



Gabriela Boggio
Guillermina Harvey
Nora Arnesi
Virginia Laura Borra

Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística

ESTUDIO DE LA CAPACIDAD PREDICTIVA DE UN MODELO DE SUPERVIVENCIA EN PACIENTES SOMETIDOS A ANGIOPLASTIA CAROTIDEA¹

Resumen:

Frecuentemente la búsqueda de un modelo estadístico está orientada a la necesidad de predecir el fenómeno bajo estudio. Este es el caso del estudio de la supervivencia de pacientes sometidos a angioplastia carotídea donde interesa poder predecir la probabilidad de supervivencia a lo largo del tiempo en futuros pacientes según sus características personales y antecedentes médicos. Debido a ello, en este trabajo se ponen a prueba diferentes medidas de capacidad predictiva desarrolladas para modelos de hazards proporcionales. Entre ellas pueden mencionarse medidas de tipo R^2 y las relacionadas con los conceptos de sensibilidad y especificidad dando lugar a las denominadas curvas ROC dependientes del tiempo. Los resultados obtenidos muestran que la edad del paciente y el denominado score SYNTAX residual predicen adecuadamente la supervivencia con valores del área bajo la curva ROC cercanas a 0.80 para diferentes períodos de tiempo pos intervención.

Palabras claves: modelo de hazards proporcionales; capacidad predictiva; curvas ROC;

Abstract:

In many occasions, the search for a statistical model is directed to the need to predict the occurrence of the phenomenon under study. In this application, we are interested in predicting the probability of survival of future patients undergoing carotid angioplasty over time according to their personal characteristics and medical history. For this purpose, different predictive ability measures developed for proportional hazards models are assessed. These include R^2 -type measures and those related to the concept of sensitivity and specificity that give rise to the so-called time-dependent ROC curves. The results obtained show that the age of the patient and the so-called residual STNTAX score successfully predict survival with values of the area under the ROC curve close to 0.80 for different periods of time after the procedure.

Keywords: proportional hazard model; predictive ability; ROC curves

¹Este trabajo se elaboró en el marco del Proyecto ECO 181 Titulado "Tópicos especiales en modelización estadística", dirigido por Gabriela Boggio.



Introducción

En muchos casos, en el campo de la investigación médica interesa no sólo indagar acerca de las características de los pacientes que propician una supervivencia mayor, sino que es de importancia detectar aquéllos factores que permiten predecir mejor la probabilidad de supervivencia de futuros pacientes. En particular, el Centro de Tratamientos Endovasculares del Sanatorio Británico y el servicio de Diagnóstico y Tratamiento Endovascular del Sanatorio Los Arroyos, ambos de la ciudad de Rosario (Argentina), realizan el seguimiento de los pacientes sometidos a angioplastia carotídea y una de las preocupaciones principales es poder predecir cuál es la probabilidad de supervivencia de futuros pacientes en base a sus características personales y antecedentes médicos.

Los modelos de regresión para datos de supervivencia y, en particular, los modelos de hazards proporcionales de Cox permiten determinar qué factores y en qué medida influyen sobre el riesgo o hazard de muerte. A la hora de elegir el "mejor" modelo es importante tener en cuenta justamente qué es lo que se pretende del mismo: explicar o predecir el fenómeno bajo estudio, ya que hay dos aspectos a tener en cuenta al evaluar un modelo. Uno de ellos es la *calibración*, es decir la habilidad del mismo para reflejar los valores observados de la respuesta a partir de las probabilidades estimadas, y la *discriminación*, que se refiere a la habilidad del modelo para separar los individuos según presenten o no el evento de interés. Cuando el modelo se construye con fines predictivos interesa fundamentalmente la capacidad de discriminación del mismo. De ahí que en este trabajo se elige un conjunto de medidas de evaluación de la capacidad predictiva de modelos de hazards proporcionales para el tiempo hasta la muerte en pacientes cardíacos sometidos a una angioplastia carotídea.

Entre estas medidas, se pueden considerar las extensiones de coeficientes tipo R^2 y las relacionadas con la especificidad y sensibilidad. En este sentido, las curvas ROC constituyen una herramienta ampliamente utilizada como medida de capacidad predictiva de un modelo de regresión logística, es decir de un modelo para la probabilidad de ocurrencia de un evento bajo estudio, que en este caso estaría dado por el evento muerte. Heagerty y Zheng (2005) proponen medidas de sensibilidad y especificidad dependientes del tiempo y las curvas ROC respectivas cuando la variable respuesta es un tiempo de supervivencia con censura utilizando conceptos establecidos para el caso de respuestas binarias. Su programación en el paquete R debida a Heagerty (2015) ha facilitado la implementación de ellas.

Los datos

El Centro de Tratamientos Endovasculares del Sanatorio Británico y el Servicio de Diagnóstico y Tratamiento Endovascular del Sanatorio Los Arroyos de la ciudad de Rosario cuentan con el seguimiento de 408 pacientes sometidos a angioplastia carotídea, intervenidos entre enero de 2010 y junio de 2016.

Para cada uno de los pacientes se cuenta con información sobre variables tanto basales como propias del procedimiento quirúrgico realizado así como también otras relacionadas con el seguimiento. En esta oportunidad se tienen en cuenta las siguientes variables: edad al momento de la angioplastia (en años); sexo (femenino, masculino); presencia de obesidad abdominal (sí, no); condición de fumador (sí, no); presencia de dislipidemia (sí, no) y presencia de diabetes (sí, no). Se considera, además, el valor del score SYNTAX residual. Este score es una medida angiográfica que trata de cuantificar el nivel alcanzado de la revascularización luego de realizada la intervención; puntuaciones más altas indican que la revascularización fue incompleta. Si



bien es de naturaleza continua, se lo utiliza categorizado en dos niveles (alto, bajo) tomando como punto de corte relevante desde el punto de vista clínico el score igual a 8.

La supervivencia de los pacientes se mide a través del tiempo (en días) desde que se realiza la angioplastia hasta la muerte o hasta la fecha en la que se contactó al paciente por última vez.

Metodología

Además de evaluar si un modelo de supervivencia ajustado está bien calibrado, es decir cuán bien las probabilidades de supervivencia ajustadas concuerdan con las correspondientes proporciones observadas, es importante medir la performance del mismo con respecto a la predicción de probabilidades de supervivencia de futuras observaciones.

Un enfoque diseñado para medir la capacidad predictiva de los modelos de supervivencia se basa en el cálculo de medidas de variación en los datos que es explicada por el modelo propuesto. Han sido propuestas un gran número de medidas diferentes de la proporción de la variación en los datos que es explicada por el modelo de regresión de Cox y desafortunadamente no existe consenso acerca de cuál es la mejor de ellas. Las que se presentan a continuación tienen propiedades deseadas: toman valores entre 0 y 1, son independientes del grado de censura, no se encuentran afectadas por la escala de los datos de supervivencia y aumentan en su valor a medida que se agregan variables explicativas al modelo.

Medidas de variación explicada

Considérese un modelo de regresión de Cox, $h_i(t) = \exp(\boldsymbol{\beta}' \mathbf{x}_i) h_0(t)$, donde \mathbf{x}_i es el vector de valores de las p variables explicativas para el i -ésimo individuo, $\boldsymbol{\beta}$ es el vector de parámetros desconocidos y $h_0(t)$ es la función hazard basal. Una de las primeras sugerencias para una estadística tipo R^2 , debida a Kent y O'Quigley (1988) es similar en su forma a la estadística R^2 para el análisis de regresión lineal. Esta Estadística se define como:

$$R_p^2 = \frac{\hat{V}_p}{\hat{V}_p + \pi^2/6}$$

donde $\hat{V}_p = \hat{\boldsymbol{\beta}}' \mathbf{S} \hat{\boldsymbol{\beta}}$ es una estimación de la variación en el score de riesgo $\eta_i = \boldsymbol{\beta}' \mathbf{x}_i$, entre los n individuos, $\hat{\boldsymbol{\beta}}$ es el vector de estimaciones de los parámetros del modelo de Cox ajustado y \mathbf{S} es la matriz de variancias y covariancias de las variables explicativas.

Otra estadística, debida también a Kent y O'Quigley (1988) se basa en una medida de la "distancia" entre el modelo de interés y el modelo nulo, es decir, el que no tiene variables explicativas. Esta estadística es

$$R_w^2 = 1 - \exp(-\tilde{W})$$

donde \tilde{W} es una estimación del valor esperado de la estadística de razón de verosimilitud, definida como $W = -2E\{\log L(0) - \log L(\hat{\boldsymbol{\beta}})\}$. Ella está dada por:



$$\tilde{W} = 2 \left[(1 - \tilde{\omega}) \Psi(1) + \log \Gamma(\tilde{\omega}) + \log \left\{ n^{-1} \sum_{i=1}^n \exp(-\tilde{\omega} z_i) \right\} \right]$$

donde $z_i = \hat{\beta}'(\mathbf{x}_i - \bar{\mathbf{x}})$, $\bar{\mathbf{x}}$ es el vector de valores medios de las p variables explicativas y $\Gamma(\cdot)$ es una función gamma. Además, $\tilde{\omega}$ es el valor de ω que satisface la ecuación no lineal

$$\Psi(1) - \Psi(\omega) + \sum_{i=1}^n \frac{\exp(-\tilde{\omega} z_i)}{\sum_{i=1}^n \exp(-\tilde{\omega} z_i)} z_i = 0$$

donde $\Psi(\omega)$ es la función digamma definida por $\Psi(\omega) = -\lambda + \sum_{j=0}^{\infty} \frac{\omega - 1}{(1 + j)(\omega + j)}$,

de manera que $\Psi(1) = -\lambda$ y λ es la constante de Euler. Esta es una estadística complicada de calcular pero ha sido incluida dado que presenta buen comportamiento (Collett, 2015) y puede obtenerse de varios paquetes computacionales de análisis de supervivencia.

Por último, Allison (2010) recomienda el uso de una simplificación de la medida recién presentada, la cual fue propuesta por Cox y Snell (1989). Esta estadística toma la forma:

$$R_L^2 = 1 - \exp(-n^{-1} \text{LRT})$$

donde

$$\text{LRT} = -2 \left\{ \log L(0) - \log L(\hat{\beta}) \right\}.$$

Tiene la ventaja de poder calcularse muy fácilmente a partir del valor de la estadística del test de razón de verosimilitud proporcionado por cualquier software.

A diferencia del caso del coeficiente R^2 del modelo lineal, no hay un consenso generalizado acerca de su interpretación como una proporción de la variación de la variable respuesta explicada por las covariables. De cualquier forma constituyen una medida de cuán bien se puede predecir la variable respuesta con el conjunto de covariables, están acotadas entre 0 y 1 y toman valores más grandes cuando las covariables están más fuertemente asociadas con la respuesta.

Un enfoque alternativo para evaluar la capacidad predictiva es calcular medidas que resuman el acuerdo o concordancia entre los rangos de los tiempos de supervivencia observados y los predichos por el modelo. Estas estadísticas resumen el potencial de un modelo ajustado para discriminar entre individuos separando aquéllos con tiempos de supervivencia mayor de aquéllos con tiempos más cortos. Al igual que las medidas de variación explicada, estas medidas toman valores entre 0 y 1, correspondiendo a la perfecta discordancia y a la perfecta concordancia respectivamente. Ellas se describen a continuación.

Medidas de capacidad discriminativa

Una primera medida de la capacidad discriminativa de un modelo descripta por Harrell, Lee y Mark (1996) es la denominada estadística C_H . Esta estadística es una estimación de la proba-



bilidad de que, para cualquier par de individuos, quien tiene el tiempo de supervivencia más corto sea el que tiene el mayor hazard de muerte. Para calcular esta estadística, se consideran todos los pares de tiempos de supervivencia, donde ambos murieron o donde un individuo del par murió antes del tiempo de supervivencia censurado del otro individuo del par. Los pares en que ambos individuos tienen tiempos de supervivencia censurados, o que el tiempo de supervivencia de uno excede el tiempo de supervivencia censurado del otro, no se incluyen. Si en un par donde ambos individuos han muerto, el tiempo de supervivencia predicho por el modelo es mayor para el individuo que vivió más tiempo, se dice que los individuos son *concordantes*. Para pares donde un individuo muere y el otro individuo tiene un tiempo de supervivencia que es censurado después del tiempo de supervivencia del primero del par, el individuo con el tiempo censurado tiene supervivencia mayor que el otro y entonces puede determinarse si los dos miembros de tal par son concordantes. La estadística C_H se obtiene dividiendo el número de pares concordantes por el total de pares considerados.

El inconveniente de no poder considerar los pares de individuos donde ambos están censurados o donde el tiempo de censura de uno es menor al tiempo en que ocurre la muerte del otro, se supera con la estadística propuesta por Gönen y Heller (2005). La medida de concordancia que estos autores desarrollaron es una estimación de la probabilidad para cualquier par de individuos, de que el tiempo de supervivencia de uno exceda el del otro, condicional a que el individuo con mayor tiempo de supervivencia tenga menor score de riesgo.

Para obtener esta estadística se considera que un par de tiempos de supervivencia (T_i, T_j) es concordante si el individuo con el menor score de riesgo tiene mayor tiempo de supervivencia. La probabilidad de concordancia es entonces:

$$K = P(T_i > T_j / \eta_i \leq \eta_j)$$

y una estimación de esta probabilidad de concordancia está dada por:

$$\hat{K} = \frac{2}{n(n-1)} \sum_{i < j} \sum \left\{ \frac{I\{(\hat{\eta}_j - \hat{\eta}_i) < 0\}}{1 + \exp(\hat{\eta}_j - \hat{\eta}_i)} + \frac{I\{(\hat{\eta}_i - \hat{\eta}_j) < 0\}}{1 + \exp(\hat{\eta}_i - \hat{\eta}_j)} \right\}$$

donde la función indicadora $I(\cdot)$ es tal que $I\{X < 0\} = 1$ si $X < 0$ y cero en otro caso.

Por último, se parte del cálculo de medidas de discriminación para el predictor $\eta_i = \beta' x_i$ con la idea de considerar los datos sobre el tiempo de supervivencia como variables binarias dependientes del tiempo con niveles "evento en t" y "evento después de t". De esta forma, la capacidad predictiva puede medirse usando conceptos establecidos para la evaluación de reglas de clasificación binarias. Es decir, para cada punto del tiempo fijo t se considera a las observaciones que presentan el evento en t como "casos" y a las observaciones que presentan el evento después de t, como "controles" y se definen la *sensibilidad* y *especificidad dependiente del tiempo* como sigue:

$$\text{sens}(c, t) = P(\eta > c / T = t) \text{ y}$$

$$\text{espec}(c, t) = P(\eta \leq c / T > t),$$

respectivamente (Heagerty y Zheng, 2005; Heagerty et al., 2000). Aquí c es un umbral de η .



Una gran sensibilidad significa que η y entonces el hazard tiende a ser grande y puede verse como una medida de acierto. En contraste, una alta especificidad significa que los hazards para $T > t$ son pequeños, indicando que los "controles" están correctamente identificados. Por definición, la sensibilidad y especificidad varían con los umbrales. En el caso extremo en el que $c \rightarrow -\infty$, la sensibilidad tiende a uno y la especificidad a cero, mientras que para $c \rightarrow \infty$ la sensibilidad se convierte en cero y la especificidad se convierte en uno.

Resumiendo, $\text{sens}(c,t)$ y $\text{espec}(c,t)$ para t fijo produce la curva ROC *dependiente del tiempo*, la cual se define como:

$$\text{ROC}(c,t) = \{1 - \text{espec}(c,t), \text{sens}(c,t)\}_{c \in \mathbb{R}}$$

La curva ROC grafica la sensibilidad versus (1- especificidad) para umbrales diferentes. En general, tiene forma cóncava conectando los puntos (0,0) y (1,1). Si η tiene un alto poder discriminativo la curva es fuertemente cóncava y tiene una gran área bajo la curva.

Dado que en muchas aplicaciones no es de interés analizar un tiempo t en particular, es útil obtener una medida global de la capacidad predictiva a través de los diferentes valores de t . Heagerty y Zheng (2005) proponen como tal medida resumen de "concordancia", C_P , un promedio ponderado del área bajo las curvas ROC - AUC(t) - a través de los diferentes tiempos específicos. En la práctica se restringe esta medida al período de seguimiento finito de manera que se interpreta como la probabilidad de que las predicciones para un par de individuos sean concordantes con sus respuestas, dado que el menor tiempo hasta la muerte ocurre durante el seguimiento.

La definición de sensibilidad y especificidad elegida responde a uno de los enfoques presentados por Heagerty y Zheng (2005) en el que la sensibilidad se relaciona con la identificación de casos *incidentes* donde $T=t$ y la especificidad con la consideración *dinámica* de los controles de manera que cada individuo juega el rol de un "control" para los tiempos anteriores a su tiempo de supervivencia y asume el rol de "caso" cuando presenta el evento.

Una de las razones de la preferencia de este enfoque es que las otras extensiones de los conceptos de sensibilidad y especificidad dependientes del tiempo, por ejemplo definiendo los "casos" acumulados donde $T \leq t$, no tienen interpretaciones probabilísticas sencillas como en el enfoque adoptado (Tutz y Schmid, 2016).

Los métodos descriptos para determinar la variación en un conjunto de datos de supervivencia que es explicada por el modelo de Cox ajustado y su capacidad discriminativa, pueden conducir a resultados sesgados si los datos han sido usados tanto para ajustar el modelo como para validarlo. Por esta razón, puede procederse a dividir los datos en una muestra de entrenamiento para el ajuste del modelo y una muestra de prueba para su validación, garantizando de este modo la validación interna del mismo. En este caso, se extrae una muestra al azar del 60% de los registros ($n=245$) para realizar el proceso de selección del modelo y se reserva el 40% restante ($n=163$) para el cálculo de las medidas de capacidad predictiva.

El recurso computacional elegido para el análisis de los datos es el entorno R (R Core Team, 2017).



Resultados

En este trabajo se analizan los datos correspondientes a 408 pacientes para los cuales se realizó un seguimiento máximo de 6 años. Sólo el 11.52% de los pacientes fallecieron (47) en el período considerado. Los pacientes analizados tienen entre 38 y 96 años, con una edad promedio de 65 años (desvío estándar: 10 años). El 81.62% son hombres, el 39.26% tiene obesidad abdominal, el 55.88% son fumadores, el 71.08% tiene dislipidemia y, por último, el 26.23% padece diabetes. El score SYNTAX residual resultó menor o igual a 8 en 355 pacientes (87%) y mayor a 8 en los 53 pacientes restantes (13%).

El proceso de selección del modelo de hazards proporcionales semiparamétrico se realiza en base al conjunto de 245 pacientes que conforman la muestra de entrenamiento. Se obtiene un modelo que contiene los efectos principales y la interacción entre las variables explicativas edad y el score SYNTAX residual dicotomizado.

Este modelo elegido ajustado al conjunto de 408 pacientes produce las estimaciones que se presentan en la Tabla 1. A partir de ellas se puede estimar la razón de hazards correspondiente al score SYNTAX residual para diferentes edades dada la significación del efecto de interacción. El riesgo de muerte es mayor para los pacientes que presentan el score SYNTAX residual alto variando la intensidad según la edad (el efecto del score alto es mayor en pacientes más jóvenes).

Tabla 1: Estimaciones de los parámetros del modelo elegido

	Coefficiente estimado	Error estándar	Probabilidad asociada
SYNTAX residual (mayor vs. menor)	7.77	2.28	0.0007
Edad	0.11	0.02	<.0001
Edad*SYNTAX residual	-0.10	0.03	0.0024

Con el objeto de evaluar la capacidad predictiva del modelo elegido, en una primera instancia se calculan en el conjunto de 168 pacientes reservados a tal fin las medidas tipo R^2 tanto para el modelo elegido como para el modelo completo incluyendo el conjunto total de variables explicativas disponible.

Las medidas de variación explicada difieren en una magnitud pequeña al comparar ambos modelos ajustados, lo cual convalida la elección del modelo reducido. Los valores de los coeficientes R_p^2 y R_w^2 tienen valores similares para ambos modelos y son sustancialmente mayores que los valores del coeficiente R_L^2 . Si bien la comparación de los valores de este último coeficiente conduce también a la elección del modelo reducido, su baja magnitud no lo hace atractivo en función de los valores usuales del R^2 en el análisis clásico de regresión lineal. (Tabla 2).



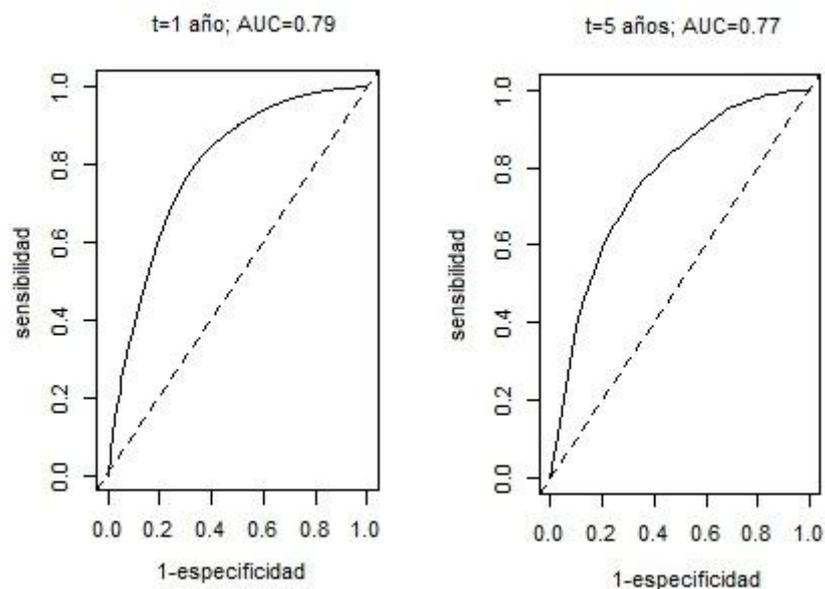
Tabla 2: Medidas de variación explicada

Coeficiente	Modelo elegido	Modelo completo
R_p^2	0.52	0.56
R_w^2	0.64	0.66
R_L^2	0.13	0.14

Con respecto a las medidas de capacidad discriminativa, el valor de la estadística de Harrell es 0.72 para el modelo seleccionado y 0.73 para el modelo completo. La medida de concordancia de Gönen y Heller resulta igual a 0.72 para el modelo seleccionado y 0.75 para el completo. En este caso las estadísticas de Harrell y de Gönen y Heller son muy similares a pesar de que se observan muchos datos censurados que no se tuvieron en cuenta en el cálculo de la estadística de Harrell. También hay poca diferencia en el valor de cada estadística para el modelo seleccionado y el completo, confirmando que la habilidad predictiva del modelo seleccionado es comparable con la del modelo completo.

La Figura 1 muestra las curvas ROC que fueron obtenidas a partir del conjunto de datos de prueba al año ($t=365$ días) y a los 5 años de seguimiento ($t=1826$ días) para el modelo elegido.

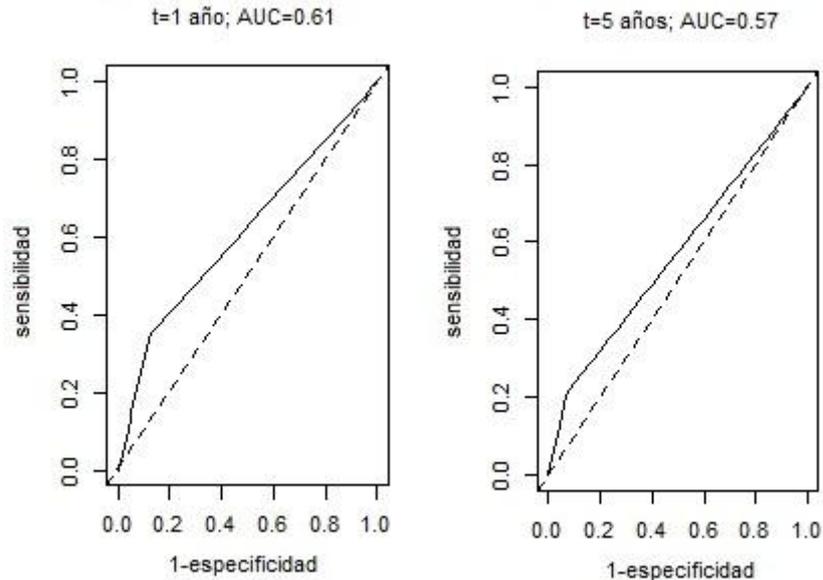
Figura 1: Curvas ROC para el modelo elegido



El valor del área bajo la curva ROC es satisfactorio tanto cuando se evalúan las predicciones al año de seguimiento ($AUC=0.79$) como también a los 5 años ($AUC=0.77$). La estimación de la medida resumen de concordancia a través de todo el período de seguimiento también presenta un valor razonable ($C_p=0.74$).



Figura 2: Curvas ROC para el modelo con sólo la covariable SYNTAX residual



Si bien es de conocimiento médico que el score SYNTAX residual influye en la supervivencia de los pacientes, un buen modelo predictor debe incluir también la edad del paciente. Se aprecia que agregar esta covariable al modelo conduce a un incremento importante en la capacidad predictiva del modelo para $t = 1$ año y $t = 5$ años ya que la curva ROC para el modelo con sólo SYNTAX residual está muy cerca de la línea diagonal en casi todos los umbrales. La línea diagonal corresponde a un modelo de supervivencia para un modelo sin covariables, de manera que cualquier modelo que prediga mejor que el azar debe resultar en una curva por encima de esta recta diagonal.

Discusión

Con el interés puesto en obtener adecuados pronósticos acerca de la supervivencia de pacientes sometidos a angioplastia carotidea se elige un modelo de hazards proporcionales de Cox y se evalúa su capacidad predictiva. En este trabajo se eligió trabajar de acuerdo al enfoque denominado sensibilidad incidente aunque pueden ser utilizadas sensibilidades acumuladas para construir las curvas.

Para evitar los sesgos que se producen si se utiliza el mismo conjunto de datos para ajustar el modelo y para medir su capacidad predictiva, se elige dividir la muestra en una muestra de entrenamiento y otra de prueba. Sin embargo no es la forma más recomendada si se cuenta con un número reducido de observaciones por lo que resultan más apropiadas las técnicas de validación cruzada, es decir, técnicas basadas en dividir múltiples veces el conjunto de datos en



subconjuntos de entrenamiento y de prueba y promediar las estimaciones de capacidad predictiva a través de ellas (Molinario et al., 2005; Kim, 2009). En este sentido, Harrell (2017) desarrolla en el entorno R un paquete para la estimación de las curvas ROC dependientes del tiempo basada en técnicas de validación cruzada y queda pendiente la puesta a punto de los resultados obtenidos a partir de él.

Sería útil, además, la realización del proceso de validación del modelo en base a un conjunto de datos independientes como podrían ser los de otra región o país de manera de garantizar la validación externa del modelo.

Por último resulta de interés distinguir el evento muerte según las causas que la producen, trabajar con modelos para riesgos competitivos y en ese caso sería interesante calcular las respectivas medidas de capacidad predictiva.

AGRADECIMIENTOS

Se agradece especialmente al Dr. Claudio Cigalini por facilitar la base de datos que permitió realizar la puesta a prueba de la metodología estudiada.

REFERENCIAS BIBLIOGRÁFICAS

- Allison, P. D. (2010). *Survival analysis using SAS. A practical guide*. 2ª edición. SAS Institute, Cary, USA.
- Collett, D. (2015). *Modelling survival data in medical research*. 3º edición. Chapman & Hall.
- Cox, D. R.; Snell, E. J. (1989). *The analysis of binary data*. 2º edición. London: Chapman & Hall.
- Gönen, M.; Heller G. (2005). Concordance probability and discriminatory power in proportional hazards regression. *Biometrics*, 92: 965-970.
- Heagerty, P. (2015). Package 'risksetROC'. Disponible en: <https://cran.r-project.org/web/packages/risksetROC/risksetROC.pdf>. Fecha de consulta: 11/2017.
- Heagerty, P. J.; Lumley, T.; Pepe, M. S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*, 56: 337-344.
- Heagerty, P. J.; Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics*, 61: 92-105.
- Harrell, F. E. (2017). Regression modeling strategies. Disponible en: <http://biostat.mc.vanderbilt.edu/wiki/Main/Rrms>. Fecha de consulta: 11/2017.
- Harrell, F. E.; Lee, K. L.; Mark, D. B. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, 15: 361-87.



- Kent, J. T.; O'Quigley, J. (1988). Measures of dependence for censored survival data. *Biometrika*, 75 (3): 525-534.
- Kim, J. (2009). Estimating classification error rate: repeated cross-validation, repeated holdout and bootstrap. *Computational Statistics and Data Analysis*, 53: 3735-3745.
- Molinaro, A. M.; Simon, R.; Pfeiffer, R. M. (2005). Prediction error estimation: a comparison of resampling methods. *Bioinformatics*, 21(15): 3301-3307.
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Disponible en: <https://www.R-project.org/>. Fecha de consulta: 11/2017.
- Tutz G., Schmid, M. (2016). Modeling *discrete time to event data*. Springer Series in Statistics, Switzerland.