



Universidad de Valladolid

Facultad de ciencias

TRABAJO FIN DE GRADO

Grado en Estadística

Predicción de Supervivencia en Pacientes
Post-TAVI: Modelo y Aplicación Web

Autor: **Diego de la Puente Alonso**

Tutora: **Cristina Rueda Sabater**

Año: **2025**

Agradecimientos

A los investigadores del Hospital Clínico Javier, Manuel e Itziar, por su confianza y apoyo en el desarrollo del proyecto.

A Cristina, mi tutora, por su orientación, dedicación y constante supervisión a lo largo de este trabajo.

A los profesores de la carrera, por todo lo que me han enseñado y aportado durante estos años de formación.

Y por último, aunque no menos importante, a mi familia y a mis compañeros de estudios, cuyo respaldo y ánimo han sido fundamentales para llegar hasta aquí.

A todos ellos, muchas gracias.

Resumen

TAVI son las siglas en inglés de Implante de la Válvula Aórtica Transcatéter. La TAVI es usada para el cambio de válvula en los pacientes con estenosis aórtica severa. Durante este proceso mínimamente invasivo se inserta una nueva válvula cardíaca sin extraer la válvula dañada anterior. La nueva válvula se coloca por dentro de la válvula dañada.[1]

En este trabajo, se busca estudiar los factores de riesgo en la mortalidad Post-TAVI. Para ello, se dispone de datos de 907 pacientes que han sido sometidos a este procedimiento, con un seguimiento de un año. El conjunto de datos incluye múltiples variables, que serán analizadas con el objetivo de identificar aquellas que se asocian significativamente con un mayor riesgo de mortalidad tras la intervención.

Además, a partir de los resultados obtenidos con este modelo, se plantea el desarrollo de una aplicación web que permita estimar el riesgo de mortalidad en pacientes candidatos al procedimiento TAVI. Esta herramienta tiene como objetivo facilitar la toma de decisiones clínicas, proporcionando una estimación individualizada del riesgo de fallecimiento.

Palabras clave: TAVI, Análisis de supervivencia, Modelo de Cox, aplicación web con Shiny.

Abstract

TAVI stands for Transcatheter Aortic Valve Implantation. It is used for valve replacement in patients with severe aortic stenosis. In this minimally invasive procedure, a new aortic valve is inserted without removing the previously damaged valve. This new valve is placed inside the old one. [1]

This study seeks to identify risk factors related to mortality after the TAVI procedure. To accomplish this, data from 907 patients who underwent the procedure with a one-year follow-up are available. The dataset includes multiple variables, which will be analyzed to identify those significantly associated with a higher risk of mortality after the intervention.

Furthermore, based on the results obtained from this model, the development of a web application is proposed. This tool aims to estimate the mortality risk for patients who are candidates for the TAVI procedure. Its objective is to support clinical decision-making by providing an individualized risk estimate.

Key words: Tavi, Survival analysis, Cox model, Shiny app.

Índice general

Lista de figuras	V
1. Introducción	1
1.1. TAVI	1
1.2. Variable respuesta	2
1.3. Objetivos	2
1.4. Estructura	2
2. Metodología	3
2.1. Notación y definiciones análisis de supervivencia	3
2.1.1. Notación básica	3
2.1.2. Funciones y sus estimadores en el análisis de supervivencia	4
2.1.3. Test de log-rank	6
2.2. Modelo de regresión de Cox	6
2.2.1. Análisis residuales	8
2.3. Selección de variables	9
2.3.1. Primer filtro de selección	9
2.3.2. Segundo filtro de selección	9
3. Análisis descriptivo	11
3.1. Variables demográficas	12
3.2. Variables clínicas	14
3.3. Variables de pruebas complementarias	14
3.3.1. Ecocardiograma	14
3.3.2. TAC cardíaco	15
3.3.3. Electrocardiograma (ECG)	15
3.3.4. Analítica sanguínea	15
3.4. Variables relacionadas con el procedimiento	16
3.5. Análisis supervivencia en un año	17
4. Resultados del modelo de Cox	20
4.1. Selección de variables	20
4.1.1. Primer filtro de selección	20
4.1.2. Segundo filtro de selección	21
4.2. Modelo obtenido	25
4.2.1. Estudio de los residuales	25

5. Aplicación web	28
5.1. Shiny	28
5.2. Descripción de la aplicación	29
5.3. Ejemplos de uso	29
6. Conclusiones	33
Bibliografía	34
A. Descripción variables	35
A.1. Variables demográficas	35
A.2. Variables clínicas	36
A.3. Variables pruebas complementarias	37
B. Herramientas de R	39
B.1. Gráficos	39
B.2. Modelo de Cox	39
B.2.1. Test para la selección de variables	39
B.2.2. Selección de variables automática y creación del modelo	40
B.3. App shiny	40

Lista de Figuras

1.1. Indicaciones para decidir la intervención de cambio de válvula HCUV	1
3.1. Distribución edades	12
3.2. Comorbilidades	13
3.3. Procedimientos por trimestre	16
3.4. Tipos de válvulas usadas en la TAVI	17
3.5. Escenario de los fallecimientos	18
3.6. Curva de supervivencia en un año Kaplan-Meier	18
3.7. Función de riesgo suavizada muerte en un año	19
4.1. Funciones de supervivencia según hemoglobina	22
4.2. Funciones de supervivencia según variables cardíacas	23
4.3. Funciones de supervivencia según el tipo de diabetes	23
4.4. Funciones de supervivencia según variables de la función renal	24
4.5. Residuos de Schoenfeld y p-valores del test de la hipótesis de riesgos proporcionales	25
4.6. Residuos deviance	26
4.7. Residuos de influencia	27
5.1. Página web tras abrirla	30
5.2. Modificaciones al deslizar los switch de las variables	30
5.3. Riesgo anual según combinaciones de variables y su percentil asociado.	31
5.4. Aplicación web con las curvas K-M escaladas.	31
5.5. Aplicación vista en dispositivos con pantallas estrecha como los móviles.	32

1 Introducción

Dado que este estudio se enmarca en el contexto de una intervención médica específica, resulta fundamental comenzar con una breve descripción del procedimiento TAVI, su finalidad clínica y su relevancia actual. Esta contextualización permite comprender mejor la motivación del análisis y la importancia de identificar los factores de riesgo asociados a la mortalidad durante el año posterior al procedimiento.

1.1. TAVI

El Implante de Válvula Aórtica Transcatéter (TAVI, por sus siglas en inglés) se ha consolidado en los últimos años como una alternativa eficaz y menos invasiva frente a la cirugía cardíaca convencional en pacientes con estenosis aórtica severa (EAO), especialmente en aquellos con alto riesgo quirúrgico [2]. Este procedimiento consiste en la implantación de una nueva válvula cardíaca dentro de la válvula aórtica deteriorada, sin necesidad de extirparla, lo que reduce significativamente los tiempos de recuperación y las complicaciones asociadas al reemplazo quirúrgico de válvula aórtica (RQVA). Sin embargo, a pesar de sus ventajas, la mortalidad post-TAVI sigue siendo un aspecto relevante a estudiar, ya que diversos factores clínicos y fisiológicos pueden influir en la evolución de los pacientes tras la intervención.

En el caso que nos compete en el HCUV (Hospital Clínico Universitario de Valladolid) se siguen estas indicaciones para la decisión de la intervención a realizar.

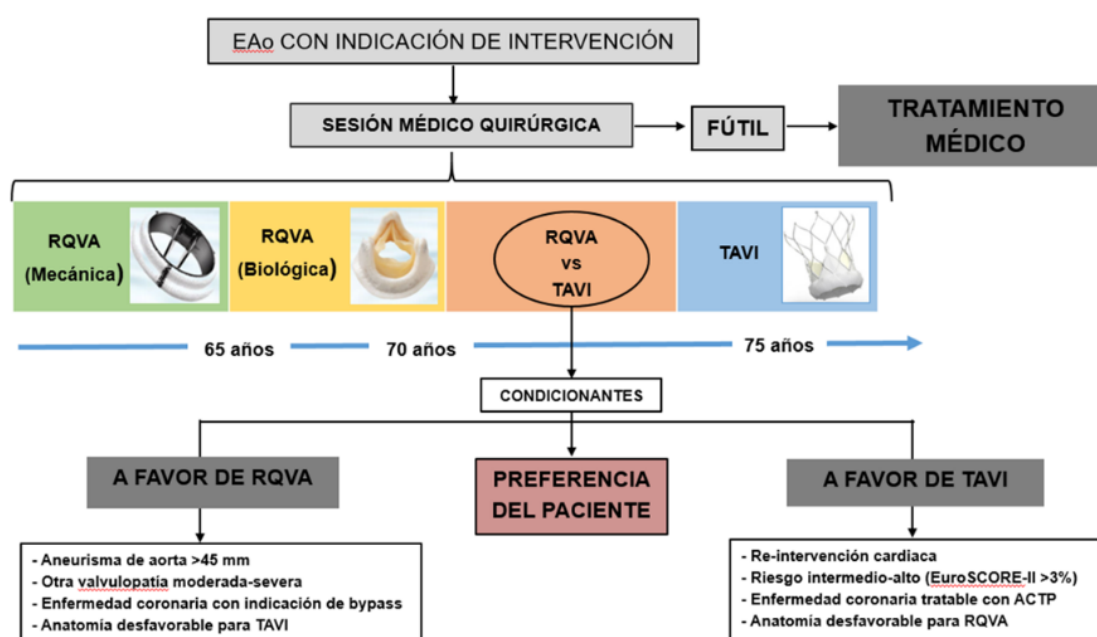


Figura 1.1: Indicaciones para decidir la intervención de cambio de válvula HCUV

Este aspecto es relevante para la selección de pacientes, ya que introduce un sesgo al priorizar a aquellos con mayor riesgo para someterse a la TAVI, con el objetivo de evitar una intervención más invasiva como el RQVA.

1.2. Variable respuesta

La variable principal del estudio es el tiempo hasta la muerte, definido como el intervalo transcurrido desde la realización del implante de válvula aórtica transcatóter (TAVI) hasta la ocurrencia del fallecimiento. Se trata de un estudio observacional retrospectivo con un período de seguimiento de un año posterior al procedimiento. Por ello, se opta por realizar un análisis de supervivencia, el cual será desarrollado a lo largo del presente documento.

1.3. Objetivos

- Analizar los factores de riesgo asociados a la mortalidad al año de haberse sometido a la intervención TAVI en el HCUV.
- Construir un modelo de regresión de Cox basado en los factores de riesgo identificados.
- Desarrollar una aplicación web que permita estimar el riesgo de mortalidad al año tras la TAVI, en función de los valores de dichos factores de riesgo.

1.4. Estructura

Este trabajo se dividirá en las siguientes partes:

- **1. Introducción.** Se presenta el problema a tratar y se definen los conceptos más importantes como la TAVI, la variable de estudio, los objetivos y la estructura del proyecto.
- **2. Metodología.** Se describen los fundamentos del análisis de supervivencia, el modelo de Cox y los métodos elegidos para la selección de variables.
- **3. Análisis descriptivo.** Este constituye la primera etapa del estudio estadístico y se busca explorar las variables más relevantes involucradas en la investigación.
- **4. Resultados modelo.** Se seleccionan las variables, categorizando las numéricas, y se crea y analiza el modelo de Cox.
- **5. Aplicación web.** Se explica cómo se crea la aplicación y se demuestra su funcionamiento.
- **6. Conclusiones.** Se resume los resultados obtenidos con sus limitaciones y se exponen las líneas de investigación futuras que surgen de este proyecto.

Finalmente, se incluirán dos apéndices. El primero contiene un estudio descriptivo de las variables presentado en formato de tabla, y en el segundo se detallan las herramientas y paquetes de R utilizados.

2 Metodología

En este estudio se emplea el análisis de supervivencia debido a la naturaleza de la variable respuesta, la cual corresponde al tiempo transcurrido hasta la ocurrencia de un suceso de interés (el fallecimiento). En este capítulo se introducen los conceptos relevantes del análisis de supervivencia y del modelo de Cox desde un enfoque teórico. Se describe la metodología empleada, el procedimiento de estimación del modelo, su validación y los distintos métodos de selección de variables. La bibliografía utilizada para este capítulo procede de los siguientes libros: [3, 4].

2.1. Notación y definiciones análisis de supervivencia

Para comprender este TFG, es necesario conocer ciertos aspectos de la teoría del análisis de supervivencia. A diferencia de otros métodos estadísticos, este se centra en el estudio del tiempo o la distancia entre dos sucesos. En nuestro caso, este tiempo corresponde al intervalo entre la intervención y el fallecimiento.

En el análisis de supervivencia, se introduce el concepto de censura, que hace referencia a aquellos casos en los que la información sobre la variable de interés es incompleta. En este estudio, únicamente se registran censuras por la izquierda si el paciente no fallece tras un año de la intervención, lo que significa que, si el suceso de interés no ha ocurrido durante el primer año posterior a la intervención, no se considera la información más allá de ese período.

2.1.1. Notación básica

Entre estos conceptos se encuentran los siguientes, que serán expuestos a continuación junto con la notación que se empleará a lo largo de este estudio.

Variable de estudio

Esta indica el tiempo desde que ocurre un suceso origen hasta el suceso final. En este estudio es el tiempo desde la implantación de la TAVI hasta que ocurre el fallecimiento. Se representa como X .

Tiempo de censura

Este es el tiempo desde el suceso origen hasta que el individuo deja de ser observado, ya sea porque ocurre el suceso final o una censura. Se denota como C .

Tiempo de supervivencia

Es el tiempo desde que sucedió el suceso origen hasta que ocurre el suceso final o una censura. Se denota como $t_i = \min(x_i, c_i)$, donde $i = 1, \dots, n$ representa el individuo i en los datos.

Censura

Esta variable será uno si en el individuo correspondiente ocurre una censura antes del suceso final. Se identifica como $\delta_i = I(x_i \leq c_i)$.

2.1.2. Funciones y sus estimadores en el análisis de supervivencia

En el análisis de supervivencia, existen diversas funciones matemáticas que permiten describir el tiempo hasta la ocurrencia de un suceso de interés. Estas funciones capturan aspectos clave sobre la probabilidad de ocurrencia del suceso. Comprender su definición e interpretación es esencial para el análisis correcto de datos de este tipo.

A partir de datos concretos, existen estimadores que nos permiten aproximarnos a estas funciones, que se verán después de describir las funciones.

Función de supervivencia

Esta función se define como la probabilidad de que el suceso final no ocurra. Dado que en muchos contextos el suceso final corresponde a la muerte, se le atribuye este nombre. Esta función nos permitirá conocer cómo evoluciona la probabilidad de supervivencia en este estudio. De una forma matemática, con la notación anterior se define como:

$$S(x) = P(X > x) = \int_x^\infty f(u)du = 1 - F(x)$$

Esta función es positiva, no decreciente, $S(0) = 1$ y $S(\infty) = 0$

Función de riesgo o tasa de fallo

La función de riesgo se define como la probabilidad instantánea de que ocurra el suceso en el siguiente instante de tiempo, condicionada a que el suceso aún no haya ocurrido. Esta función es clave para el modelo de Cox. Matemáticamente se describe de la siguiente forma:

$$h(x) = \lim_{\Delta x \rightarrow 0} \frac{p(x \leq X \leq x + \Delta x | X \geq x)}{\Delta x}$$

Esta es no negativa y nos proporciona información sobre los momentos en los que el riesgo de que ocurra el suceso final es mayor. Es más útil que la función de supervivencia para observar cómo varía la probabilidad de que ocurra el suceso de interés a lo largo del tiempo.

Estimador de Kaplan-Meier

Es una de las herramientas más comunes para estimar la función de supervivencia. Siendo $t_1 < \dots < t_D$ los distintos tiempos en los que se observa al menos un suceso en los datos, Y_j el número de individuos a riesgo en t_j con $j = 1, \dots, D$ y d_j el número de sucesos que se dan en el momento j . El estimador se define de esta forma:

$$\hat{S}_{KM}(t) = \prod_{j=1}^k \left(1 - \frac{d_j}{Y_j}\right); t_k \leq t \leq t_{k-1}$$

Este estimador de la función de supervivencia sólo cambia en los momentos que se observa al menos un suceso final y $\hat{S}_{KM}(t)$ representa la probabilidad estimada de supervivencia hasta el tiempo t .

Para poder acompañar la estimación de la función de supervivencia con un intervalo de confianza, es necesario calcular la varianza de la función, la cual se estima mediante la fórmula de Greenwood, definida como:

$$\hat{V}[\hat{S}_{KM}(t)] = [\hat{S}_M(t)]^2 = \left(\sum_{j=1}^k \frac{d_j}{Y_j(Y_j - d_j)}\right); t_k \leq t \leq t_{k-1}$$

Con esta estimación de la varianza, se puede obtener un intervalo de confianza asintótico de la siguiente manera:

$$[\hat{S}_{KM}(t) \pm z_{1-\alpha/2} \sqrt{\hat{V}[\hat{S}_{KM}(t)]}]$$

Donde $z_{1-\alpha/2}$ es el cuantil de una normal estándar para el nivel de confianza $1 - \alpha/2$.

Estimador de Nelson Aalen

Este se utiliza para estimar la función de riesgo acumulada, siguiendo la misma notación usada que en la definición del estimador de Kaplan-Meier, se define como:

$$\hat{H}_{NA}(t) = \sum_{j=1}^k \frac{d_j}{Y_j}; t_k \leq t \leq t_{k-1}$$

Esta función es no decreciente, solo varía cuando se observa un suceso y nos ayuda a conocer dónde es más probable que ocurran los sucesos de interés.

2.1.3. Test de log-rank

También conocido como la prueba de Mantel-Cox, es un test no paramétrico que sirve para contrastar varias funciones de supervivencia. La hipótesis nula plantea que las curvas de supervivencia de los distintos grupos son iguales, mientras que la hipótesis alternativa sostiene que existen diferencias entre ellas.

Siendo Y_{ij} el número de individuos a riesgo en el grupo j en el momento t_i , d_{ij} el número de sucesos observados en el grupo j en el momento t_i , $Y_i = \sum_{j=1}^J Y_{ij}$ y $d_i = \sum_{j=1}^J d_{ij}$ el número de individuos a riesgo y sucesos respectivamente de todos los grupos en el momento t_i y $W(t_i)$ una función que define los pesos que comparten los grupos. Para dos grupos, este test se calcula de la siguiente manera:

$$Z = \frac{\sum_{i=1}^D W(t_i) \left(d_{i1} - Y_{i1} \frac{d_i}{Y_i} \right)}{\sum_{i=1}^D \frac{Y_{i1}}{Y_i} \left(1 - \frac{Y_{i1}}{Y_i} \right) \left(\frac{Y_i - d_i}{Y_i - 1} \right) d_i}$$

Esta Z sigue una χ^2 asintóticamente y si se añaden más grupos usando la fórmula adecuada seguiría un χ^2_{K-1} donde K es el número de grupos. Para el test de log-rank $W(t_i) = 1$, si se realizan cambios en la estimación de esta función se llegan a otros test que también comparan las curvas de supervivencia.

2.2. Modelo de regresión de Cox

El modelo de Cox, también conocido como modelo de regresión de Cox o modelo de riesgos proporcionales, permite estimar el efecto de las covariables sobre el tiempo hasta la ocurrencia de un suceso. En este estudio, se utilizará dicho modelo para identificar los factores de riesgo que influyen significativamente, según el test log-rank, en la probabilidad de fallecimiento tras la intervención. Una característica clave del modelo es que asume que el efecto de cada covariable, representado por su coeficiente de riesgo, es constante a lo largo del tiempo. Siendo Z el vector de covariables, β' el vector de los coeficientes de regresión y $h_0(t)$ la función de riesgo base. El modelo se expresa de la siguiente manera:

$$H(t|Z) = h_0(t) \exp(\beta' Z)$$

El objetivo de este modelo es estimar el resto de parámetros sin tener que estimar $h_0(t)$, ya que lo que nos interesa es cómo las distintas covariables afectan a la función de riesgo base, aumentando o disminuyendo el riesgo, sin depender del tiempo. El modelo de Cox asume censura no informativa, independencia de los tiempos de supervivencia y cumplimiento de la hipótesis de riesgos proporcionales, es decir, razones de riesgo constantes en el tiempo.

Los coeficientes β' se estiman en el modelo maximizando la función de verosimilitud parcial. Siendo t_j los distintos tiempos en los que ocurre un suceso, R_j los individuos a riesgo en el tiempo t_j y Z_j el valor de las covariables del individuo que experimenta el suceso en el tiempo t_j . La función de verosimilitud parcial se define como:

$$L_p(\beta) = \prod_{j=1} \left(\frac{\exp(\beta Z_j)}{\sum_{l \in R_j} \exp(\beta Z_l)} \right)$$

Hazard ratio

El hazard ratio, o cociente de riesgos, es la proporción del riesgo asociado con unas covariables concretas comparado con un modelo en el que todas las covariables tienen un valor de referencia. Siendo Z y Z^* dos valores distintos de las covariables Z , su cálculo se expresa de la siguiente manera:

$$HR = \frac{H(t|Z)}{H(t|Z^*)} = \frac{h_0(t)\exp(Z\beta)}{h_0(t)\exp(Z^*\beta)} = \exp((Z - Z^*)\beta)$$

El hazard ratio, como se observa en la fórmula, es independiente del tiempo, lo cual debe validarse. Este valor refleja el cambio en el riesgo asociado con la modificación de las covariables, lo que lo convierte en una medida clave para evaluar cómo sus variaciones influyen en el riesgo de que ocurra el suceso. Por esto último, el hazard ratio es un valor clave en nuestro estudio.

Concordance

El C-index, también conocido como índice de concordancia, es una métrica comúnmente utilizada para evaluar la capacidad de un modelo para discriminar entre pacientes con distintos niveles de riesgo. Dado un par de observaciones (i, j) con los valores (x_i, \hat{x}_i) , donde x_i es el tiempo entre el evento origen y final observado y \hat{x}_i es el valor predicho, se considera que las observaciones son concordantes si tanto la predicción como los datos siguen la misma dirección, es decir, si $(x_i > x_j, \hat{x}_i > \hat{x}_j)$ o $(x_i < x_j, \hat{x}_i < \hat{x}_j)$. La concordancia mide la proporción de pares concordantes en los datos y se estima como $p(x_i > x_j | \hat{x}_i > \hat{x}_j)$. Este estadístico puede calcularse de diversas formas, asignando distintos pesos a los empates en x ($x_i = x_j$) o en \hat{x} ($\hat{x}_i = \hat{x}_j$). Para calcular los pares concordantes en el análisis de supervivencia, se utiliza la siguiente fórmula:

$$\sum_{i=1}^n \delta_i \left(\sum_{t_j > t_i} \text{sign}(x_i - x_j) \right)$$

Esta forma de calcular los pares concordantes incluye la variable censura (δ), ya que si t_i es una censura, todas las observaciones con $t_j \geq t_i$ no son comparables porque no se puede saber cuál de los dos sobrevive más tiempo [5].

2.2.1. Análisis residuales

En esta sección se analizarán los distintos residuos del modelo, abarcando la validación de la hipótesis de riesgos proporcionales, la identificación de posibles valores atípicos y la robustez del modelo.

Residuos de Schonfield

Se utilizan para verificar la hipótesis de riesgos proporcionales en el modelo de Cox. Estos residuos se calculan como la diferencia entre el valor observado de una covariable z_{ij} para el individuo i y su valor esperado \bar{z}_{t_i} en el tiempo t_i .

$$\hat{r}_{ij} = z_{ij} - \bar{z}_{t_i}$$

Los residuos de Schoenfeld solo se definen para los individuos que no están censurados. Si la hipótesis de riesgos proporcionales es válida, un gráfico de estos residuos mostrará puntos distribuidos en el eje y alrededor del valor 0. Para probar esta hipótesis, se puede realizar una regresión lineal de los residuos de Schoenfeld y verificar si la pendiente es cero. Si esta hipótesis se rechaza, es necesario recurrir a modelos alternativos o realizar transformaciones para que no se rechace.

Residuos de la deviance

Estos sirven para identificar valores atípicos (outliers), los cuales se evidencian en los valores más extremos de estos residuos. Además, la suma de sus cuadrados corresponde al estadístico de la prueba de razón de verosimilitud.

Sea m_i los residuos martingala definidos como:

$$m_i = \delta_i - \hat{H}(t_i)e^{z_i'\hat{\beta}}$$

Donde $\hat{H}(t_i)$ es la estimación de la función de riesgo acumulada hasta el tiempo t_i . Estos residuos se definen de la siguiente manera:

$$d_i = \text{sign}(m_i) (-2(m_i + \delta_i \log(\delta_i - m_i)))^{1/2}$$

Residuos de influencia

Los residuos de influencia o DFBETA son una aproximación al cambio que se produciría al valor estimado de los coeficientes (β_i) si un individuo no estuviese en el modelo. Sirve para comprobar si algún individuo es muy influyente para la estimación de algún coeficiente. Estos nos indican la robustez del modelo.

2.3. Selección de variables

Para construir un modelo de regresión de Cox robusto y clínicamente interpretable, es fundamental aplicar criterios adecuados de selección de variables. Por lo que se necesita un mecanismo para escoger variables candidatas y, dentro de ellas, seleccionar las más adecuadas.

2.3.1. Primer filtro de selección

Primero se realiza una evaluación preliminar para identificar las variables asociadas con el suceso de interés. Para ello, se utilizan los p-valores de distintos tests aplicados a cada variable según su naturaleza, dividiendo la muestra en dos grupos según si el individuo fallece o si se corresponde a una censura. A continuación, explicamos el test que se aplica según sea la variable:

- Para las variables categóricas, se aplica el test exacto de Fisher cuando tienen dos categorías o el test de Fisher-Freeman-Halton si presentan más de dos categorías y frecuencias esperadas menores a cinco. Cuando las frecuencias esperadas son mayores, se emplea el test chi-cuadrado sin corrección de Yates.
- Para las variables numéricas, si no siguen una distribución normal según el gráfico QQ y el test de Kolmogórov-Smirnov, se utiliza el test de Wilcoxon. Si la normalidad no se rechaza, se aplica el test t adecuado según se rechace o no que las varianzas sean iguales.

Además de considerar las variables cuyos p-valores en estos tests sean menores a 0,1, también se incluyen aquellas que presentan un p-valor inferior a 0,1 en el test de log-rank.

A continuación, se proponen aquellas variables que, además de cumplir con el umbral estadístico, tienen relevancia clínica o están respaldadas por la literatura científica. Esta selección asegura su pertinencia dentro del contexto del estudio, priorizando las variables que son más relevantes por parte de los investigadores del HCUV.

2.3.2. Segundo filtro de selección

Tras un primer filtrado en el que puede haber varias variables candidatas a ser usadas en el modelo, se aplica un segundo filtro de selección automática para refinar aún más el conjunto de variables. Antes de este segundo filtro, se decidirá cómo categorizar las variables numéricas para facilitar su interpretación.

Durante esta etapa se aplican métodos estadísticos que, de forma automatizada, seleccionan o descartan covariables para determinar el subconjunto más apropiado. A continuación, se detallarán estos métodos de selección de variables.

Métodos de selección

Existen tres métodos comunes de selección de variables, que se diferencian en la forma en que avanzan paso a paso hacia el modelo final considerado más adecuado. Estos son los siguientes:

- **Backward:** Se comienza con un modelo que incluye todas las variables explicativas y, en cada paso, se elimina la variable menos influyente según un criterio predefinido. El proceso continúa hasta que no sea posible eliminar más variables sin incumplir dicho criterio.
- **Forward:** Se parte de un modelo vacío, y progresivamente se incorporan las variables más influyentes en función del criterio establecido. El procedimiento termina cuando ninguna variable candidata cumple el criterio de entrada.
- **Stepwise:** Combina elementos de los métodos forward y backward. En cada paso, se evalúa tanto la posibilidad de añadir nuevas variables como de eliminar variables presentes en el modelo, siempre de acuerdo con el criterio seleccionado. El procedimiento concluye cuando no es posible realizar más cambios que mejoren el modelo. En nuestro caso, usaremos este método para la selección de variables.

Criterio de selección

Existen diversos criterios para guiar el proceso de selección de variables. En este trabajo, se utilizará el p-valor asociado al test de log-rank.

Se establecerán umbrales específicos para la inclusión y la exclusión de variables. Una variable podrá entrar en el modelo si su p-valor es inferior a 0,1 y será eliminada si su p-valor supera dicho umbral en etapas posteriores.

3 Análisis descriptivo

Para iniciar el análisis descriptivo, se presenta una caracterización de los datos utilizados. Estos incluyen 907 individuos a los que se les ha aplicado la TAVI entre los años 2013 y 2023 y su seguimiento a 1 año. Este es un estudio observacional y retrospectivo en el que se recogen datos de los implantes de TAVI realizados en el HCUV.

Entre los datos recogidos por el HCUV sobre estos individuos se encuentran diversas variables que pueden clasificarse en las siguientes categorías:

- **Demográficas.** Incluyen información básica sobre los pacientes, como edad, peso, altura, así como antecedentes médicos relevantes como hipertensión o diabetes, entre otras.
- **Clínicas.** Agrupan datos relacionados con el estado de salud del paciente.
- **Derivadas de pruebas complementarias.** Corresponden a los resultados obtenidos a partir de diversas pruebas diagnósticas. Estas pruebas son: ecocardiogramas, electrocardiogramas, tomografías computarizadas (TAC) cardíacas y análisis de sangre.
- **Relacionadas con el procedimiento.** Comprenden variables como las fechas de las intervenciones y detalles específicos del procedimiento, como el tipo de válvula implantada.
- **De seguimiento.** Corresponden a la información obtenida tras la intervención, como las pruebas de control realizadas, la duración de la estancia hospitalaria o la fecha de fallecimiento, si aplica.

Para construir el modelo, se analizarán únicamente las variables demográficas, clínicas y derivadas de pruebas complementarias. Es importante señalar que no se considerarán como posibles variables para el modelo aquellas que no estén disponibles antes de la realización del procedimiento, ya que no serían útiles para predecir el riesgo de mortalidad de futuros pacientes antes de aplicar el procedimiento. En cuanto a las variables de seguimiento, solo se empleará la información sobre si el paciente falleció durante el año posterior a la intervención y, en caso afirmativo, la fecha de fallecimiento, con el fin de calcular el tiempo de supervivencia a partir de la fecha del procedimiento.

En las siguientes páginas, se describen brevemente las variables de cada grupo, destacando las más relevantes. Todas las variables se presentan en una tabla que incluye una breve descripción numérica para cada una, diferenciada según los grupos de fallecimiento y censura. Dicha tabla se encuentra en el anexo A.

3.1. Variables demográficas

Se analizarán las variables demográficas más relevantes con el objetivo de comprender mejor el perfil de los pacientes incluidos en este estudio observacional y de identificar aquellas variables que pudieran ser utilizadas en el modelo.

Edad

La edad del paciente en el momento del implante es un factor determinante en la selección de candidatos para la TAVI, ya que en individuos de edad avanzada suele haber mayores argumentos clínicos y quirúrgicos para preferir esta técnica sobre el reemplazo quirúrgico de válvula aórtica.

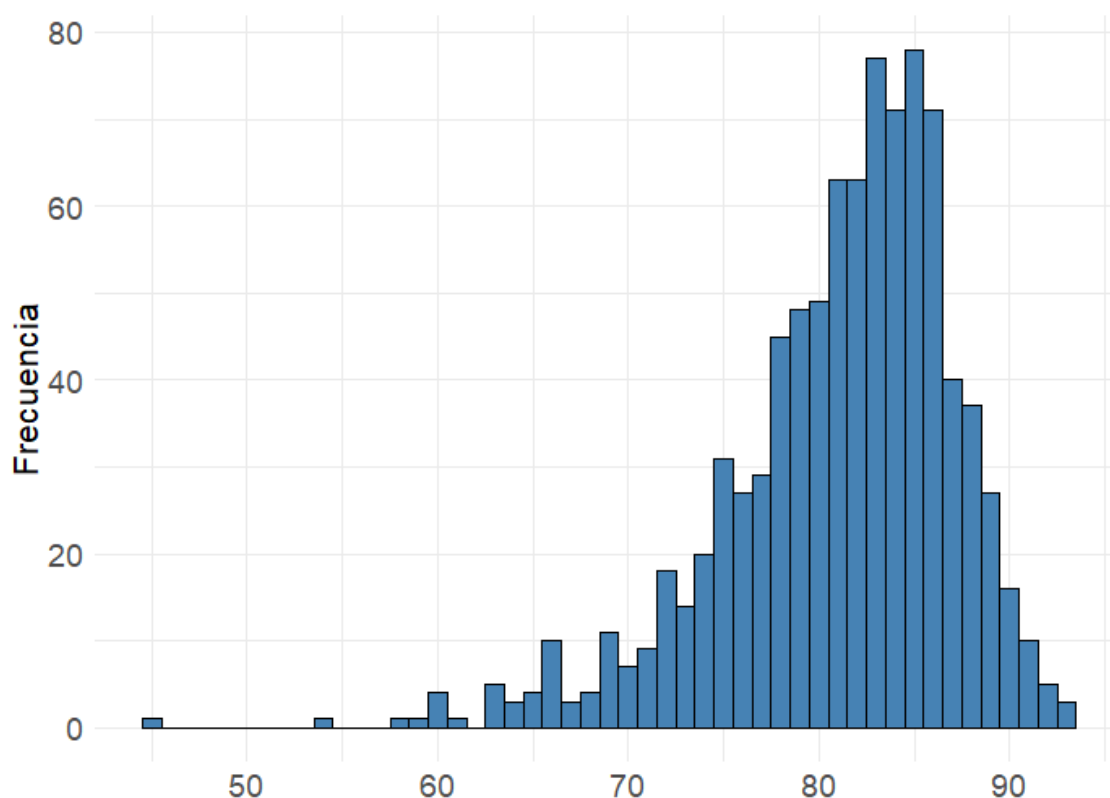


Figura 3.1: Distribución edades

Como se observa en la figura 3.1 la mayoría de las edades están concentradas entre los 75 y los 90 años y la media de estas edades es de 81,07.

Sexo

Otra variable demográfica relevante es el sexo. En esta se observa que en el estudio hay una ligera mayor cantidad de mujeres que de hombres, representando éstas 55,2 % de los individuos y los hombres un 44,8 %.

Comorbilidades

Una comorbilidad es la presencia de una o más enfermedades o trastornos adicionales que ocurren junto con una enfermedad principal, en nuestro caso la estenosis aórtica severa. Estas condiciones pueden influir en la evolución y el tratamiento de la enfermedad principal, por lo que son claras candidatas al modelo desde un punto de vista clínico. Con respecto a la TAVI haremos una breve exploración de las comorbilidades más comunes observadas. Entre las que se encuentran: hipertensión arterial, diabetes mellitus, de estos los que son insulino dependientes (DMID), enfermedad coronaria previa (ECP), enfermedad arterial periférica (EAP), fibrilación auricular (FA), insuficiencia renal crónica (IRC) y enfermedad pulmonar obstructiva crónica (EPOC).

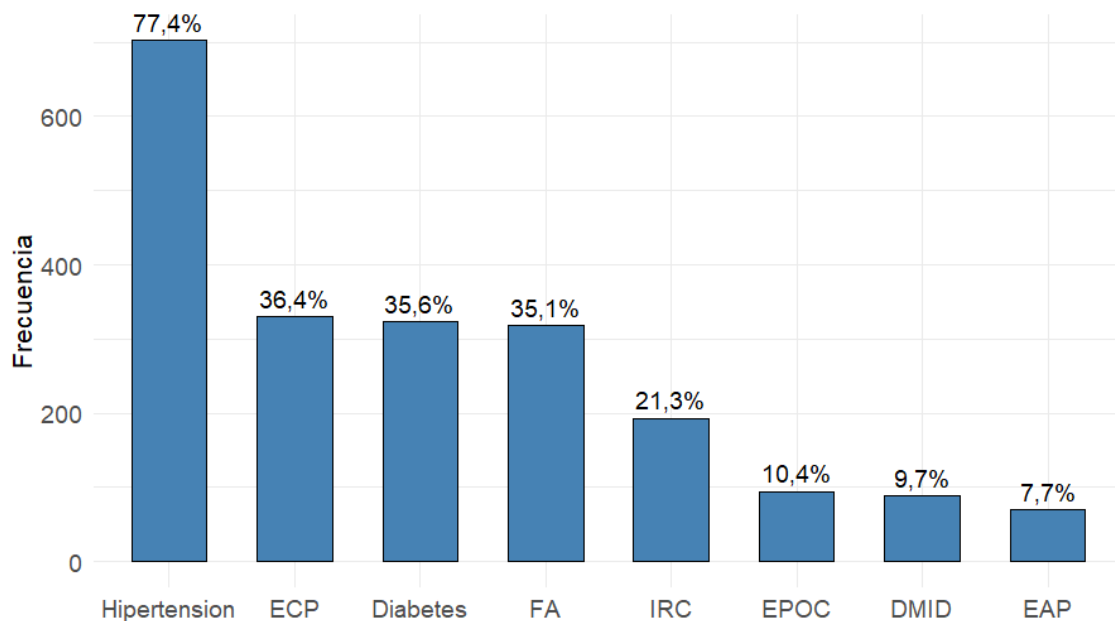


Figura 3.2: Comorbilidades

Dentro de las variables demográficas se incluyen otras como son índice de masa corporal (IMC: peso Kg / talla m²), procedencia de esta sacaremos los remitidos de otros hospitales (hospitales distintos del HCUV), tabaquismo, insuficiencia cardíaca, intervención coronaria previa, infarto de miocardio previo dividido según la distancia a la operación, cirugía cardíaca previa, cirugía valvular, bypass coronario previo, ictus previo, marcapasos o desfibrilador automático implantable, fragilidad, aorta en porcelana, cáncer, placa aterosclerótica compleja, puntuación de las escalas CHADS₂, CHA₂DS₂-VASc, EuroSCORE I, EUROSCORE II, STS y TAVI-2 score.

Las variables que puntúan el riesgo vascular según diferentes escalas, en principio, no se deberían incluir en el modelo, ya que son combinaciones de otras variables que serán consideradas individualmente.

3.2. Variables clínicas

Estas variables recogen información sobre el estado de salud del paciente previo al procedimiento, incluyendo los síntomas que presenta y la clasificación según la escala NYHA, utilizada para evaluar la gravedad de la insuficiencia cardíaca.

Síntomas

Respecto a estos, los más comunes en el estudio son la disnea (90 %), la angina (22,4 %) y el síncope (9,2 %). Además, muy pocos pacientes son asintomáticos, representando estos solo el 1,8 % de los pacientes.

Escala NYHA

La escala NYHA [6] (New York Heart Association) es una clasificación funcional utilizada para evaluar la gravedad de los síntomas en pacientes con insuficiencia cardíaca. Divide a los pacientes en cuatro clases (I a IV) según la limitación que presentan en la actividad física:

- Clase I. Sin limitación de la actividad física.
- Clase II. Ligera limitación, causando la actividad física ordinaria fatiga, palpitaciones o disnea.
- Clase III. Limitación marcada, menor actividad de lo habitual provoca síntomas.
- Clase IV. Incapacidad para realizar cualquier actividad física sin malestar, los síntomas pueden aparecer incluso en reposo.

En esta variable, existen 32 (3,5 %) NA. De los valores disponibles, la mayoría se concentra en las clases II y III, representando 797 (87,8 %) casos.

3.3. Variables de pruebas complementarias

Antes del procedimiento, se llevan a cabo diversas pruebas clínicas destinadas a evaluar el estado general del paciente y sus problemas de salud. Estas incluyen análisis de sangre, estudios de imagen y valoraciones del funcionamiento cardíaco. A continuación, se describirán las variables más relevantes obtenidas a partir de cada una de estas pruebas.

3.3.1. Ecocardiograma

Es una prueba fundamental para evaluar la gravedad de la estenosis aórtica, una enfermedad que, en sus casos más severos, requiere el reemplazo de la válvula aórtica del paciente. Aunque se recogen múltiples variables, en este apartado se destacan dos que resultan especialmente útiles para identificar posibles alteraciones en la función cardíaca.

Presión sistólica de la arteria pulmonar (PSAP)

La PSAP representa la presión ejercida por la sangre en las arterias pulmonares durante la sístole y se utiliza como un indicador indirecto de hipertensión pulmonar, una condición con importantes implicaciones clínicas. Cabe destacar que, en ciertos casos, no es posible obtener una medición cuando los valores de presión son muy bajos.

Fracción de eyección del ventrículo izquierdo (FEVI)

La FEVI representa el porcentaje de sangre que el ventrículo izquierdo expulsa con cada contracción. Se trata de un parámetro clave en la evaluación de la función sistólica del corazón y se utiliza habitualmente para diagnosticar y clasificar la insuficiencia cardíaca.

Además de estas dos variables, el ecocardiograma proporciona otras como área valvular aórtica, gradiente medio y máximo transvalvular, velocidad máxima transvalvular, insuficiencia aórtica concomitante, y los grados de insuficiencia mitral y tricúspide.

3.3.2. TAC cardíaco

La tomografía computarizada cardíaca es una técnica de imagen que utiliza rayos X para obtener representaciones detalladas del corazón y sus vasos sanguíneos. Esta prueba permite medir múltiples variables anatómicas y funcionales. Entre ellas se encuentran: el diámetro máximo y mínimo del anillo aórtico, área y perímetro del anillo aórtico, altura de origen del tronco coronario izquierdo y de la arteria coronaria derecha, diámetros máximo y mínimo del seno aórtico, puntuación de calcio valvular mediante el score de Agatston, presencia de válvula aórtica bicúspide o de aorta en porcelana. Estas dos últimas se incluyen dentro de las variables demográficas.

3.3.3. Electrocardiograma (ECG)

Es una prueba médica que registra la actividad eléctrica del corazón a lo largo de un periodo de tiempo. Esta actividad eléctrica es fundamental para que el corazón pueda contraerse y bombear sangre de manera adecuada. Entre las variables recogidas en esta prueba se encuentran el bloqueo de rama derecha e izquierda, sin contar los individuos con ritmo cardíaco que han sido clasificados en pacientes con necesidad de marcapasos, fibrilación auricular, flutter auricular y ritmo electroestimulado.

Entre todas estas, destaca la fibrilación auricular (FA), una de las arritmias cardíacas más comunes, que puede evolucionar hacia insuficiencia cardíaca o incrementar el riesgo de ictus, por lo que tiene especial relevancia clínica.

3.3.4. Analítica sanguínea

Mediante el análisis de sangre se obtienen diversas variables bioquímicas y hematológicas. Entre ellas se incluyen hemoglobina, leucocitos, plaquetas, creatinina, NT-proBNP, aclaramiento de creatinina, filtrado glomerular e insuficiencia renal crónica.

De estas variables destacamos las siguientes: la hemoglobina, que se considera un marcador del estado hematológico general, y distintas variables relacionadas con la función renal, como la creatinina, el aclaramiento de creatinina y el filtrado glomerular. Estas dos últimas son variables que miden el funcionamiento renal mediante los métodos de Cockcroft-Gault y CKD-EPI respectivamente. La función renal es considerada un aspecto relevante en pacientes con enfermedad cardiovascular, por lo que probablemente alguna de estas variables acabe en el modelo de Cox.

3.4. Variables relacionadas con el procedimiento

Se analizarán las variables que reflejan el momento del procedimiento, el acceso empleado y el tipo de válvula utilizada en la TAVI. Todo con el fin de comprender mejor cuándo se realiza el procedimiento y el modo en el que se llevó a cabo.

Fechas procedimientos

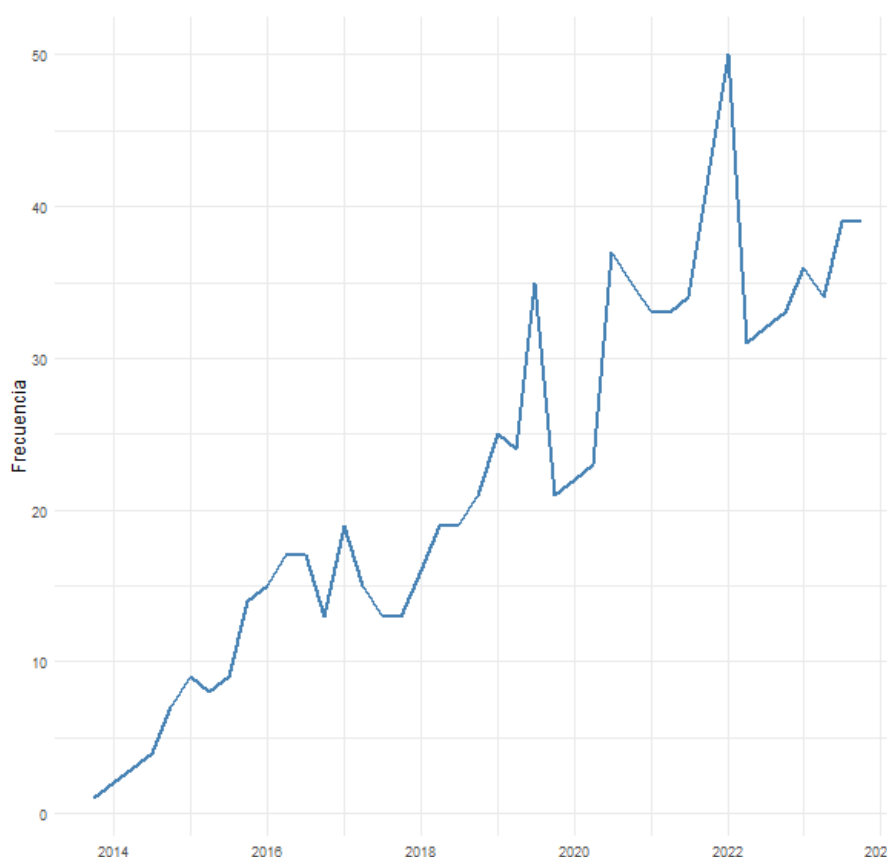


Figura 3.3: Procedimientos por trimestre

En la Figura 3.3 se observa que, en los últimos años, el número de intervenciones para la implantación de una TAVI ha aumentado de forma progresiva, con un descenso puntual alrededor del año 2020, probablemente debido a la pandemia de COVID-19. Como consecuencia de este aumento, una gran proporción de los pacientes incluidos en el estudio fue sometida al procedimiento entre 2021 y 2023.

Acceso a la valvula aórtica

El acceso principal para la inserción de la TAVI es el transfemoral, utilizado en el 93,4 % de los pacientes. En menor medida, se emplea el acceso transaxilar en un 5 %, mientras que otros tipos de acceso representan el 1,6 %.

Tipo de válvulas

En la figura 3.4 se observan los distintos tipos de válvulas implantadas mediante la TAVI en los pacientes con sus distintas categorías. Representando las válvulas autoexpansibles un 70,8 % de las usadas, las expansibles con balón el 28,9 % y las mecánicamente expansibles el 0,3 %.

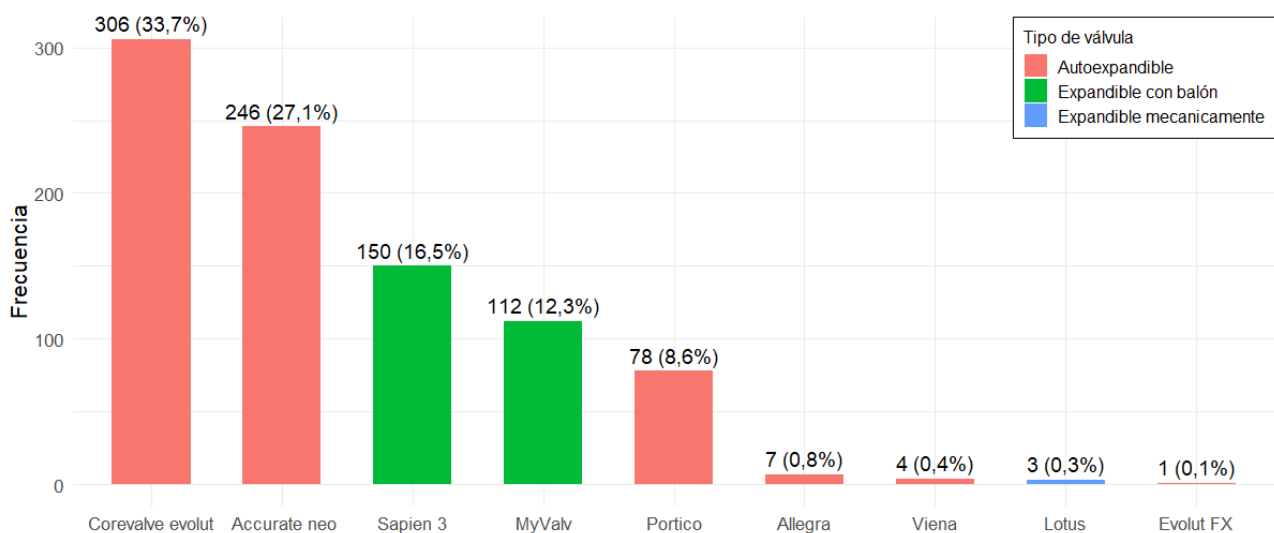


Figura 3.4: Tipos de válvulas usadas en la TAVI

3.5. Análisis supervivencia en un año

Se empezará estudiando la variable de fallecimiento al año, habiendo 90 (9,92 %) fallecimientos entre los 907 pacientes del estudio. Por tanto, el resto de los pacientes, 817 (90,08 %) han sobrevivido al estudio y son censuras. En el siguiente gráfico se muestran las censuras y los fallecimientos, incluyendo el escenario en el que estos se producen (procedimiento, hospital y muertes tras el alta hospitalaria).

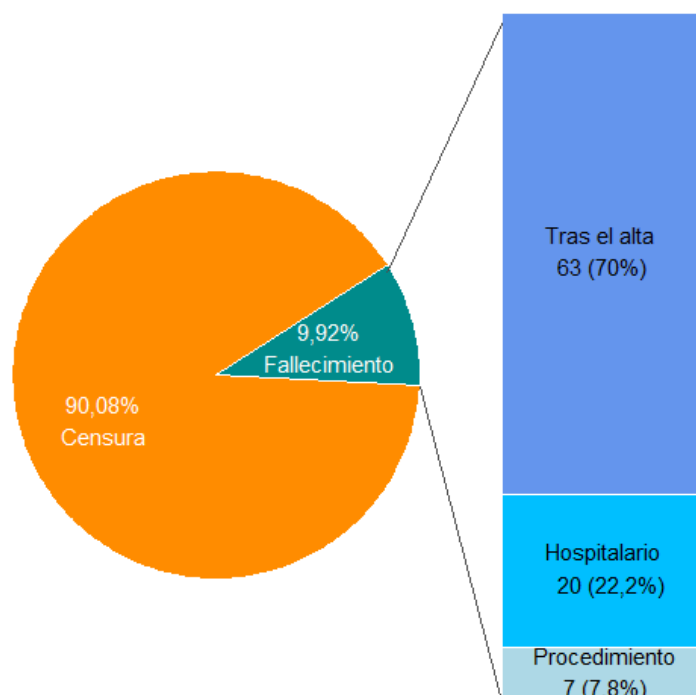


Figura 3.5: Escenario de los fallecimientos

Se observa que en los primeros momentos en el procedimiento y en el hospital mueren un 30 % de los fallecidos del estudio.

Estos fallecimientos en los primeros momentos del estudio se notan en la curva de supervivencia de Kaplan-Meier, explicada en la metodología.

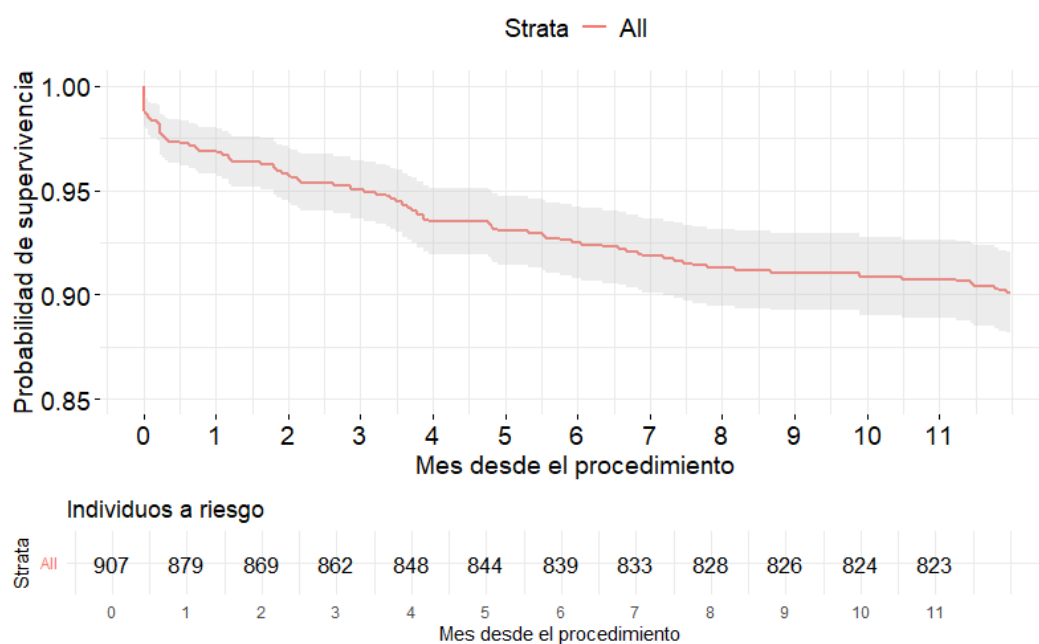


Figura 3.6: Curva de supervivencia en un año Kaplan-Meier

También, se pueden apreciar mejor los momentos en los que el riesgo de fallecimiento es mayor en la siguiente función de riesgo suavizada.

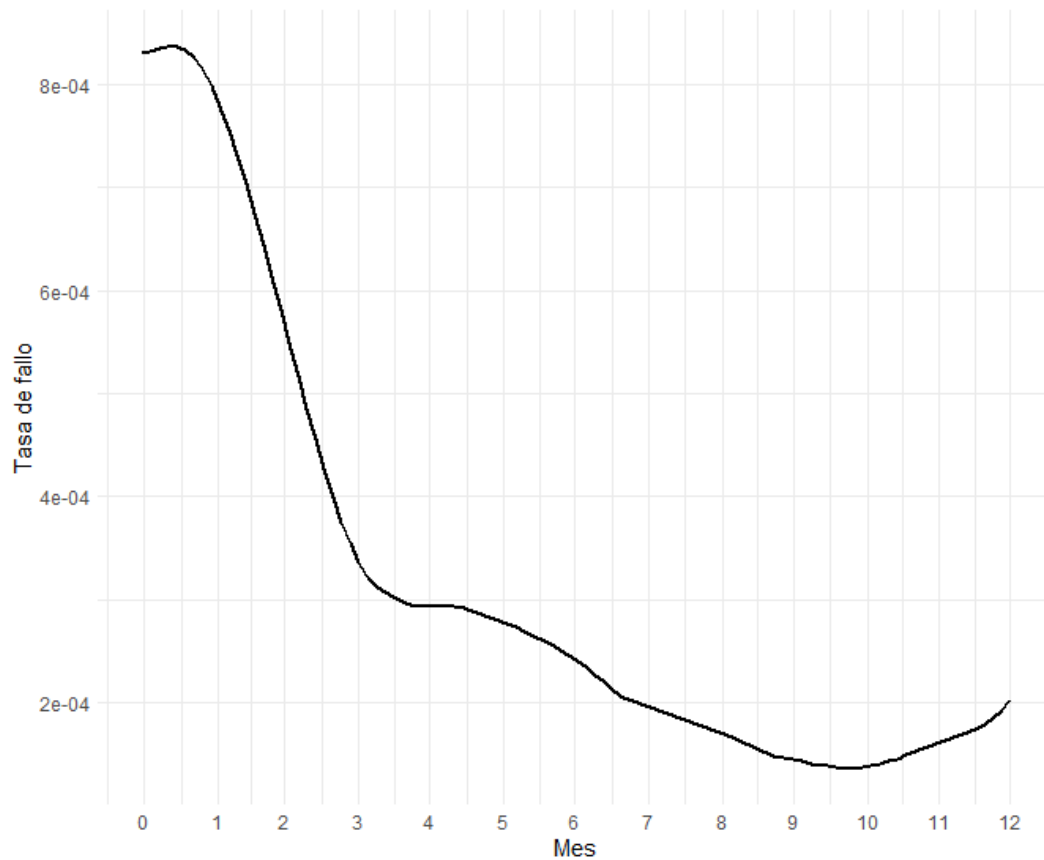


Figura 3.7: Función de riesgo suavizada muerte en un año

Para crear la gráfica se usa suavizado global con núcleo gaussiano y sin corrección de bordes. En este, se observa con mayor claridad que en la curva de supervivencia que los momentos críticos para la supervivencia ocurren en los primeros meses.

4 Resultados del modelo de Cox

Se realizó la selección de variables según lo establecido en la metodología. Posteriormente, se analizaron los resultados obtenidos con el modelo ajustado por dichas variables, evaluando la validez de la hipótesis de riesgos proporcionales, los posibles outliers y la robustez del modelo.

4.1. Selección de variables

Se aplicaron los dos filtros de selección de variables descritos en la metodología, con el objetivo de identificar las variables que serán utilizadas en el modelo.

4.1.1. Primer filtro de selección

La siguiente tabla muestra las variables con p-valor $<0,1$ tras aplicar los tests descritos en la metodología. Para las variables categóricas se indica el total de casos, así como el número de vivos y muertos. Para las numéricas, se presenta la media \pm desviación estándar si no se rechaza normalidad, o la mediana con los percentiles 25 y 75 si se rechaza.

Variable	Total(N=907)	Vivos(N=817)	Muertos(N=90)	P
Insuficiencia renal crónica	193 (21,3 %)	167(20,4 %)	26 (28,9 %)	0,063
Enfermedad arterial periférica	70 (7,7 %)	58 (7,1 %)	12 (13,3 %)	0,035
Fibrilación auricular	318 (35,1 %)	279 (34,1 %)	39 (43,3 %)	0,083
Marcapasos o DAI previo	75 (8,3 %)	63 (7,7 %)	12 (13,3 %)	0,066
Angina	203 (22,4 %)	191 (23,4 %)	12 (13,3 %)	0,030
Disnea	819 (90,3 %)	729 (89,2 %)	90 (100 %)	0,001
Síncope	74 (8,2 %)	67 (8,2 %)	7 (7,8 %)	0,089
Insuficiencia mitral moderada/severa	199 (22 %)	173 (21,2 %)	26 (28,9 %)	0,095
PSAP (mmHg)	40 [33-51]	40 [32-50]	45,5 [37,3-55]	0,008
Mínimo diámetro del anillo aórtico	215 [196-234]	213 [196-233]	227 [198-243]	0,072
Hemoglobina (mg/dl)	12,3 \pm 1,9	12,3 \pm 1,9	11,8 \pm 1,8	0,004
Creatinina (mg/dl)	1 [0,82-1,3]	1 [0,82-1,29]	1,1 [0,86-1,41]	0,084
Grado1 aclaramiento \geq 90	57 (6,3 %)	51 (6,2 %)	6 (6,7 %)	0,003
Grado2 aclaramiento[60-90)	239 (26,4 %)	218 (26,7 %)	21 (23,3 %)	
Grado3a aclaramiento[45-60)	274 (30,2 %)	252 (30,8 %)	22 (24,4 %)	
Grado3b aclaramiento[30-45)	249 (26,4 %)	222 (27,2 %)	27 (30 %)	
Grado4 aclaramiento[15-30)	69 (7,6 %)	63 (7,8 %)	6 (6,7 %)	
Grado5 aclaramiento $<$ 15	17 (1,8 %)	9 (1,1 %)	8 (8,9 %)	

Continúa en la siguiente página

Variable	Total(N=907)	Vivos(N=817)	Muertos(N=90)	P
Filtración glomerular	60,8 [45,7-76,6]	61,2 [46,1-76,7]	57,6 [40,7-70,8]	0,094
Fragilidad	156 (17,2 %)	134 (16,4 %)	22 (24,4 %)	0,068
Euroscore II	3,4 [2,64-4,94]	3,4 [2,64-5,29]	3,8 [2,25-6,3]	0,057
STS score	3,7 [2,64-5,32]	3,7 [2,64-5,29]	4,2 [2,89-6,1]	0,086

Entre estas variables se incluyó el aclaramiento de creatinina, calculada con el método de Cockcroft, agrupada por los intervalos recomendados por la literatura médica. A estas variables se les añadió dos cuyo p-valor en el test de log-rank es menor que 0,1.

Variable	Total(N=907)	Vivos(N=817)	Muertos(N=90)	P log-rank
Diabéticos insulino dependientes	88 (9,7 %)	74 (9,1 %)	14 (15,5 %)	0,04
Ateroma complejo	11 (1,2 %)	8 (1 %)	3 (3,3 %)	0,02

Tabla 4.2: Test de log-rank con p-valores menores que 0,1 de variables no seleccionadas

Antes del segundo filtro, se descartaron varias variables por distintos motivos, como la falta de relevancia clínica, escasa frecuencia de casos positivos (como el ateroma complejo), problemas de separación completa (como en disnea, donde todos los fallecidos presentan esta condición) o por ser combinaciones de otras variables, como las escalas de riesgo EuroSCORE II y STS Score.

4.1.2. Segundo filtro de selección

En este, se identificaron las variables seleccionadas en función de los resultados del primer filtro y de su relevancia clínica, determinada por los investigadores del HCUV. A continuación, se describe cómo han sido categorizadas las variables numéricas. Posteriormente, se realizó un análisis mediante sus curvas de Kaplan-Meier. Finalmente, se aplicó el método de selección automática descrito en la metodología. Las variables que han sido usadas en este filtro son las siguientes:

- Enfermedad arterial periférica (EAP)
- Fibrilación auricular (FA)
- Insuficiencia mitral moderada o severa
- Presión sistólica de la arteria pulmonar (PSAP)
- Hemoglobina (mg/dl)
- Creatinina (mg/dl)
- Aclaramiento creatinina (mg/dl)
- Filtración glomerular
- Insuficiencia renal crónica
- Diabéticos insulino dependientes

Las variables numéricas se dividieron en distintas categorías para facilitar la interpretación de los resultados del modelo. Estas categorías son las siguientes:

- La PSAP se categorizó en dos grupos según considerando presiones menor de 30 mmHg como bajas [7]. Los valores que no pudieron ser medidos debido a presiones demasiado bajas se incorporaron al grupo menor de 30 mmHg.
- La hemoglobina ha sido categorizada según la presencia o ausencia de anemia. Aunque no exista un valor único y universalmente aceptado para definir la anemia, los puntos de corte suelen oscilar entre 11 y 13 g/dL [8]. Para esta separación, se utilizará un valor de 11,8 g/dL, ya que tras haber probado con varios es el que mejores resultados aportó y es clínicamente razonable.
- La creatinina se categorizó utilizando un punto de corte de 0,6. Valores inferiores a este han sido considerados creatinina baja.
- El aclaramiento de creatinina se agrupó en tres categorías según los grados de esta variable. La primera incluye los grados 1 y 2, la segunda los grados 3a y 3b, y la tercera los grados 4 y 5.
- El filtrado glomerular se dividió en dos considerando los valores menores de 30 como bajos, estos valores se consideran como enfermedad renal (<15) o disminución grave de la función renal [15-30] [9] .

Análisis de curvas Kaplan-Meier para cada variable

Se analizaron las curvas de Kaplan-Meier con el objetivo de identificar las variables que presentan mayor relevancia en la supervivencia. Asimismo, en algunas de estas variables se evaluó la pertinencia de considerar posibles modificaciones o categorizaciones alternativas que pudieran mejorar la interpretación de los resultados.

Empezamos representando la hemoglobina con dos cortes según se considere que hay anemia o no.

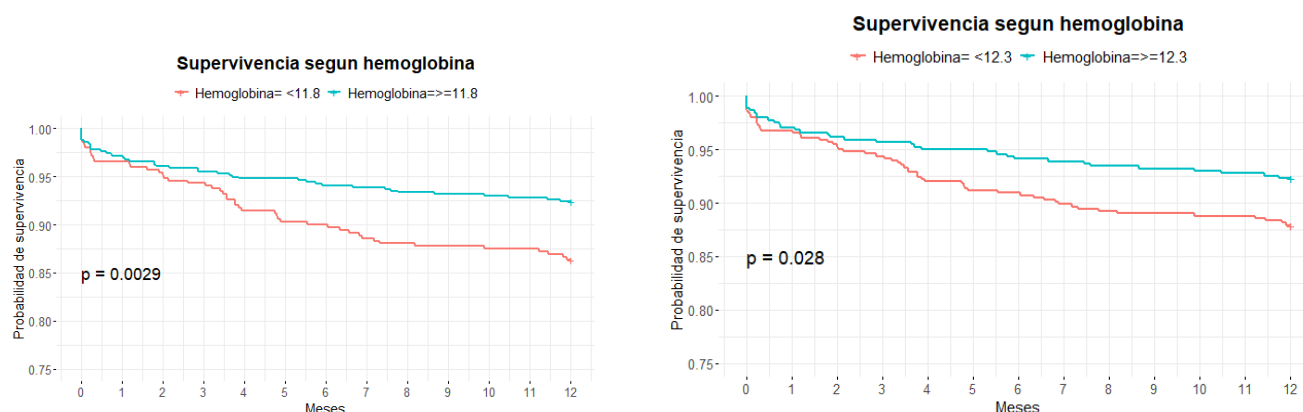


Figura 4.1: Funciones de supervivencia según hemoglobina

Se observa que el punto de corte 11,8 discrimina mejor que el 12,3.

Mostramos las cuatro variables relacionadas con problemas cardíacos.

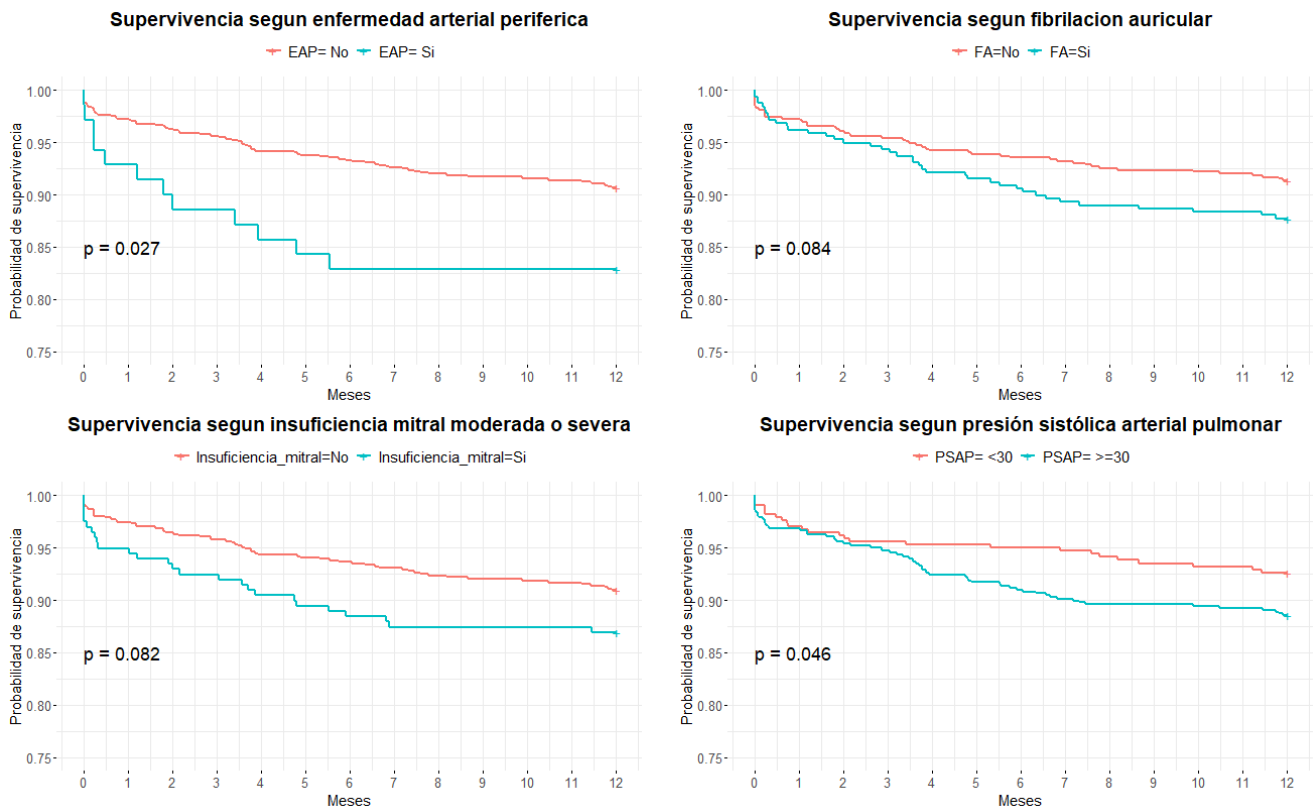


Figura 4.2: Funciones de supervivencia según variables cardíacas

Se exploró la posibilidad de separar a los pacientes diabéticos en insulino dependientes y no insulino dependientes para el análisis de supervivencia. No obstante, las curvas de Kaplan-Meier no evidenciaron diferencias relevantes en la supervivencia de los no diabéticos y los diabéticos no insulino dependientes.

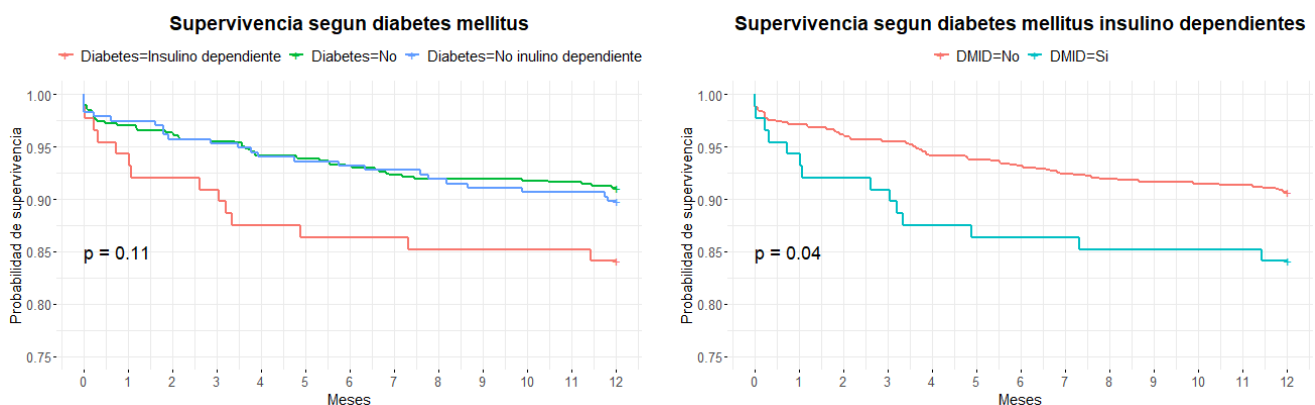


Figura 4.3: Funciones de supervivencia según el tipo de diabetes

Para concluir con el análisis de las funciones de supervivencia, se exploraron las variables asociadas a problemas renales.

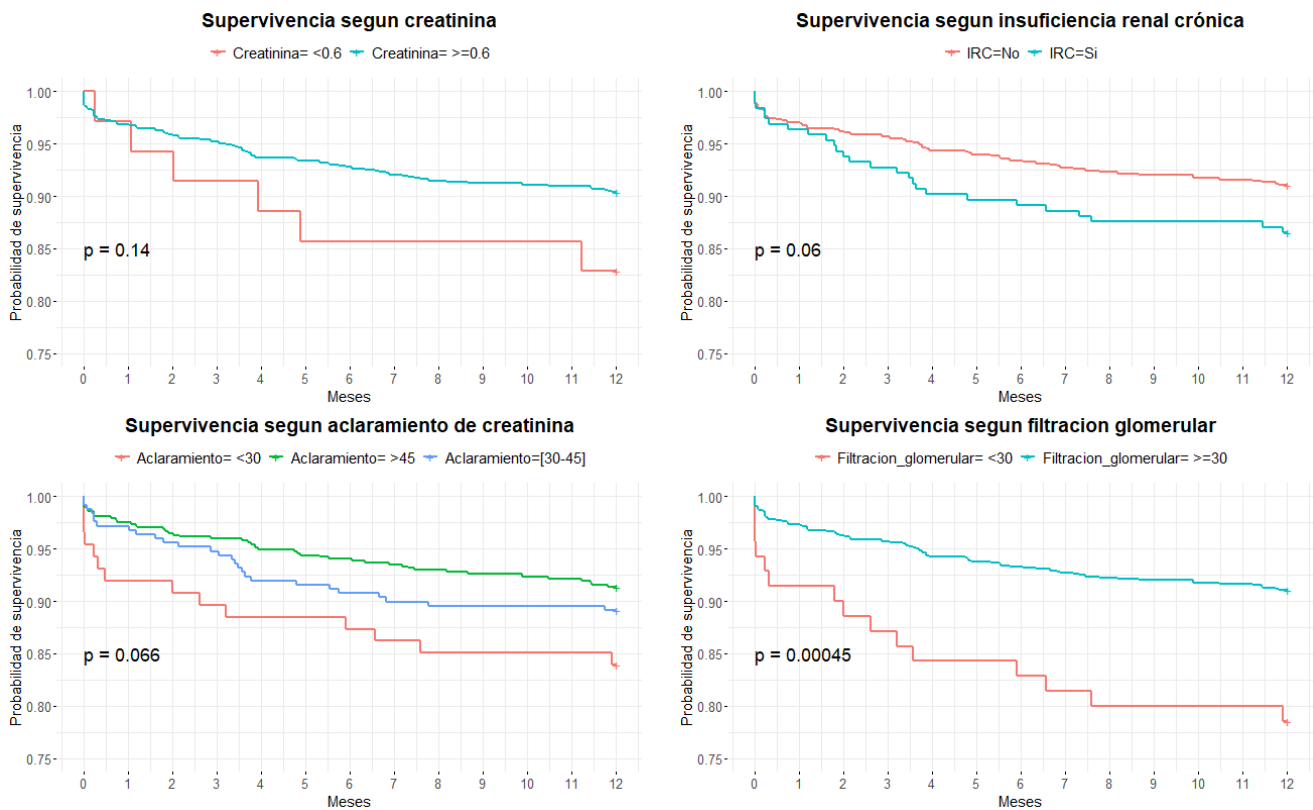


Figura 4.4: Funciones de supervivencia según variables de la función renal

Aunque existan diferencias marcadas en las curvas de Kaplan-Meier, ello no garantiza que una variable sea adecuada para su inclusión en un modelo de Cox. Es imprescindible, además, que dicha variable cumpla la hipótesis de riesgos proporcionales, condición que será evaluada en el modelo final.

Variables seleccionadas

Tras haber aplicado la selección automática stepwise basada en el p-valor del test de log-rank (con umbrales de entrada y salida de 0,1) explicada en la metodología. Se obtuvieron las siguientes variables:

- Filtración glomerular separada en dos grupos <30 y ≥30.
- Hemoglobina dividida en dos grupos <11,8 y ≥11,8.
- Enfermedad arterial periférica.
- Presión sistólica arterial pulmonar dividida en dos grupos <45 y ≥45.

4.2. Modelo obtenido

Con las variables seleccionadas, se obtuvo un modelo de Cox con una concordancia de 0,616, lo cual indica una capacidad predictiva moderada. El intervalo de confianza del 95 %, basado en la aproximación normal, fue de 0,557 a 0,675.

Variable	Hazard ratio	IC Hazard ratio	p-valor log rank
Filtrado glomerular <30	2,195	(1,251 – 3,851)	0,0061
Enfermedad arterial periférica	1,951	(1,060 – 3,591)	0.0317
Hemoglobina <11,8	1,698	(1,115 – 2,584)	0,0136
Presión sistólica arterial pulmonar ≥ 30	1,522	(0,958 – 2,418)	0,0756

Tabla 4.3: Resultados modelo de Cox

4.2.1. Estudio de los residuales

Residuos de Schoenfeld

A partir de estos, se comprobó que no se rechaza la hipótesis de riesgos proporcionales para ninguna variable a un nivel inferior al 0,1.

Global Schoenfeld Test p: 0.413

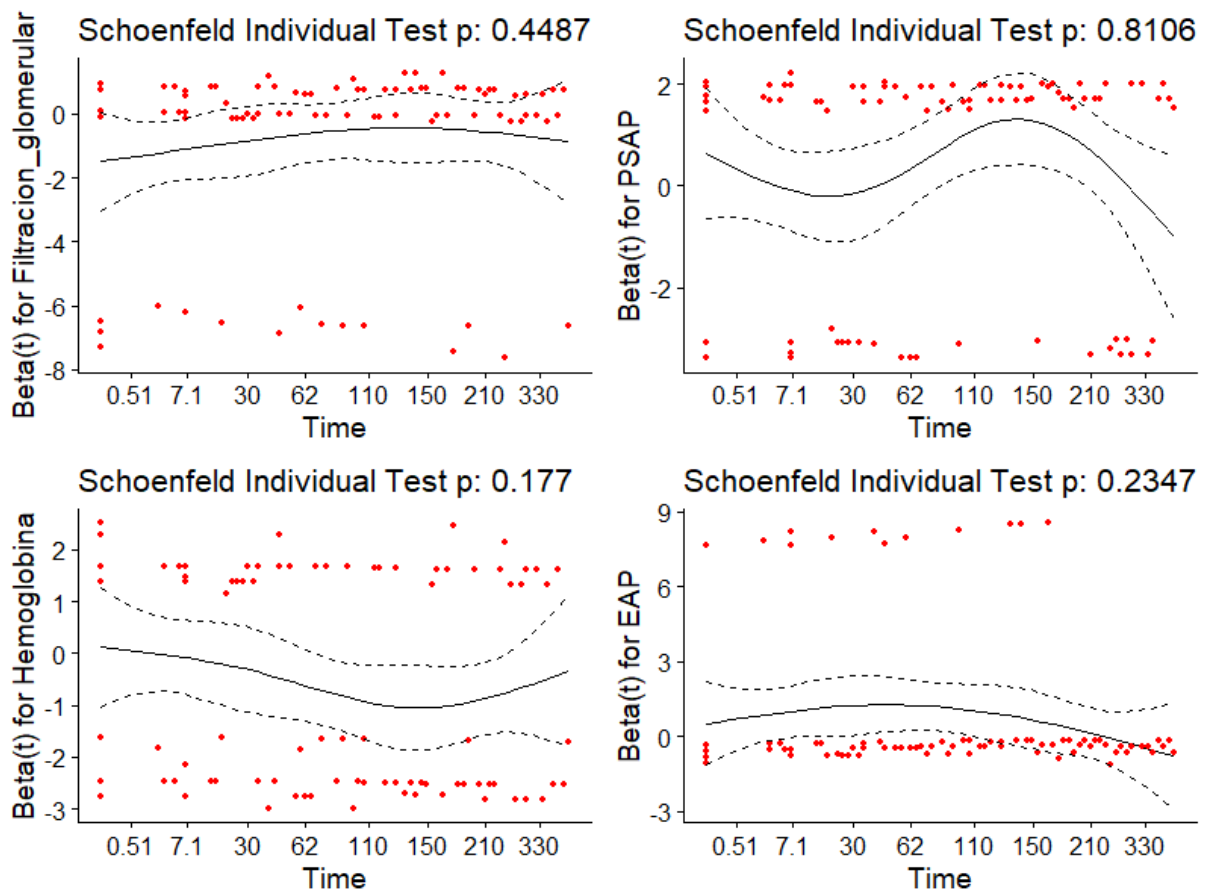


Figura 4.5: Residuos de Schoenfeld y p-valores del test de la hipótesis de riesgos proporcionales

Residuos de deviance

Estos reflejan el grado de discrepancia entre el tiempo de supervivencia observado y el predicho por el modelo para cada individuo.

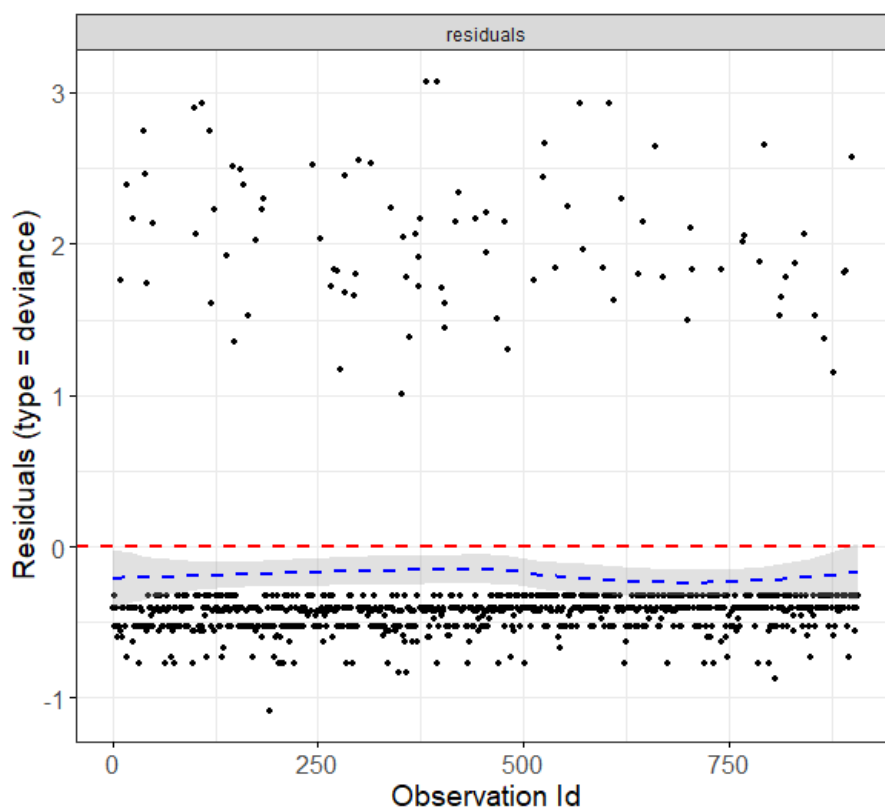


Figura 4.6: Residuos deviance

Los valores negativos de estos residuos se corresponden a observaciones censuradas o a individuos con mayor tiempo de supervivencia del previsto por el modelo. En nuestro caso, estos residuos negativos se debieron a censuras y no parecen preocupantes, ya que todos se encuentran por debajo de 1 en valor absoluto, lo cual sugiere un buen ajuste para esas observaciones.

Por otro lado, los residuos positivos por ser mayores en valor absoluto merecen más atención. Estos indican que el modelo subestima el riesgo, es decir, predice que el fallecimiento ocurrirá más tarde de lo que realmente ocurre. Valores mayores a 2 pueden considerarse preocupantes, y valores por encima de 3 indican una elevada discrepancia que podría sugerir la presencia de observaciones atípicas o un ajuste deficiente del modelo para ciertos individuos. Estos valores elevados de residuos se deben a que, en esas observaciones, los factores de riesgo no se alinean con la ocurrencia del fallecimiento antes de un año, pero el suceso ocurre. Por ejemplo, en los dos casos con estos residuos superiores a 3, los pacientes presentan niveles de hemoglobina $\geq 11,8$ y filtración glomerular ≥ 30 , indicadores vinculados a menor riesgo, aunque en estos casos los pacientes fallecieron antes del primer año.

Residuos de influencia

En cuanto a los residuos de influencia (DFBETA), que miden el cambio en los coeficientes del modelo al eliminar una observación individual, no se identificaron valores notablemente atípicos. Los valores más extremos se hallaron en las 12 observaciones vinculadas con la variable enfermedad arterial periférica (EAP), correspondientes a casos con valores positivos de EAP en los que, además, se produjo el fallecimiento. Aunque estos valores mostraron cierta influencia sobre el coeficiente asociado a EAP, su magnitud es moderada (todos son menores a 0,08) y se mantienen ligeramente por encima del umbral habitualmente considerado preocupante, que se estima en $\frac{2}{\sqrt{n}} \approx 0,066$. Por tanto, si bien podrían generar ligeras variaciones en dicho coeficiente al ser eliminados, no comprometen de forma relevante la estabilidad general del modelo.

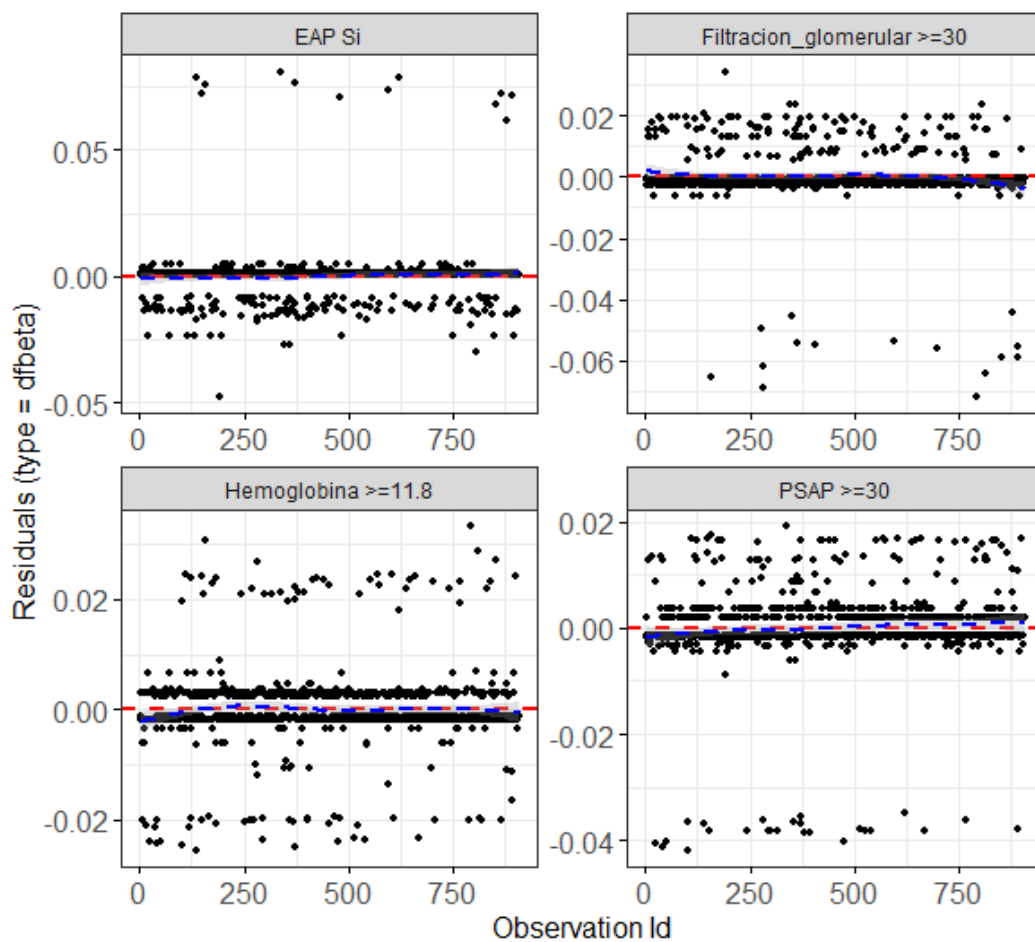


Figura 4.7: Residuos de influencia

5 Aplicación web

Una vez obtenido el modelo de Cox, se desarrolló una aplicación web que permite su exploración de forma interactiva. Esta herramienta facilita la visualización de cómo varía la curva de supervivencia en función de diferentes combinaciones de los valores de las variables, así como la estimación del percentil de riesgo de fallecimiento asociado a un paciente según los valores introducidos. Todo ello proporciona un recurso útil para los profesionales sanitarios, al permitirles analizar con mayor precisión los casos con baja probabilidad de supervivencia y contar con un apoyo adicional en la toma de decisiones clínicas.

Aunque inicialmente podría considerarse excluir del procedimiento TAVI a los pacientes con mayor riesgo estimado de fallecimiento, lo más adecuado es identificar las causas subyacentes que elevan ese riesgo. De esta forma, se busca primero corregir o mitigar dichos factores, disminuyendo el riesgo y permitiendo que más pacientes se beneficien de la TAVI en condiciones más seguras. Para facilitar la identificación de estos pacientes de manera gráfica y sencilla, se desarrollará esta aplicación.

Para su creación, usaremos el paquete Shiny [10] de R y desplegaremos la app a través shinyapps.io [11]. Se podrá acceder a la misma mediante el siguiente enlace: [app web Tavi](#)

5.1. Shiny

Este es un paquete de R que facilita la creación de aplicaciones web interactivas. En **shiny**, el desarrollo de una aplicación web se divide en dos componentes principales: la interfaz de usuario (UI) y el servidor (server). La UI representa la parte visible de la aplicación, donde se definen elementos como botones, textos y la estructura general de la interfaz. Por otro lado, el server contiene la lógica que gestiona cómo debe reaccionar la aplicación ante las acciones del usuario, generando respuestas dinámicas.

Estas respuestas se logran mediante el sistema de reactividad, un mecanismo central en **shiny**, que permite que la aplicación responda automáticamente a los cambios realizados por el usuario. Cuando un input, como un botón o un switch, se modifica, el server detecta ese cambio y actualiza de forma dinámica los outputs asociados, sin necesidad de recargar la página. Así, la reactividad actúa como el puente que conecta las acciones del usuario con las actualizaciones en la interfaz.

Tras realizar la aplicación en **shiny**, la subiremos a la nube mediante shinyapps.io . Esta plataforma permite publicar y compartir aplicaciones Shiny de una forma sencilla, sin necesidad de configurar un servidor propio.

5.2. Descripción de la aplicación

La aplicación tendrá que permitir visualizar los resultados del modelo sobre distintas combinaciones de los valores de las variables de una forma intuitiva e interactiva. Para ello, dividiremos la app en tres partes según la función que cumpla:

- **Inputs.** En la parte superior se encuentran cuatro botones tipo switch, los cuales permiten configurar los valores de las variables predictoras del modelo. Inicialmente, cada switch representa un valor negativo para la variable correspondiente, al deslizarlo el valor se convierte en positivo. Para hacer más intuitivo el efecto de las variables en el modelo, se considera el valor negativo de las variables como protector y al modificarlo el switch se vuelve rojo indicando que la probabilidad de supervivencia disminuye.
- **Gráficos.** En la parte central se mostrará un gráfico interactivo que puede alternarse mediante botones entre dos opciones. La primera opción presenta la función de supervivencia generada a partir de las predicciones del modelo de Cox con los valores ingresados por el usuario, junto con la función correspondiente al conjunto completo de datos. En la segunda opción, se visualizará el riesgo de fallecimiento a un año asociado a los valores actuales de las variables predictoras, así como el percentil de riesgo correspondiente en relación con el conjunto de datos. Los gráficos se generarán con el paquete `plotly`, lo que permitirá añadir funcionalidades interactivas, como se ilustrará más adelante.
- **Tabla de supervivencia.** En la zona inferior se mostrará una tabla con la probabilidad de supervivencia para cada mes, expresada como porcentaje, según las predicciones del modelo de Cox basadas en los valores introducidos. También se incluirá la probabilidad observada en la muestra del estudio, lo que permitirá comparar ambas de forma rápida y sencilla.

5.3. Ejemplos de uso

En esta sección se presentará el funcionamiento de la aplicación web de forma práctica, mostrando cómo varían las distintas partes de la interfaz gráfica al interactuar con ella. Esta se realizará a través de los switches, los botones y el gráfico interactivo, como se podrá observar en las siguientes figuras.

Al abrir la aplicación, se presentan las tres secciones descritas previamente, incluyendo en la zona de gráficos las curvas de supervivencia de Kaplan-Meier.

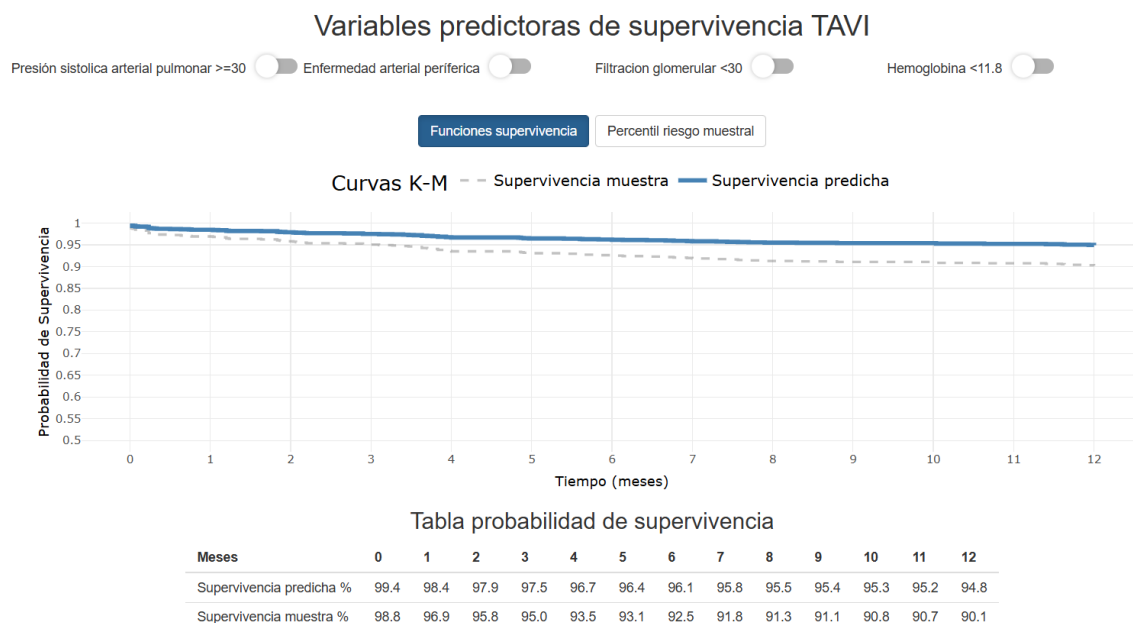


Figura 5.1: Página web tras abrirla

Al modificar alguno de los valores de las variables, se actualizan los resultados reflejados en el gráfico y en la tabla de supervivencia.

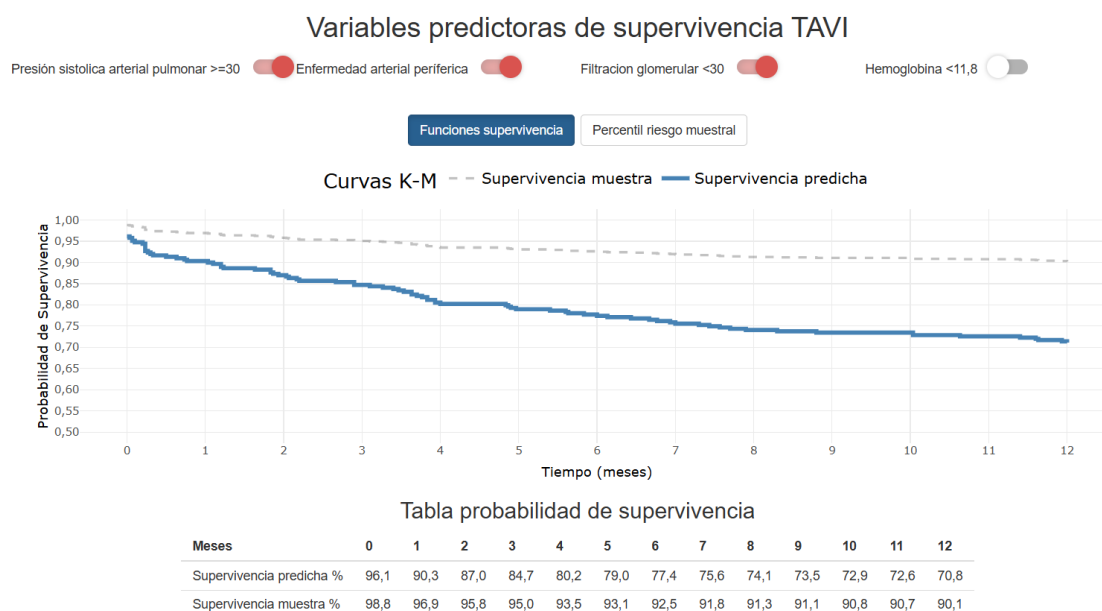


Figura 5.2: Modificaciones al deslizar los switch de las variables

Al seleccionar los distintos botones del tipo de gráfico, la visualización se actualiza automáticamente según la opción elegida. Por ejemplo, al pulsar el botón *Percentil riesgo muestral*, se genera un gráfico que muestra el riesgo predicho en función del percentil de riesgo de fallecimiento dentro de la muestra.

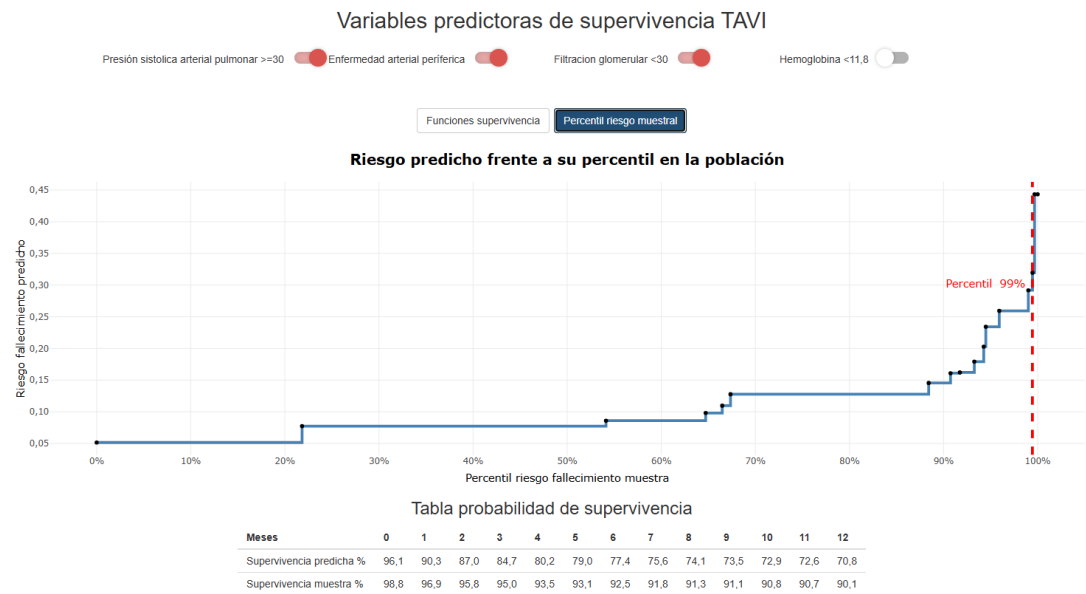


Figura 5.3: Riesgo anual según combinaciones de variables y su percentil asociado.

Los gráficos son interactivos, lo que permite visualizar los valores exactos al situar el cursor (o el dedo, en dispositivos táctiles) sobre ellos. Además, se puede reescalar las curvas K-M utilizando el botón Autoscale y restaurar las dimensiones originales pulsando en el ícono de la casa. Estos elementos se visualizan al pasar el cursor sobre el gráfico, o bien se muestran por defecto en dispositivos que no cuentan con ratón.

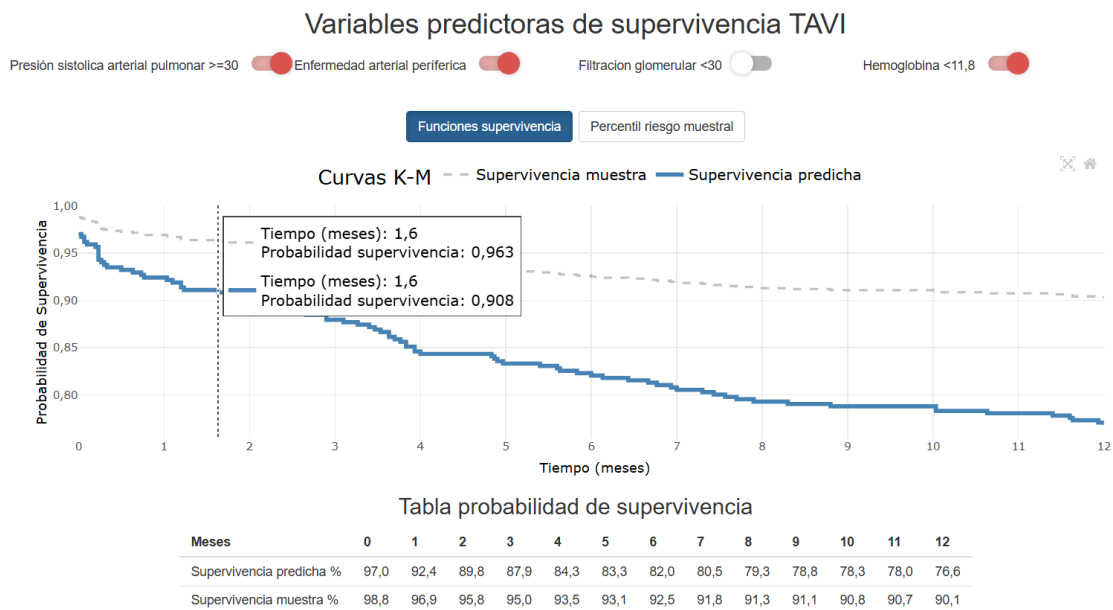


Figura 5.4: Aplicación web con las curvas K-M escaladas.

La aplicación cuenta con un diseño responsivo, lo que garantiza una visualización adecuada para distintos tamaños de pantalla. En caso de que el ancho disponible sea reducido, la tabla de riesgo se transpone automáticamente para mantener su legibilidad, los gráficos ocuparán todo el ancho de la pantalla y los switches aparecerán en distintas filas.

variables predictoras de supervivencia TAVI

Presión sistólica arterial pulmonar ≥ 30 ☒

Enfermedad arterial periférica ☐

Filtración glomerular < 30 ☒

Hemoglobina $< 11,8$ ☐

Funciones supervivencia

Percentil riesgo muestral

Curvas K-M -- Supervivencia muestra

— Supervivencia predicha

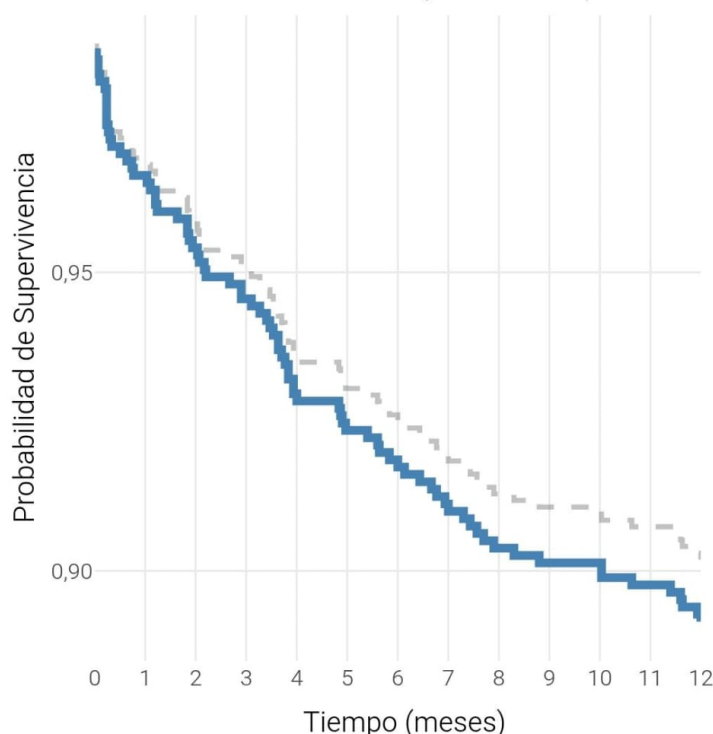


Tabla probabilidad de supervivencia

Figura 5.5: Aplicación vista en dispositivos con pantallas estrecha como los móviles.

6 Conclusiones

Este Trabajo de Fin de Grado ha permitido identificar, mediante un análisis de supervivencia, los principales factores asociados a la mortalidad en pacientes sometidos a la TAVI. Utilizando datos de 907 pacientes del Hospital Clínico Universitario de Valladolid con un seguimiento de un año, se ha ajustado un modelo de regresión de Cox que ha revelado como variables predictoras significativas la presencia de enfermedad arterial periférica, anemia, filtrado glomerular menor de 30 y presión sistólica en la arteria pulmonar elevada (mayor o igual a 30). Todas estas son variables que aumentan el riesgo de fallecimiento.

El modelo desarrollado, con un índice de concordancia moderado (0,616), permite estimar el riesgo individualizado de fallecimiento en el primer año tras la intervención. Además, se ha implementado una aplicación web interactiva, mediante `shiny`, que facilita la consulta y visualización de estas estimaciones, lo que representa una herramienta potencialmente útil para la práctica clínica.

Sin embargo, el estudio presenta algunas limitaciones, como su carácter retrospectivo, la posible existencia de variables mal registradas y la falta de validación externa del modelo. Además, resulta llamativo que la edad no se haya identificado como un factor de riesgo significativo. Una posible explicación es que los pacientes más jóvenes incluidos en el estudio podrían haber presentado un estado de salud más comprometido, lo que justificaría su indicación para TAVI frente a opciones quirúrgicas más invasivas como la RQVA, tradicionalmente reservadas a personas con menor riesgo.

Como líneas futuras, sería conveniente ampliar el período de seguimiento y validar el modelo con datos de otros centros. También sería interesante valorar la utilidad práctica de la herramienta desarrollada en entornos reales.

Bibliografía

- [1] American Heart Association. URL: https://www.heart.org/-/media/Files/Health-Topics/Answers-by-Heart/Answers-by-Heart-SPANISH/What-is-TAVR-Spanish.pdf?sc_lang=en.
- [2] David H. Adams et al. “Transcatheter Aortic-Valve Replacement with a Self-Expanding Prosthesis”. En: *New England Journal of Medicine* 370.19 (2014), págs. 1790-1798. DOI: 10.1056/NEJMoa1400590. URL: <https://doi.org/10.1056/NEJMoa1400590>.
- [3] Dirk F. Moore. “Model Diagnostics”. En: *Applied Survival Analysis Using R*. Springer International Publishing, 2016. ISBN: 978-3-319-31243-9. DOI: 10.1007/978-3-319-31245-3_7.
- [4] John P. Klein y Melvin L. Moeschberger. “Survival Analysis: Techniques for Censored and Truncated Data”. En: 2nd. New York: Springer, 2003. ISBN: 978-0-387-95399-1.
- [5] Terry M. Therneau y Elizabeth Atkinson. *Concordance*. Vignette del paquete survival. 2024. URL: <https://cran.r-project.org/web/packages/survival/vignettes/concordance.pdf>.
- [6] American Heart Association. *Classes and Stages of Heart Failure*. 2023. URL: <https://www.heart.org/en/health-topics/heart-failure/what-is-heart-failure/classes-of-heart-failure>.
- [7] José Ramón de Berrazueta Fernández. “Hipertensión pulmonar. Aclarando conceptos”. En: *Anales de la Real Academia Nacional de Medicina* 138.2 (2021), págs. 137-142. DOI: 10.32440/ar.2021.138.02.doc01. URL: https://analesranm.es/revista/2021/138_02/13802_doc01#:~:text=La%20Hipertensi%C3%B3n%20Arterial%20pulmonar%20es,media%20superior%20a%2025%20mmHg..
- [8] O. Y. Addo et al. “Evaluation of Hemoglobin Cutoff Levels to Define Anemia Among Healthy Individuals”. En: *JAMA Network Open* 4.8 (2021), e2119123. DOI: 10.1001/jamanetworkopen.2021.19123. URL: <https://doi.org/10.1001/jamanetworkopen.2021.19123>.
- [9] National Kidney Foundation. *Filtración glomerular estimada (FGe)*. 2025. URL: <https://www.kidney.org/es/kidney-topics/filtracion-glomerular-estimada-fge>.
- [10] PBC Posit. *Shiny by Posit: Easy web applications in R*. 2024. URL: <https://shiny.posit.co/>.
- [11] PBC Posit. *shinyapps.io: Deploy Shiny apps to the web*. 2024. URL: <https://www.shinyapps.io/>.

Apéndice A

Descripción variables

A partir de las variables incluidas en el análisis descriptivo, se elaborará una tabla resumen que presentará los p-valores correspondientes a los test descritos en la metodología. Además, para las variables categóricas, se mostrará el total de casos, así como la distribución entre pacientes vivos y fallecidos. En el caso de las numéricas, se mostrará la media \pm desviación estándar cuando no se rechace la hipótesis de normalidad; en caso contrario, se expondrá la mediana junto con los percentiles 25 y 75. En aquellas variables que puedan ser categorizadas de manera natural, se analizará su comportamiento en cada grupo. Por ejemplo, el índice de masa corporal se analiza para las categorías: bajo peso ($<18,5$), peso normal [$18,5-30$], sobrepeso [$30-40$] y obesidad (>40).

A.1. Variables demográficas

Variable	Total (N=907)	Vivos (N=817)	Muertos (N=90)	P
Edad	82 [78-85]	82 [78-85]	82 [78-85]	0,84
Mujeres	408 (45 %)	372 (45,5 %)	36 (40 %)	0,317
Enfermedad coronaria previa	330 (36,4 %)	298 (36,5 %)	32 (35,5 %)	0,863
Diabetes mellitus	323 (35,6 %)	285 (34,9 %)	38 (42,2 %)	0,168
Insulino dependientes	88/323(27,2 %)	74/285(26 %)	14/38(36,8 %)	0,157
Hipertensión arterial	702(0 77,4 %)	633 (77,5 %)	69 (76,7 %)	0,861
ACTP previa	222 (24,5 %)	200 (24,5 %)	22 (24,4 %)	1
IRC	193 (21,3 %)	167 (20,4 %)	26 (28,9 %)	0,063
EPOC	94 (10,4 %)	83 (10,2 %)	11 (12,2 %)	0,542
Enfermedad arterial periférica	70 (7,7 %)	58 (7,1 %)	12 (13,3 %)	0,035
FA	318 (35,1 %)	279 (34,1 %)	39 (43,3 %)	0,083
IMC (kg/m ²)	26,7 [24,2-29,7]	26,6 [24,2-29,7]	27,3 [24,3-31]	0,2
IMC ≥ 40 (kg/m ²)	15 (1,6 %)	13 (1,6 %)	2 (2,7 %)	0,63
IMC 30-39,9(kg/m ²)	203 (22,4 %)	179 (21,9 %)	24 (26,7 %)	
IMC 18,5-29,9(kg/m ²)	682(74,4 %)	618(75,6 %)	64(71,1 %)	
IMC $<18,5$ (kg/m ²)	7 (0,8 %)	7 (0,8 %)	0 (0 %)	
Remitidos de otros hospitales	537 (59,2 %)	482 (59 %)	55 (61,1 %)	0,698
Insuficiencia cardíaca	174 (19,12 %)	158 (19,3 %)	16 (17,8 %)	0,383

Continúa en la siguiente página

Variable	Total(N=907)	Vivos(N=817)	Muertos(N=90)	P
Tabaquismo				
No	678 (74,8 %)	614 (75,2 %)	64 (71,1 %)	0,448
Activo	51 (5,6 %)	47 (5,7 %)	4 (4,4 %)	
Extabaquismo	178(19,6 %)	156 (19,1 %)	22 (24,4 %)	
Intervención coronaria previa	236 (26,0 %)	211 (25,8 %)	25 (27,8 %)	0,689
Infarto de miocardio previo	122 (13,5 %)	113 (13,8 %)	9 (10 %)	0,312
>90 días	114 (12,6 %)	105 (12,9 %)	9 (10 %)	1
<30 días	5 (0,6 %)	5 (0,6 %)	0 (0 %)	
30-90 días	3 (0,3 %)	3 (0,4 %)	0 (0 %)	
Cirugía cardiaca previa	60 (6,6 %)	56 (6,9 %)	4 (4,4 %)	0,383
Cirugía valvular previa	38 (4,2 %)	35 (4,3 %)	3 (3,3 %)	1
Bypass coronario previo	31 (3,4 %)	28 (3,4 %)	3 (3,3 %)	1
Ictus previo	80 (8,8 %)	70 (8,6 %)	10 (11,1 %)	0,419
Marcapasos o DAI previo	75 (8,3 %)	63 (7,7 %)	12 (13,3 %)	0,066
Fragilidad	156 (17,2 %)	134 (16,4 %)	22 (24,4 %)	0,068
Aorta en porcelana	27 (3 %)	24 (2,9 %)	3 (3,3 %)	0,74
Cáncer	145 (16 %)	129 (15,8 %)	16 (17,8 %)	0,469
Placa aterosclerótica compleja	11 (1,2 %)	8 (1 %)	3 (3,3 %)	0,034
Euroscore I logistic	11,4 [8,16-17,5]	11,4 [8,1-17,3]	11,9 [8,48-19,9]	0,36
Euroscore II	3,4[2,64-4,94]	3,4 [2,64-5,29]	3,8 [2,25-6,3]	0,057
STS score	3,7 [2,64-5,32]	3,7 [2,64-5,29]	4,2 [2,89-6,1]	0,086
CHADSVASC	4 [3-5]	4 [3-5]	4 [3-5]	0,29
CHADSVA	3 [3-4]	3 [3-4]	4 [3-4]	0,11
Tavi2 Score	2 [1-2]	2 [1-2]	2 [1-2]	0,52

A.2. Variables clínicas

Variable	Total(N=907)	Vivos(N=817)	Muertos(N=90)	P
Angina	203 (22,4 %)	191 (23,4 %)	12 (13,3 %)	0,030
Disnea	816 (90 %)	729 (89,2 %)	87 (96,7 %)	0,026
Síncope	74 (8,2 %)	67 (8,2 %)	7 (7,8 %)	0,089
Asintomáticos	16 (1,8 %)	16 (2 %)	0 (0 %)	0,39
NYHA I	56 (6,2 %)	52 (6,4 %)	2 (2,2 %)	0,42
NYHA II	394 (43,4 %)	354 (43,3 %)	40 (44,4 %)	
NYHA III	403 (44,4 %)	358 (43,8 %)	45 (50 %)	
NYHA IV	24 (2,6 %)	22 (2,7 %)	2 (2,2 %)	

A.3. Variables pruebas complementarias

Variable	Total(N=907)	Vivos(N=817)	Muertos(N=90)	P
FEVI	60 [52-65]	60 [52-65]	60 [50-65]	0,966
FEVI \leq 40 %	116 (12,8 %)	103 (12,6 %)	13 (14,4 %)	0,82
FEVI 40-49,9 %	76 (8,4 %)	68 (8,3 %)	8 (8,8 %)	
FEVI \geq 50 %	715 (78,8 %)	646 (79,1 %)	69 (76,7 %)	
PSAP(mmHg)	40 [33-51]	40 [32-50]	45,5 [37,3-55]	0,008
Área valvular aórica(cm ²)	0,7 [0,6-0,9]	0,7 [0,6-0,9]	0,75 [0,6-0,87]	0,982
Gradiente máximo transvalvular(mmHg)	72 [61,4-86]	72 [61,9-86]	70 [61-91]	0,901
Gradiente medio transvalvular(mmHg)	45 [39-55]	45 [39-55]	43 [38,8-58]	0,668
Velocidad máxima transvalvular	4,2 [3,9-4,7]	4,2 [3,9-4,7]	4,2 [3,9-4,8]	0,762
Insuficiencia aórtica moderada o severa	200 (22,1 %)	182 (22,3 %)	18 (20 %)	0,617
Insuficiencia mitral moderada o severa	199 (22 %)	173 (21,2 %)	26 (28,9 %)	0,095
Insuficiencia tricúspide moderada o severa	138 (15,2 %)	122 (14,9 %)	16 (17,8 %)	0,482
Máximo diámetro del anillo aórtico	271 [250-291]	272 [250-290]	270 [250-294]	0,736
Mínimo diámetro del anillo aórtico	215 [196-234]	213 [196-233]	227 [198-243]	0,072
Área del anillo aórtico	454,5 [394-527]	454 [393-521]	458 [410-555]	0,232
Perímetro del anillo aórtico	78 [72-83]	77 [72-83]	78 [73-85,5]	0,4115
Altura hasta el tronco coronario izquierdo	13,3 [11,5-15,4]	13,4 [11,6-15,4]	13 [11,5-15,1]	0,76
Altura hasta ostium coronaria derecha	14,5 \pm 3,5	14,4 \pm 3,5	14,7 \pm 3,8	0,549
Máximo diámetro de aorta sinusal	33 [30-35,8]	33 [30-36]	33 [30,2-25]	0,908
Válvula aórtica bicúspide	20 (2,2 %)	19 (2,32 %)	1 (1,1 %)	1
LBBB sin ritmo marcapasos	96/852 (11,3 %)	83/768 (10,8 %)	13/84 (15,5 %)	0,20
RBBB sin ritmo marcapasos	113/852 (13,2 %)	102/768 (13,3 %)	11/84 (13,1 %)	0,961
Ritmo sinusal	620 (68,4 %)	567 (69,4 %)	53 (58,9 %)	0,109
FA / flutter	232 (25,6 %)	201 (24,6 %)	31 (34,4 %)	
Electroestimulado	55 (6,1 %)	49 (6 %)	6 (6,7 %)	

Continúa en la siguiente página

Variable	Total(N=907)	Vivos(N=817)	Muertos(N=90)	P
Score Agaston	2726[1811-3994]	2697[1806-3993]	2912[1830-4131]	0,448
Hemoglobina (mg/dl)	12,3 \pm 1,9	12,3 \pm 1,9	11,8 \pm 1,8	0,004
Leucocitos	6970 [5680-8585]	6960 [5670-8580]	7000 [5708-8683]	0,524
Plaquetas x1000	181 [147-231]	181 [147-226]	186 [154-254]	0,111
Creatinina (mg/dl)	1 [0,82-1,3]	1 [0,82-1,29]	1,1 [0,86-1,41]	0,084
Aclaramiento de creatinina Cockcroft	50,6 [40,2-64,8]	51,1 [40,5-64,7]	47,6 [38,5-64,9]	0,21
Grado1 aclaramiento \geq 90	57 (6,3 %)	51 (6,2 %)	6 (6,7 %)	0,003
Grado2 aclaramiento[60-90)	239 (26,4 %)	218 (26,7 %)	21 (23,3 %)	
Grado3a aclaramiento[45-60)	274 (30,2 %)	252 (30,8 %)	22 (24,4 %)	
Grado3b aclaramiento[30-45)	249 (26,4 %)	222 (27,2 %)	27 (30 %)	
Grado4 aclaramiento[15-30)	69 (7,6 %)	63 (7,8 %)	6 (6,7 %)	
Grado5 aclaramiento< 15	17 (1,8 %)	9 (1,1 %)	8 (8,9 %)	
Filtración glomerular	60,8 [45,7-76,6]	61,2 [46,1-76,7]	57,6 [40,7-70,8]	0,094
NT-proBNP	2355 [988-5417]	2303 [955-5312]	2638 [1372-7661]	0,126

Apéndice B

Herramientas de R

En este se describirán brevemente los paquetes utilizados en R y se explicará el enfoque técnico empleado, sin incluir ni detallar todo el código. Esto permitirá mejorar la claridad del texto y evitar la inclusión de extensos fragmentos de código poco relevantes.

B.1. Gráficos

Para la representación gráfica del análisis descriptivo se utilizaron los paquetes `ggplot2`, `survminer` y `muhaz`. En concreto, `ggplot2` se empleó para las visualizaciones generales, las curvas de supervivencia se generaron mediante la función `ggsurvplot` del paquete `survminer`, y la curva de riesgo suavizada se obtuvo con la función `muhaz` de la librería homónima.

En cuanto a los resultados del modelo, las curvas de Kaplan-Meier también se han representado con la función `ggsurvplot`, y el análisis de los residuos se ha realizado mediante la función `ggcoxdiagnostics`, pertenecientes al mismo paquete `survminer`.

B.2. Modelo de Cox

Explicaremos cómo se hizo la creación y selección de variables del modelo mediante R.

B.2.1. Test para la selección de variables

Para llevar a cabo los tests descritos en la metodología, en los cuales los individuos se dividen en dos grupos según si fallecen o son censurados, se emplean las siguientes funciones:

- Para las variables categóricas, se utilizaron las pruebas correspondientes entre: `fisher.test()`, `chisq.test(correct=FALSE)` y `FisherFreemanHalton_asymptotic_test_rxc()` del paquete `contingencytables`.
- Para las numéricas, se evaluó la normalidad mediante el test de Kolmogórov-Smirnov (`ks.test()`), en caso de duda se complementa con un gráfico Q-Q (`qqnorm()`) para estudiar la normalidad. Si se rechaza la normalidad se aplica el test de Wilcoxon (`wilcox.test()`), si no se rechaza, se utilizó el test t de Student (`t.test()`), siempre que tampoco se rechazase la igualdad de varianzas entre grupos, evaluada mediante el test de F (`var.test()`). En nuestro caso, cuando no se rechazó la normalidad, tampoco se rechazó la igualdad de varianzas.

Para aplicar el test de log-rank, primero se construyó un objeto de supervivencia mediante la función `Surv(time = tiempo_supervivencia, event = suceso)`. A continuación, se utilizó la función `survdifff()` sobre este objeto y se obtuvo el p-valor correspondiente. Perteneciendo ambas funciones al paquete `survival`

B.2.2. Selección de variables automática y creación del modelo

Se utilizó la función `stepwise(type="cox", strategy = "bidirection", metric = "SL", sle = 0.1, sls = 0.1)` del paquete `StepReg`, aplicándola sobre una fórmula que incluye el objeto `Surv()` y las variables que superaron la fase inicial de selección. A partir de las variables seleccionadas por el procedimiento `stepwise`, se construyó el modelo de regresión de Cox utilizando la función `coxph()` del paquete `survival`.

B.3. App shiny

Esta sección complementa lo expuesto anteriormente en el capítulo dedicado a la aplicación web, centrándose en aspectos técnicos específicos.

Se utilizó CSS para asegurar que las dimensiones de los tres elementos principales (inputs, gráfico y tabla) sean adecuadas y que la aplicación se adapte correctamente según el tamaño de la pantalla. Esto incluye ampliar la anchura del gráfico y disponer los switches en varias filas cuando la pantalla es demasiado pequeña. Además, se empleó JavaScript para cambiar el color de los botones seleccionados, transponer la tabla según el tamaño de la pantalla y ajustar las dimensiones al recargar la página si varían. Para lograr una interfaz responsiva, se utilizaron los componentes `FluidPage` y `FluidRow` de Shiny.

En cuanto a la interactividad, los elementos de la interfaz se actualizan dinámicamente en función de la manipulación de los switches. Esto se implementó mediante un elemento reactivo cuyos valores se modifican al cambiar el estado de los switches, lo que a su vez actualiza los componentes de la UI en tiempo real.

Finalmente, se utilizó el paquete `shinyWidgets`, específicamente el componente `materialSwitch` para los switches y se integraron gráficos interactivos mediante la combinación de `plotly` y `ggplot2`.