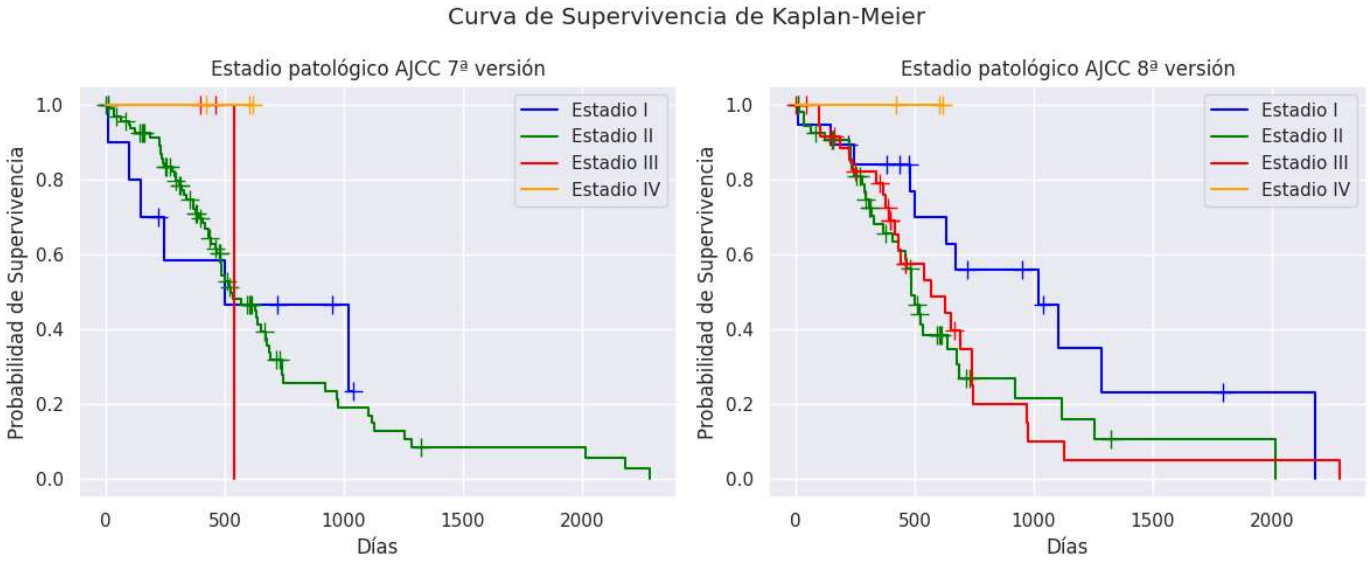


Figura 19: Curvas de supervivencia en función del estadio patológico

En la figura 19 se muestra la curva de supervivencia según el estadio patológico con dos grupos. Se aprecia una separación de las curvas significativa y constante. Los tiempos medios de supervivencia son de 1021 días y 525 días. Si se implementa una regresión de Cox se obtiene una reducción del riesgo de muerte del 49% con p=0.05, estadísticamente significativo.

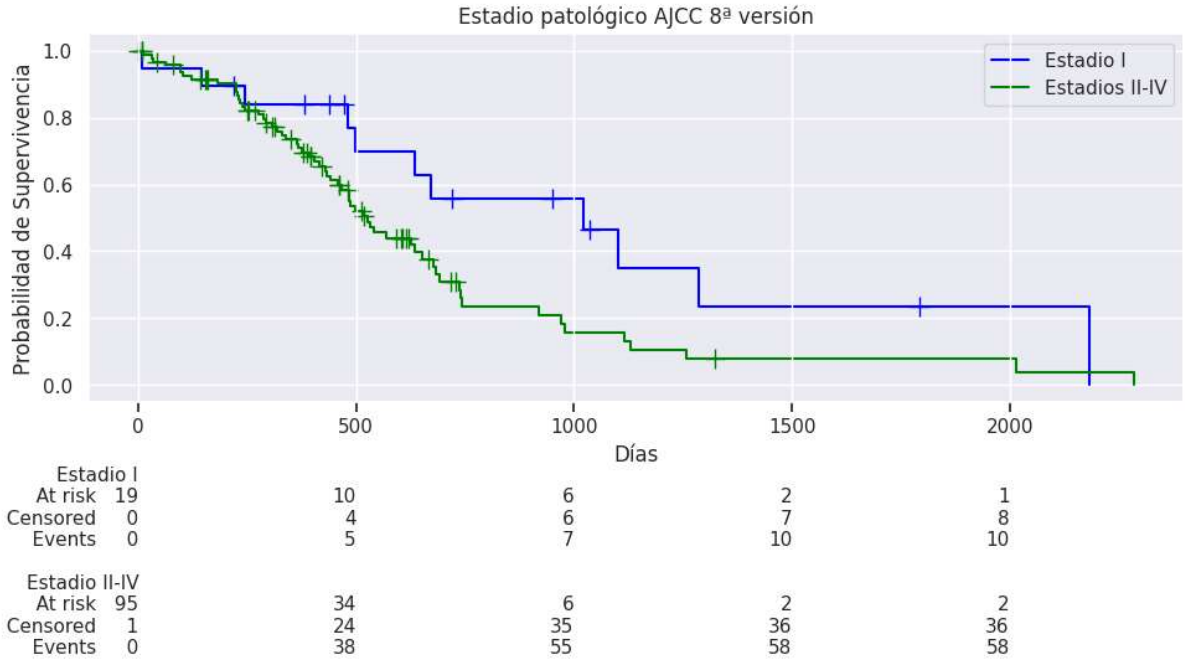


Figura 20: Curva de supervivencia en función del estadio patológico (2)

En la figura 20 se muestra la curva de supervivencia según el grado del tumor. Los tiempos medios de supervivencia son de 627 días y 541 días. Si se implementa una regresión de Cox se obtiene una reducción del riesgo de muerte del 24% con p=0.52, no significativo. Se utilizará esta variable en el modelo de regresión.

En la figura 21 se muestra la curva de supervivencia en función del margen de resección. Los tiempos medios de supervivencia son de 635 días y 484 días. Si se implementa una regresión de Cox se obtiene una reducción del riesgo de muerte del 35% con p=0.1, no significativo. Se utilizará esta variable en el modelo de regresión.

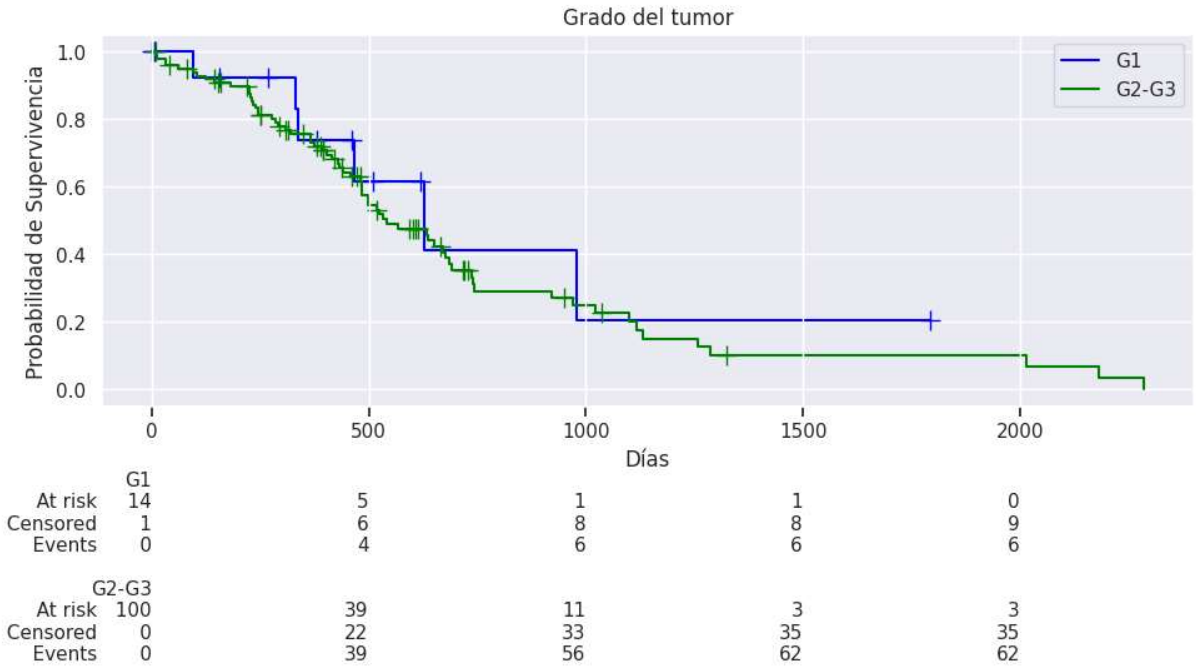


Figura 21: Curva de supervivencia en función del grado del tumor

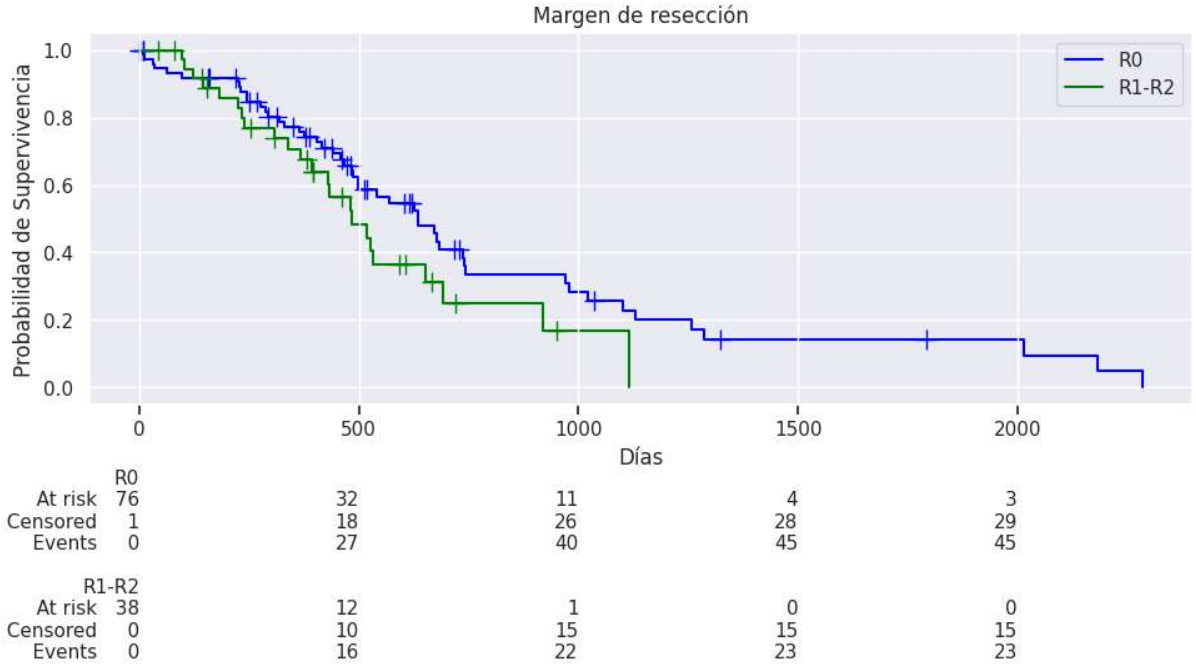


Figura 22: Curva de supervivencia en función del margen de resección

En la figura 22 se muestra la curva de supervivencia según el lugar de participación. Los tiempos medios de supervivencia son de 525 días y 676 días. Si se implementa una regresión de Cox se obtiene un aumento del riesgo de muerte del 61% con p=0.14, no significativo. Estos resultados contradicen lo descrito en la literatura, dónde los tumores localizados en la cabeza del páncreas tienen un mejor pronóstico. Se utilizará esta variable en el modelo de regresión.

En la figura 23 se muestra la curva de supervivencia en función de la presencia de metástasis en el hígado. Los tiempos medios de supervivencia son de 627 días y 541 días. Si se implementa una regresión de Cox se obtiene un aumento del riesgo de muerte del 11% con p=0.72, no significativo. En este caso, no se puede tener en cuenta este resultado, ya que, las curvas se cruzan y no mantienen las proporciones Se utilizará esta variable en el modelo de regresión.

En la figura 24 se muestra la curva de supervivencia según la presencia de mutaciones en el gen *KRAS*. Los tiempos medios de supervivencia son de 627 días y 568 días. Si se implementa una regresión de Cox se obtiene un aumento del riesgo de muerte del 3% con p=0.92, no significativo. Se utilizará esta variable en el modelo de regresión.

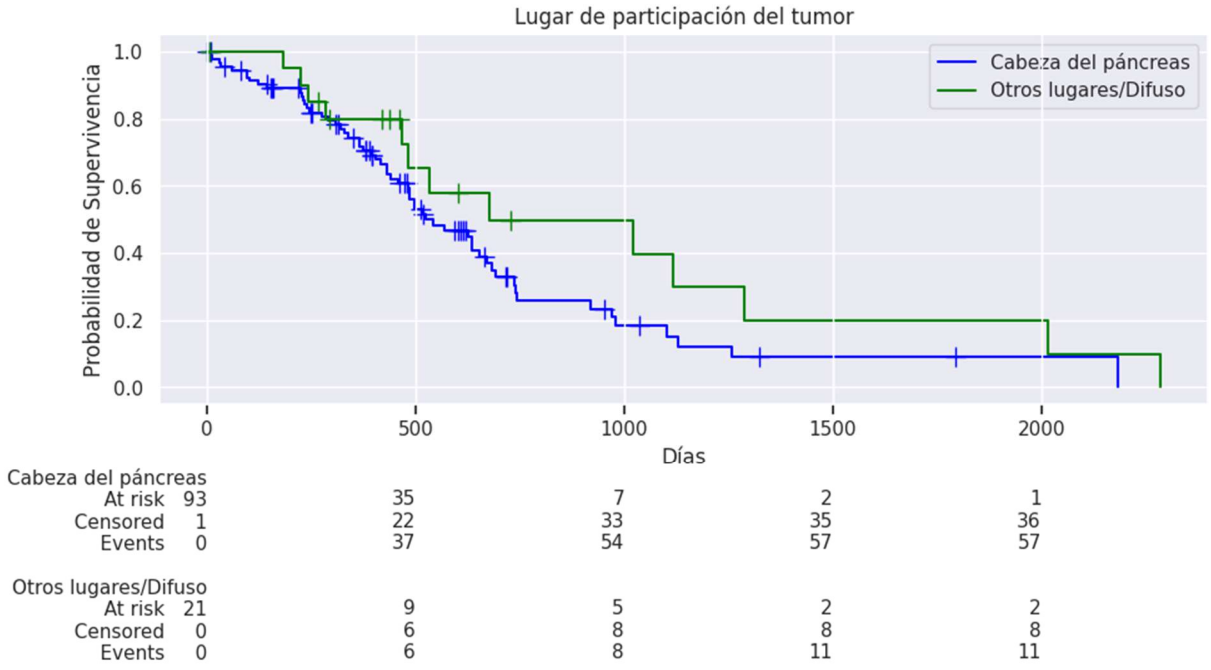


Figura 23: Curva de supervivencia en función del lugar de participación

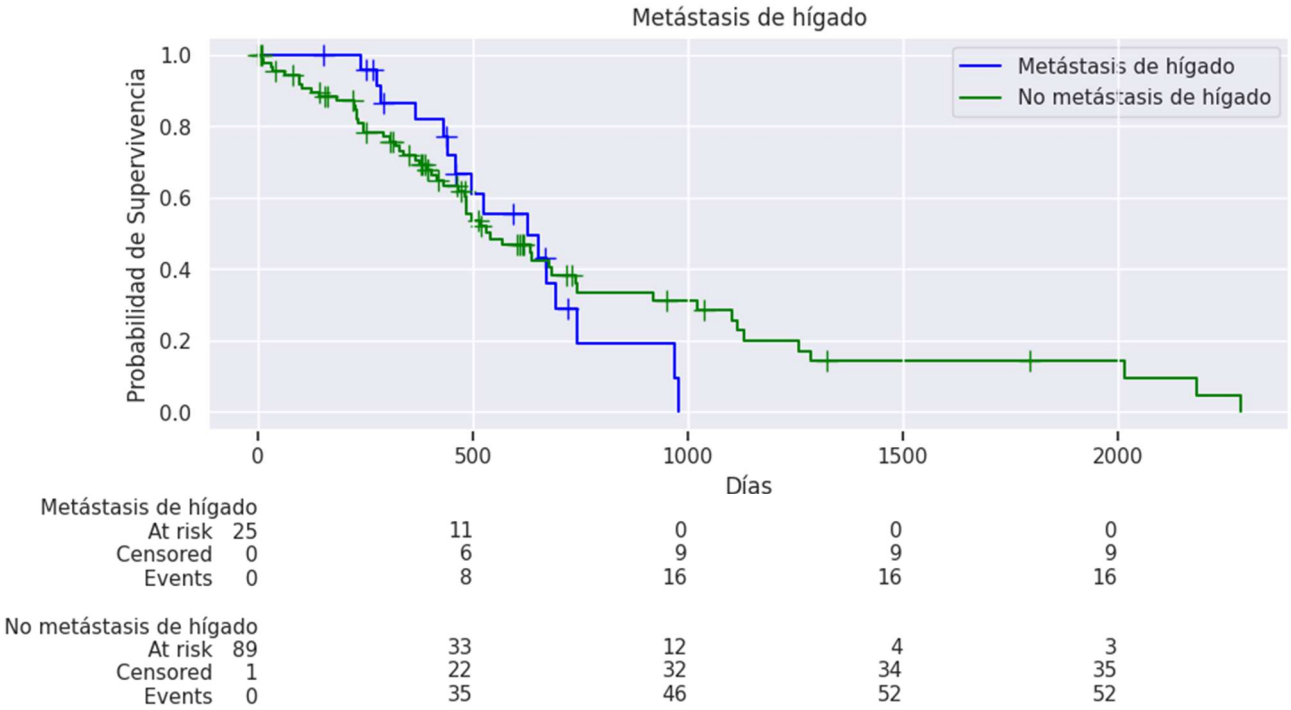


Figura 24: Curva de supervivencia en función de la presencia de metástasis en el hígado

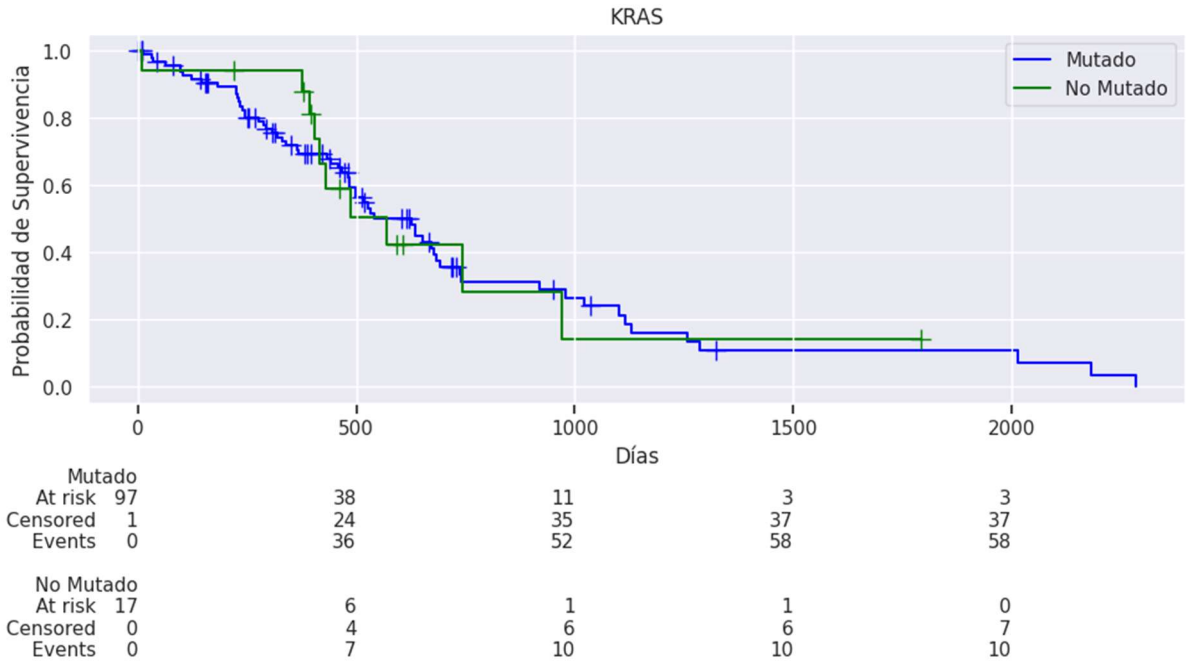


Figura 25: Curva de supervivencia en función de la mutación *KRAS*

En la figura 25 se muestra la curva de supervivencia según la presencia de mutaciones en el gen *TP53*. Los tiempos medios de supervivencia son de 525 días y 969 días. Si se implementa una regresión de Cox se obtiene un aumento del riesgo de muerte del 90% con p=0.03, significativo. Por lo tanto, este marcador tiene una capacidad pronóstica muy importante Se utilizará esta variable en el modelo de regresión.

En la figura 26 se muestra la curva de supervivencia según la presencia de mutaciones en el gen *CDKN2A*. Los tiempos medios de supervivencia son de 684 días y 541 días. Si se implementa una regresión de Cox se obtiene una reducción del riesgo de muerte del 31% con p=0.23, no significativo. Se utilizará esta variable en el modelo de regresión.

En la figura 27 se muestra la curva de supervivencia según la presencia de mutaciones en el gen *SMAD4*. Los tiempos medios de supervivencia son de 920 días y 525 días. Si se implementa una regresión de Cox se obtiene una reducción del riesgo de muerte del 21% con p=0.43, no significativo. No podemos tener completamente en cuenta este resultado, ya que, se produce un cruce de las curvas. Este resultado contradice lo descrito en la literatura, donde está asociado con una peor supervivencia y mayor riesgo de metástasis (44). Se utilizará esta variable en el modelo de regresión.

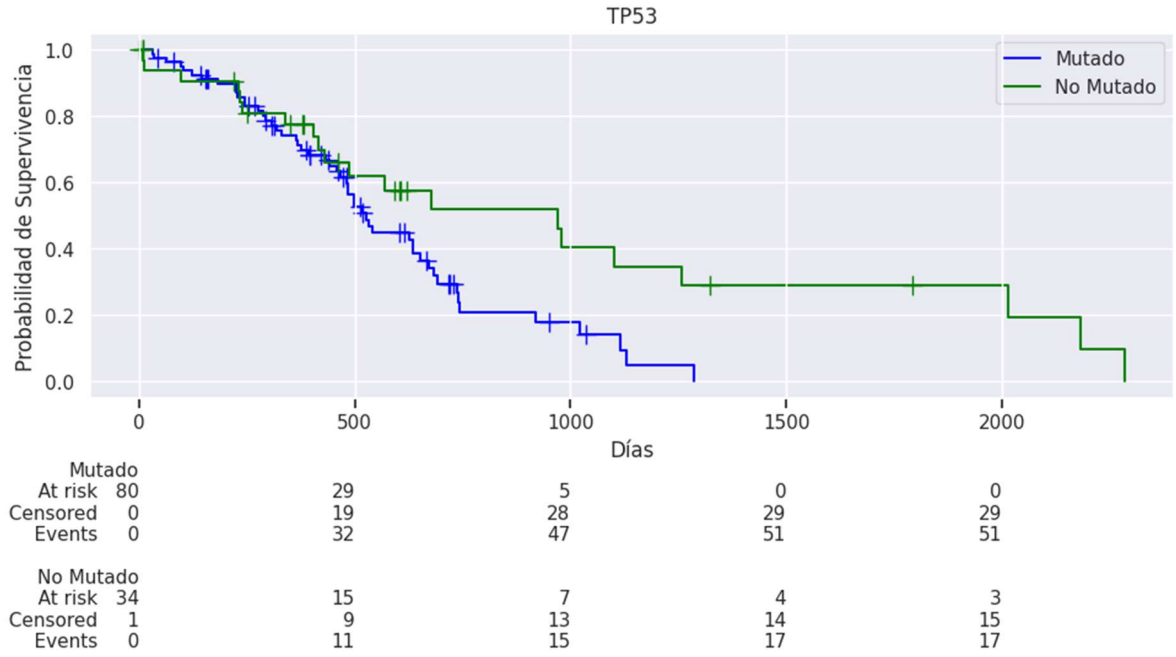


Figura 26: Curva de supervivencia en función de la mutación *TP53*

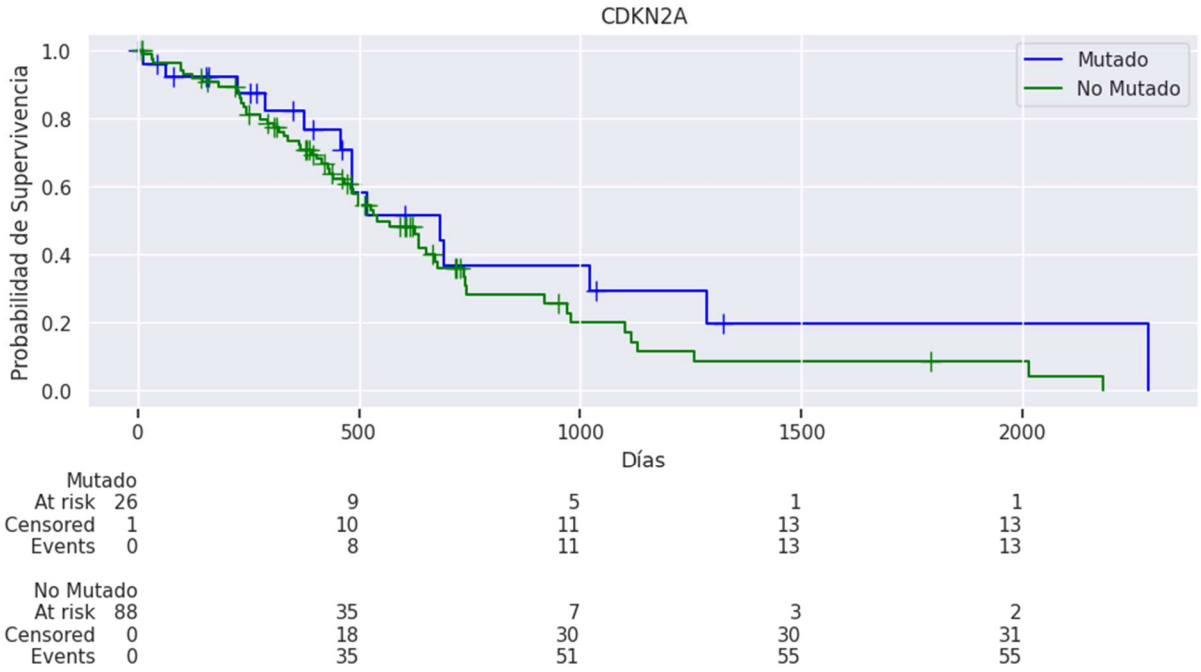


Figura 27: Curva de supervivencia en función de la mutación *CDKN2A*

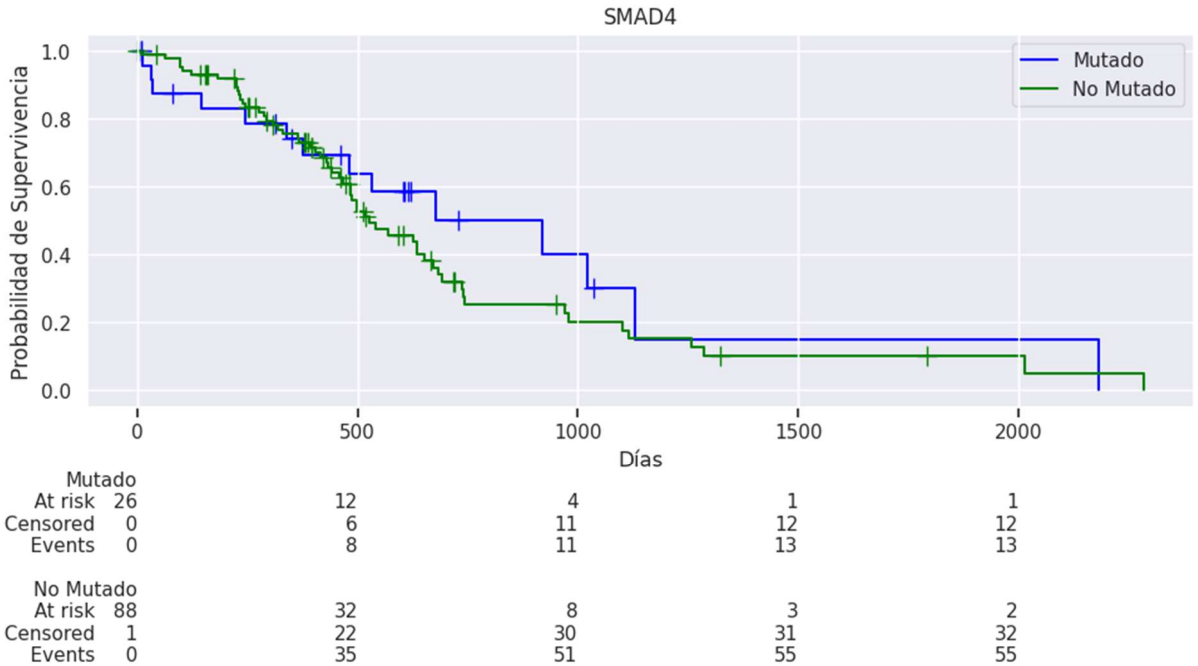


Figura 28: Curva de supervivencia en función de la mutación *SMAD4*

4.7. Modelo de regresión de supervivencia

El modelo implementado obtiene un valor promedio del índice de concordancia de alrededor de 0.6 con máximos alrededor de 0.7. Estos valores se pueden considerar positivos dados los pocos datos disponibles para entrenar el modelo.

La figura 28 muestra el gráfico SHAP del modelo de regresión de supervivencia entrenado. En el se puede observar la importancia de cada variable y la dirección de cada variable. Por ejemplo, la edad es la variable con mayor peso en el modelo y valores altos de edad suponen mayor riesgo de muerte. En segundo lugar, se encuentra el valor del margen de resección. Si el valor es R0, es decir R0 es igual a 1, es beneficioso para la supervivencia. La mutación con mayor peso es la del gen *TP53*, como se había descrito anteriormente.

Todas las variables coinciden con lo descrito en el análisis de supervivencia. Además, el orden de importancia también coincide con la magnitud del valor de las regresiones de Cox implementadas.

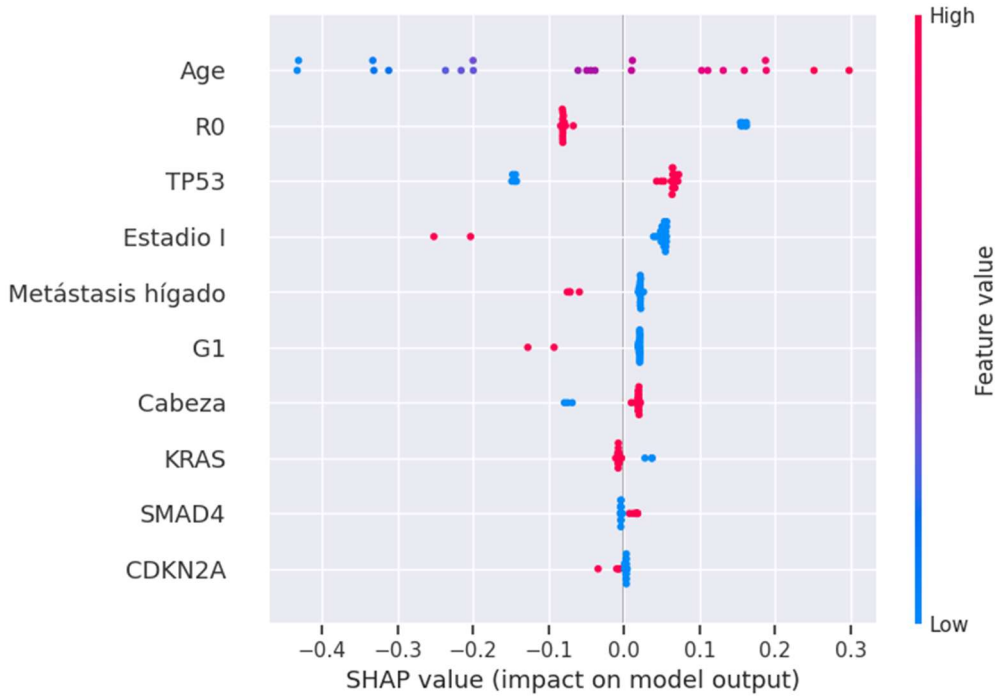


Figura 29: SHAP del modelo de regresión de supervivencia

La figura 29 muestra la predicción de dos curvas de supervivencia individuales y el promedio de las curvas para el conjunto de validación. De los pacientes que conforman el conjunto de validación se muestran las curvas de supervivencia del que tiene mayor riesgo y del que tiene menor riesgo. Se observa una diferencia significativa entre ambos, la supervivencia media del de riesgo bajo es casi el doble de la del riesgo alto, indicando un rendimiento aceptable del modelo.

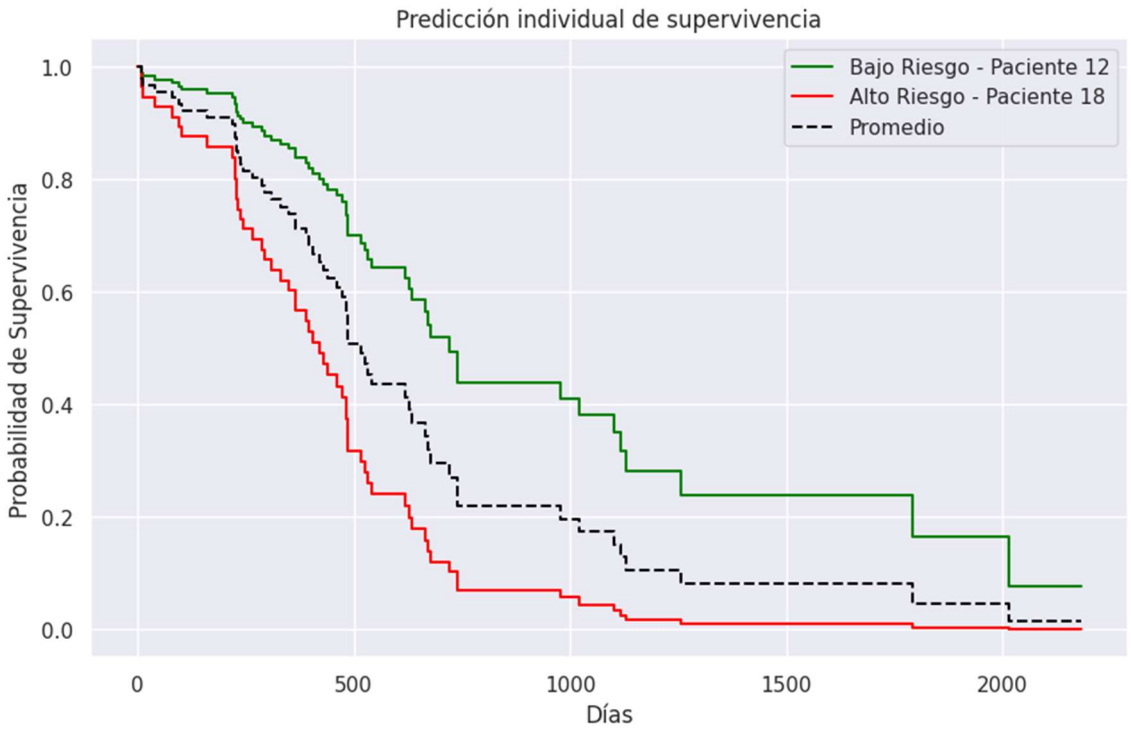


Figura 30: Predicción individual de supervivencia

5. Conclusiones y trabajos futuros

Primero de todo, se discute el cumplimiento de las tareas técnicas propuestas en los objetivos del trabajo:

- Se han extraído y procesado los datos requeridos satisfactoriamente permitiendo extraer el máximo de conocimiento posible de ellos.
- Se ha realizado un EDA que ha permitido explorar la información, plantear hipótesis y aportar conocimiento sobre los datos.
- Se ha podido integrar esta nueva información en los análisis de supervivencia y explorar las diferencias en las capacidades pronósticas.

A continuación, se discute el cumplimiento de los objetivos científicos marcados al inicio de este trabajo:

- Se ha establecido una relación estadísticamente significativa entre la dosis total de los tratamientos de RHE y la respuesta de los pacientes al tratamiento.
- Se ha comprobado que los datos obtenidos en este estudio son representativos de la variedad de tratamientos de quimioterapia y permiten observar la evolución que estos tuvieron.
- Se han analizado las principales mutaciones genómicas implicadas en este cáncer y las diferentes co-mutaciones que presentan.
- Se ha realizado un extenso análisis de supervivencia comprobando los factores que tenían mayor capacidad pronóstica para ser utilizados en el posterior modelo.
- Se ha implementado un modelo de regresión de supervivencia con unos resultados bastante aceptables dadas las limitaciones presentes.

Se pude concluir que el cáncer de páncreas es una enfermedad muy heterogénea y diversa. Estos factores contribuyen a su mal pronóstico. Es necesario continuar desgranando la enfermedad para encontrar tratamientos cada vez más efectivos.

La mayoría de los resultados que se han obtenido durante el trabajo coinciden con los descritos en la literatura. No obstante, algunos contradicen lo descrito en la literatura. Esto puede ser debido a que la muestra es muy pequeña y pequeñas variaciones pueden tener un gran impacto. Por eso, ha sido muy complicado encontrar diferencias o comparativas que sean estadísticamente significativas.

El modelo de regresión de supervivencia desarrollado, aunque muy simple y con pocos datos, consigue recrear lo descrito previamente. Consigue detectar aquellas variables que tienen un mayor peso a la hora de determinar el riesgo de un paciente y en que magnitud lo hacen. Este hecho demuestra el potencial de las redes neuronales.

La dedicación al proyecto ha sido algo irregular. En cuanto a horas dedicadas al proyecto se podría decir que han sido las adecuadas pero su distribución en el tiempo ha sido incorrecta, concentrando la mayoría de los esfuerzos en el último mes y medio de trabajo. También, el balance entre las fases de preprocesamiento y descripción de los datos y la de análisis puede haber estado algo desequilibrado, dedicando más tiempo y espacio del debido a esas primeras fases.

No han aparecido nuevos impactos ético-sociales, de sostenibilidad y de diversidad. Aquellos impactos previamente han sido analizados dentro del apartado correspondiente. Su impacto ha sido menor.

5.1. Limitaciones y trabajo futuro

La principal línea de trabajo futura sería la integración de más capas de información disponible. Esto permitía desarrollar un análisis más completo e implementar modelos de aprendizaje automático más robustos que pudiesen identificar relaciones más profundas.

La principal limitación del trabajo es el limitado tamaño de la muestra. Este hecho solo permite establecer conclusiones generales y no permite explorar subgrupos de pacientes. Debido al pequeño número de pacientes disponibles si se intentan explorar relaciones secundarias, las conclusiones que se pueden extraer carecen de una gran solidez y la variación del resultado de unos pocos pacientes tiene un gran efecto.

Para resolver esta limitación se debería buscar otra fuente de datos e integrarlos conjuntamente. Esto podría ser una solución parcial, ya que, la integración de datos de pacientes de múltiples estudios también puede traer consigo problemas como, por ejemplo, la armonización de los datos.

6. Bibliografía

1. Survival Rates for Pancreatic Cancer [Internet]. [citado 19 de diciembre de 2025]. Disponible en: https://www.cancer.org/cancer/types/pancreatic-cancer/detection-diagnosis-staging/survival-rates.html

2. Cancer Over Time [Internet]. [citado 20 de diciembre de 2025]. Disponible en: https://gco.iarc.fr/overtime

3. Chen L, Linden HM, Anderson BO, Li CI. Trends in 5-year survival rates among breast cancer patients by hormone receptor status and stage. Breast Cancer Res Treat. octubre de 2014;147(3):609-16.

4. Lehtonen M, Kellokumpu-Lehtinen PL. The past and present of prostate cancer and its treatment and diagnostics: A historical review. SAGE Open Med. 1 de diciembre de 2023;11:20503121231216837.

5. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229-63.

6. Rawla P, Sunkara T, Gaduputi V. Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors. World J Oncol. febrero de 2019;10(1):10-27.

7. Nikšić M, Minicozzi P, Weir HK, Zimmerman H, Schymura MJ, Rees JR, et al. Pancreatic cancer survival trends in the US from 2001 to 2014: a CONCORD-3 study. Cancer Commun. 9 de noviembre de 2022;43(1):87-99.

8. The Cancer Genome Atlas Program (TCGA) - NCI [Internet]. 2022 [citado 18 de diciembre de 2025]. Disponible en: https://www.cancer.gov/ccg/research/genome-sequencing/tcga

9. Raphael BJ, Hruban RH, Aguirre AJ, Moffitt RA, Yeh JJ, Stewart C, et al. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. Cancer Cell. 14 de agosto de 2017;32(2):185-203.e13.

10. portal.gdc.cancer.gov/projects/TCGA-PAAD [Internet]. [citado 6 de diciembre de 2025]. Disponible en: https://portal.gdc.cancer.gov/projects/TCGA-PAAD

11. LinkedOmics :: Data Download [Internet]. [citado 4 de diciembre de 2025]. Disponible en: https://www.linkedomics.org/data_download/TCGA-PAAD/

12. Ansari D, Althini C, Ohlsson H, Andersson R. Early-onset pancreatic cancer: a population-based study using the SEER registry. Langenbecks Arch Surg. 2019;404(5):565-71.

13. Herremans KM, Riner AN, Winn RA, Trevino JG. Diversity and Inclusion in Pancreatic Cancer Clinical Trials. Gastroenterology. diciembre de 2021;161(6):1741-1746.e3.

14. krysti. Diversity Lacking in Pancreatic Cancer Trials [Internet]. Let's Win Pancreatic Cancer. 2021 [citado 20 de diciembre de 2025]. Disponible en: https://letswinpc.org/research/diversity-lacking-in-pancreatic-cancer-trials/

15. CDC. Diabetes. 2024 [citado 20 de diciembre de 2025]. National Diabetes Statistics Report. Disponible en: https://www.cdc.gov/diabetes/php/data-research/index.html

16. Etiology and pathogenesis of chronic pancreatitis in adults - UpToDate [Internet]. [citado 20 de diciembre de 2025]. Disponible en: https://www.uptodate.com/contents/etiology-and-pathogenesis-of-chronic-pancreatitis-in-adults

17. Lifetime Risk of Developing or Dying From Cancer [Internet]. [citado 20 de diciembre de 2025]. Disponible en: https://www.cancer.org/cancer/risk-prevention/understanding-cancer-risk/lifetime-probability-of-developing-or-dying-from-cancer.html

18. Risks and causes of pancreatic cancer [Internet]. [citado 20 de diciembre de 2025]. Disponible en: https://www.cancerresearchuk.org/about-cancer/pancreatic-cancer/risks-causes

19. Naudin S, Wang M, Dimou N, Ebrahimi E, Genkinger J, Adami HO, et al. Alcohol intake and pancreatic cancer risk: An analysis from 30 prospective studies across Asia, Australia, Europe, and North America. PLOS Med. 20 de mayo de 2025;22(5):e1004590.

20. Pandol SJ, Apte MV, Wilson JS, Gukovskaya AS, Edderkaoui M. The Burning Question: Why is Smoking a Risk Factor for Pancreatic Cancer? Pancreatol Off J Int Assoc Pancreatol IAP Al. 2012;12(4):344-9.

21. Amin S, McBride R, Kline J, Mitchel EB, Lucas AL, Neugut AI, et al. Incidence of Subsequent Pancreatic Adenocarcinoma in Patients with a History of Non-Pancreatic Primary Cancers. Cancer. 1 de marzo de 2012;118(5):1244-51.

22. He X, Li Y, Su T, Lai S, Wu W, Chen L, et al. The impact of a history of cancer on pancreatic ductal adenocarcinoma survival. United Eur Gastroenterol J. julio de 2018;6(6):888-94.

23. Shin DW, Kim J. The American Joint Committee on Cancer 8th edition staging system for the pancreatic ductal adenocarcinoma: is it better than the 7th edition? Hepatobiliary Surg Nutr. febrero de 2020;9(1):98-100.

24. Puckett Y, Garfield K. Pancreatic Cancer. En: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 [citado 22 de diciembre de 2025]. Disponible en: http://www.ncbi.nlm.nih.gov/books/NBK518996/

25. Tummers WS, Groen JV, Sibinga Mulder BG, Farina-Sarasqueta A, Morreau J, Putter H, et al. Impact of resection margin status on recurrence and survival in pancreatic cancer surgery. Br J Surg. julio de 2019;106(8):1055-65.

26. Wasif N, Ko CY, Farrell J, Wainberg Z, Hines OJ, Reber H, et al. Impact of Tumor Grade on Prognosis in Pancreatic Cancer: Should We Include Grade in AJCC Staging? Ann Surg Oncol. 2010;17(9):2312-20.

27. Pancreatic Cancer Types [Internet]. 2025 [citado 22 de diciembre de 2025]. Disponible en: https://www.hopkinsmedicine.org/health/conditions-and-diseases/pancreatic-cancer/pancreatic-cancer-types

28. Sarantis P, Koustas E, Papadimitropoulou A, Papavassiliou AG, Karamouzis MV. Pancreatic ductal adenocarcinoma: Treatment hurdles, tumor microenvironment and immunotherapy. World J Gastrointest Oncol. 15 de febrero de 2020;12(2):173-81.

29. Takayama Y, Murayama R, Tanaka S, Sato K, Goto K, Honda G, et al. Rare pancreatic ductal adenocarcinoma variants and other malignant epithelial tumors: a comprehensive clinical and radiologic review. Jpn J Radiol. 1 de agosto de 2025;43(8):1239-60.

30. Survival Rates for Pancreatic Neuroendocrine Tumor [Internet]. [citado 22 de diciembre de 2025]. Disponible en: https://www.cancer.org/cancer/types/pancreatic-neuroendocrine-tumor/detection-diagnosis-staging/survival-rates.html

31. Pancreas Basics - Pancreatic Cancer | Johns Hopkins Pathology [Internet]. [citado 22 de diciembre de 2025]. Disponible en: https://pathology.jhu.edu/pancreas/basics

32. Barreto SG, Shukla PJ, Shrikhande SV. Tumors of the Pancreatic Body and Tail. World J Oncol. abril de 2010;1(2):52-65.

33. Agarwal T, Mohanraj K, Pai A. Prolonged survival in metastatic pancreatic cancer: a case of multimodal therapy. Clin Med. 1 de julio de 2025;25(4, Supplement):100449.

34. Wu L, Zhu L, Xu K, Zhou S, Zhou Y, Zhang T, et al. Clinical significance of site-specific metastases in pancreatic cancer: a study based on both clinical trial and real-world data. J Cancer. 18 de enero de 2021;12(6):1715-21.

35. Boyle J, Czito B, Willett C, Palta M. Adjuvant radiation therapy for pancreatic cancer: a review of the old and the new. J Gastrointest Oncol. agosto de 2015;6(4):436-44.

36. Wachtel MS, Xu KT, Zhang Y, Chiriva-Internati M, Frezza EE. Pancreas Cancer Survival in the Gemcitabine Era. Clin Med Oncol. 29 de abril de 2008;2:405-13.

37. Conroy T, Desseigne F, Ychou M, Bouché O, Guimbaud R, Bécouarn Y, et al. FOLFIRINOX versus Gemcitabine for Metastatic Pancreatic Cancer. N Engl J Med. 12 de mayo de 2011;364(19):1817-25.

38. Hoff DDV, Ervin T, Arena FP, Chiorean EG, Infante J, Moore M, et al. Increased Survival in Pancreatic Cancer with nab-Paclitaxel plus Gemcitabine. N Engl J Med. 31 de octubre de 2013;369(18):1691-703.

39. Wainberg ZA, Melisi D, Macarulla T, Cid RP, Chandana SR, Fouchardière CDL, et al. NALIRIFOX versus nab-paclitaxel and gemcitabine in treatment-naive patients with metastatic pancreatic ductal adenocarcinoma (NAPOLI 3): a randomised, open-label, phase 3 trial. The Lancet. 7 de octubre de 2023;402(10409):1272-81.

40. Neuzillet C, Hentic O, Rousseau B, Rebours V, Bengrine-Lefèvre L, Bonnetain F, et al. FOLFIRI regimen in metastatic pancreatic adenocarcinoma resistant to gemcitabine and platinum-salts. World J Gastroenterol WJG. 7 de septiembre de 2012;18(33):4533-41.

41. Zaanan A, Trouilloud I, Markoutsaki T, Gauthier M, Dupont-Gossart AC, Lecomte T, et al. FOLFOX as second-line chemotherapy in patients with pretreated metastatic pancreatic cancer from the FIRGEM study. BMC Cancer. 14 de junio de 2014;14(1):441.

42. Idachaba S, Dada O, Abimbola O, Olayinka O, Uma A, Olunu E, et al. A Review of Pancreatic Cancer: Epidemiology, Genetics, Screening, and Management. Open Access Maced J Med Sci. 14 de febrero de 2019;7(4):663-71.

43. Khan A, Khan I, Maqbool S, Tangri A, Ihtesham A, Gohari P. Trends and persistent disparities in incidence, mortality, and survival of pancreatic cancer: A Surveillance, Epidemiology, and End Results program–based retrospective cohort study (2000–2021). J Clin Oncol. junio de 2025;43(16_suppl):e16360-e16360.

44. Stefanoudakis D, Frountzas M, Schizas D, Michalopoulos NV, Drakaki A, Toutouzas KG. Significance of *TP53*, *CDKN2A*, *SMAD4* and *KRAS* in Pancreatic Cancer. Curr Issues Mol Biol. 23 de marzo de 2024;46(4):2827-44.

7. Anexo A: Pruebas de normalidad

7.1. Prueba de normalidad de la edad de los pacientes

	W	pval	normal
Age	0.984399	0.037477	Falso

7.2. Prueba de normalidad de los tratamientos de RHE

	W	pval	normal
Resultado			
Progresión	0.748257	0.005175	Falso
Respuesta completa	0.692878	0.000312	Falso
Enfermedad estable	NaN	NaN	Falso