



UNIVERSIDAD DE GRANADA

**MÁSTER EN ESTADÍSTICA APLICADA
DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA
TRABAJO DE FIN DE MÁSTER
2022**

**ANÁLISIS DE TIEMPO DE RECURRENCIA DE CÁNCER DE COLON EN
ESTADIO B/C EN ENSAYOS EXITOSOS DE QUIMIOTERAPIA ADYUVANTE
CON EL USO DEL LEVAMISOL Y EL 5-FU**

TUTORES

**JUAN ELOY RUIZ CASTRO
CHRISTIAN J. ACAL GONZALES**

ESTUDIANTE

GERSON JOEL GRANADOS YANEZ

RESUMEN

El cáncer es la segunda causa de muerte a nivel mundial, solo superada por las muertes cardiovasculares. El cáncer de colon es el tercer cáncer de mayor incidencia, donde al igual que en otros tipos de cáncer, la cirugía es en la mayoría de los casos la forma de combatirlo. Según las estadísticas, la recurrencia es del 20 a 30% de los casos.

En este trabajo se presenta la metodología de análisis de supervivencia aplicada al problema de recurrencia del cáncer de colon en estadio B/C en pacientes en los que se les aplicó quimioterapia adyuvante con el uso del Levamisol y Fluorouracilo. En este trabajo se considerarán metodologías clásicas no paramétricas (Kaplan-Meier) y el modelo de regresión de Cox (1972).

Se trata de un estudio de cohorte retrospectivo con datos censurados, en el que el evento a estudiar es el tiempo de recurrencia. Se han construido modelos de supervivencia con múltiples covariables, endógenas y exógenas, identificando cuáles de ellas son significativas en la recurrencia de la enfermedad.

Los resultados de este trabajo son útiles para la implementación de la medicación en el tratamiento postcirugía, así como para identificar datos relevantes a tomar en cuenta desde su diagnóstico. Este trabajo también es un primer estudio para el análisis de la metodología y posterior aplicación en otras investigaciones.

Para el estudio se ha considerado la base de datos dada en Laure, Estados Unidos (1989) y se han tratado con el programa SPSS y el paquete *survival* de R.

ABSTRACT

Cancer is the second leading cause of death globally, second only to cardiovascular deaths. Colon cancer is the third cancer with the highest incidence, where as in other types of cancer, surgery is in most cases the way to fight it. According to statistics, recurrence is 20 to 30% of cases.

This paper presents the survival analysis methodology applied to the problem of recurrence of stage B/C colon cancer in patients in whom adjuvant chemotherapy was applied with the use of Levamisole and Fluorouracil. This paper will consider classical non-parametric methodologies (Kaplan-Meier) and the Cox regression model (1972).

This is a retrospective cohort study with censored data, in which the event to be studied is the time of recurrence. Survival models have been constructed with multiple covariates, endogenous and exogenous, identifying which of them are significant in the recurrence of the disease.

The results of this work are useful for the implementation of medication in post-surgery treatment, as well as to identify relevant data to be taken into account from its diagnosis. This work is also a first study for the analysis of the methodology and subsequent application in other research.

For the study, the database given in Laure, United States (1989) has been considered and they have been treated with the SPSS program and the R *survival* package.

CONTENIDO

I INTRODUCCIÓN	7
II MARCO TEÓRICO	10
2.1 CÁNCER COLORRECTAL.....	11
2.2 TIPOS DE CÁNCER EN EL COLON Y EN EL RECTO.....	12
2.3 ETAPAS DEL CÁNCER COLORRECTAL	13
2.4 TRATAMIENTO.....	13
2.5 RECURRENCIA DEL CÁNCER COLORRECTAL	15
2.6 QUIMIOTERAPIA NEOADYUVANTE Y QUIMIOTERAPIA ADYUVANTE.....	15
III JUSTIFICACIÓN	17
IV OBJETIVOS	19
4.1 Objetivos Generales.....	19
4.2 Objetivos Específicos	19
V ANÁLISIS DE SUPERVIVENCIA	20
5.1 VARIABLE TIEMPO DE VIDA.....	20
5.2 ELEMENTOS BASICOS.....	21
5.3 MODELOS.....	23
5.3.1 ESTIMADO DE KAPLAN-MEIER.....	23
5.3.2 VARIANZA DEL ESTIMADOR DE KAPLAN-MEIER.....	25
5.3.3 MEDIA MUESTRAL Y VARIANZA.....	28
5.3.4 ESTIMACION DE CUANTILES Y VARIANZA.....	29
5.3.5 COMPARACION FUNCIONES DE SUPERVIVENCIA	30
5.4 MODELO DE REGRESION DE COX.....	32
5.4.1 ESTIMACION DE PARAMETROS MODELO DE COX.....	34
5.5 CENSURA	40
5.5.1 TIPOS DE CENSURA	40
5.6 HERRAMIENTAS ESTADISTICAS.....	42
VI METODOLOGÍA	44
6.1 TIPO DE ESTUDIO.....	44
6.2 BASE DE DATOS.....	44
6.3 CARACTERISTICAS Y ATRIBUTOS DE LOS DATOS.....	45
VII RESULTADOS	47
7.1 TRATAMIENTO.....	47

7.2 MODELO COX (1972).....	50
7.3 VARIABLE NODE4.....	53
7.4 VARIABLE EXTENT.....	54
7.5 VARIABLE NODES.....	58
7.6 REGRESION DE COX.....	59
7.8 HAZARD RATIO.....	69
IX REFERENCIA.....	73
X ANEXOS.....	75

LISTA DE TABLAS

Tabla 1.	Tabla para la prueba de igualdad de la función de supervivencia de dos grupos.	30
Tabla 2.	Tabla para la prueba de igualdad de la función de supervivencia de M grupos.....	31
Tabla 3.	Tabla resumen de la variable tratamiento (rx).....	47
Tabla 4.	Tabla comparaciones globales para la variable tratamiento (rx).....	48
Tabla 5.	Tabla para las medias y medianas del tiempo de supervivencia variable tratamiento (rx).....	49
Tabla 7.	Tabla comparaciones globales variable node4.....	54
Tabla 8.	Tabla para las medias y medianas del tiempo de supervivencia variable node4.....	54
Tabla 9.	Tabla comparaciones globales variable extent.....	54
Tabla 10.	Tabla para las medias y medianas del tiempo de supervivencia variable extent.....	55
Tabla 14.	Tabla comparaciones globales variable nodes	58
Tabla 15.	Tabla resumen regresión de Cox	59
Tabla 16.	Tabla codificación variables categóricas.....	60
Tabla 17.	Tabla variables de la ecuación	61

I INTRODUCCIÓN

La Organización Mundial de la Salud (OMS) define que el cáncer es la multiplicación rápida de células anormales que se extienden más allá de sus límites habituales y pueden invadir partes adyacentes del cuerpo o propagarse a otros órganos, en un proceso que se denomina metástasis (Organización Mundial de la Salud, 2022).

Según Our World In Data, 56 millones de personas murieron en 2017. De ellas, casi la mitad (49%) eran mayores de 70 años; sólo el 1% eran mayores de cinco años y menores de 14; y casi el 10% eran niños menores de cinco años. El mayor asesino, con diferencia, fue la enfermedad cardiovascular (ECV), responsable de más de 17 millones de muertes, alrededor de un tercio de todas las muertes de ese año. El cáncer **mató a más de 9,5 millones de personas, alrededor de una de cada seis personas, lo que lo convierte en la segunda causa de muerte en el mundo.** (World Economic Forum, 2020)

Según la organización británica Cancer Research UK, en los últimos 40 años ha habido pocos cambios en los tipos de cáncer que se diagnostican con mayor frecuencia en el mundo.

Los cuatro tipos más comunes -pulmón, mama, colon y próstata- son responsables de 4 de cada 10 diagnósticos (42% de todos los casos de cáncer globales), según el Proyecto Globocan, una base de datos de relevamiento de cáncer a cargo de la Asociación Internacional de Registros del Cáncer (IACR, por sus siglas en inglés).

Para el año 2012, el cáncer colorrectal llegó a ser la tercera neoplasia en hombres con 746.000 casos (10 % del total) y la segunda en mujeres con 614.000 casos (9,2 % del total); el 52 % de ellos en países menos desarrollados, con menor sobrevivencia y un aumento de casos en la última década. (International Agency for Research on Cancer of World Health Organization, 2012).

Al igual que otros tumores, algunos de los tratamientos más utilizados en el cáncer de colon son las intervenciones quirúrgicas, la radioterapia y la terapia sistémica (quimioterapia, tratamientos hormonales, tratamientos biológicos dirigidos).

En el tratamiento de quimioterapia, la quimioterapia neoadyuvante se utilizó inicialmente en el cáncer de mama para el tratamiento de la enfermedad localmente avanzada e inoperable. Este tratamiento sistémico logró hacer que algunos pacientes inoperables fueran elegibles para la cirugía, de ahí su importancia. El término “neoadyuvante” se refiere al uso de quimioterapia antes de la cirugía, mientras “adyuvante” se utiliza para la quimioterapia realizada postratamiento, es decir, después de una intervención, que podría ser la cirugía. **El objetivo del tratamiento con quimioterapia adyuvante es erradicar la enfermedad micro metastásica y prevenir la recurrencia a distancia** (Korde, 2021, p.2).

El cáncer recurrente es aquel que, tras un período (intervalo libre de enfermedad) durante el cual no es detectado, reaparece en el mismo sitio del tumor original (primario) y se denomina recidiva, o en otro lugar, después de que había desaparecido. **La recurrencia del cáncer de colon se presenta entre el 10 y el 30 % de los pacientes operados con intención curativa, y la mayoría de éstos fallecen a consecuencia de ésta** (Revista cubana de cirugía, 2009).

De lo mencionado anteriormente, se puede observar que **el cáncer de colon y recto es el tercero más diagnosticado a nivel mundial**, solo superado por el de pulmón y el de mama, **con un alto porcentaje de recurrencia**. Se desea analizar si la quimioterapia adyuvante o postratamiento con el uso de dos medicamento como lo son Levamisol y el 5-FU, o combinación de ambos, aseguran un mayor tiempo de no recurrencia en este tipo de cáncer. Por otro lado, deseamos abordar si las características de un paciente a otro en las diferencias de las variables endógenas como es el sexo, el número de pólipos encontrado en su diagnóstico, etc., es determinante en que el riesgo de volver a padecer la enfermedad sea mayor.

Se propone entonces realizar un análisis de la recurrencia del cáncer de colon a través del análisis de supervivencia que permite calcular el periodo en el que es más probable dicha recurrencia después de haber sometido a los pacientes a una quimioterapia adyuvante con el uso de los medicamentos mencionados anteriormente. Además se estudia identificar y cuantificar las causas que aumentan el riesgo de recurrencia.

La información que permitirá realizar los análisis propuestos corresponde a un ensayo clínico de quimioterapia adyuvante en pacientes con cáncer de colon en estadio B/C en Laurie (1989). El análisis se realizará sobre una base de datos que contiene la información referente a estos pacientes, información desde el momento de su diagnóstico, así como el tiempo en que se presentó nuevamente el tumor después de haberse sometido al tratamiento.

En cuanto al procedimiento de análisis, inicialmente se realizarán análisis no paramétricos (Kaplan-Meier), para cada tipo de tratamiento con el fin de extraer algunas conclusiones interesantes; sobre todo en lo referente a la diferenciación de los pacientes que tuvieron recurrencia en la enfermedad versus los que no presentaron tal recurrencia. Este mismo análisis se llevará a cabo con todas las variables de la base de datos. Después de haber realizado la fase anterior se procede a construir modelos semi-paramétricos (Regresión de Cox), obteniéndose formas funcionales para la variable recurrencia en sus diferentes variables. Para realizar los cálculos referentes al modelamiento estadístico se utilizarán los paquetes estadísticos SPSS Statistics 21 y RStudio 1.4.1717, en los cuales se hará todo lo concerniente al modelo de Kaplan Meier y al modelo de Cox.

II MARCO TEÓRICO

El cáncer se produce cuando células normales se transforman en células tumorales a través de un proceso en varias etapas que suele consistir en la progresión de una lesión precancerosa a un tumor maligno. Esas alteraciones son el resultado de la interacción entre factores genéticos de la persona afectada y tres categorías de agentes externos: carcinógenos físicos, como las radiaciones ultravioletas e ionizantes; carcinógenos químicos, como el amianto, sustancias contenidas en el humo de tabaco, las aflatoxinas que contaminan los alimentos y el arsénico presente en el agua de bebida; y carcinógenos biológicos, como determinados virus, bacterias y parásitos (Organización Mundial de la Salud, 2022).

Según las estadísticas de la OMS se atribuyeron a esta enfermedad casi 10 millones de defunciones, lo que lo convierte en la segunda causa de muerte a nivel mundial, solo superada por las muertes cardiovascular (Organización Mundial de la Salud, 2022).

En los últimos reportes de la Organización Mundial de la Salud muestra que más del 60% de todos los nuevos casos de cáncer se registra en África, Asia y América Central y del Sur. Las tasas más altas en América, después de Estados Unidos y Canadá, están en Barbados, Uruguay, Cuba, Argentina y Puerto Rico, todos países que figuran en la lista de los 50 más afectados por el cáncer en el mundo (BBC, 2016).

Entre los cánceres que causaron más fallecimientos para el año 2018 tenemos: de pulmón (1,8 millones de defunciones); colorrectal (916 000 defunciones); hepático (830 000 defunciones); gástrico (769 000 defunciones); y de mama (685 000 defunciones), (OMS, 2022). El cáncer de pulmón es el tipo más común entre hombres, mientras que la recurrencia del de mama es la mayor entre mujeres. El número de muertes por tumores es mayor para los hombres que para mujeres, según lo muestra el grafico de la CIE-10 (Clasificación Internacional de Enfermedades, *10ma edición*, 2018, Anexo).

El cáncer colorrectal es la tercera neoplasia en hombres con 746.000 casos (10 % del total) y la segunda en mujeres con 614.000 casos (9,2 % del total), según Globocan, lo que lo convierte en el tercer tipo de cáncer más diagnosticado a nivel mundial.

2.1 CÁNCER COLORRECTAL

El cáncer colorrectal es el que se origina en el colon o en el recto. A estos cánceres también se les puede llamar cáncer de colon o cáncer de recto (rectal) dependiendo del lugar donde se originen. El cáncer de colon y el cáncer de recto a menudo se agrupan porque tienen muchas características comunes.

La mayoría de los cánceres colorrectales comienza como un crecimiento en el revestimiento interno del colon o del recto, estos crecimientos son referidos como pólipos. Algunos tipos de pólipos pueden convertirse en cáncer con el paso del tiempo (por lo general muchos años), pero no todos los pólipos se convierten en cáncer, la probabilidad de que un pólipo se vuelva canceroso depende del tipo de pólipo.

Los dos tipos principales de pólipos son:

- Pólipos adenomatosos (adenomas): estos pólipos algunas veces se transforman en cáncer. Debido a esto, los adenomas se denominan afecciones precancerosas.
- Pólipos inflamatorios y pólipos hiperplásicos: estos pólipos son más frecuentes, pero en general no son precancerosos.

Entre otros factores que pueden hacer que un pólipo sea propenso a contener cáncer o que aumenten el riesgo de que una persona llegue a tener cáncer colorrectal se incluye: si se descubre un pólipo mayor a un centímetro; si se descubre que hay más de dos pólipos; si después de la extirpación de un pólipo se descubre la presencia de displasia, la cual es otra afección precancerosa. Esto implica la existencia de un área del pólipo o del revestimiento del colon o del recto en el que las células tienen un aspecto anormal, pero sin llegar a tener la apariencia de células cancerosas verdaderas (American Cancer Society, 2018).

Si se forma el cáncer dentro de un pólipo, con el tiempo puede crecer hacia la pared del colon o del recto. La pared del colon y del recto está compuesta por muchas capas. El cáncer colorrectal se origina en la capa más interna (la mucosa) y puede crecer hacia el exterior a través de algunas o de todas las demás capas (American Cancer Society, 2018).

Cuando las células cancerosas se encuentran en la pared, éstas pueden crecer hacia los vasos sanguíneos o los vasos linfáticos (canales diminutos que transportan material de desecho y líquido). Desde allí, las células cancerosas pueden desplazarse a los ganglios linfáticos cercanos o a partes distantes del cuerpo (American Cancer Society, 2018).

La etapa (extensión) de un cáncer colorrectal depende de cuán profundamente crece en la pared y si se ha extendido fuera del colon o del recto (American Cancer Society, 2018).

2.2 TIPOS DE CÁNCER EN EL COLON Y EN EL RECTO

Los adenocarcinomas representan alrededor del 96% de los cánceres colorrectales. Estos cánceres se originan de las células que producen mucosidad para lubricar el interior del colon y del recto. Cuando los médicos hablan de cáncer colorrectal, casi siempre se refieren a este tipo de cáncer. Algunos subtipos de adenocarcinoma, como el adenocarcinoma con células en anillo de sello, pueden tener un peor pronóstico (perspectiva) (American Cancer Society, 2018).

Otros tipos de tumores mucho menos comunes también pueden comenzar en el colon y en el recto. Entre estos se incluye: tumores carcinoides, tumores estromales gastrointestinales, linfomas y sarcomas.

2.3 ETAPAS DEL CÁNCER COLORRECTAL

Los cánceres colorrectales en etapas más tempranas se identifican como etapa 0 (un cáncer muy precoz), y luego van desde etapas I (1) a IV (4). Por regla general, mientras más bajo sea el número, menos se ha propagado el cáncer. Un número más alto, como la etapa IV, significa una mayor propagación del cáncer. Además, dentro de una etapa, una letra menor significa una etapa menos avanzada. El sistema de estadificación que se emplea con más frecuencia para el cáncer colorrectal es el sistema TNM del American Joint Committee on Cancer (AJCC) que se basa en tres piezas clave de información: la extensión (tamaño) del tumor (T); la propagación a los ganglios (nódulos) linfáticos adyacentes (N); la propagación (metástasis) a sitios distantes (M).

El sistema AJCC es el más reciente, ya que entró en vigor en enero de 2018. Los números y las letras después de la T, N y M proporcionan más detalles sobre cada uno de estos factores. Los números más altos significan que el cáncer está más avanzado. Una vez que se han determinado las categorías T, N y M de una persona, esta información se combina en un proceso llamado agrupación por etapas para asignar una etapa general.

Por ejemplo, IIB:T4a,N0,M0, nos diría que el cáncer ha crecido a través de la pared del colon o del recto, pero no ha crecido hacia otros tejidos u órganos adyacentes (T4a). El cáncer aún no se ha propagado a los ganglios linfáticos adyacentes (N0) ni a sitios distantes (M0).

2.4 TRATAMIENTO

Los investigadores siempre están buscando mejores formas de tratar el cáncer colorrectal. Entre los tratamientos se tiene:

-Cirugía: Los cirujanos siguen trabajando en mejorar las operaciones que se usan para el cáncer colorrectal. Es común la investigación enfocada a los beneficios de las operaciones laparoscópicas y robóticas en comparación con las cirugías abiertas (en donde se hacen cortes de mayor tamaño en la piel). La cirugía a través del ano para el cáncer rectal, sin incurrir en cortes en la piel, también está siendo estudiada.

-Quimioterapia: Se han probado diferentes métodos en estudios clínicos, incluyendo probar nuevos medicamentos de quimioterapia o medicamentos que ya se estaban usando contra otros cánceres; buscar nuevas maneras de combinar los medicamentos que ya se sabe que actúan contra el cáncer colorrectal a fin de aumentar su eficacia; estudiar mejores maneras de combinar la quimioterapia con la radioterapia, las terapias dirigidas y/o con la inmunoterapia.

Otras áreas de interés en la investigación incluyen la búsqueda de mejores formas de identificar, prevenir y tratar los efectos secundarios de la quimioterapia.

-Terapia dirigida: Los medicamentos de terapia dirigida funcionan de distinta manera que los medicamentos de la quimioterapia convencional, ya que afectan partes específicas de las células cancerosas que las diferencian de las células normales. Se están llevando a cabo investigaciones para ver si el uso de terapia dirigida junto con quimioterapia en casos de cáncer en etapa inicial puede ayudar a reducir el riesgo de recurrencia.

-Radioterapia: Las radiaciones son generadas y emitidas por máquinas de gran tamaño, capaces de realizar tratamientos de muy alta precisión. Actúa únicamente en el lugar donde se encuentra el tumor, preservan y minimizan los efectos secundarios en los tejidos sanos.

-Inmunoterapia: Tratamiento que usa el propio sistema inmunitario del cuerpo para combatir el cáncer.

2.5 RECURRENCIA DEL CÁNCER COLORRECTAL

Los datos de incidencia del cáncer colorrectal son alarmantes; la recidiva es la causa de la mayoría de las muertes tras cirugía con intención curativa. Los factores asociados a la recidiva locoregional se relacionan fundamentalmente con las características histopatológicas y el grado de invasión del tumor. El factor cirujano no debe aparecer más entre estos factores, pues se evita con una adecuada preparación. En el cáncer de recto, esta preparación pasa por la técnica de escisión del mesorrecto, un adecuado margen circunferencial, y la radioquimioterapia neoadyuvante selectiva. Tras la cirugía curativa, los enfermos deben ser controlados para intentar detectar la recidiva asintomática. La recurrencia local aislada sucede en el 20-30% de los casos, pero incluso con metástasis hepáticas o pulmonares se puede intentar la cirugía curativa. El éxito depende de una correcta valoración preoperatoria multidisciplinaria. Cuando el diagnóstico se realiza en una fase incurable, la misión es mejorar la calidad de vida (Cirugía Española, 2005).

2.6 QUIMIOTERAPIA NEOADYUVANTE Y QUIMIOTERAPIA ADYUVANTE

El término neoadyuvante se refiere al uso de quimioterapia antes de la cirugía. La quimioterapia neoadyuvante se utilizó inicialmente en el cáncer de mama para el tratamiento de la enfermedad localmente avanzada e inoperable. Sobre la base del hallazgo de que el tratamiento sistémico podría hacer que algunos pacientes inoperables fueran elegibles para la cirugía, hubo interés en utilizar la terapia sistémica preoperatoria para reducir el alcance y la morbilidad de la cirugía curativa. Múltiples estudios de quimioterapia y terapia endocrina han demostrado que el tratamiento neoadyuvante puede aumentar la probabilidad de cirugía conservadora, estableciendo el tratamiento neoadyuvante como una opción viable en pacientes con enfermedad operable. (Korde, 2021, p.2).

El término adyuvante se utiliza para la quimioterapia realizada postratamiento, es decir, después de una intervención, que podría ser la cirugía. **El objetivo del tratamiento con quimioterapia adyuvante es erradicar la enfermedad micro metastásica y prevenir la recurrencia a distancia** (Korde, 2021, p.2).

La quimioterapia ha sido y es uno de los tratamientos más usados en cualquier tipo de cáncer. Es de vital importancia, dar lugar al uso de nuevos medicamentos o combinación de estos con el fin de obtener mejores resultados. La quimioterapia adyuvante ha demostrado disminuir el riesgo a recurrencia y muerte por lo que se considera un estándar de tratamiento (Arce, 2006, p.177).

III JUSTIFICACIÓN

El levamisol es un compuesto de baja toxicidad utilizado anteriormente para tratar infestaciones de gusanos en animales. Es activo contra nematodos intestinales y parece actuar como vermífugo, al paralizar los vermes susceptibles que son luego eliminados desde el intestino. En particular, es efectivo contra el *Ascaris lumbricoide* y uncinarias. El levamisol parece actuar como agente inmunorestaurador en presencia de inmunosupresión como resultado de una operación de quimioterapia reciente, aunque no estimula una respuesta inmune hacia niveles más altos que los normales. Puede estar relacionada con la activación y proliferación de las células T, aumento de los monocitos y actividad macrófaga (se incluyen la fagocitosis y quimiotaxis) y un incremento en la movilidad de neutrófilos, adherencia y quimiotaxis. **Usado como tratamiento adyuvante en el cáncer colorrectal, no presenta efectos citotóxicos.** También presenta efectos colinérgicos, aumento del estado de ánimo, mientras que como efectos adversos tenemos que en altas dosis puede ocasionar convulsiones, suele provocar náuseas, vómito, dolores abdominales, diarrea, alteraciones del gusto, fiebre, síndrome parecido a la influenza y otros más. (INFOMED, 2020).

El Fluorouracilo o comúnmente llamado 5-FU se usa en el tratamiento de cáncer colorrectal como en otros tipos de cáncer. En los regímenes de tratamiento usados comúnmente se combinan con ácido folínico y otros agentes quimioterápicos como el irinotecan (FOLFIRI y FLIRI), oxaliplatino (FOLFOX) o ambos, irinotecan y oxaliplatino (FOLFIRINOX). El rango de dosis de fluorouracilo más usado se encuentra entre 200-600 mg/m² de superficie corporal. La dosis también puede variar según si la administración es por bolus intravenoso o perfusión intravenosa continua. La dosis de fluorouracilo puede repetirse semanal, bimensual o mensualmente. El número de ciclos varía según el régimen de tratamiento usado y depende también de la decisión clínica basada en el éxito del tratamiento y la tolerancia.

El perfil de toxicidad del 5-fluorouracilo puede ser mejorado o modificado por el ácido folínico. Las manifestaciones más comunes son leucopenia, mucositis, estomatitis y/o diarrea que pueden ser dosis limitantes. Cuando ácido folínico y 5-fluorouracilo son utilizados en

combinación, la dosis de fluorouracilo debe reducirse más en casos de toxicidad que cuando fluorouracilo se utiliza solo. Las toxicidades observadas en pacientes tratados con la combinación son cualitativamente similares a las observadas en pacientes tratados con 5-fluorouracilo solo.

Se observa toxicidad gastrointestinal más comúnmente y puede ser más grave o incluso potencialmente mortal (particularmente la estomatitis y la diarrea) (Agencia Española de Medicamentos y Productos Sanitarios, 2020, p.7) . En casos severos, 5-fluorouracilo y ácido folínico deben ser retirados e iniciarse una terapia intravenosa de apoyo.

Por lo antes expuesto, es relevante poder estimar el tiempo de recurrencia en los pacientes en los que se les administre cualquiera de estos tratamientos o combinación de ambos en la quimioterapia adyuvante. Podemos observar que uno de ellos presenta menos toxicidad por lo que se vuelve pertinente la investigación.

IV OBJETIVOS

4.1 Objetivos Generales

- Hacer una breve descripción de la metodología clásica en análisis de supervivencia.
- Analizar el tiempo de recurrencia de cáncer de colon en estadio B/C en ensayos exitosos de quimioterapia adyuvante con el uso del Levamisol y el 5-FU, utilizando las técnicas clásicas y software apropiado en el análisis de supervivencia.

4.2 Objetivos Específicos

- Explicar el efecto del Levamisol y el 5-FU, quimioterapia adyuvante, para el estudio de recurrencia en el cáncer de colon.
- Analizar el tiempo de recurrencia con otras covariables, endógenas y exógenas, como son, el sexo, edad o el número de ganglios linfáticos encontrados en su diagnóstico, entre otras.
- Describir los aspectos más importantes de los paquetes SPSS y survival de R-cran e implementarlos para dar un estudio lo más completo posible de la base de datos.

V ANÁLISIS DE SUPERVIVENCIA

La supervivencia viene a ser una medida del tiempo que transcurre hasta una respuesta de interés: recaída, alta médica, manifestación de los primeros síntomas de una determinada enfermedad, etc.

Cuando se trata de enfermedades graves, el análisis de supervivencia deriva en la estimación de probabilidades de mantenerse con vida durante determinados períodos de tiempo, informando así del grado de severidad de la enfermedad, algo que a menudo se desconoce.

El análisis estadístico de tiempos de vida tiene un área de trabajo muy amplio, especialmente en los campos de biomedicina e ingeniería.

5.1 VARIABLE TIEMPO DE VIDA

Dependiendo del campo donde se haga el estudio, se hablará de tiempo de vida, tiempo de supervivencia o tiempo de fallo. En un estudio de supervivencia cuando se hace mención a los datos o a la variable que los modeliza se habla de tiempo de vida, aunque las observaciones no siempre tienen que referirse literalmente al concepto de tiempo de vida. Algunas veces hablamos de tiempo de vida cuando nos referimos a la longitud de vida hasta un tiempo de muerte desde un punto prefijado. En otras ocasiones el término tiempo de vida se usa en sentido figurado. Matemáticamente el tiempo de vida es una variable no negativa de probabilidad. Algunos ejemplos de aplicación son los siguientes: en el campo de la ingeniería se puede hacer un estudio de la duración de una componente electrónica o mecánica; en el campo de la ingeniería en un sistema reparable se puede analizar el tiempo de funcionamiento y el de reparación; en medicina es de interés el estudio del tiempo desde diagnóstico hasta muerte, pérdida o curación del enfermo, etc.

La finalidad de un estudio de supervivencia es obtener información sobre la variable tiempo de vida; tiempo hasta la muerte, curación o recurrencia de una enfermedad, probabilidad de fallo en cada instante, riesgo de fallo, etc. Todo ello pudiendo considerar distintos factores endógenos o exógenos que intervengan en el comportamiento del tiempo de vida, así como la consideración de distintos grupos de riesgo.

La característica principal de la variable tiempo de vida, en el sentido que estamos mencionando, es que forzosamente es una variable no negativa.

5.2 ELEMENTOS BASICOS

En análisis de supervivencia es de suma importancia la variable aleatoria T , que representa el tiempo de duración hasta la ocurrencia de un evento (muerte, deserción, fallo de un sistema, entre otros). La variable aleatoria T , posee una función de densidad de probabilidad $f(t)$, y una función de distribución de probabilidad acumulada $F(t) = P(T \leq t)$, que representa la probabilidad de que el evento de interés ocurra en un tiempo menor o igual a t . En análisis de supervivencia es de especial interés la función de supervivencia $S(t) = 1 - F(t) = P(T > t)$, la cual expresa la probabilidad de que el evento de interés ocurra en un tiempo mayor a t .

Otra función de interés y fundamental para este análisis es la llamada función de riesgo, $h(t)$, que representa la razón instantánea de que un evento ocurra en un intervalo infinitesimal de tiempo $(t, t + \Delta t)$, dado que no ha ocurrido hasta el tiempo t . Se puede escribir como,

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T > t)}{\Delta t}.$$

Desarrollando la probabilidad condicional de la expresión anterior se obtiene:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \frac{P(t \leq T < t + \Delta t)}{P(T > t)}.$$

Del cálculo de probabilidades y del hecho de que $P(T > t) = S(t)$, se tiene que:

$$h(t) = \frac{1}{S(t)} \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t}.$$

De la definición de derivada se tiene que esta expresión se convierte en:

$$h(t) = \frac{f(t)}{S(t)}.$$

Se tiene entonces una expresión que relaciona la función de supervivencia y la función de riesgo. Integrando se obtiene la función de riesgo acumulada,

$$H(t) = \int_0^t h(s) ds = \int_0^t \frac{f(s)}{S(s)} ds,$$

que se puede escribir como: $H(t) = \int_0^t \frac{f(s)}{1-F(s)} ds$, haciendo la sustitución $u = 1 - F(s)$,

y $du = -f(s) ds$, la expresión se convierte en $H(t) = - \int_1^{1-F(t)} \frac{du}{u}$, de donde resulta

que:

$$H(t) = -\text{Ln } S(t);$$

$$S(t) = e^{-H(t)};$$

$$h(t) = -\frac{d}{dt} (\text{Ln } S(t));$$

$$f(t) = -\frac{d}{dt} S(t).$$

5.3 MODELOS

Un estudio de análisis de supervivencia puede ser abordado por diferentes enfoques: paramétrico, no paramétrico y métodos de regresión semi-paramétricos. El primero de ellos consiste en considerar distribuciones paramétricas para analizar la evolución del tiempo de vida. Las funciones de distribución más habituales son la Exponencial, log-normal y la Weibull. Dada una muestra, se estima el modelo y las medidas asociadas. El análisis de supervivencia paramétrico tiene la rigidez de tener que ajustarse al modelo. En el caso de las técnicas de análisis de supervivencia no paramétricas la más usada y popular es, sin lugar a duda, el estimador de Kaplan-Meier de la función de supervivencia. En el campo de la regresión consideraremos el modelo de Cox (1972). (Royal Stastical Society, Vol 21).

5.3.1 ESTIMADO DE KAPLAN-MEIER

El estimador de Kaplan-Meier (1958), de la función de supervivencia, es el estimador más común entre los paquetes estadísticos. El método de construcción del estimador es un método no paramétrico ya que no asume ninguna estructura para la función de distribución de probabilidad de tiempo de vida. (Universidad de Washington, 1992).

El estimador de Kaplan-Meier utiliza toda la información disponible, casos censurados y no censurados, para realizar la estimación de la función de supervivencia. El estimador en

cualquier instante de tiempo es obtenido de la multiplicación de una secuencia de probabilidades condicionales de supervivencia estimadas. Cada probabilidad condicional estimada se obtiene del número de casos observados en riesgo y el número de “muertes” en un instante del tiempo, y se calcula como,

$$\hat{S}(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i};$$

$d_i = \text{numero de sucesos en cada } t_i;$

$n_i = \text{numero de individuos en riesgo antes de } t_i.$

Al hacer la gráfica de este estimador no paramétrico de la función de supervivencia, cuando tenemos un valor censurado, es decir, que ha salido del estudio o que el estudio ha terminado y todavía este individuo no le había sufrido el suceso estudiado, lo representamos como un salto, lo que significa que su valor se asocia a valores próximos, lo mismo sucede con valores repetidos, se calculan valores del próximo, pero evidentemente, los demás quedan asociados a este último valor.

En análisis de supervivencia uno de los problemas estadísticos más frecuentes e interesantes es la comparación de curvas de supervivencia. El problema surge cuando tenemos dos o más grupos, cada uno con su muestra, y queremos ver si tenemos igualdad de curvas de supervivencia. Para ello, debemos comprobar si el estimador de Kaplan-Meier aplicado a cada una de esas muestras nos permite deducir que las diferencias que vemos entre ellas son estadísticamente significativas o, por el contrario, pueden ser atribuibles al azar. La hipótesis nula será, en este caso, igualdad de curvas. La hipótesis alternativa será que hay diferencia entre esas curvas y que, por lo tanto, el suceso analizado se produce con ritmo distinto a lo largo del tiempo en los diferentes grupos analizados.

La técnica más usual para realizar esta comparación de curvas de supervivencia es el Test Log-Rank. Es, de hecho, una técnica ji-cuadrado, donde se realiza una comparación entre un observado (las curvas de supervivencia muestrales) y un esperado (reuniendo todos los valores en una única muestra).

5.3.2 VARIANZA DEL ESTIMADOR DE KAPLAN-MEIER

Adicional a la estimación de la función de supervivencia es de gran interés construir intervalos de confianza para los estimadores de la función de supervivencia, así como de estimaciones puntuales y de intervalo para algunos percentiles.

Aunque existen varios enfoques para tratar este problema, en este trabajo se tratara el enfoque del método delta el cual está basado en la expansión en series de Taylor de primer orden (Pereira, 2018).

Método Delta

La idea básica del método es usar una aproximación en series de Taylor para obtener una función lineal que aproxime una función más complicada como es el caso del estimador de la función de supervivencia. Para poder aplicar el método delta es necesaria que la función que se considere sea de clase uno.

Sea X una variable aleatoria y $f(X)$ una función de dicha variable. El desarrollo en serie de Taylor de primer orden de dicha función centrado en la media de X , u , viene dado por,

$$f(x) \cong f(u) + (x - u)f'(u).$$

De la ecuación anterior se tiene que la varianza de la función $f(x)$ es aproximadamente igual a,

$$\text{Var}(f(x)) \cong [f'(u)]^2 \sigma^2,$$

donde σ^2 es la varianza de la variable aleatoria X . Como ejemplo se muestra la varianza para la función logaritmo natural y la función exponencial.

Logaritmo natural o neperiano

$$\text{Ln}(X) \cong \text{Ln}(u) + (X - u) \frac{1}{u};$$

$$\widehat{\text{Var}}(\text{Ln}(X)) \cong \frac{1}{\hat{u}^2} \hat{\sigma}^2.$$

Función exponencial

$$\exp(X) \cong \exp(u) + (X - u) \exp(u); \quad (1)$$

$$\widehat{\text{Var}}(\exp(X)) \cong \hat{\sigma}^2 (\exp(\hat{u}))^2. \quad (2)$$

Varianza función de supervivencia

Para calcular el estimador de la varianza a través del método delta del estimador Kaplan-Meier, primero se calcula el estimador delta del logaritmo natural así:

$$\hat{S}(t) = \prod_{t_i \leq t} \frac{n_i - d_i}{n_i};$$

$$\text{Ln}(\hat{S}(t)) = \sum_{t_i \leq t} \text{Ln} \frac{n_i - d_i}{n_i};$$

Haciendo $\hat{p}_i = \frac{n_i - d_i}{n_i}$, se tiene que la ecuación se vuelve:

$$\text{Ln}(\hat{S}(t)) = \sum_{t_i \leq t} \text{Ln}(\hat{p}_i).$$

Aplicando el operador varianza a ambos lados de la ecuación, se tiene que:

$$\widehat{\text{Var}}(\text{Ln}(\hat{S}(t))) = \sum_{t_i \leq t} \text{Var}(\text{Ln}(\hat{p}_i)). \quad (3)$$

Lo anterior debido a la suposición de independencia de las variables \hat{p}_i . Si además suponemos que las \hat{p}_i se distribuyen como variables aleatorias de Bernoulli con probabilidad constante p_i . Se tiene entonces como estimador de p_i a \hat{p}_i , y como estimador de la varianza la siguiente expresión $\frac{\hat{p}_i(1-\hat{p}_i)}{n_i}$, ahora utilizando ambas ecuaciones se tiene que:

$$\widehat{\text{Var}}(\text{Ln}(\hat{p}_i)) \cong \frac{1}{\hat{p}_i^2} \frac{\hat{p}_i(1-\hat{p}_i)}{n_i}.$$

Utilizando el hecho de que $\hat{p}_i = \frac{n_i - d_i}{n_i}$, la ecuación se convierte en:

$$\widehat{\text{Var}}(\text{Ln}(\hat{p}_i)) \cong \frac{d_i}{n_i(n_i - d_i)}. \quad (4)$$

Reemplazando (4) en (3), se tiene que:

$$\widehat{\text{Var}}(\text{Ln}(\hat{S}(t))) \cong \sum_{t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}. \quad (5)$$

Esta expresión da una estimación de la varianza del logaritmo natural de la función de supervivencia, lo que se debe hacer ahora es aplicar nuevamente el método delta sobre la función exponencial (como se hizo en (1) y (2)) para eliminar el logaritmo natural de la expresión (5).

Utilizando las ecuaciones (1) y (2) y el hecho de que $\hat{S}(t) = \exp(Ln(\hat{S}(t)))$, se tiene:

$$\widehat{Var}(\exp(Ln(\hat{S}(t)))) \cong \left\{ \exp[Ln(\hat{S}(t))] \right\}^2 \sum_{t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}. \quad (6)$$

Simplificando:

$$\widehat{Var}(\hat{S}(t)) \cong [\hat{S}(t)]^2 \sum_{t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}.$$

La desviación estándar se calcula como la raíz cuadrada de la varianza dada en la expresión (6).

5.3.3 MEDIA MUESTRAL Y VARIANZA

(Pereira, 2018) El estimador usado para la media es debido a un resultado de la teoría de la probabilidad sobre variables aleatorias no negativas. La esperanza matemática de una variable aleatoria no negativa es igual a la integral definida (área bajo la curva) desde cero hasta infinito de la función de supervivencia como se muestra a continuación:

$$u = \int_0^{\infty} tf(t) dt = \int_0^{\infty} S(t) dt \quad t \geq 0 \text{ y } t < t_l.$$

Debido a que la variable t no esta definida para valores mayores a t_l ,

$$u(t_l) = \int_0^{t_l} S(t) dt.$$

El valor del estimador de la media es entonces el área de la función escalonada definida por el estimador de Kaplan-Meier de la función de supervivencia así,

$$\hat{u} = \sum_{i=0}^{k-1} \hat{S}(t_i)(t_{i+1} - t_i) \quad \text{si } t_l = t_k,$$

donde t_l es el tiempo observado más grande, ahora si $t_l > t_k$ entonces el cálculo de la media se hace con:

$$\hat{u} = \sum_{i=0}^{k-1} \hat{S}(t_i)(t_{i+1} - t_i) + \hat{S}(t_k)(t_l - t_k).$$

Cuando no se tienen casos censurados, la varianza de la media muestral se calcula como la varianza muestral dividida sobre el tamaño de la muestra. Cuando se presentan censuras el estimador de la varianza de la media muestral se calcula como:

$$\widehat{Var}(\hat{u}) = \frac{d}{d-1} \sum_{i=1}^{k-1} \frac{\{\sum_{i=0}^{k-1} \hat{S}(t_i)(t_{i+1} - t_i)\}^2 d_i}{n_i(n_i - d_i)} \quad \text{si } t_l = t_k,$$

Donde $d = \sum d_i$, la cantidad total de muertes. En el caso en el que $t_l > t_k$, la varianza se calcula como:

$$\widehat{Var}(\hat{u}) = \frac{d}{d-1} \sum_{i=1}^{k-1} \frac{\{\sum_{i=0}^{k-1} \hat{S}(t_i)(t_{i+1} - t_i) + \hat{S}(t_k)(t_l - t_k)\}^2 d_i}{n_i(n_i - d_i)}.$$

5.3.4 ESTIMACION DE CUANTILES Y VARIANZA

(Pereira, 2018) La fórmula para el cálculo de los cuantiles se muestra a continuación:

$$\hat{t}_p = \inf\{t: \hat{S}(t) \leq p\}.$$

La varianza del estimador del p – *esimo* cuantil se estima con:

$$\widehat{Var}(\hat{t}_p) = \frac{\widehat{Var}(\hat{S}(\hat{t}_p))}{[\hat{f}(\hat{t}_p)]^2}.$$

En la anterior expresión el numerador se calcula con la ecuación (6), y el denominador que es la estimación de la función de densidad de la distribución de los tiempos de supervivencia se calcula así:

$$\hat{f}(\hat{t}_p) = \frac{\hat{S}(\hat{u}_p) - \hat{S}(\hat{l}_p)}{\hat{l}_p - \hat{u}_p}.$$

Donde $\hat{u}_p = \sup(t: \hat{S}(t) \geq p + 0.05)$ y $\hat{l}_p = \inf(t: \hat{S}(t) \leq p - 0.05)$, aunque se pudieran usar números inferiores al 0.05.

5.3.5 COMPARACION FUNCIONES DE SUPERVIVENCIA

Es necesario contar con pruebas estadísticas que permitan identificar si las diferencias observadas entre las funciones de supervivencia son significativas. Las pruebas estadísticas están basadas en su mayoría en tablas de contingencia de grupo por estado de cada tiempo de supervivencia observado, t_i , como se muestra en la siguiente tabla:

Tabla 1. Tabla para la prueba de igualdad de la función de supervivencia de dos grupos.

<i>Evento</i>	<i>Grupo 1</i>	<i>Grupo 0</i>	<i>Total</i>
<i>Muerte</i>	d_{1i}	d_{0i}	d_i
<i>No muerte</i>	$n_{1i} - d_{1i}$	$n_{0i} - d_{0i}$	$n_i - d_i$
<i>En riesgo</i>	n_{1i}	n_{0i}	n_i

El estadístico de prueba se construye con el número esperado de muertes y la varianza del número de muertes para alguno de los dos grupos, por ejemplo, para el grupo 1 se tiene que el estimador del número de muertes es,

$$\hat{e}_{1i} = \frac{n_{1i}d_i}{n_i}.$$

Y la varianza suponiendo una distribución hipergeométrica:

$$\hat{v}_{1i} = \frac{n_{1i}n_{0i}d_i(n_i - d_i)}{n_i^2(n_i - 1)}.$$

Ya con los estimadores del número de muertes y su varianza se construye el estadístico de prueba Q.

$$Q_1 = \frac{\sum_{i=1}^k w_i (d_{1i} - \hat{e}_{1i})^2}{\sum_{i=1}^k w_i^2 \hat{v}_{1i}}.$$

Bajo la hipótesis nula de que las dos funciones de supervivencias son iguales y que el número observado de eventos es grande se puede demostrar que Q se distribuye como una chi-cuadrado con un grado de libertad ($p = P(\chi^2(1)) \geq Q$). Dependiendo del valor de w_i , se tienen diferentes tipos de contrastes, como por ejemplo,

$$w_i = 1 \quad \text{Mantel y Haenzel}$$

$$w_i = n_i \quad \text{Breslow}$$

$$w_i = \sqrt{n_i} \quad \text{Tarone y Ware}$$

El caso de más de dos grupos se trata en forma similar al caso ya visto de solo dos grupos. Se muestra a continuación la tabla asociada a varias funciones de supervivencia base para el cálculo de los estadísticos de prueba para la igualdad de las funciones:

Tabla 2. Tabla para la prueba de igualdad de la función de supervivencia de M grupos.

<i>Evento</i>	<i>1</i>	<i>2</i>	...	<i>m</i>	...	<i>M</i>	<i>Total</i>
<i>Muerte</i>	d_{1i}	d_{2i}	...	d_{mi}	...	d_{Mi}	d_i
<i>No muerte</i>	$n_{1i} - d_{1i}$	$n_{2i} - d_{2i}$...	$n_{mi} - d_{mi}$...	$n_{Mi} - d_{Mi}$	$n_i - d_i$
<i>En riesgo</i>	n_{1i}	n_{2i}	...	n_{mi}	...	n_{Mi}	n_i

Inicialmente se debe estimar el número esperado de muertes de cada grupo bajo la suposición de igual función de supervivencia, es decir:

$$\hat{e}_{ki} = \frac{n_{ki} d_i}{n_i}, \quad k = 1, \dots, m.$$

Ya con varios grupos lo que se calcula ya no es una varianza sino una matriz de covarianzas del vector d_i . La matriz de covarianzas es obtenida suponiendo que el número observado de eventos sigue una distribución hipergeométrica central multivariada. Los términos de la diagonal de la matriz denotada por \hat{V}_i , se calculan de la siguiente manera:

$$\hat{v}_{mmi} = \frac{n_{mi}(n_i - n_{mi})d_i(n_i - d_i)}{n_i^2(n_i - 1)}, \quad m = 1, 2, \dots, M - 1.$$

Los elementos fuera de la diagonal se calculan como:

$$\hat{v}_{lmi} = \frac{n_{li}n_{mi}d_i(n_i - d_i)}{n_i^2(n_i - 1)}, \quad l, m = 1, 2, \dots, M - 1, \quad l \neq m.$$

Con los estimadores del número de muertes y la varianza se construye el siguiente estadístico,

$$Q = \left[\sum_{i=1}^k W_i(d_i - \hat{e}_i) \right]' \left[\sum_{i=1}^k W_i \hat{V}_i W_i \right]^{-1} \left[\sum_{i=1}^k W_i(d_i - \hat{e}_i) \right],$$

donde W_i es una matriz diagonal que tiene los pesos w_i , d_i es el vector que contiene la cantidad de muertes para cada grupo en el tiempo t_i , \hat{e}_i es el vector que contiene el número esperado de muertes para cada grupo en el tiempo t_i . Al igual que en el caso de dos grupos se tiene que Q para el caso de varios grupos, bajo la hipótesis nula de la igualdad de funciones de supervivencia, se distribuye también como una chi-cuadrada con $M - 1$ grados de libertad ($p = P(\chi^2(1)) \geq Q$).

5.4 MODELO DE REGRESION DE COX

Supóngase que se quiere comparar la supervivencia de dos grupos, el análisis de Kaplan-Meier no es suficiente debido a su imposibilidad de tener en cuenta otras “covariables”. Lo que se ahora es plantear un modelo para el riesgo $h(t)$, en función del tiempo y de variables explicativas o covariables.

La presentación inicialmente se realizará sobre el supuesto de una sola covariable y después se generalizará a p variables. El modelo de regresión que se trata en este trabajo se puede escribir como sigue,

$$h(t, x, \beta) = h_0(t)r(x, \beta). \quad (7)$$

Desde (7) se puede ver que la razón de riesgos de dos sujetos con valor de covariable x_1 y x_2 respectivamente viene dado por,

$$RR(t, x_1, x_2) = \frac{h(t, x_1, \beta)}{h(t, x_2, \beta)}, \quad (8)$$

utilizando (7) y reemplazando en (8) se obtiene:

$$RR(t, x_1, x_2) = \frac{r(x_1, \beta)}{r(x_2, \beta)}.$$

De la expresión anterior se tiene que la razón de funciones de riesgo de dos individuos depende solo de la función relacionada con la covariable $r(x, \beta)$. Si adicionalmente se supone una forma exponencial a la función de la covariable estamos ante el modelo de la regresión de Cox o modelo de riesgos proporcionales debido a que la razón de riesgo depende solamente del valor que tome la covariable. El modelo de Cox se puede escribir así,

$$h(t, x, \beta) = h_0(t)e^{\beta x}. \quad (9)$$

En otras palabras, el modelo de Cox es un modelo semi-paramétrico en el cual se supone una forma conocida para la función de las covariables y una desconocida para la función que depende solo del tiempo de supervivencia. La razón de funciones de riesgo se escribe como,

$$RR(t, x_1, x_2) = e^{\beta(x_1 - x_2)}. \quad (10)$$

Uno de los atractivos del modelo de Cox planteado es lo que resulta de la ecuación (10), por ejemplo, si la variable X indicara el género del paciente codificado con X=1 si es hombre y X=0 si es mujer, se tendría que la razón de riesgo entre hombres y mujeres sería igual a,

$$RR(t, x_1, x_2) = e^{\beta}.$$

Ahora si de los cálculos resulta que $\beta = \ln(3)$, quiere decir que los hombres tienen un riesgo de fallar tres veces mayor que el riesgo de las mujeres.

La cuestión ahora será calcular la función de supervivencia para un modelo que parte de la función de riesgo. De la ecuación $S(t) = e^{-H(t)}$, y con la función de riesgo dada en (9) se obtiene:

$$S(t, x, \beta) = e^{-H(t,x,\beta)}, \quad (11)$$

donde $H(t, x, \beta)$ es la función de riesgo acumulada que se puede calcular como,

$$H(t, x, \beta) = \int_0^t h(u, x, \beta) du = r(x, \beta) \int_0^t h_0(u) du = r(x, \beta) H_0(t), \quad (12)$$

sustituyendo (12) en (11) se tiene,

$$S(t, x, \beta) = e^{-r(x,\beta)H_0(t)}, \quad (13)$$

de donde se sigue que,

$$S(t, x, \beta) = [e^{-H_0(t)}]^{r(x,\beta)} = [S_0(t)]^{r(x,\beta)},$$

donde $S_0(t)$ es la función de supervivencia base. Ahora bajo las suposiciones del modelo de Cox la expresión (13) se convierte en

$$S(t, x, \beta) = [S_0(t)]^{e^{\beta x}}.$$

5.4.1 ESTIMACION DE PARAMETROS MODELO DE COX

El procedimiento más utilizado para realizar la estimación de los parámetros de un modelo de regresión es el procedimiento relacionado con la verosimilitud que consiste en estimar los parámetros que maximice el logaritmo natural de la verosimilitud. En el caso del modelo de Cox este procedimiento no funciona debido al desconocimiento de la función de riesgo base que aparece en el modelo. (Pereira, 2018).

Cox desarrolla un método que permite estimar los parámetros sin necesidad de asumir una forma particular para la función de riesgo base. Propone entonces usar una expresión llamada “función de verosimilitud parcial” que depende solamente de los parámetros de interés. La función de verosimilitud parcial se escribe como,

$$l_p = \prod_{i=1}^k \frac{e^{\beta x(i)}}{\sum_{j \in R(t_i)} e^{\beta x_j}}, \quad (14)$$

donde el producto se hace sobre k tiempos de supervivencia distintos y $x_{(i)}$ corresponde con el valor de la covariable del individuo que ha fallado en el tiempo t_i , y $R(t_i)$ se refiere a todos los individuos en riesgo en el tiempo t_i , y corresponde a todos los individuos con tiempos de supervivencia o censura superiores a t_i .

Calculando el logaritmo natural de la función de verosimilitud parcial se tiene,

$$L_p = \sum_{i=1}^k \left\{ \beta x_{(i)} - \ln \left[\sum_{j \in R(t_i)} e^{\beta x_j} \right] \right\}.$$

Derivando con respecto a β se obtiene:

$$\frac{\partial L_p(\beta)}{\partial \beta} = \sum_{i=1}^k \left\{ x_{(i)} - \frac{\sum_{j \in R(t_i)} x_j e^{\beta x_j}}{\sum_{j \in R(t_i)} e^{\beta x_j}} \right\}. \quad (15)$$

La ecuación (15) se debe igual a cero y resolverse, pero como se debe notar la ecuación resultante, debido a su complejidad, debe ser resuelta a través de un método numérico como el de Newton-Raphson con algún criterio de parada,

$$\sum_{i=1}^k \left\{ x_{(i)} - \frac{\sum_{j \in R(t_i)} x_j e^{\beta x_j}}{\sum_{j \in R(t_i)} e^{\beta x_j}} \right\} = 0.$$

El estimador de la varianza del estimador del parámetro es obtenido de la segunda derivada. Tomando la derivada en (15) se tiene

$$\frac{\partial^2 L_p(\beta)}{\partial \beta \partial \beta} = - \sum_{i=1}^k \left\{ \frac{[\sum_{j \in R(t_i)} e^{\beta x_j}][\sum_{j \in R(t_i)} x_j^2 e^{\beta x_j}] - [\sum_{j \in R(t_i)} x_j e^{\beta x_j}]^2}{[\sum_{j \in R(t_i)} e^{\beta x_j}]^2} \right\}. \quad (16)$$

El negativo del lado derecho de la ecuación (16) es llamada información observada (cuando se presenta más de una covariable se llama matriz de información observada) y se denota por,

$$I(\beta) = - \frac{\partial^2 L_p(\beta)}{\partial \beta \partial \beta}. \quad (17)$$

El estimador de la varianza del coeficiente estimado es el inverso de (17) evaluado para $\hat{\beta}$ y se calcula como

$$\widehat{Var}(\hat{\beta}) = I^{-1}(\hat{\beta}). \quad (18)$$

La desviación estándar $\widehat{SE}(\hat{\beta})$, se calcula como la raíz cuadrada de la varianza dada en (18). Para determinar la significación del coeficiente se presentan las tres estadísticas más usadas en la práctica: el test de razón de verosimilitudes, el test de Wald y el test de puntajes.

El test de la razón de verosimilitudes denotada por G, se calcula como dos veces la diferencia entre el logaritmo natural de la verosimilitud parcial del modelo que contiene la variable y el logaritmo de la verosimilitud parcial del modelo sin la variable, es decir:

$$G = 2\{L_p(\hat{\beta}) - L_p(0)\}, \quad (19)$$

donde

$$L_p(0) = - \sum_{i=1}^k \ln(n_i),$$

y n_i denota el numero de individuos en riesgo en el tiempo de supervivencia observado t_i . Bajo la hipótesis nula que el coeficiente es igual a cero, el estadístico G se distribuye chi-cuadrado con un grado libertad.

El test de Wald se calcula como la razón entre el coeficiente estimado y la desviación estándar de ese coeficiente como se muestra a continuación:

$$Z = \frac{\hat{\beta}}{\widehat{SE}(\hat{\beta})}.$$

Bajo la hipótesis nula que el coeficiente es igual a cero, el estadístico de Wald se distribuye normal estándar.

El test de puntajes se calcula como la razón de la derivada del logaritmo natural de la verosimilitud parcial y la raíz cuadrada de la información observada evaluadas las dos en $\beta = 0$ como se muestra abajo:

$$Z = \frac{\partial L_p / \partial \beta}{\sqrt{I(\beta)}}, \quad \beta = 0.$$

Bajo la hipótesis nula que el coeficiente es igual a cero, el estadístico de puntaje se distribuye normal estándar.

Los intervalos de confianza se construyen de la siguiente manera:

$$\hat{\beta} \pm Z_{1-\alpha/2} \widehat{SE}(\hat{\beta}).$$

Lo anterior visto corresponde al caso en que existe solo una covariable, en adelante se generalizara el procedimiento visto al caso de múltiples variables.

Sean p covariables medidas para el individuo i que deban ser incluidas en el modelo de regresión y denotadas por $x'_i = (x_{i1}, x_{i2}, \dots, x_{ip})$. Estas covariables pueden ser de tipo continuo, nominales o dummy. El nuevo modelo multivariante se construye de manera similar que el del caso univariado, solo que ahora se deben estimar p parámetros. En la expresión (9) se sustituye la variable unidimensional x por la variable multidimensional X , quedando así:

$$h(t, X, \beta) = h_0(t) e^{\beta X'}, \quad (20)$$

La función de verosimilitud parcial se construye de la misma manera como se construyó el modelo, simplemente se reemplaza en (14), la variable unidimensional x por la variable multidimensional X , quedando así:

$$l_p = \prod_{i=1}^k \frac{e^{\beta x'(i)}}{\sum_{j \in R(t_i)} e^{\beta x'(j)}}.$$

Después de aplicar logaritmo natural a (18) y calcular p derivadas parciales, una por cada covariable en el modelo, se tienen p ecuaciones como se muestra a continuación:

$$\frac{\partial L_p(\beta)}{\partial \beta_m} = \sum_{i=1}^k \left\{ x_{(im)} - \frac{\sum_{j \in R(t_i)} x_{jm} e^{\beta x'_{ij}}}{\sum_{j \in R(t_i)} e^{\beta x'_{ij}}} \right\} \quad m = 1, \dots, p. \quad (21)$$

Entonces el estimador de máxima verosimilitud parcial se denota por $\hat{\beta}' = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_m)$.

La matriz de información es obtenida de la matriz Hessiana de la siguiente forma

$$I(\beta) = -\frac{\partial^2 L_p(\beta)}{\partial \beta^2},$$

donde los elementos de la diagonal vienen dados por:

$$\frac{\partial^2 L_p(\beta)}{\partial \beta_m \partial \beta_m} = -\sum_{i=1}^k \sum_{j \in R(t_i)} \frac{e^{\beta x'_{ij}}}{\sum_{j \in R(t_i)} e^{\beta x'_{ij}}} \left[x_{jm} - \frac{\sum_{j \in R(t_i)} \frac{e^{\beta x'_{ij}}}{\sum_{j \in R(t_i)} e^{\beta x'_{ij}}} x_{jm} \right]^2.$$

Los elementos fuera de la diagonal vienen dados por:

$$\frac{\partial^2 L_p(\beta)}{\partial \beta_l \partial \beta_m} = -\sum_{i=1}^k \sum_{j \in R(t_i)} \frac{e^{\beta x'_{ij}}}{\sum_{j \in R(t_i)} e^{\beta x'_{ij}}} \left[x_{jl} - \frac{\sum_{j \in R(t_i)} \frac{e^{\beta x'_{ij}}}{\sum_{j \in R(t_i)} e^{\beta x'_{ij}}} x_{jl} \right] \left[x_{jm} - \frac{\sum_{j \in R(t_i)} \frac{e^{\beta x'_{ij}}}{\sum_{j \in R(t_i)} e^{\beta x'_{ij}}} x_{jm} \right].$$

El estimador de la matriz de covarianzas de los estimadores de máxima verosimilitud parcial de los parámetros del modelo vienen dados por:

$$\widehat{Var}(\hat{\beta}) = I^{-1}(\hat{\beta}).$$

Las pruebas estadísticas vistas para el caso univariante se pueden generalizar para el caso multivariante realizando unos ajustes.

La prueba de la razón de verosimilitud G, se calcula igual que al caso univariante, como dos veces la diferencia entre el logaritmo natural de la verosimilitud parcial del modelo que contiene todas las variables y el logaritmo de la verosimilitud parcial del modelo sin las variables, es decir:

$$G = 2\{L_p(\hat{\beta}) - L_p(0)\}, \quad (22)$$

donde $L_p(0)$ representa la verosimilitud del modelo con cero variables iguales que en (19).

En este caso G también se distribuye, bajo la hipótesis nula de que los coeficientes son iguales

a cero, como una chi-cuadrada con p grados de libertad (un grado por cada variable en el modelo).

Para calcular los estadísticos de Wald y de puntuación estadística implican cálculos matriciales. Si se denota el vector de primeras derivadas parciales de la función parcial de verosimilitud evaluando en 0 como $U(0) = U(\beta)|_{\beta=0}$, se tiene que bajo la hipótesis nula de que todos los coeficientes son iguales a cero y algunas otras condiciones de la función de verosimilitud parcial, este vector se distribuye normal multivariado con media igual a cero y matriz de covarianzas dada por la matriz de información evaluada en el vector 0, $I(0) = I(\beta)|_{\beta=0}$. La matriz de información es obtenida de la expresión (21).

La prueba de puntajes para el caso multivariante se calcula entonces como:

$$U'(0)[I(0)]^{-1}U(0). \quad (23)$$

Bajo la hipótesis nula de que todos los coeficientes son iguales a cero, la expresión (23) se distribuye chi-cuadrado con p grados de libertad (un grado por cada variable en el modelo).

La prueba de Wald se obtiene del hecho de que $\hat{\beta}$, el estimador de los coeficientes se distribuirá asintóticamente normal con vector de medias igual a cero y una matriz de covarianzas dada por la ecuación $I(\beta) = -\frac{\partial^2 L_p(\beta)}{\partial \beta \partial \beta}$. El estadístico de Wald se escribe como:

$$\hat{\beta}'I(\hat{\beta})\hat{\beta} \quad (24)$$

Bajo la hipótesis nula de que todos los coeficientes son iguales a cero la expresión (24) se distribuye chi-cuadrado con p grados de libertad (un grado por cada variable en el modelo).

5.5 CENSURA

Un dato censurado representa, pues, un individuo que desaparece, que lo hemos tenido pero que, antes de haberse producido el suceso que analizamos, ha desaparecido, sea porque lo perdemos del estudio o sea porque el estudio ha terminado y no se ha producido el suceso. Un dato censurado no es un dato que no nos dé información. De hecho, nos la da, pero parcial. Hay que saberlo aprovechar. Prescindir de él, sin más, sería desaprovechar información.

Formalmente se dice que una observación está **censurada a la derecha** en L si el valor exacto de la observación es desconocido, pero se sabe que es mayor o igual que L . **Se observa un individuo hasta un tiempo L en el que se pierde el contacto con él por distintos motivos.** De forma análoga, una observación está censurada a la izquierda en L si se sabe que la observación es menor o igual a L . Lo más común es que los datos presenten censura a la derecha. Finalmente, se habla de censura aleatoria cuando cada individuo tiene un tiempo de vida y uno de censura aleatorios.

5.5.1 TIPOS DE CENSURA

-Censura tipo I (en tiempo): En este caso se observan n items terminando el estudio en un tiempo L prefijado de antemano. Este tipo de censura se conoce con el nombre de en tiempo. En este caso el número de fallos hasta el tiempo L prefijado es aleatorio, no fijo como en el caso de censura tipo II.

-Censura tipo II (en número de fallos): Una muestra censurada tipo II es aquella en la que sólo se consideran las r primeras observaciones habiendo n items en estudio, con $1 \leq r \leq n$. En este caso r es un valor fijo, determinado previamente por la naturaleza del estudio u otros motivos.

-Multicensura: Es frecuente en algunos ensayos de fiabilidad asignar a cada ítem i tiempo de censura fijo y tiempo de vida o fallo aleatorio. A estos tiempos los notamos por L_i y T_i respectivamente, para $1 \leq i \leq n$.

-Censura Aleatoria: Si avanzamos un poco en el caso anterior y se considera que la censura es aleatoria estamos en el caso de muestreo con censura aleatoria. Suponemos que el individuo i tiene un tiempo de vida T_i y un tiempo de censura L_i , variables aleatorias independientes entre ellas y para todo i con funciones de supervivencia $S(t)$ y $G(t)$ respectivamente.

-Datos Perdidos: Cuando se realiza un estudio de supervivencia haciendo un seguimiento de n individuos durante un periodo de tiempo determinado, es posible perder la información de alguno de ellos antes de finalizar el período. En un estudio con censura tipo I es sabido que el tiempo de supervivencia supera al de censura, pero en datos perdidos no se sabe lo que ha ocurrido.

Podemos concluir en esta sección que con el modelo de Cox encontramos la función de riesgo $h(t)$ en función del tiempo t y de un conjunto de covariables / variables explicativas / predictores / factores de riesgo / variables de confusión, es decir, estas son todas las variables que son significativas en el estudio de la variable de vida. Con la función de riesgos podemos calcular la probabilidad de recurrencia de un individuo en un tiempo dado y en base a todas las variables que le afectan, también con estas covariables podemos obtener una función de supervivencia global. También con el análisis de estas curvas (supervivencia por el estimador de Kaplan-Meire) junto con la regresión de Cox nos proporciona tres resultados de interés: la mediana de tiempo en el que han presentado el suceso el 50% de los participantes, el tiempo medio en el que se produce el evento y la tasa de riesgos instantáneos, más conocida por su nombre en inglés, hazard ratio (HR). Esta tasa de riesgo va variando a lo largo de la duración del estudio, por lo que el modelo calcula una hipotética tasa instantánea media cuando el periodo de tiempo tiende a cero. La HR no es más que el cociente de las tasas

instantáneas entre el grupo de intervención y el control. Por lo tanto, la HR final calculada por el modelo de regresión es un promedio de las razones de tasas instantáneas en cada momento durante la duración del estudio y representa la relación de probabilidades de producción del suceso en tratados frente a no tratados en un momento dado, lo que representa cuánto más probable es que se produzca el suceso a que no se produzca en tratados frente a controles.

5.6 HERRAMIENTAS ESTADISTICAS

Para el procesamiento estadístico, carga de datos, construcción de modelos, graficas de funciones de supervivencia y demás análisis exploratorio de los datos, se realiza por medio de SPSS Statistics 21 del paquete de IBM SPSS Statistics y de RStudio 1.4.1717

SPSS Statistics 21 es una potente plataforma de software estadístico. Ofrece una interfaz fácil de usar y un sólido conjunto de características que permite extraer rápidamente información procesable de los datos. Los procedimientos estadísticos avanzados ayudan a garantizar una toma de decisiones de alta precisión y calidad. Incluyen todas las facetas de los análisis, desde la preparación y gestión de datos hasta el análisis y la generación de informes. Específicamente utilizaremos las gráficas de supervivencias que obtenemos con el estimador de Kaplan-Meier de la pestaña de analizar y la opción de supervivencia, así como también el modelo de regresión de Cox.

R es un software libre para el análisis estadístico de datos. Nace en la segunda mitad de los años noventa y ha ganado popularidad ya que: su adquisición es gratuita; se pueden llevar a cabo los mismos análisis estadísticos que con S+; estadísticos de todo el mundo contribuyen con paquetes que permiten realizar análisis cada vez más específicos y sofisticados; y posee una versatilidad gráfica única destacando su variedad y facilidad de adaptación. Funciona por comandos, lo que requiere introducir instrucciones que se deben conocer previamente.

RStudio es un interfaz que permite acceder a todas las herramientas de R, facilitando tanto la tarea de uso interactivo, como la programación de scripts mediante el orden y la visualización de los procesos que son llevados a cabo en R de manera simultánea.

Existe una gran variedad de paquetes destinados a realizar análisis de supervivencia en R. En particular, en este trabajo utilizaremos el paquete “survival” que está disponible en R-cran. Entre las múltiples funciones que contiene este paquete, destacan:

survfit

Calcula una estimación de una curva de supervivencia para datos censurados utilizando el método de Kaplan-Meier o Fleming-Harrington o calcula la función de supervivencia prevista para un modelo de riesgos proporcionales de Cox.

ggsurvplot (del paquete survminer)

Es una función genérica para trazar curvas de supervivencia. Envoltura de las `ggsurvplot_xx()` funciones familiares. Puede trazar uno o una lista de objetos `survfit` generados por las funciones `survfit.formula ()` y `surv_fit`.

coxph

Se ajusta a un modelo de regresión de riesgos proporcionales de Cox. Las variables dependientes del tiempo, los estratos dependientes del tiempo, los eventos múltiples por sujeto y otras extensiones se incorporan utilizando la formulación del proceso de conteo de Andersen y Gill.

VI METODOLOGÍA

Una vez se cuenta con la información básica sobre el tema en cuestión y con los objetivos en mente, se procede a realizar el respectivo análisis. Para ello, se describe el tipo de estudio, la base de datos utilizada, las aplicaciones informáticas estadísticas y comandos utilizados en el estudio de supervivencia.

6.1 TIPO DE ESTUDIO

El presente estudio se enmarca como tipo cuantitativo y descriptivo. Es de tipo cuantitativo ya que nos basamos en puntuaciones numéricas ponderadas para determinar la relación entre el tiempo de recurrencia y factores como tratamiento, primeramente, pero también con otros factores como puede ser la edad, el sexo, el número de ganglios linfáticos encontrados en su diagnóstico, etc. Es descriptivo porque analiza el contexto de una variable concreta, nuestro estudio se basa en el tiempo de recurrencia. Un análisis descriptivo permite, entre otras cosas, detallar como se manifiesta un fenómeno o evento y buscan especificar las propiedades y/o características de este mediante técnica de análisis científico (Yannis y Nikolaos, 2018).

6.2 BASE DE DATOS

El origen de los datos con los que se trabajó en esta investigación corresponde a los estudios realizados sobre ensayos clínicos de quimioterapia adyuvante en pacientes con cáncer de colon en estadio B/C en Laurie, Estados Unidos dados a conocer en 1989.

Este fue un ensayo intergrupar nacional que fue patrocinado por el Instituto Nacional del Cáncer e involucró al Grupo cooperativo de Oncología del Este, el NCCTG, el Grupo de Oncología del Suroeste y la Clínica Mayo. La inscripción de pacientes se inició en marzo de 1984, cuando un análisis preliminar del estudio NCCTG indicó la probabilidad de una ventaja

del tratamiento para el levamisol más fluorouracilo y para el levamisol solo, con respecto al tiempo hasta la recurrencia. La inscripción se completó en octubre de 1987.

El informe principal se encuentra en Moertel (1990), el cual es el más cercano al del informe final de Moertel (1991), y se pueden encontrar de manera pública a través del enlace <https://dmkd.cs.vt.edu/projects/survival/data/>

6.3 CARACTERISTICAS Y ATRIBUTOS DE LOS DATOS

Los datos contienen registro de 888 personas sometidas al ensayo de quimioterapia adyuvante. Cada registro tiene un total de 15 variables en estudio y algunas de estas con atributos específicos como describimos a continuación (según el orden de la base de datos):

- time:** Tiempo hasta el evento o censura.
- status:** Indicador de recurrencia/censura (recurrencia=1, vivo=0).
- study:** 1 para todos los pacientes.
- rx:** Tratamiento -1 = Obs(Observación), 2 = Lev(Levamisol), 3 = Lev(Levamisol)+5-FU.
- sex:** Sexo del paciente (1 = masculino, 0 = femenino).
- age:** Edad en años del paciente.
- obstruct:** Obstrucción del colon por tumor (1 = si, 0 = no).
- perfor:** Perforación del colon (1 = si, 0 = no).
- adhere:** Adherencia a los órganos cercanos (1 = si, 0 = no).
- nodes:** Número de ganglios linfáticos con cáncer detectable.
- differ:** Diferenciación del tumor (1=bien, 2=moderado, 3=pobre).

- extent:** Extensión de la diseminación local (1=submucosa, 2=músculo, 3=serosa, 4=estructuras contiguas).
- surg:** Tiempo desde la cirugía hasta el registro (0 = corto, 1 = largo).
- node4:** Más de 4 ganglios linfáticos positivos (1 = si, 2 = no).
- etype:** Tipo de evento (1 = recurrencia, 2 = muerte).

Podemos observar por la descripción anterior de nuestras variables que nuestro estudio se centra en el tiempo (variable time) en base a la recurrencia del tumor (variable status, 1 = recurrencia), aunque se le haya practicado la quimioterapia adyuvante y la asociaremos a cualquiera de las otras variables y en especial al tratamiento (variable rx con sus tres atributos).

Nuestra variable status tendrá como evento a estudiar la recurrencia, atributo 1, donde el número de fallos es aleatorio en el tiempo, y tendrá como censura el atributo 0 = vivo. Observamos una censura por la derecha, puesto que se atribuye que todos los elementos censurados son por estar vivo. Concluimos que se determinó este atributo al terminar el estudio por lo que sería una censura tipo I o comúnmente llamada en tiempo.

VII RESULTADOS

En el presente apartado se exponen los resultados finales obtenidos. Como se mencionó anteriormente en la metodología, hemos analizado el tiempo con respecto a la recurrencia de volver a desarrollar cáncer de colon en los pacientes del estudio, aun cuando ya se le practicó quimioterapia adyuvante. Centramos nuestro análisis en este tiempo de recurrencia asociándolo con las demás variables proporcionadas. Para comenzar, mostraremos si existe relación del tiempo de recurrencia con respecto al tipo de tratamiento. La variable rx tiene los atributos Observación, Levamisol y Lev+5-FU. Para analizar esta relación utilizamos el estimador Kaplan-Meier con el test Log-Rank en el software SPSS, dándonos la gráfica respectiva de las funciones de supervivencia junto con la información necesaria:

7.1 TRATAMIENTO

De la tabla que se muestra a continuación, podemos observar que a 305 individuos se les aplicó Levamisol, de los cuales 172 tuvieron una recurrencia del tumor, es decir, el 56%, y una censura de 133, el 44% aproximadamente, que significa que algunos de éstos no volvieron a desarrollar cáncer de colon y otros se salieron del estudio. A 289 se les practicó la quimioterapia adyuvante con una combinación del Levamisol y el 5-FU, de los cuales, 113 volvieron a tener recurrencia, un 39%, mientras que 176 de los que algunos no volvieron a desarrollar el tumor y otros fueron censurados, un 61% aproximadamente. De estos datos vemos mejores porcentajes con respecto a la no recurrencia del cáncer de colon a los pacientes a los que se les aplicó la terapia adyuvante con una combinación de los medicamentos Levamisol y el 5-FU.

Tabla 3. Tabla resumen de la variable tratamiento (rx).

Resumen del procesamiento de los casos				
rx	Nº total	Nº de eventos	Censurado	
			Nº	Porcentaje
Observacion	294	161	133	45.2%
Levamisol	305	172	133	43.6%
Lev+5-FU	289	113	176	60.9%
Global	888	446	442	49.8%

También podemos notar cómo tercer atributo de la variable rx, la observación, la cual podría contener personas que se sometieron con otros tipos de medicamentos a la terapia adyuvante. Análogamente podemos hacer el mismo análisis para esta tercera componente, al analizar las curvas de supervivencias llegaremos a comparar estos tres atributos.

Tabla 4. Tabla comparaciones globales para la variable tratamiento (rx).

Comparaciones globales			
	Chi-cuadrado	gl	Sig.
Log Rank (Mantel-Cox)	21.950	2	.000

Prueba de igualdad de distribuciones de supervivencia para diferentes niveles de rx.

El cuadro de comparaciones globales para la variable rx, con el test Log-Rank da un p -valor menor a 0.05 por lo que, sí existe significación en el tratamiento con respecto al tiempo de recurrencia, es decir, el tratamiento sí influye en el tiempo de recurrencia del tumor, tienen significación estadística. Por lo que se vuelve pertinente analizar nuestras curvas de supervivencia.

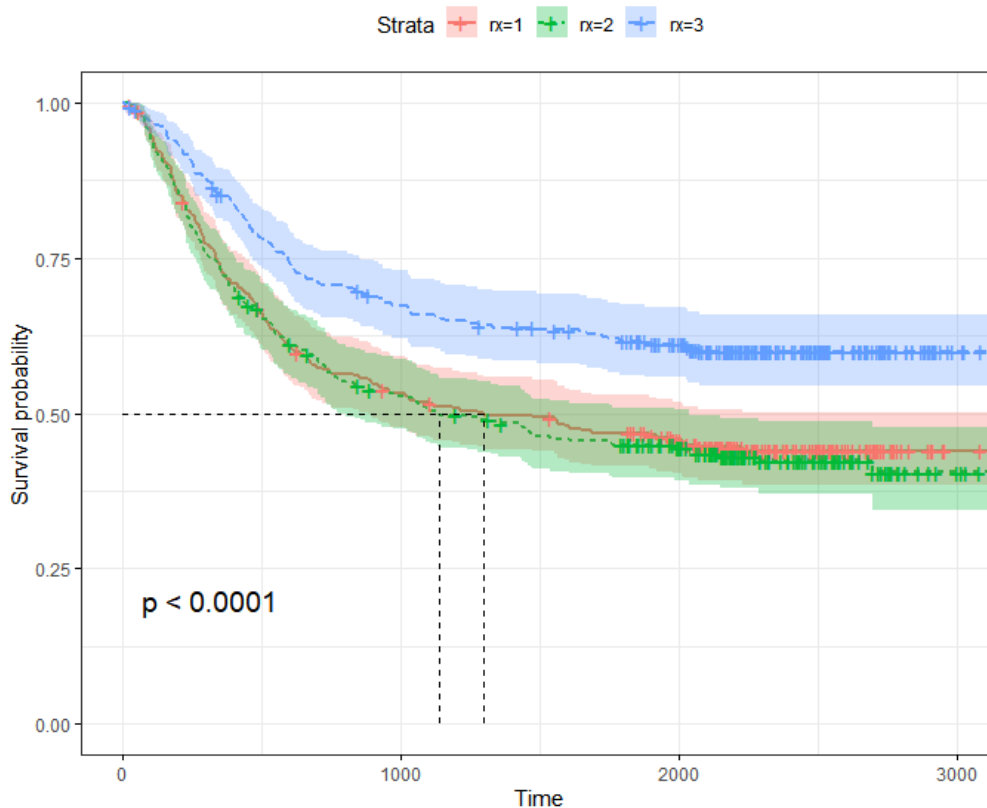


Gráfico 1 Gráfico función de supervivencia para la variable tratamiento

Notablemente en las curvas de supervivencias, la quimioterapia adyuvante con la combinación de los medicamentos Levamisol y el 5-FU permiten una media mayor de supervivencia en comparación a los otros dos atributos de la variable, de hecho, se puede observar que con el Levamisol y el atributo Observación, sus medias de supervivencia son casi iguales.

Comprobamos estas afirmaciones con el cuadro de la media y mediana del tiempo de supervivencia. En él se puede observar que el mejor de los tres tratamientos es en el que se hizo una combinación entre los medicamentos Levamisol con el 5-FU, pues la media de supervivencia es de 2227 días mayor a los otros tratamientos. El Levamisol por si solo tiene una media de supervivencia de 1671 días, que incluso es menor a la media de supervivencia del tercer tratamiento que es de 1777 días. Con el tratamiento Levamisol y 5-FU se espera que la media poblacional de supervivencia se encuentre entre 2069 y 2386 días, con un 95% de confianza

Tabla 5. Tabla para las medias y medianas del tiempo de supervivencia variable tratamiento (rx)

rx	Medias y medianas del tiempo de supervivencia							
	Media ^a				Mediana			
	Estimación	Error típico	Intervalo de confianza al 95%		Estimación	Error típico	Intervalo de confianza al 95%	
			Límite inferior	Límite superior			Límite inferior	Límite superior
Observacion	1776.889	84.065	1612.122	1941.656	1298.000	322.397	666.101	1929.899
Levamisol	1670.898	78.370	1517.292	1824.504	1139.000	275.859	598.317	1679.683
Lev+5-FU	2227.384	80.770	2069.075	2385.692
Global	1910.458	48.303	1815.784	2005.133	2028.000	.	.	.

a. La estimación se limita al mayor tiempo de supervivencia si se ha censurado.

De lo expuesto hasta este momento, según la información arrojada, el 39% de los que los que se sometieron a la quimioterapia adyuvante con una combinación del Levamisol y el 5-FU desarrollaron la enfermedad, mientras que en el 61% restante se encuentran aquellos que no desarrollaron la enfermedad o son datos censurados del estudio. Para aquellos que sí la

desarrollaron, su media, es decir, el tiempo en días en el que se les volvió a desarrollar fue mucho más largo en comparación a los otros tratamientos.

Utilizamos la prueba de Log-Rank para contrastar que entre los tratamientos Observación (1) y Levamisol (2) no hay diferencia significativa ($p\text{-valor}=0.712>0.05$). Esto ya lo habíamos intuido por medio de las curvas de supervivencia.

Tabla 6 Tabla comparaciones globales modalidades variable tratamiento

Comparaciones globales			
	Chi-cuadrado	gl	Sig.
Log Rank (Mantel-Cox)	.136	1	.712

Prueba de igualdad de distribuciones de supervivencia para diferentes niveles de rx.

Con esto corroboramos que la variable tratamiento (rx) es significativa en el estudio debido al atributo (3) de esta.

7.2 MODELO COX (1972)

Aplicando el modelo Cox en RStudio podemos observar no solo la significancia en el tiempo de recurrencia con respecto al tratamiento, sino también con cada una de las variables proporcionadas.

```
call:
coxph(formula = surv(colon$time, colon$status) ~ colon$study +
      colon$rx + colon$sex + colon$age + colon$obstruct + colon$perfor +
      colon$adhere + colon$nodes + colon$differ + colon$extent +
      colon$surg + colon$node4 + colon$type, data = colon)

n= 888, number of events= 446

      coef exp(coef) se(coef)      z Pr(>|z|)
colon$study      NA      NA 0.000000      NA      NA
colon$rx      -0.221807  0.801070  0.058781 -3.773 0.000161 ***
colon$sex      -0.139191  0.870062  0.095963 -1.450 0.146929
colon$age      -0.003001  0.997004  0.004064 -0.738 0.460281
colon$obstruct  0.211443  1.235460  0.119315  1.772 0.076370 .
colon$perfor    0.194404  1.214587  0.256875  0.757 0.449167
colon$adhere    0.158365  1.171593  0.130044  1.218 0.223309
colon$nodes     0.035841  1.036491  0.015102  2.373 0.017630 *
colon$differ    0.157323  1.170374  0.097642  1.611 0.107132
colon$extent    0.451318  1.570380  0.120132  3.757 0.000172 ***
colon$surg     0.262144  1.299714  0.103814  2.525 0.011565 *
colon$node4    0.608975  1.838546  0.141296  4.310 1.63e-05 ***
colon$type      NA      NA 0.000000      NA      NA
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ilustración 1 Modelo de Cox

Podemos ver la significación en la variable rx que se estudió previamente, pero también en las variables nodes (Número de ganglios linfáticos con cáncer detectable), extent (Extensión de la diseminación local: 1=submucosa, 2=músculo, 3=serosa, 4=estructuras contiguas), surg (Tiempo desde la cirugía hasta el registro: 0 = corto, 1 = largo) y node4 (Más de 4 ganglios linfáticos positivos: 1 = sí, 2 = no). Entre las variables con mayor significancia tenemos rx (0.000161) y node4 (0.0000163), incluso hay mayor significancia en la variable node4, lo que significa que el tiempo de recurrencia se ve bastante afectado cuando el número de ganglios linfáticos positivos encontrados en su diagnóstico eran 4 o más. La variable extent tiene casi el mismo nivel de significancia que la variable rx y en menor significancia tenemos las variables nodes y surg. El resto de las variables no tienen significancia en el modelo, es decir, variables como el sexo o como la edad no afectan el tiempo de recurrencia en la enfermedad.

Aplicamos nuevamente el modelo de Cox con las variables más significativas, estas son: rx, extent y node4. En esta ocasión estudiamos los atributos de estas variables, con la finalidad de determinar la significancia sobre la recurrencia en la enfermedad que tiene las subdivisiones de estas variables. Para ello definimos estas, como variables dummy, de la siguiente manera:

-rx_Levamisol: 0 (no), 1 (sí)

-rx_Lev_mas_FU: 0 (no), 1 (sí)

-rx_submucosa: 0 (no), 1 (sí)

-extent_musculo: 0 (no), 1 (sí)

-extent_serosa: 0 (no), 1 (sí)

-extent_est_contigua: 0 (no), 1 (sí)

Los resultados se muestran a continuación:

```

> res.cox<-coxph(formula=Surv(cancer_de_colon_1$time,cancer_de_colon_1$status)~cancer
_de_colon_1$rx_Levamisol+cancer_de_colon_1$rx_Lev_mas_FU+cancer_de_colon_1$extent_sub
mucosa+cancer_de_colon_1$extent_musculo+cancer_de_colon_1$extent_serosa+cancer_de_col
on_1$extent_est_contigua+cancer_de_colon_1$node4,data=cancer_de_colon_1)
> summary(res.cox)
Call:
coxph(formula = Surv(cancer_de_colon_1$time, cancer_de_colon_1$status) ~
      cancer_de_colon_1$rx_Levamisol + cancer_de_colon_1$rx_Lev_mas_FU +
      cancer_de_colon_1$extent_submucosa + cancer_de_colon_1$extent_musculo +
      cancer_de_colon_1$extent_serosa + cancer_de_colon_1$extent_est_contigua +
      cancer_de_colon_1$node4, data = cancer_de_colon_1)

n= 888, number of events= 446

              coef exp(coef) se(coef)      z Pr(>|z|)
cancer_de_colon_1$rx_Levamisol    0.04684  1.04795  0.10995  0.426 0.670114
cancer_de_colon_1$rx_Lev_mas_FU  -0.46617  0.62740  0.12291 -3.793 0.000149
cancer_de_colon_1$extent_submucosa -1.37886  0.25187  0.54164 -2.546 0.010906
cancer_de_colon_1$extent_musculo  -1.08243  0.33877  0.26730 -4.050 5.13e-05
cancer_de_colon_1$extent_serosa    -0.50746  0.60203  0.21120 -2.403 0.016273
cancer_de_colon_1$extent_est_contigua NA      NA      0.00000    NA      NA
cancer_de_colon_1$node4           0.83508  2.30500  0.09905  8.431 < 2e-16

cancer_de_colon_1$rx_Levamisol
cancer_de_colon_1$rx_Lev_mas_FU      ***
cancer_de_colon_1$extent_submucosa   *
cancer_de_colon_1$extent_musculo     ***
cancer_de_colon_1$extent_serosa      *
cancer_de_colon_1$extent_est_contigua
cancer_de_colon_1$node4              ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
cancer_de_colon_1$rx_Levamisol      1.0480    0.9542  0.84480  1.3000
cancer_de_colon_1$rx_Lev_mas_FU      0.6274    1.5939  0.49308  0.7983
cancer_de_colon_1$extent_submucosa    0.2519    3.9704  0.08712  0.7281
cancer_de_colon_1$extent_musculo      0.3388    2.9518  0.20063  0.5720
cancer_de_colon_1$extent_serosa       0.6020    1.6611  0.39796  0.9107
cancer_de_colon_1$extent_est_contigua NA      NA      NA      NA
cancer_de_colon_1$node4              2.3050    0.4338  1.89829  2.7989

Concordance= 0.649 (se = 0.013 )
Likelihood ratio test= 115.3 on 6 df,  p=<2e-16
Wald test               = 117 on 6 df,  p=<2e-16
score (logrank) test = 124.3 on 6 df,  p=<2e-16

```

Ilustración 2

Nuevamente se logra ver la significancia que existe en la variable rx y node4, específicamente el atributo (3), uso de Levamisol y 5FU, para el caso de la variable rx. También se observa que en la variable extent, el atributo musculo (2) tiene mayor significancia que en los otros atributos.

Como habíamos observado anteriormente la variable tratamiento es significativa por el atributo (3), el tratamiento Levamisol(2) no es determinante para una mayor supervivencia con respecto a la recurrencia de la enfermedad, por lo que se puede omitir para nuestro

estudio. En el caso de la variable extent no se observa significancia para el atributo estructuras contiguas, esto se debe a que tenemos pocos datos para este atributo y seguramente todos o en su mayoría censurados, para el análisis de nuestros datos la omitimos, pero en caso de tomar mas datos en otra investigación, este atributo si puede ser significativo.

Es importante resaltar que la variable node4 sigue conservando la máxima significancia con respecto al resto de las demás variables. Mostraremos las curvas de supervivencia con el estimador de Kaplan-Meier para estas variables significativas y luego mostraremos la función de riesgo con el modelo de Regresión de Cox.

7.3 VARIABLE NODE4

Para la variable node4, que tiene los atributos 1 si en su diagnóstico tenía 4 o más ganglios linfáticos positivo y 2 si no era este el caso, la gráfica de supervivencia nos queda:

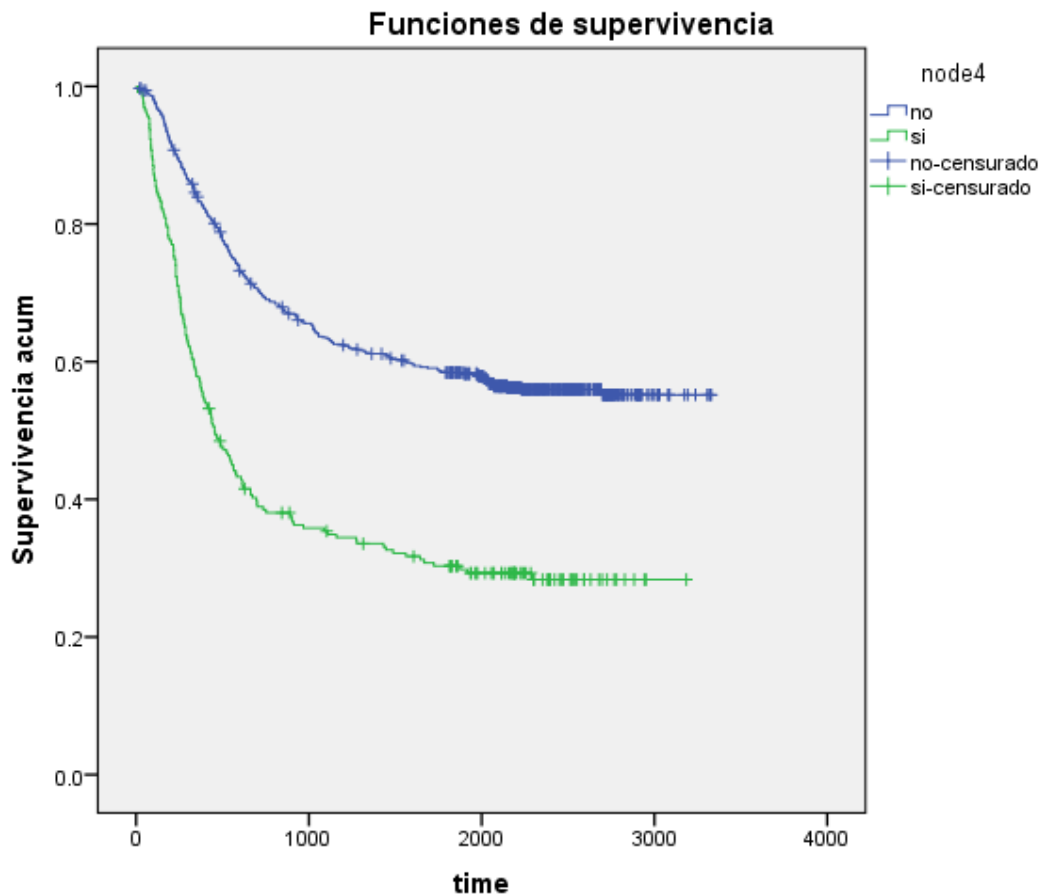


Grafico 2 Gráfico función de supervivencia variable node4

Tabla 7. Tabla comparaciones globales variable node4

Comparaciones globales			
	Chi-cuadrado	gl	Sig.
Log Rank (Mantel-Cox)	82.934	1	.000

Prueba de igualdad de distribuciones de supervivencia para diferentes niveles de node4.

Tabla 8 Tabla para las medias y medianas del tiempo de supervivencia variable node4

Medias y medianas del tiempo de supervivencia								
node4	Media ^a				Mediana			
	Estimación	Error típico	Intervalo de confianza al 95%		Estimación	Error típico	Intervalo de confianza al 95%	
			Límite inferior	Límite superior			Límite inferior	Límite superior
no	2145.032	54.387	2038.433	2251.631
si	1220.345	85.321	1053.115	1387.574	454.000	53.773	348.604	559.396
Global	1910.458	48.303	1815.784	2005.133	2028.000	.	.	.

a. La estimación se limita al mayor tiempo de supervivencia si se ha censurado.

Observamos significancia como determinamos con Cox. Evidentemente vemos en la curva una mayor supervivencia en los casos donde los pacientes no se les detectó más de 4 ganglios linfáticos positivos con el cáncer y con una media de supervivencia de 2145 días, muy favorable con comparación a la media de aquellos en los que sí se les detectó ganglios linfáticos positivos con una media de 1220 días. También un intervalo de confianza del 95% para la media más favorable de 2038 a 2251 días para la media poblacional.

7.4 VARIABLE EXTENT

Con respecto a la variable extent (Extensión de la diseminación local: 1=submucosa, 2=músculo, 3=serosa, 4=estructuras contiguas), los resultados son los siguientes:

Tabla 8. Tabla comparaciones globales variable extent

Comparaciones globales			
	Chi-cuadrado	gl	Sig.
Log Rank (Mantel-Cox)	25.770	3	.000

Prueba de igualdad de distribuciones de supervivencia para diferentes niveles de extent.

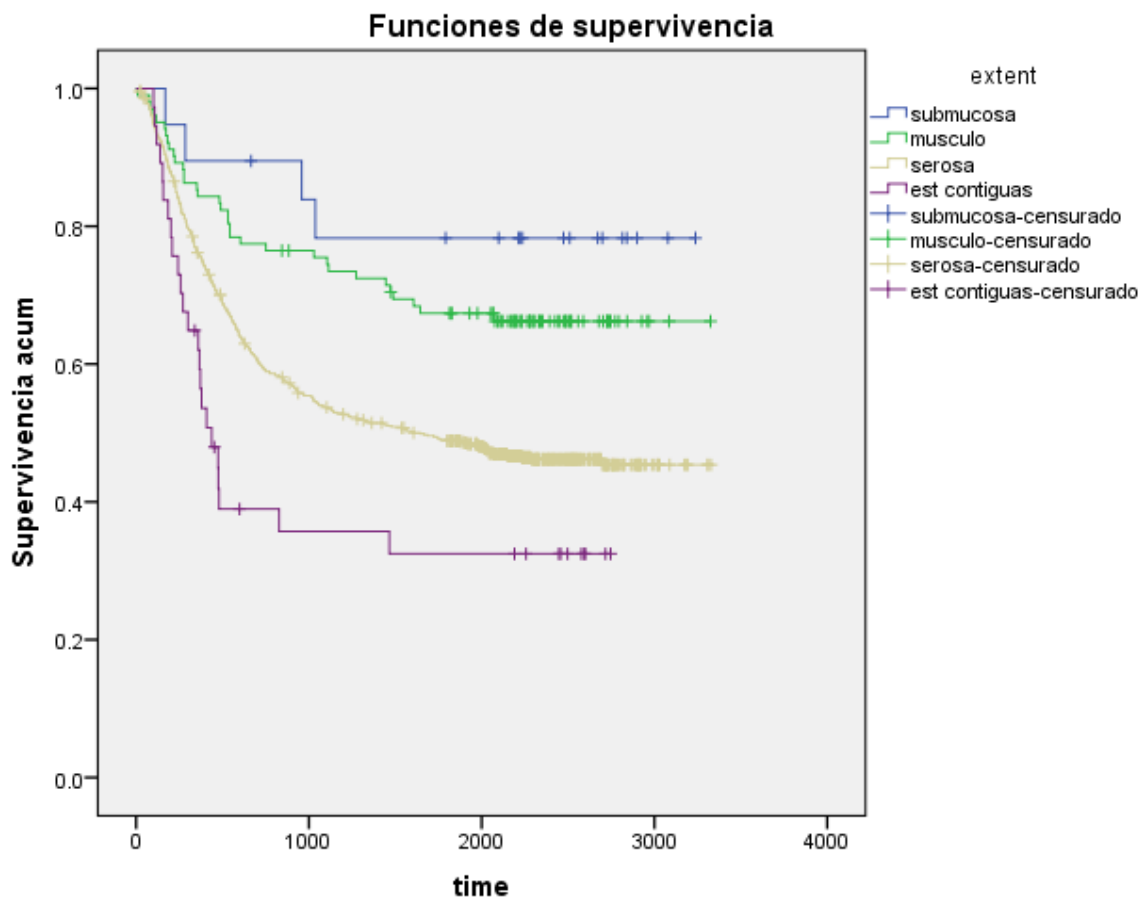


Grafico 3 Gráfico función de supervivencia variable extent

Tabla 9. Tabla para las medias y medianas del tiempo de supervivencia variable extent

Medias y medianas del tiempo de supervivencia

extent	Media ^a				Mediana			
	Estimación	Error típico	Intervalo de confianza al 95%		Estimación	Error típico	Intervalo de confianza al 95%	
			Límite inferior	Límite superior			Límite inferior	Límite superior
submucosa	2670.526	253.886	2172.909	3168.143
musculo	2422.695	129.724	2168.435	2676.955
serosa	1844.320	53.195	1740.059	1948.582	1647.000	.	.	.
est contiguas	1140.523	192.953	762.334	1518.711	437.000	73.911	292.135	581.865
Global	1910.458	48.303	1815.784	2005.133	2028.000	.	.	.

a. La estimación se limita al mayor tiempo de supervivencia si se ha censurado.

Podemos observar que dependiendo de dónde se extendió el tumor, también incide en la recurrencia de éste. La gráfica nos muestra la curva más baja de supervivencia cuando el cáncer se extendió hasta las estructuras contiguas y la supervivencia más alta cuando apenas llegó a la submucosa, donde también observamos mayor media de la supervivencia de 2761 días, es decir, si vuelve a recurrir la enfermedad lo hace en un lapso de tiempo más largo. Por el contrario cuando el tumor llegó a las estructuras contiguas tiene una media, un lapso de recurrencia más corto (1141 días).

Para este grupo de riesgo, se muestra la curva de supervivencia para el atributo musculo, en donde hemos determinado mayor significancia al momento de la recurrencia. El gráfico nos queda:

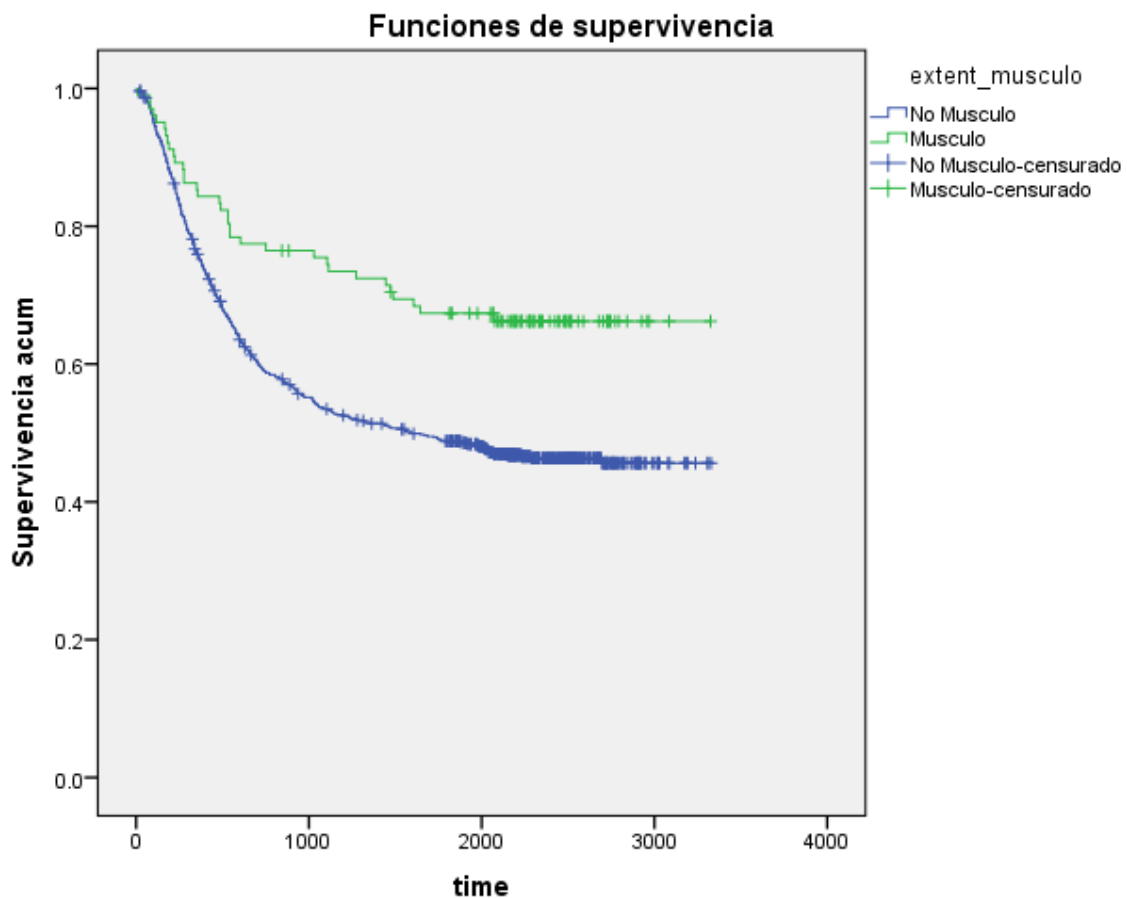


Gráfico 4 Gráfico curva de supervivencia variable extent_musculo

Tabla 11 Tabla resumen para extent_musculo

Resumen del procesamiento de los casos				
extent_musculo	Nº total	Nº de eventos	Censurado	
			Nº	Porcentaje
No Musculo	786	412	374	47.6%
Musculo	102	34	68	66.7%
Global	888	446	442	49.8%

Tabla 12 Tabla media y mediana para extent_musculo

Medias y medianas del tiempo de supervivencia								
extent_musculo	Media ^a				Mediana			
	Estimación	Error típico	Intervalo de confianza al 95%		Estimación	Error típico	Intervalo de confianza al 95%	
			Límite inferior	Límite superior			Límite inferior	Límite superior
No Musculo	1842.495	51.434	1741.684	1943.306	1606.000	.	.	.
Musculo	2422.695	129.724	2168.435	2676.955
Global	1910.458	48.303	1815.784	2005.133	2028.000	.	.	.

a. La estimación se limita al mayor tiempo de supervivencia si se ha censurado.

Tabla 13 Tabla comparaciones globales para extent_musculo

Comparaciones globales			
	Chi-cuadrado	gl	Sig.
Log Rank (Mantel-Cox)	12.712	1	.000

Prueba de igualdad de distribuciones de supervivencia para diferentes niveles de extent_musculo.

El *p-value* de Log-Rank, nuevamente nos confirma la significancia de este atributo en la recurrencia de la enfermedad, observando en nuestra grafica que tiene una mayor supervivencia una persona cuya lesión haya llegado al musculo en comparación a otros tejidos, una media de supervivencia de 2423 días.

7.5 VARIABLE NODES

Para la variable nodes (Número de ganglios linfáticos con cáncer detectable) tenemos:

Tabla 14. Tabla comparaciones globales variable nodes

Comparaciones globales			
	Chi-cuadrado	gl	Sig.
Log Rank (Mantel-Cox)	129.743	22	.000

Prueba de igualdad de distribuciones de supervivencia para diferentes niveles de nodes.

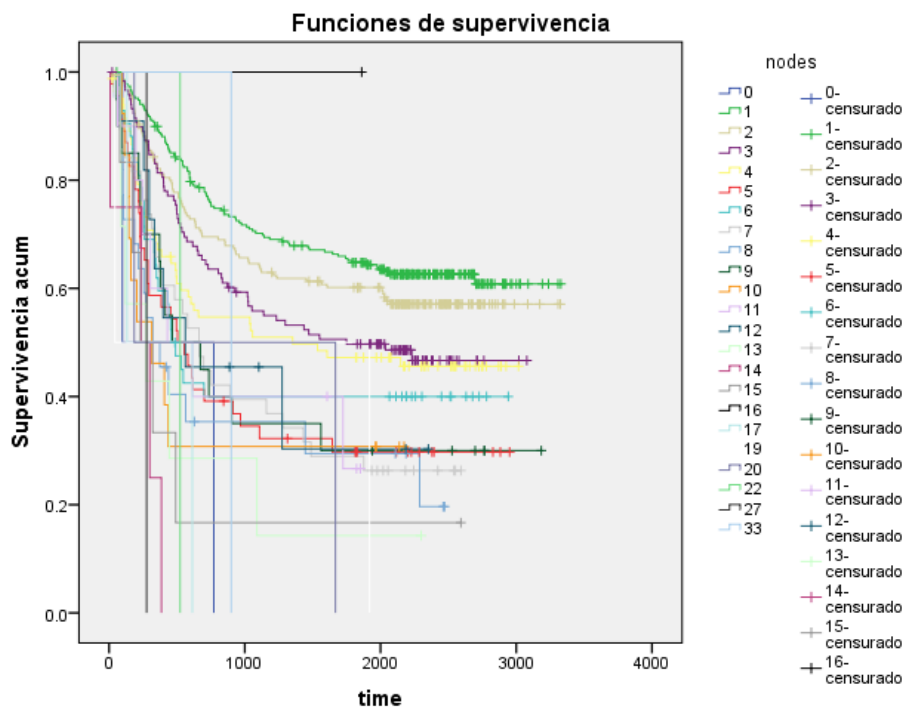


Gráfico 4 Gráfico función de supervivencia variable nodes

Advertencia

No se ha calculado ningún estadístico porque se han censurado todos los casos.

Cuanto mayor es el número de ganglios linfáticos detectados de manera positiva, la curva de supervivencia es mucho más baja. Esta variable está estrechamente relacionada con la variable node4, es decir, ya se había analizado que cuando un paciente se le había detectado 4 o más ganglios linfáticos positivos en su diagnóstico se observaba una mayor recurrencia

del tumor y con esta grafica nos comprueba que efectivamente la recurrencia se ve muy afectada por estas variantes.

Por todo lo antes expuesto, vemos en sí, que la recurrencia del fenómeno no se ve afectada solo por el tratamiento, sino por cuatro variables más, que tanto se extendió el tumor, el número de ganglios linfáticos infectados, etc., por lo que es necesario que con el modelo de Cox podamos generar nuestra curva de riesgos en base a todas las variables que afectan la recurrencia.

7.6 REGRESION DE COX

En esta sección, se genera primero los coeficientes para la función de riesgo, en forma general sobre las variables estudiadas, así como la curva de supervivencia. Introducimos luego nuestras variables dummy, para un estudio más detallado de esta curva de supervivencia en una segunda gráfica y los coeficientes para la función de riesgo.

Tabla 15. Tabla resumen regresión de Cox

Resumen del proceso de casos		N	Porcentaje
Casos disponibles en el análisis	Evento ^a	446	50.2%
	Censurado	442	49.8%
	Total	888	100.0%
Casos excluidos	Casos con valores perdidos	0	0.0%
	Casos con tiempo negativo	0	0.0%
	Casos censurados antes del evento más temprano en un estrato	0	0.0%
	Total	0	0.0%
Total		888	100.0%

a. Variable dependiente: time

Tabla 16. Tabla codificación variables categóricas

Codificaciones de variables categóricas^{a,c,e,f,g,h,i,j}

		Frecuencia	(1) ^d
study ^b	1	888	
sex ^b	0=femenino	428	1
	1=masculino	460	0
obstruct ^b	0=no	717	1
	1=si	171	0
perfor ^b	0=no	861	1
	1=si	27	0
adhere ^b	0=no	760	1
	1=si	128	0
surg ^b	0=corto	650	1
	1=largo	238	0
node4 ^b	0=0	653	1
	1=si	235	0
etype ^b	1=recurrencia	888	

- a. Variable de categoría: study
- b. Codificación de parámetros de indicador
- c. Variable de categoría: sex
- d. Se ha recodificado la variable (0,1), de manera que sus coeficientes no serán los mismos que para la codificación del indicador (0,1).
- e. Variable de categoría: obstruct
- f. Variable de categoría: perfor
- g. Variable de categoría: adhere
- h. Variable de categoría: surg
- i. Variable de categoría: node4
- j. Variable de categoría: etype

De la tabla que a continuación se presenta, vemos paso a paso el cálculo de los coeficientes que determinan la función de riesgo. En el paso 5 tenemos el coeficiente 0.812 para la variable rx, 1.038 para la variable nodes, 1.675 para la variable extent, 0.769 para la variable surg y 0.546 para la variable node4. Una vez obtenidos nuestros coeficientes, estamos listos para hacer una gráfica general de supervivencia con todas las variables que afectan a la recurrencia.

Tabla 17. Tabla variables de la ecuación

		Variables en la ecuación						95.0% IC para Exp(B)	
		B	ET	Wald	gl	Sig.	Exp(B)	Inferior	Superior
Paso 1	nodes	.084	.009	88.197	1	.000	1.088	1.069	1.107
Paso 2	nodes	.080	.009	75.627	1	.000	1.083	1.064	1.103
	extent	.507	.116	19.067	1	.000	1.660	1.322	2.084
Paso 3	nodes	.039	.015	6.712	1	.010	1.040	1.010	1.071
	extent	.509	.118	18.604	1	.000	1.663	1.320	2.096
	node4	-.586	.142	17.093	1	.000	.556	.421	.735
Paso 4	rx	-.209	.058	12.986	1	.000	.811	.724	.909
	nodes	.036	.015	5.814	1	.016	1.037	1.007	1.067
	extent	.507	.119	18.151	1	.000	1.660	1.315	2.096
	node4	-.604	.141	18.408	1	.000	.547	.415	.720
Paso 5	rx	-.209	.058	12.805	1	.000	.812	.724	.910
	nodes	.037	.015	6.084	1	.014	1.038	1.008	1.069
	extent	.516	.120	18.548	1	.000	1.675	1.325	2.118
	surg	-.262	.104	6.417	1	.011	.769	.628	.942
	node4	-.606	.141	18.444	1	.000	.546	.414	.719

Anteriormente, observamos la supervivencia por cada una de las variables, pero ahora analizamos la curva de supervivencia tomando en cuenta todas las variables que la afectan:

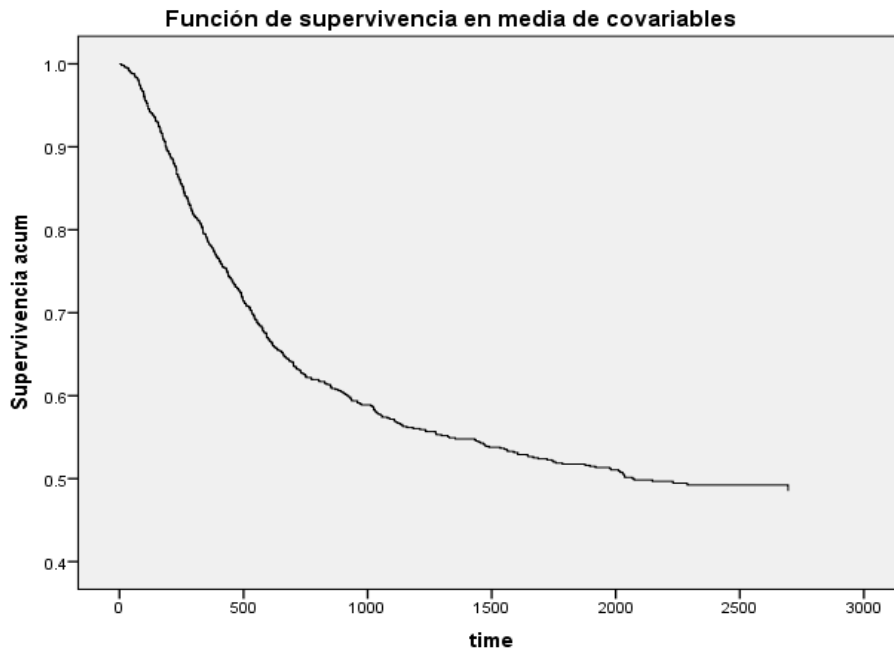


Gráfico 5 Gráfico función de supervivencia con las variables categóricas

De esta grafica podemos determinar la probabilidad de que un individuo no haya tenido una recurrencia en un determinado tiempo, por ejemplo, ¿Cuál es la probabilidad que una persona no haya tenido una recurrencia en la enfermedad a los 1000 días? Esta interrogante podemos contestarla observando el porcentaje acumulado cuando el valor de la abscisa (variable time) es igual a 1000.

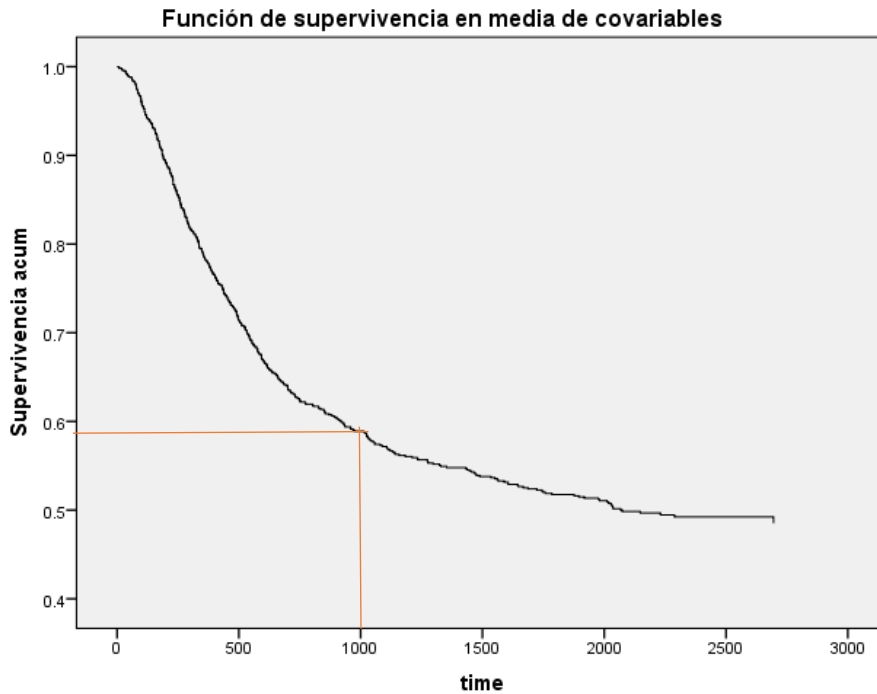


Gráfico 6 Grafico análisis de la función de supervivencia con las variables categóricas

La probabilidad que una persona en 1000 días no haya tenido una recurrencia en el cáncer de colon es de aproximadamente un 58%.

La función de riesgo nos queda:

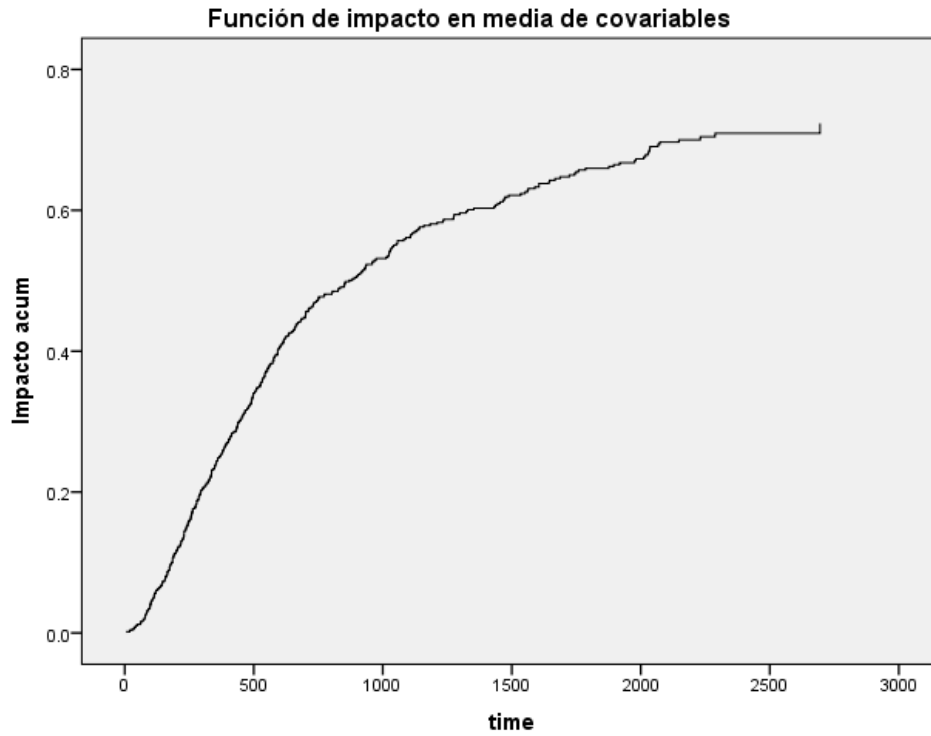


Grafico 7 Grafico función de riesgo con las variables categóricas

Procedemos ahora a obtener nuestros coeficientes para la función de riesgo y nuestra curva de supervivencia, introduciendo las modalidades de nuestras variables significativas como variables dummy.

Tabla 18 Tabla variables de la ecuacion incluyendo las dummy

		Variables en la ecuación					
		B	ET	Wald	gl	Sig.	Exp(B)
Paso 1	nodes	.084	.009	88.197	1	.000	1.088
Paso 2	rx_Lev_mas_FU	.480	.109	19.321	1	.000	1.616
	nodes	.082	.009	84.817	1	.000	1.085
Paso 3	node4	-.602	.138	18.920	1	.000	.548
	rx_Lev_mas_FU	.492	.109	20.355	1	.000	1.636
	nodes	.042	.014	8.651	1	.003	1.043
Paso 4	node4	-.603	.140	18.591	1	.000	.547
	extent_musculo	.554	.179	9.575	1	.002	1.741
	rx_Lev_mas_FU	.496	.109	20.715	1	.000	1.643
	nodes	.039	.015	7.062	1	.008	1.040
	node4	-.614	.141	19.114	1	.000	.541
Paso 5	extent_musculo	.532	.179	8.786	1	.003	1.702
	extent_est_contigua	-.502	.211	5.653	1	.017	.606
	rx_Lev_mas_FU	.494	.109	20.551	1	.000	1.639
	nodes	.038	.015	6.472	1	.011	1.038
	node4	-.614	.141	19.005	1	.000	.541
Paso 6	extent_musculo	.540	.180	9.041	1	.003	1.716
	extent_est_contigua	-.526	.211	6.200	1	.013	.591
	rx_Lev_mas_FU	.481	.109	19.376	1	.000	1.617
	nodes	.039	.015	6.714	1	.010	1.039
	surg	-.248	.104	5.718	1	.017	.780

Tenemos como coeficientes para la función de riesgo: 0.541 para la variable node4, 1.716 para la variable extent_musculo, 0.591 para la variable extent_est_contigua, 1.617 para la variable rx_Lev_mas_FU, 1.039 para la variable nodes y 0.780 para la variable surg. La curva de supervivencia y la función de riesgo tomando en cuenta estas modalidades nos quedan de la siguiente manera:

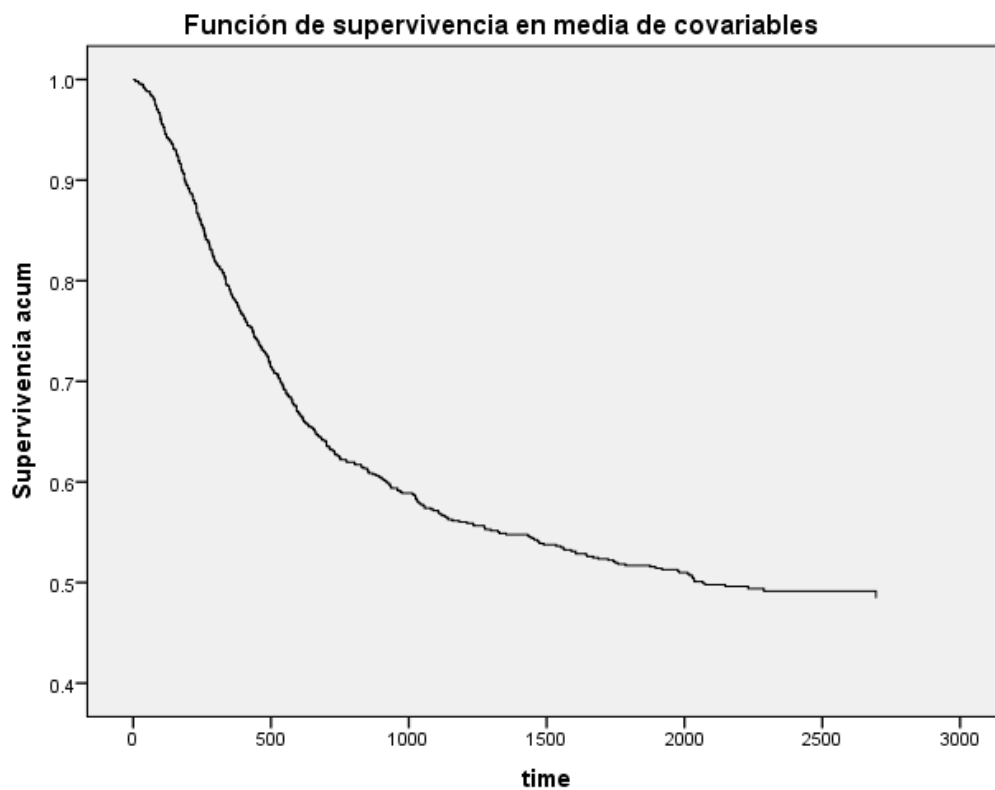


Gráfico 8 Función se supervivencia global con las variables dummy

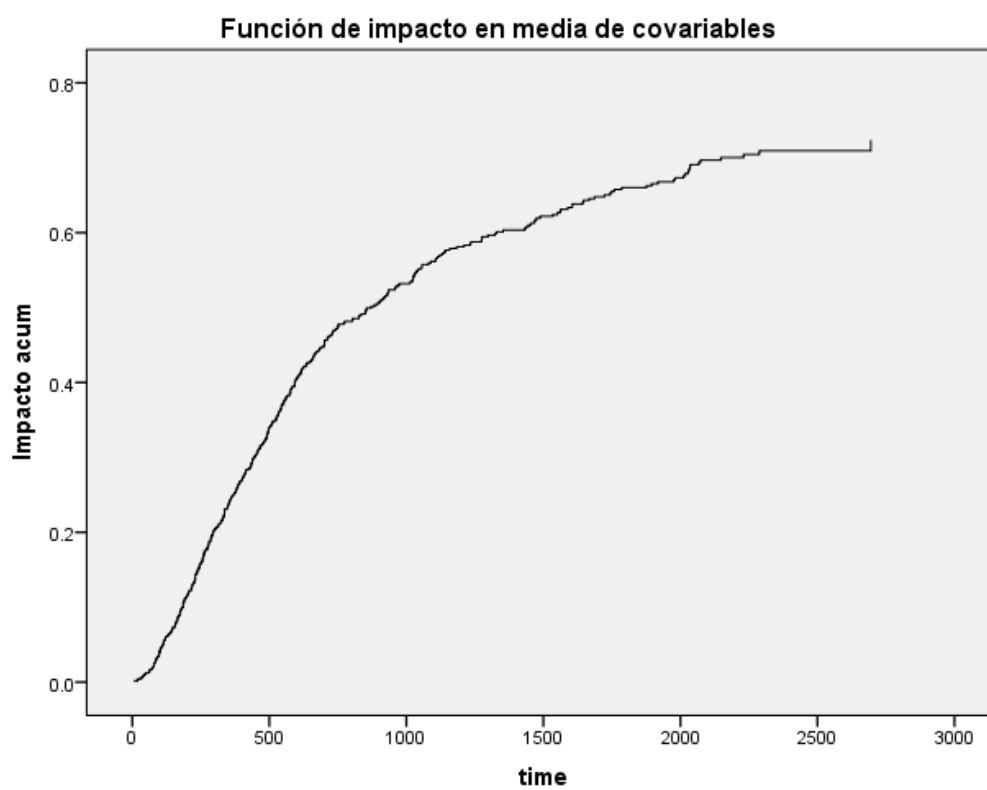


Gráfico 9 Función de riesgo global con las variables dummy

Es de hacer notar que tenemos las mismas graficas. Esto es porque en realidad son las mismas variables, aunque las estudiemos según sus modalidades, serán las mismas variables significativas que intervienen en ésta.

Con las cinco variables más significativas, el modelo y las funciones nos quedan:

Tabla 19 Tabla método regresión de Cox con las cinco variables más significativas

Codificaciones de variables categóricas^{a,d,e,f,g}		Frecuencia	(1)^c
node4 ^b	0=0	653	1
	1=Si	235	0
extent_submucosa ^b	0=No submucosa	869	1
	1=Submucosa	19	0
extent_musculo ^b	0=No Musculo	786	1
	1=Musculo	102	0
extent_serosa ^b	0=No Serosa	158	1
	1=Serosa	730	0
rx_Lev_mas_FU ^b	0=No Lev+5FU	599	1
	1=Lev+5FU	289	0

a. Variable de categoría: node4

b. Codificación de parámetros de indicador

c. Se ha recodificado la variable (0,1), de manera que sus coeficientes no serán los mismos que para la codificación del indicador (0,1).

d. Variable de categoría: extent_submucosa

e. Variable de categoría: extent_musculo

f. Variable de categoría: extent_serosa

g. Variable de categoría: rx_Lev_mas_FU

Tabla 19 Tabla de los coeficientes por el modelo

Pruebas omnibus sobre los coeficientes del modelo^d

Paso	-2 log de la verosimilitud	Global (puntuación)			Cambio desde el paso anterior			Cambio desde el bloque anterior		
		Chi-cuadrado	gl	Sig.	Chi-cuadrado	gl	Sig.	Chi-cuadrado	gl	Sig.
1 ^a	5683.865	82.876	1	.000	70.690	1	.000	70.690	1	.000
2 ^b	5661.280	104.158	2	.000	22.586	1	.000	93.276	2	.000
3 ^c	5648.810	115.127	3	.000	12.470	1	.000	105.745	3	.000

- a. Variables introducidas en el paso número 1: node4
- b. Variables introducidas en el paso número 2: rx_Lev_mas_FU
- c. Variables introducidas en el paso número 3: extent_musculo
- d. Bloque inicial número 1. Método = Por pasos hacia adelante (Razón de verosimilitud)

Tabla 20 Tabla variables de la ecuación

Variables en la ecuación

		B	ET	Wald	gl	Sig.	Exp(B)
Paso 1	node4	-.871	.099	77.942	1	.000	.419
Paso 2	rx_Lev_mas_FU	.499	.109	20.961	1	.000	1.647
	node4	-.868	.099	77.444	1	.000	.420
Paso 3	rx_Lev_mas_FU	.501	.109	21.160	1	.000	1.651
	node4	-.849	.099	73.938	1	.000	.428
	extent_musculo	.581	.179	10.575	1	.001	1.788

Al final las funciones de supervivencia y riesgo para la recurrencia se ven definidas y afectadas sobre todo por estas tres variables según los datos estudiados.

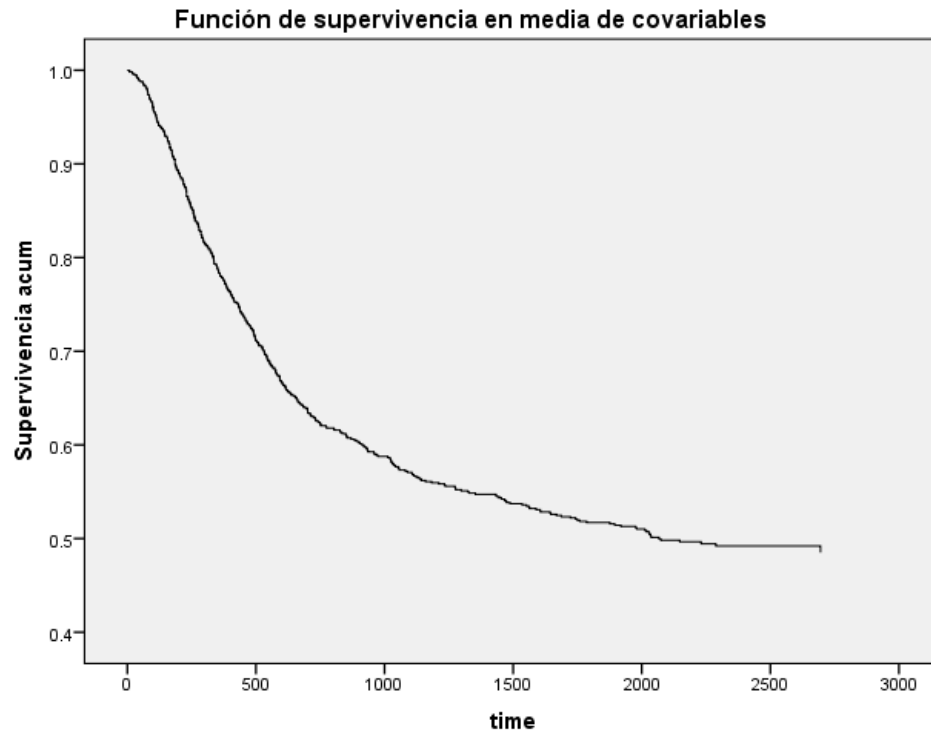


Gráfico 10 Grafico función de supervivencia modelo de regresión de Cox con las variables significativas

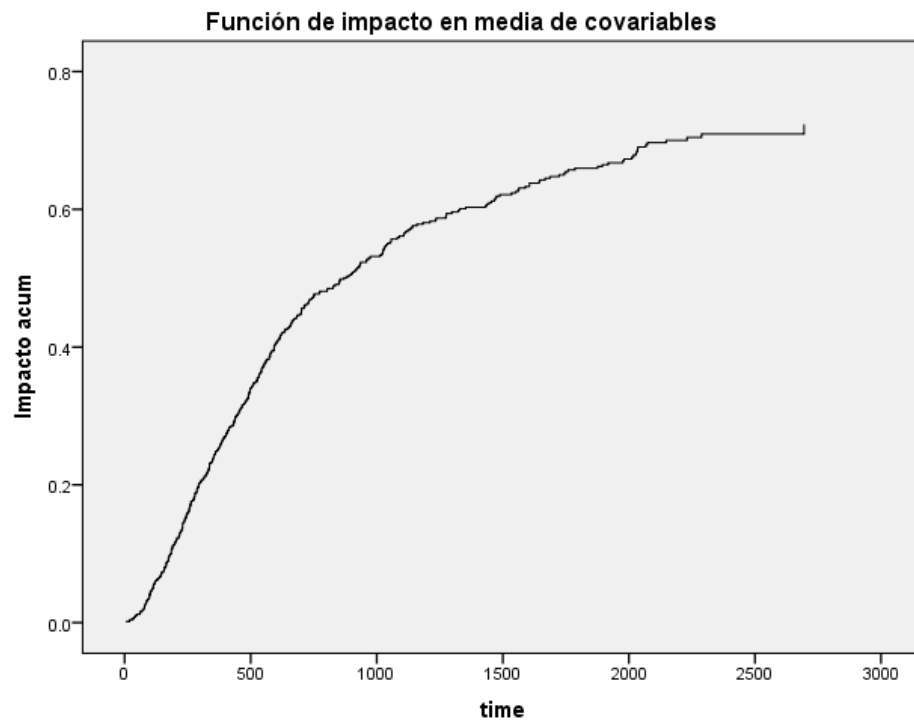


Gráfico 11 Grafico función de riesgo modelo de regresión de Cox con las variables significativas

7.8 HAZARD RATIO

Finalmente se va a realizar un análisis al gráfico de los Hazard Ratio.

Podemos comprobar del gráfico, que las variables como sexo o edad se encuentran en la zona de menos riesgo en la recurrencia del fenómeno, y la variable rx en el atributo Lev+5-FU es la que se encuentra como la de menor riesgo. Esto significa que efectivamente este tratamiento nos da casi una garantía de menos riesgo en la ocurrencia de volver a revivir el cáncer. Por otro lado, tenemos entre las variables en la zona de riesgo a la variable node4, extent, surg, etc., que son las que ya habíamos concluidos anteriormente que eran significativas, y de entre todas las variables node4 que es la que habíamos determinado con mayor significancia. Agregado a esto la variable extent que es una de las mas significativas en HR, dentro de sus modalidades, la más notable es extent_musculo, es decir, según los datos estudiados si a una persona el cancer se extendió hasta el musculo, aumenta el tiempo de supervivencia en comparación a los otros atributos.

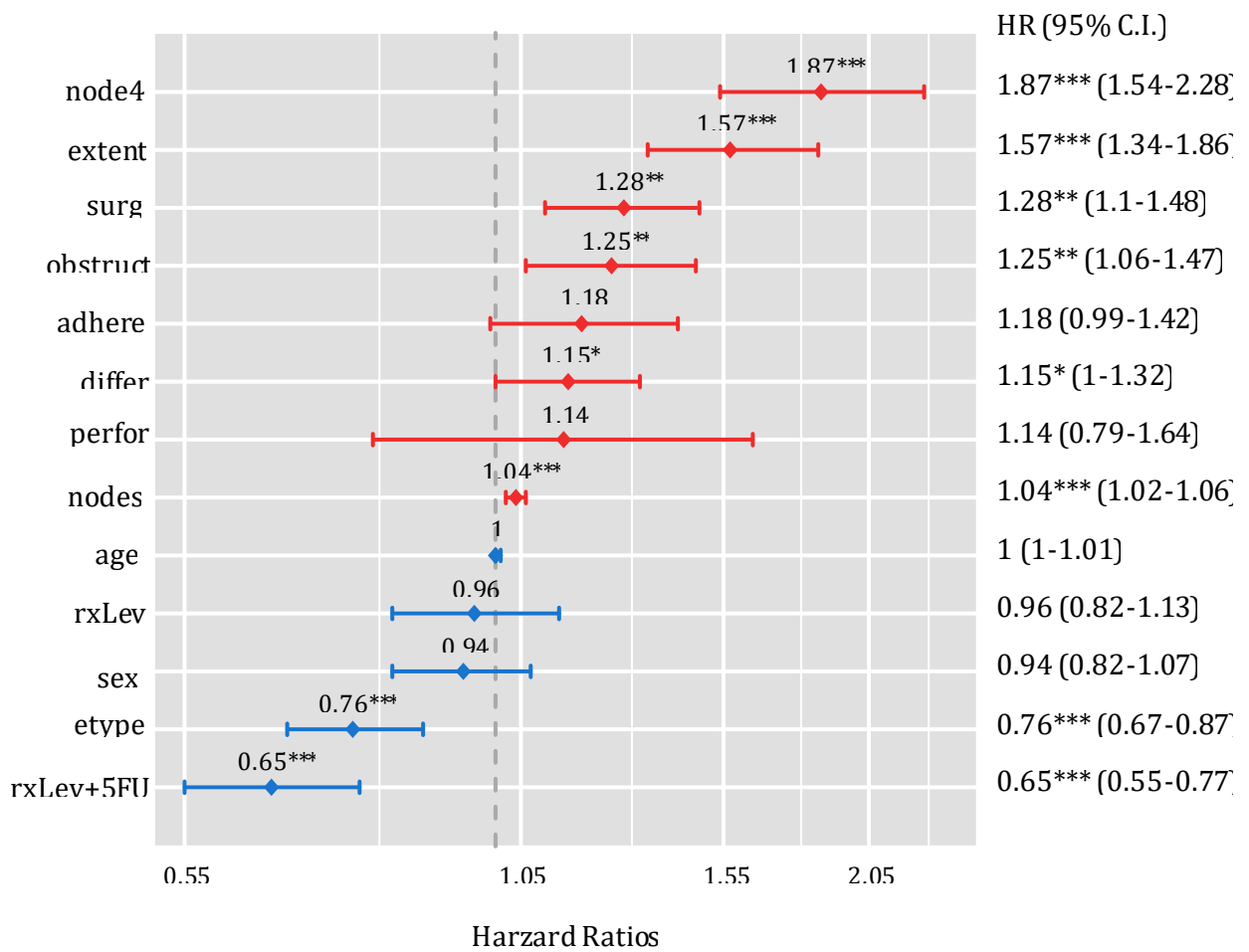


Grafico 8 Gráfico Harzard Ratios

VIII CONCLUSIONES

1. Es de vital importancia emplear metodologías estadísticas en estudios de temas de medicina, y de otras áreas, como la biología, física y mecánica. Como se observó en este trabajo, el análisis de supervivencia permitió hacer un estudio completo de todas las variables involucradas en la temática, así como ir obteniendo conclusiones relevantes para el área al cual está enfocada la investigación.

2. En esta era de las ciencias computacionales, es necesario la implementación e invención de software estadístico, que nos permitan manejar grandes flujos de información y bases de datos con múltiples variables. Para nuestro caso, los softwares RStudio y SPSS nos permitieron analizar de una manera rápida y objetiva nuestros datos.

3. Dentro de los tratamientos vistos en el estudio, se nota una clara superioridad en términos de supervivencia, del uso Levamisol en combinación con el 5-FU, es decir, esta combinación de químicos usados en las quimioterapias adyuvantes efectivamente logró un mejor porcentaje en que la enfermedad no volviera, y en los casos de recurrencia, el tiempo medio de recurrencia era mayor en comparación a los otros tratamientos.

4. La variable node4 es la variable más significativa del estudio, es decir, existe una mayor probabilidad de recurrencia del cáncer de colon, si se detectan más de 4 ganglios linfáticos infectados en su diagnóstico.

5. La función de riesgo se ve afectada por las variables: node4 y nodes (números de nodos detectados como infectados en su diagnóstico), extent (hasta donde se diseminaron las células cancerígenas), tratamiento y surg (tiempo desde la cirugía hasta el registro).

6. Variables como el sexo o la edad no son determinantes para una recurrencia de la enfermedad según el estudio.

7. Existe una mayor recurrencia de la enfermedad si el cáncer se extendió a estructuras contiguas al colon. En cambio la curva de supervivencia es mayor para los casos en los que el cáncer llegó a la submucosa. También, si la enfermedad llegó hasta el musculo logra una mayor supervivencia en comparación a si llegó a otras áreas, incluyendo la submucosa.

8. Es determinante la detección pronta del cáncer, ya que pudimos observar que en aquellos casos en los que fue más tardío su diagnóstico, el cáncer avanzó hacia otros órganos, afectando a más ganglios linfáticos; lo que se traduce en un aumento en la probabilidad de recurrencia.

IX REFERENCIA

Organización Mundial de la Salud. (2 de febrero de 2022). Cáncer. <https://www.who.int/es/news-room/fact-sheets/detail/cancer>

World Economic Forum. (30 de septiembre de 2020). <https://es.weforum.org/agenda/2020/09/a-medida-que-el-numero-de-muertos-de-covid-19-supera-el-millon-como-se-compara-con-otras-causas-de-muerte/>

Paris: International Agency for Research on Cancer of World health Organization; 2012 [actualizado al 2018; Citado 09 Jul 2018] Disponible en: <http://globocan.iarc.fr/>

Sociedad Americana de Oncología Clínica, (2021). Quimioterapia neoadyuvante, terapia endocrina y terapia dirigida para el cáncer de mama: guía de la ASCO.

Revista Cubana de Cirugía (2009). Cancer de colon y recto. [Cáncer colorrectal y anal en Sancti Spíritus en un quinquenio | Albiza Sotomayor | Revista Cubana de Cirugía \(sld.cu\)](#)

BBC (4 de febrero del 2016). https://www.bbc.com/mundo/noticias/2016/02/160203_cancer_graficos_impacto_men

Cancer.org. (21 de febrero del 2018). A cerca del cáncer colorrectal. American Cancer Society. <https://www.cancer.org/es/cancer/cancer-de-colon-o-recto/acerca/que-es-cancer-de-colon-o-recto.html>

Cirugía Española (Diciembre 2005). Recidiva local en el cancer de colon y recto. [Recidiva local en el cáncer de colon y recto | Cirugía Española \(elsevier.es\)](#)

Arce C, Martinez-Tlahuel J, Lara F (2006), Quimioterapia adyuvante en cáncer de mama, Instituto Nacional de Cancerología.

INFOMED. Formulario nacional de medicamentos. Levamisol-150.
<http://fnmedicamentos.sld.cu/index.php?P=FullRecord&ID=90>

Agencia Española de Medicamentos y Productos Sanitarios. (2020). Ficha Técnica.

Royal Statistical Society, Vol 21, (1972). The Analysis of Multivariate Binary Data.

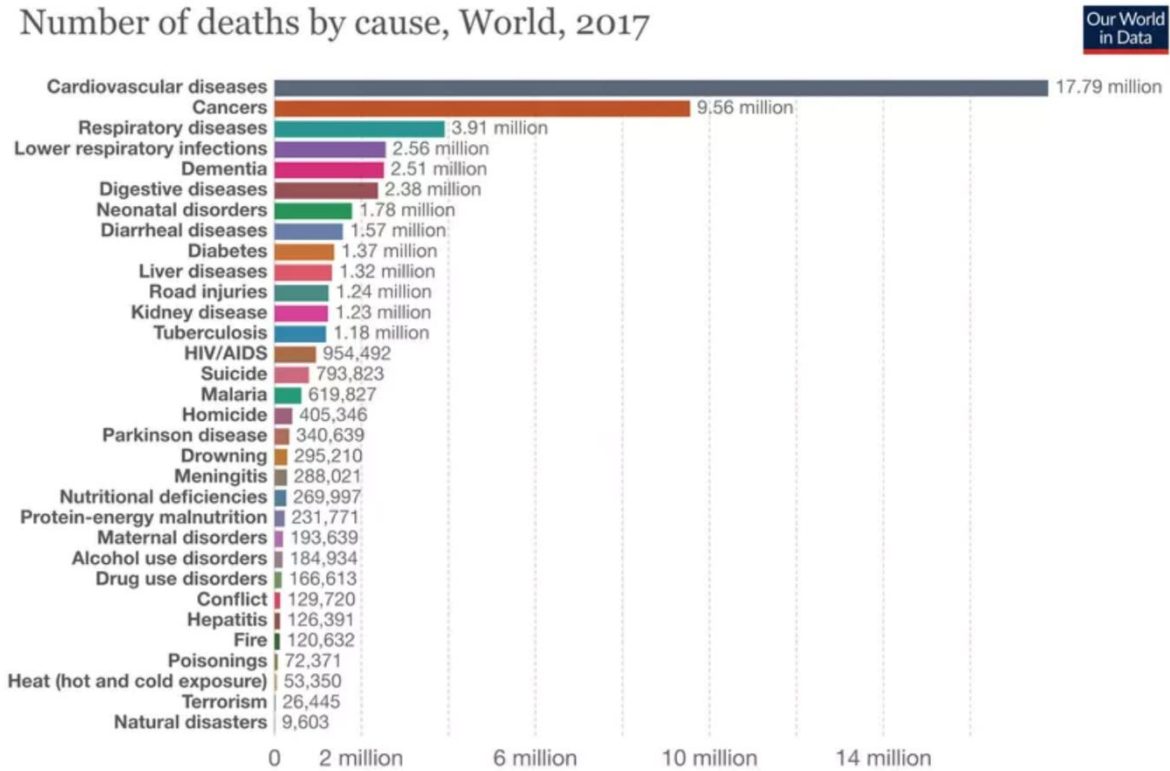
Universidad de Washington (1992). Introduction to Kaplan and Meier (1958) Nonparametric Estimation from Incomplete Observations

Universidad Tecnológica de Pereira (2018). Análisis de supervivencia aplicado al problema de la deserción estudiantil.

X ANEXOS

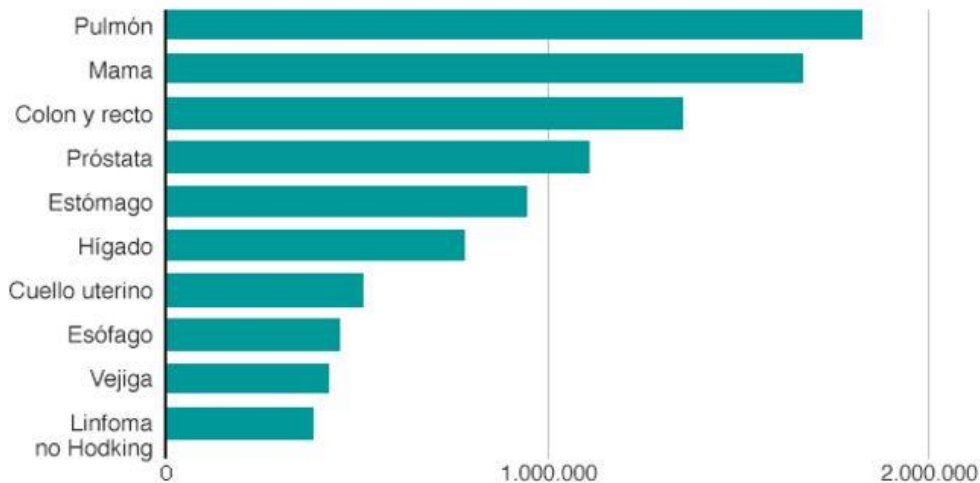
Grafica causa de defunciones a nivel mundial, año 2017:

Number of deaths by cause, World, 2017



Grafica de los canceres más diagnosticados:

10 tipos de cáncer diagnosticados más comúnmente



Fuente: Cancer Research UK, Globocan/ Cálculos de 2012

Grafica comparación de defunciones por género:

