

Resultados Fore-Ictus: Incidencia del tabaquismo en el ictus

Juan Alejandro Farauste y José Ramón Torres Martín

Resumen

A lo largo del proceso de análisis de los datos de ingreso de pacientes registrados en el hospital Virgen del Rocío, los resultados más interesantes se han obtenido entorno al factor de riesgo del tabaquismo. A continuación, nos dispondremos a exponer concisamente en qué condiciones se ha tenido en cuenta como factor de riesgo y cómo afecta a través de su media de repetición en el tiempo con la mediana.

1. Introducción.

Partimos de una base de datos que hemos limpiado y simplificado a través de técnicas de manipulación de datos quedándonos con un total de 14808 pacientes distintos que han sufrido al menos un ictus.

En este estudio el nuevo criterio general a seguir ha sido filtrar solamente aquellos registros que tengan alguno de estos códigos ictus como *primer, segundo o tercer diagnóstico*. No obstante, nuestra intención es partir de una situación inicial en la que sólo se consideren los ictus de primer diagnóstico y a continuación ir introduciendo registros con códigos ictus en diagnóstico segundo en los que no quepa ninguna duda de que se han codificado de esa manera debido a que el diagnóstico inicial esté intrínsecamente relacionado con el ictus. Tras esto el estudio se ha centrado en las siguientes codificaciones de ictus:

Códigos Ictus (CIE-9)	Patología
431 y 432	Ictus hemorrágico
433	Estenosis de arteria precerebral
434	Ictus isquémico
435	Isquemia cerebral transitoria
436	Código ictus en desuso (ambiguo)

Tabla 1. Códigos ictus tomados en el criterio de elección

Por lo general el estudio se centra en el fenómeno de recurrencia o repetición de ictus que consideramos entre la fecha de ingreso hospitalario registrada de dos ictus consecutivos. No obstante, también hemos considerado la posibilidad de obtener información del propio fenómeno de la aparición del ictus (y en concreto su primera aparición) en un paciente, lo cual nos ha permitido reafirmar nuestras hipótesis de los resultados.

2. Resumen de la metodología aplicada.

Inicialmente recibimos los datos clínicos

(previamente encriptados) a través de una columna que denominaremos *Diagnóstico*, donde quedan recogidos todos los códigos patológicos por orden de prioridad que fueron descritos en el diagnóstico registrado. A través diferentes paquetes de **R** hemos retirado los signos sobrantes y hemos ordenado la información; dejando por un lado el código **NUHS** y por otros sus distintos diagnósticos y las fechas de ingreso y alta.

Tras realizar una apertura a tabla ancha para trabajar ahora con el orden de prioridad del código en un diagnóstico, como una variable más flexible, hemos determinado los factores de riesgo más frecuentes (superiores a un 10% de frecuencia) respecto al total de los registros. Posteriormente se han fusionado las dos bases de datos que nos fueron cedidas (una desde el año 1994 en codificación **CIE-9** y otra más posterior desde el 2018 en **CIE-10**) con otra que nos ha permitido incorporar así datos referentes a la edad y el sexo de los pacientes.

Una vez conseguido esto se han filtrado los códigos ictus y los correspondientes a factores de riesgo con base en los criterios previamente establecidos. Esto nos ha permitido obtener las funciones de distribución empírica además de las densidades de probabilidad (haciendo uso de las medianas) y finalmente sus representaciones gráficas para distintas casuísticas de repetición o aparición de ictus (por factores de riesgo, sexo y edad). En esta sección hemos podido plantear ciertas hipótesis basándonos en la tendencia de un factor de riesgo (el tabaco) a repetir y a aparecer más precozmente y su relación con la mujer y su edad.

Por último, se ha llevado a cabo un ajuste no lineal numérico sobre la tendencia que siguen las recurrencias del ictus en la función de distribución acumulada con cierta precisión, lo cual permite comenzar a ahondar en la posibilidad de tratar de modelar el fenómeno más analíticamente.

3. Análisis de los datos

Vamos a discutir ahora las técnicas utilizadas en el estudio. Nuestro planteamiento ha consistido en determinar el comportamiento estadístico de los datos usando la potencia de computación de R mediante algoritmos centrados en los propios datos de manera empírica.

Para ello hemos hecho uso de la función `ecdf()`, que permite describir una función de distribución empírica a partir de una muestra. Esta expresión matemática funciona como estimación de la función de distribución teórica que nuestra muestra debería seguir. El resultado de esta función es un objeto que se comporta de forma similar a un modelo predictivo; recibe un vector de observaciones y devuelve su probabilidad acumulada.

Por otro lado hemos hecho uso del paquete `ggplot2` para tener una interfaz cómoda en la que observar la fluctuación de las curvas según los filtros tomados para cada una de ellas. Cuantos más datos se incluyen tras el filtrado más “suaves” son las curvas al hacer uso de un mayor número de datos.

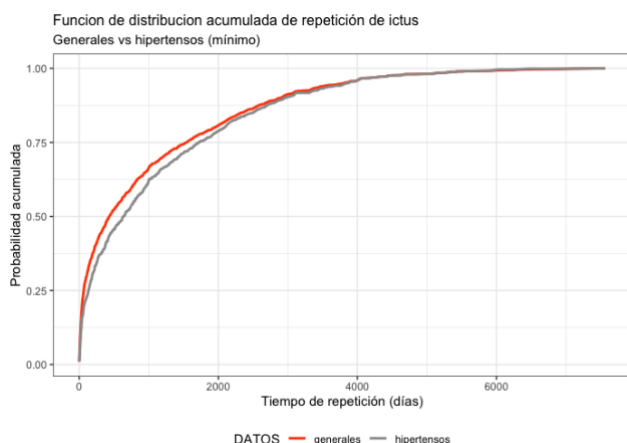


Figura 1. Ejemplo de representación de la función de probabilidad acumulada empírica. En este caso, pacientes de la muestra total frente a pacientes que, al menos, tienen hipertensión.

Cuando la curva llega al 0.5 de probabilidad acumulada, se alcanza el tiempo correspondiente a la mediana, lo que da información fiable e importante, ya que en ella la media se dispara hacia valores extremos. De esa manera la intención es obtener información interesante observando cómo se desplazan en el tiempo de repetición las curvas según los pacientes han padecido uno u otro factor de riesgo.

Se han seguido dos vías de análisis principales: por un lado, prestando atención a la repetición del ictus y su dispersión a lo largo del tiempo y por otro, el estudio de los primeros accidentes cerebrovasculares de cada paciente. Por lo general, los datos más esclarecedores han

aparecido cuando se escoge cada factor de riesgo de manera disjunta, es decir, se excluyen todos los restantes, frente a los pacientes sin factores de riesgo. No obstante, esto último es así en el caso de las repeticiones. Si sólo se observan los ictus en total, o los primeros ictus aparecen fuertes tendencias diferenciadas en los pacientes fumadores incluso si tienen otros factores de riesgo además del tabaco. Se entiende que tomando esta dirección de filtrado se retiran registros irrelevantes incidiendo así en la información importante subyacente que puede ser interpretable.

4. Resultados

Los factores de riesgo más frecuentes son los mostrados a continuación:

Patología	Frecuencia (%)
Hipertensión	54.61%
Diabetes	24.56%
Fibrilación auricular	18.51%
Hiperlipidemia	16.83%
Tabaquismo	19.20% y 12.81% (dos códigos)

Tabla 2. Frecuencia de aparición de los factores de riesgo en la muestra total de los pacientes.

Nº de pacientes según FR (excluyendo)	N
Sin factores de riesgo	3549
Hipertensión	2965
Diabetes	555
Fibrilación auricular	688
Tabaquismo	1075
Hiperlipidemia	346
Total	14808

Tabla 3. Número de pacientes con repeticiones según factores de riesgo excluyendo los demás.

Nº de pacientes con repetición según FR (excluyendo)	N
Sin factores de riesgo	439
Hipertensión	366
Diabetes	90
Fibrilación auricular	79
Tabaquismo	119
Hiperlipidemia	69
Total	2014

Tabla 4. Número de pacientes con repeticiones según factores de riesgo excluyendo los demás.

Nº de repeticiones según FR (excluyendo)	N
Sin factores de riesgo	546
Hipertensión	424
Diabetes	106
Fibrilación auricular	85
Tabaquismo	155
Hiperlipidemia	77
Total	2628

Tabla 5. Número de repeticiones según factores de riesgo excluyendo todos los demás.

Lo que parece observarse es que cuando se excluyen los factores de riesgo en las distintas representaciones, en cada uno de ellos se tiende a atrasar la mediana de repetición del ictus respecto a la muestra total y a la muestra de pacientes sin factores de riesgo, algo en principio contra intuitivo. Esto puede observarse en la *Figura 2* que hemos escogido como ejemplo ilustrativo del resultado:

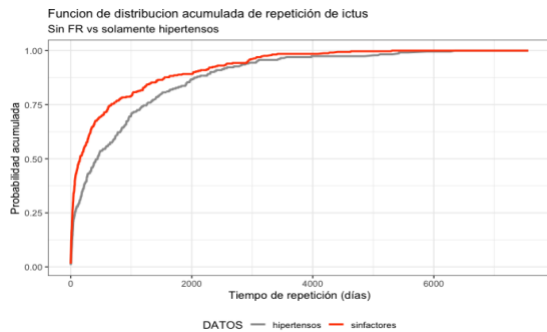


Figura 2. Representación de la función de probabilidad acumulada empírica. Se muestran ahora pacientes sin factores de riesgo frente a pacientes que solamente tienen hipertensión.

No obstante, el comportamiento de la curva se modifica drásticamente cuando el factor de riesgo que se toma es el **tabaquismo**. En esas circunstancias (*Figura 3* y *Figura 4*) observamos que no sólo no atrasa la mediana de repetición, si no que la adelanta incluso respecto a los pacientes sin factores de riesgo. Ante este dato, nuestra interpretación inicial ha sido suponer que a la detección de cualquiera de los factores de riesgo estudiado, le sigue la confección de un tratamiento adecuado que permita regular dicha patología y por ende debe de tender a impedir que se de una repetición de ictus. Sin embargo, en el caso del tabaquismo, esto no es así por motivos evidentes. Por ello sospechamos que refleja una incidencia crucial en la repetición precoz de ictus, o al menos algo a tener en cuenta. Para contrastar esto hemos seguido otra línea de análisis centrándonos en los casos en los que se da el accidente cerebrovascular por primera vez.

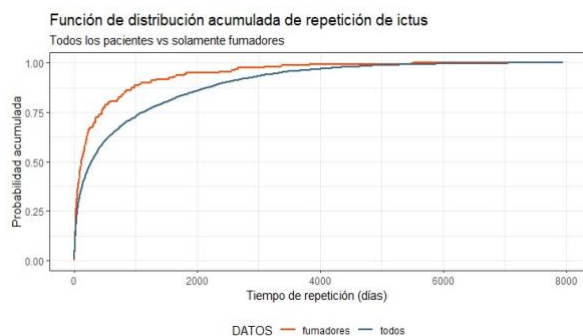


Figura 3. Representación de la función de probabilidad acumulada empírica; pacientes sin factores de riesgo frente a pacientes que solamente son fumadores..

Usando como descripción la densidad de probabilidad empírica en función de la edad del paciente (*Figura 5*) se vuelve a confirmar esta sospecha. En este caso, el valor medio de la función de densidad se sitúa entorno al intervalo de tiempo

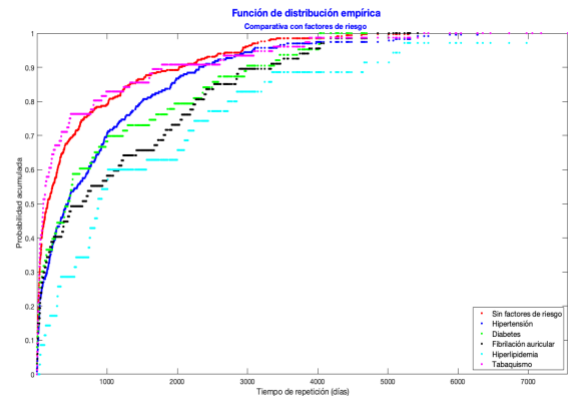


Figura 4. Representación de la función de probabilidad acumulada empírica para todos los casos excluidos y su desplazamiento.

correspondiente al pico de cada curva. Tanto en el caso de exclusión de otros FR (**Tabaquismo Exc**) como en el no se descarta la existencia de otras patologías (**Tabaquismo**), se observa que la media de repetición se adelanta respecto a los demás casos y con una clara tendencia.

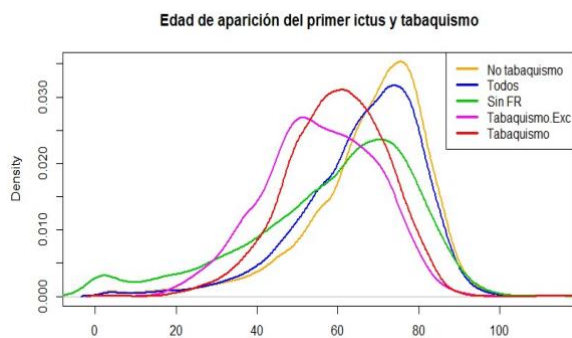


Figura 5. Función de densidad de probabilidad para los distintos filtros frente a la edad (en años).

Esta nueva perspectiva vuelve a confirmar de nuevo nuestra hipótesis sobre el efecto más que importante del tabaco en la aparición de un ictus o su repetición.

Por otro lado, también hemos observado que el ictus afecta de manera distinta a mujeres y a hombres. Cuando se ha visualizado su función de distribución acumulada a lo largo del tiempo desde su ingreso, no hemos observado ningún rasgo diferenciador que mostrara un comportamiento especial. No obstante, cambian los resultados al modificar la variable sobre la que representábamos la probabilidad de que se de un ictus. Al usar ahora los datos de la edad de los

pacientes, se observa claramente a través de la densidad de probabilidad (Figura 6), que el ictus se da a una edad más temprana (aproximadamente 10 años) en hombres que en mujeres.

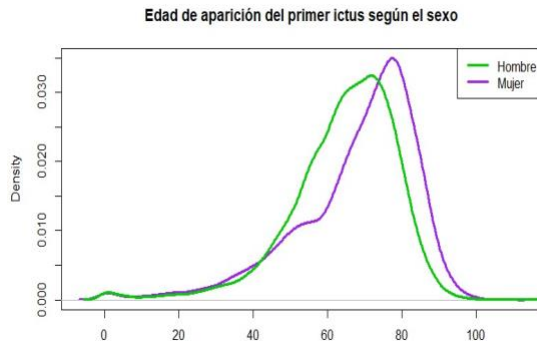


Figura 6. Función de densidad de probabilidad en el primer ictus para hombres y mujeres frente a la edad (en años).

Esto al fin y al cabo es algo que con base en la experiencia médica se sabía o al menos intuía. Sin embargo, al introducir el factor de riesgo del tabaco en la combinación, los resultados son muy distintos (Figura 7). El accidente cerebrovascular tiende a adelantarse en las mujeres cuando son fumadoras. Esta observación se ha advertido en los 3 tipos de muestras (aparición, primera aparición y repetición) lo cual aporta solidez a la hipótesis.

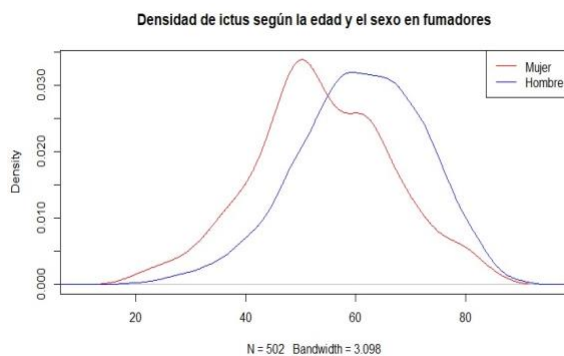


Figura 7. Función de densidad de probabilidad de aparición de ictus para personas fumadoras frente a la edad (en años).

Con la idea de dilucidar si esta tendencia se da con otros factores de riesgo, hemos representado también las distintas curvas para todos ellos (hiperlipidemia, fibrilación auricular e hipertensión entre otros). Lo que se observa de nuevo y de una manera más acentuada es que el tabaco vuelve a aparecer como un catalizador en la aparición precoz del ictus en las mujeres diferenciado claramente respecto de los demás factores de riesgo (Figura 8). Cabe destacar que en esta figura los FR han sido escogidos de manera no excluyente y no hemos considerado

necesario usar un color distinto en cada FR distinto del tabaco porque todos tienen un comportamiento diferenciado (al tabaco) y semejante entre sí.

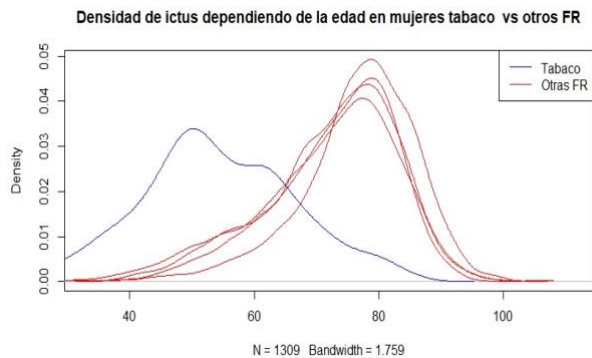


Figura 8. Densidad de probabilidad de aparición de ictus para fumadoras y mujeres con otros FR frente a la edad (en años).

Sorprendentemente los datos indican que el ictus en las pacientes que fuman tiende a aparecer varias décadas antes que si al menos sufren alguno de los otros FR. Quedamos satisfechos al poder haber contrastado nuestra hipótesis a través de distintas vías de análisis y observando una incidencia concreta en las mujeres.

Por último y como comentamos en el resumen, también hemos obtenido con Matlab a través de la herramienta de análisis no lineal *CFTOOL*, un ajuste numérico con un coeficiente de determinación (r) relativamente alto de la función de distribución acumulada empírica para el caso de la muestra total de datos de repetición de ictus. Su tendencia es la de una distribución exponencial

$$f(x) = 1 - e^{-bx^c}$$

Cuyos parámetros de ajuste han sido estos:

Parámetro	Valor
b	0.02206
c	0.57
r^2	0.9919
SSE	0.4219
RMSE	0.02136

Tabla 6. Parámetros de ajuste numérico

A continuación se puede observar una comparativa entre la curva empírica y la ajustada para apreciar la desviación relativa de ambas:

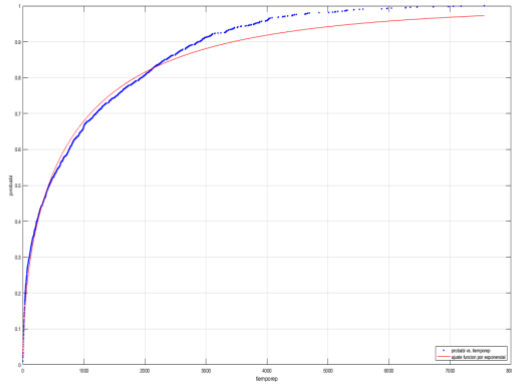


Figura 9. Comparativa entre función de distribución empírica y su ajuste numérico

A pesar de que la curva se desvía respecto a los datos acercándose al final, se confirma que es un buen ajuste ya que cerca del origen es tremendamente preciso, lo cual es importante ya que a medida que pasa el tiempo el fenómeno de la repetición del ictus se hace más incierto (sobre todo a partir de los 5 años). Este tipo de procedimientos abren una nueva vía a la posibilidad de tabular estas distribuciones, de tal manera que sea posible desarrollar un método con el que no predecir, pero sí en cierta medida, se pueda estimar la posibilidad de que un paciente experimente una recurrencia de ictus en función a sus factores de riesgo y de la extensión de su tratamiento o proceso de recuperación.