

Supervivencia del injerto a largo plazo: métodos estadísticos de medida

Víctor Abraira

Unidad de Bioestadística Clínica, Hospital Universitario Ramón y Cajal, IRYCIS, Madrid
CIBER de Epidemiología y Salud Pública (CIBERESP)

Nefrología Sup Ext 2015;6(2):25-36

INTRODUCCIÓN

El diseño de los estudios para estudiar la supervivencia a largo plazo consiste en seguir en el tiempo a un grupo de pacientes, desde un momento inicial, tal como el diagnóstico, el inicio del tratamiento (el trasplante en este caso) o la aleatorización en un ensayo clínico, hasta que ocurre el evento de interés (muerte del paciente o pérdida del injerto) o hasta que acaba la recogida de datos.

El análisis de este tipo de variable, «tiempo hasta la ocurrencia de un evento», suele presentar 2 dificultades que impiden realizarlo con las técnicas estadísticas generales como la prueba t de Student, el análisis de la varianza o los modelos de regresión lineal. En primer lugar, en la mayor parte de los estudios, la variable tiempo no tiene la distribución normal que esas técnicas exigen, más bien suele tener una distribución asimétrica, en la que típicamente los eventos ocurren con mayor frecuencia al principio que al final del período de seguimiento y, aunque podrían intentarse transformaciones que la normalizaran, existe una segunda y más importante dificultad, y es que, en general, al final del período de seguimiento siempre hay sujetos en los que no ha ocurrido el evento y, por lo tanto, en los que el tiempo hasta su ocurrencia es desconocido, si bien se tiene una información parcial sobre él: es mayor o igual que el tiempo observado. A este fenómeno se le denomina censura y es el que exige un análisis específico para esta

variable, que permita aprovechar la información contenida en los tiempos censurados.

Existen 3 motivos por los que pueden aparecer las censuras. En primer lugar, por fin del estudio. Para estudiar la supervivencia del injerto, por ejemplo, se sigue durante 10 años a una cohorte de pacientes trasplantados y se registra el tiempo en el que ocurre el fallo. Al final del estudio puede haber pacientes con el injerto funcionando. La segunda causa es la pérdida del seguimiento, en el ejemplo anterior, algunos de los sujetos pueden desaparecer del estudio en algún momento de este por diversos motivos: cambio de domicilio, falta de interés, etc. Una última causa de censura es la ocurrencia de otro evento que impida la observación del evento de interés, por ejemplo muerte por una causa ajena al estudio.

Todos estos tiempos censurados infraestiman el tiempo hasta el evento, que siempre ocurrirá, asumiendo que ocurra, en un tiempo posterior. Es importante fijarse en la última causa de censura que suele plantear un problema en el diseño de estos estudios. Por ejemplo, en un estudio de supervivencia al trasplante, un paciente muere por accidente de tráfico, hay autores que prefieren definir el evento de interés como muerte por una causa relacionada con la enfermedad en estudio, por ejemplo Opelz y Döhler¹, y por tanto censuran las muertes por causas no relacionadas, argumentando, con cierto sentido, que una muerte por accidente de tráfico no se debería tener en cuenta si se está evaluando la supervivencia al trasplante, mientras que otros prefieren definirlo como muerte por todas las causas, por ejemplo Costa et al². A favor de esta última opción está la dificultad que suele tener la clasificación de una muerte

Correspondencia: Víctor Abraira

Unidad de Bioestadística Clínica,
Hospital Universitario Ramón y Cajal, IRYCIS, Madrid.
victor.abraira@gmail.com

como relacionada o no con la enfermedad en estudio y su tratamiento y, por lo tanto, la potencial inclusión de sesgos por mala clasificación. En general, la muerte por todas las causas se considera un evento más robusto que la muerte por causas específicas³ y sería por ello el preferido. En el caso del trasplante, y estudiando la supervivencia del injerto, se plantea el problema de cómo considerar la muerte con injerto funcionando, si como evento o como censura, y a veces se realizan los análisis para las 2 definiciones del evento, como Curran et al⁴. A la hora de comparar estudios, es muy importante fijarse en las definiciones de evento y censura usadas, aunque, desgraciadamente con excesiva frecuencia, en las publicaciones no quedan claras^{5,6}.

Al conjunto de técnicas estadísticas que permiten estudiar la variable «tiempo hasta que ocurre un evento» y su dependencia de otras posibles variables explicativas, teniendo en cuenta la información parcial contenida en las censuras, se le denomina análisis de supervivencia y aunque el nombre sugiere que el evento es muerte, también puede y debe usarse para analizar el tiempo hasta cualquier otro evento, como por ejemplo curación, aparición de la enfermedad o rechazo del injerto. En trasplante se suelen estudiar sobre todo la muerte, por todas las causas o por causas específicas, rechazo del injerto, probado por autopsia o clínico y fallo de injerto. Si bien los primeros análisis de este tipo aplicados a la supervivencia humana aparecen en el siglo XVII⁷, los procedimientos clave se incorporaron en la segunda mitad del siglo pasado y desde entonces se han usado intensamente en la investigación clínica para responder a una gran variedad de preguntas, desde evaluación de tratamientos a identificación y cuantificación de factores pronósticos. En otros ámbitos, como procesos de control de calidad, se estudia con las mismas técnicas, aunque con otro nombre (análisis del tiempo de fallo), el tiempo hasta que un cierto producto falla o el tiempo de espera hasta recibir un servicio (análisis de tiempo de espera), etc.

FUNCIONES DE SUPERVIVENCIA Y RIESGO

Para describir y modelizar la variable tiempo hasta que ocurre el evento suelen usarse 2 funciones de probabilidad relacionadas entre sí: las de supervivencia y la de

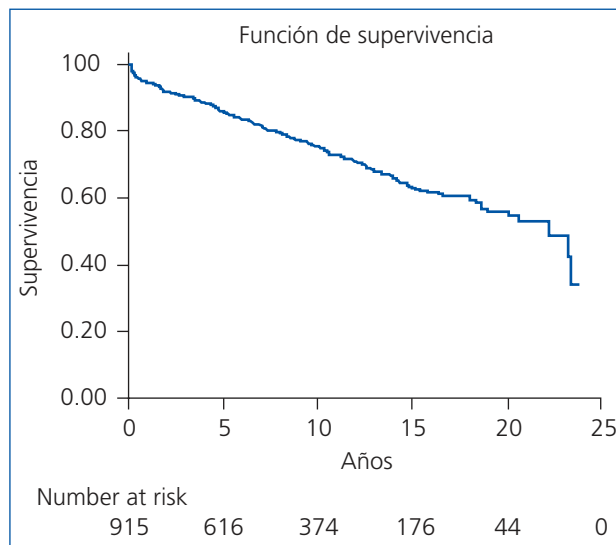


Figura 1. Curva de supervivencia de los enfermos trasplantados en el hospital Ramón y Cajal desde 1979 hasta 2002. Debajo del eje del tiempo figura el número de pacientes que siguen en observación en cada tiempo.

riesgo⁸, que resultan especialmente útiles para datos con observaciones censuradas.

La probabilidad de supervivencia, representada por la denominada función de supervivencia $S(t)$ es la probabilidad de que, en un sujeto de la cohorte, el evento ocurra en un tiempo igual o mayor que t (si el evento es muerte, sobreviva al menos t). Para ilustrar los métodos descritos se usará como ejemplo los datos de supervivencia al trasplante renal de los enfermos trasplantados en el hospital Ramón y Cajal desde el 10 marzo de 1979 hasta el 31 de diciembre de 2002, con seguimiento hasta el 1 de junio de 2003⁹, considerando como evento la muerte por todas las causas. En la figura 1 se muestra la función de supervivencia en la que se observa su característica forma monótona no creciente. En ella, por ejemplo, $S(10) = 0,75$ es la probabilidad de que un enfermo sobreviva 10 años o más, $S(20) = 0,55$ la probabilidad de que sobreviva al menos 20 años, etc. Es, por lo tanto, una manera de resumir la información de forma muy relevante en estos problemas y que, para facilitar su lectura, suele presentarse gráficamente. En ocasiones, por ejemplo Curran et al⁴, se presenta, en lugar de la función de supervivencia, la de incidencia acumulada calculada como

$1 - S(t)$ y que es, para cada tiempo, la probabilidad de que el evento haya ocurrido antes.

El riesgo (*hazard* en inglés), representado por la función de riesgo $h(t)$, es la probabilidad por unidad de tiempo de que un sujeto que está en observación en el tiempo t (ha sobrevivido hasta ese tiempo) tenga el evento justo en ese tiempo, es decir, representa la tasa instantánea de eventos en cada tiempo o, dicho con otras palabras, el potencial instantáneo de ocurrencia de eventos por unidad de tiempo, condicionado a que el evento no ocurrió antes¹⁰. En nuestro ejemplo, la función de riesgo a los 20 años es la densidad de probabilidad de que el paciente fallezca a los 20 años del trasplante, condicionada a que ha sobrevivido hasta entonces.

Obsérvese que la función de supervivencia $S(t)$ se enfoca en la no ocurrencia del evento, mientras que la de riesgo $h(t)$ lo hace en la ocurrencia; además $S(t)$ es una probabilidad, mientras que $h(t)$ es una «densidad de probabilidad» por unidad de tiempo; ello hace más intuitivo el manejo de $S(t)$, por ejemplo, los valores de $S(t)$ están comprendidos entre 0 y 1, mientras que $h(t)$ no tiene cota superior. Sin embargo, ambas funciones están relacionadas y dada una de ellas se puede obtener la otra⁸.

Como ya se ha mencionado, los tiempos hasta el evento suelen tener distribuciones asimétricas, en las que los tiempos mayores que los típicos son menos frecuentes y están más dispersos que los tiempos menores. En esta situación, la media no es una buena medida de centralización de la variable porque la pueden agrandar unos pocos valores atípicamente grandes y, en consecuencia, la medida de centralización aconsejada es la mediana. La mediana de supervivencia, también denominada tiempo de semivida o vida-media (*half-life* en inglés), es el tiempo en que la probabilidad de supervivencia es igual a 0,5 o, dicho en términos de frecuencia, el tiempo al que sobrevivirá la mitad de la población y que, por lo tanto, se obtiene fácilmente a partir de la función de supervivencia. Gráficamente se obtiene trazando una línea horizontal desde el valor 0,5 del eje vertical hasta que corte a la curva de supervivencia y desde este punto de corte, una línea vertical hasta el eje horizontal; este eje será cortado en la mediana. Haciendo esto en la

figura 1 se observa que la mediana de supervivencia en nuestro ejemplo es aproximadamente de 22 años.

ESTIMACIÓN DE LA FUNCIÓN DE SUPERVIVENCIA (MÉTODO DE KAPLAN-MEIER)

La función de supervivencia se suele estimar por el método de Kaplan-Meier, que tiene en cuenta la información contenida en las censuras, asumiendo, y es importante resaltarlo porque es un punto crítico para la validez de la estimación, que las censuras son no informativas o, dicho de otro modo, que los sujetos censurados tienen el mismo riesgo de ocurrencia del evento que los que siguen en observación más allá de ese tiempo. La importancia de esta asunción es tanto mayor cuanto mayor sea el número de censuras. Nótese que esta asunción no es trivial y se violaría, por ejemplo, si los pacientes que abandonan el estudio están más graves, o menos, que los que continúan o, en otro ejemplo, si el evento fuera progresión de la enfermedad, la muerte sin progresión seguramente no cumple con esa asunción. En el estudio del fallo del injerto, la muerte con el injerto funcionando ¿la cumple? Aunque excede los objetivos de este artículo profundizar en ello, hay que distinguir entre censura (circunstancia que impide la observación del evento, pero se supone que este ocurrirá en el futuro), que es la situación que contempla el método descrito aquí, y evento competitivo (circunstancia que impide que el evento ocurra), que requiere un tratamiento estadístico distinto¹¹ y que, aunque apenas usado en la literatura clínica, sería el adecuado en estas situaciones.

En la figura 1 se pone claramente de manifiesto como el efecto de la información contenida en las censuras aumenta la supervivencia estimada en los tiempos posteriores a ellas. Por ejemplo, a los 10 años del trasplante solo permanecen en el estudio 374 pacientes (el 40,9% de los 915 iniciales) y, sin embargo, la probabilidad de supervivencia estimada es de 0,75. Otra manera de entender la asunción fundamental del método es que se considera que las observaciones censuradas tienen la misma mediana de supervivencia que las observaciones que siguen en el estudio; es decir, a la estimación de la supervivencia en cualquier tiempo están contribuyendo no solo los pacientes que se observa que sobreviven hasta ese tiempo, sino también los censurados pre-

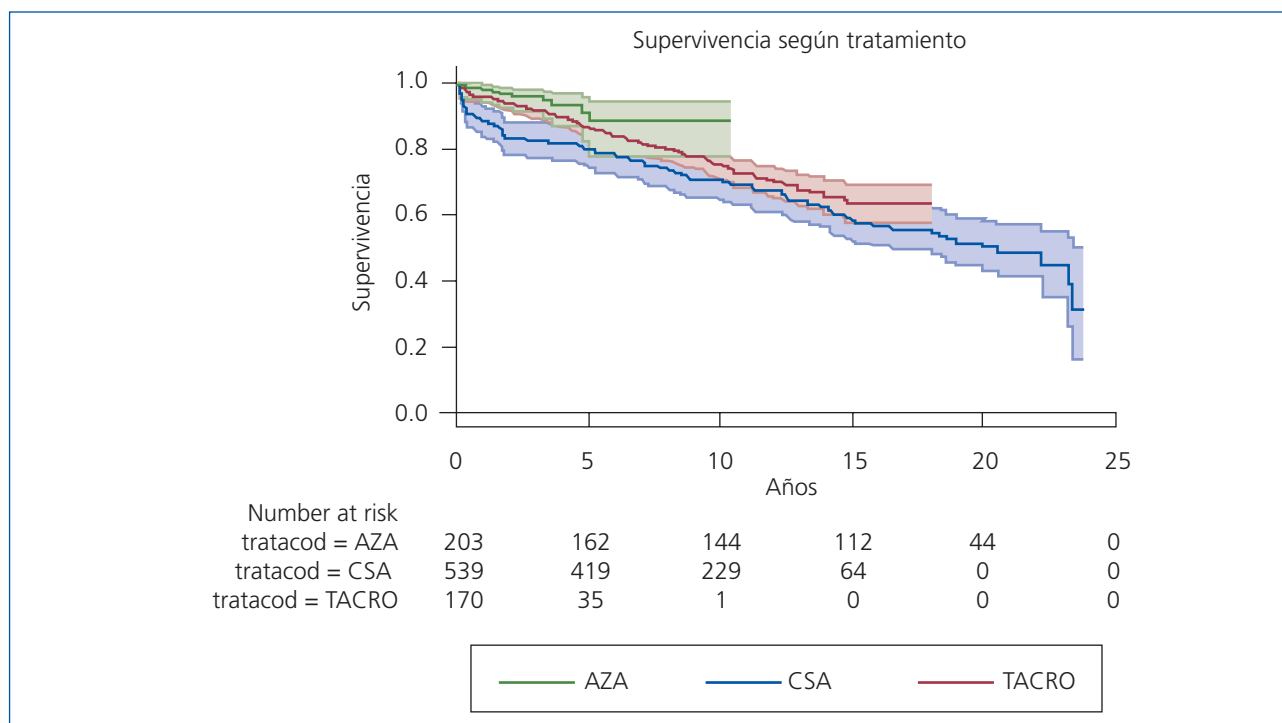


Figura 2. Curvas de supervivencia con sus intervalos de confianza del 95%, para los 3 grupos de tratamiento. Obsérvese como la anchura de los intervalos aumenta según avanza el tiempo y como la figura llega a ser engorrosa al representar más de una curva. Se presenta también, como alternativa, el número de individuos que permanece en el estudio en cada rama en distintos tiempos.

AZA: azatioprina; CsA: ciclosporina; TACRO: tacrolimus.

viamente, en la asunción de que si no se hubieran censurado se comportarían como los que sí siguen. Nótese el impacto que puede tener esta asunción y, por lo tanto, el cuidado con el que debe considerarse en el diseño del estudio.

PRECISIÓN DE LAS ESTIMACIONES

Es importante destacar también que estas estimaciones se hacen a partir de muestras, es decir, un subconjunto de los pacientes de interés en lugar de todos ellos, por tanto presentan una imprecisión que se debe cuantificar por el intervalo de confianza —IC— (intervalo dentro del cual se puede asegurar, con una probabilidad determinada, habitualmente el 95%, que se encontraría el verdadero valor si se hubieran estudiado todos los pacientes, asumiendo que la muestra sea aleatoria). En el caso de estas curvas, la anchura del IC aumenta según avanza el tiempo como

consecuencia de la disminución del número de pacientes. En algunos estudios, la anchura al final del tiempo de seguimiento es tan grande, aunque con demasiada frecuencia no mostrada⁶, que se hace difícil la interpretación de la curva. Una de las razones por las que no siempre se muestran los IC es que cuando se presentan varias curvas en la misma gráfica, esta puede llegar a ser engorrosa. Un modo alternativo de presentar la precisión en estos casos, como hicieron Curran et al⁴, es incluir debajo del eje del tiempo el número de pacientes que permanece en el estudio en cada tiempo. Como ayuda a la interpretación de cómo afecta este número a la precisión se puede usar la recomendación de Pocock et al¹² de no presentar las curvas, en cada rama, a partir del momento en que queden menos del 20-10% de los pacientes que empezaron el estudio. En la figura 2 se muestran las curvas con sus IC del 95% para los pacientes del ejemplo según el tratamiento recibido y se observa el aumento de la anchura de los intervalos se-

gún avanza el tiempo y el emborronamiento de la gráfica, al mostrar varias curvas con sus IC. También se presentan el número de pacientes que permanece en el estudio.

OBJETIVOS DEL ANÁLISIS DE SUPERVIVENCIA

En un análisis de supervivencia se pueden establecer 3 objetivos y un estudio puede tener los 3, o solo alguno de ellos:

1. *Estimar e interpretar* las curvas de supervivencia y/o riesgo, más frecuentemente la de supervivencia, aunque la de riesgo tiene mayor utilidad para buscar modelos teóricos, por ejemplo, un riesgo constante a lo largo del tiempo indica un modelo exponencial, un riesgo creciente, o decreciente, puede modelizarse mediante la función de Weibull, etc.^{3,8,13}. Este es el objetivo de lo que Hemingway et al¹⁴ denominan investigación fundamental en pronóstico, en la que se trata de estudiar el curso natural, o con el cuidado habitual, de la enfermedad y sus variaciones.
2. *Comparar* curvas de supervivencia, por ejemplo, entre 2 tratamientos en los ensayos clínicos, o entre 2 grupos de pacientes establecidos con respecto a cualquier otro criterio en la denominada investigación en factores pronósticos¹⁴, en la que se trata de identificar factores específicos (como biomarcadores, fármacos, otras enfermedades, signos, etc.) asociados con el pronóstico.
3. *Evaluar la relación* de la supervivencia con otras, más de una, variables pronósticas en la investigación sobre modelos pronósticos, que trata de desarrollar, validar y evaluar el impacto de modelos estadísticos para predecir el riesgo individual de un futuro evento¹⁴.

En los próximos apartados se muestran los métodos para cumplir con estos objetivos.

COMPARACIÓN DE CURVAS DE SUPERVIVENCIA

La prueba más usada para comparar curvas de supervivencia es la conocida como prueba del *log-rank* y consiste

en calcular, en cada tiempo y para cada grupo, el número de eventos que se esperarían encontrar si no hubiera diferencia entre los grupos, para construir un estadístico con la suma de las diferencias entre eventos observados y esperados al cuadrado, divididas por los esperados, que se distribuye como una χ^2 con $g - 1$ grados de libertad (siendo g el número de grupos). A partir de este estadístico se calcula el valor p correspondiente a la comparación de las supervivencias de los grupos. Este método se puede usar para comparar 2 o más de 2 grupos.

La comparación mediante la prueba del *log-rank* para las curvas de la figura 2 produce un valor $p = 0,0149$, es decir, con la convención habitual, la diferencia es estadísticamente significativa.

ANÁLISIS MULTIVARIABLE: SU NECESIDAD

Comparar las curvas con este test es muy útil en la investigación clínica, sobre todo en los ensayos clínicos en los que la asignación aleatoria de los pacientes a los tratamientos hace que inicialmente todos los grupos tengan igual pronóstico, pero en los estudios observacionales es insuficiente, porque hay que realizar las comparaciones teniendo en cuenta el efecto de otras variables distintas a la que se utiliza para definir los grupos. En el ejemplo de los pacientes del hospital Ramón y Cajal se encontró que la supervivencia es distinta según el tratamiento recibido, pero como los tratamientos no se asignaron al azar sino con el mejor criterio clínico en cada paciente, teniendo en cuenta los tratamientos disponibles en cada momento, podría ocurrir que los pacientes asignados a cada tratamiento tuvieran distintos valores en otras variables que influyan en el pronóstico, como la edad, el sexo, etc., lo que dificultaría la interpretación de ese resultado. En la comparación con el método del *log-rank* se ignora este efecto de otras variables. Por otro lado, el *log-rank* ofrece un valor p para la diferencia entre los grupos, pero no estima el *tamaño del efecto* (y recuérdese que el valor p no depende solo del tamaño del efecto, sino, y sobre todo, del tamaño muestral)¹⁵; es decir hace una evaluación estadística, pero no clínica, de la diferencia entre los grupos.

El análisis multivariable, en particular los modelos de regresión, permite soslayar ambas limitaciones, es decir,

Tabla 1. Coeficientes (β), riesgos relativos (RR) y sus intervalos de confianza (IC) estimados mediante modelos de Cox univariable y multivariable para los pacientes trasplantados en el Hospital Ramón y Cajal

Variable	Análisis univariable			Análisis multivariable		
	β	RR (IC del 95%)	p	β	RR (IC del 95%)	p
Sexo	-0,29	0,74 (0,57-0,98)	0,031	-0,24	0,79 (0,60-1,03)	0,083
Edad (años)	0,04	1,04 (1,03-1,05)	0,000	0,04	1,04 (1,02-1,05)	0,000
Tratamiento			0,010			0,083
AZA	0,92	2,50 (1,27-4,91)	0,008	0,76	2,13 (1,08-4,19)	0,028
CsA	0,68	1,97 (1,02-3,78)	0,043	0,62	1,85 (0,96-3,56)	0,065

AZA: azatioprina; CsA: ciclosporina; IC: intervalo de confianza.

evaluar el efecto de una variable, teniendo en cuenta (*controlando por*, en la terminología clinicoepidemiológica) otras variables y estimar el tamaño del efecto. También permite desarrollar modelos para predecir el riesgo individual de un futuro evento (objetivo 3). Además, y a diferencia de los métodos de Kaplan-Meier y *log-rank* que requieren que la variable sea categórica, los modelos de regresión permiten estudiar variables continuas (p. ej., edad, o tiempo de isquemia fría), binarias (p. ej., sexo), categóricas (p. ej., histología o tratamiento) u ordinales (p. ej., estadio de un tumor o estatus funcional).

MODELO DE RIESGO PROPORCIONAL O DE COX

Aunque hay varios modelos de regresión propuestos, el que se usa casi en exclusiva en la investigación clínica⁶, por su sencillez y facilidad para interpretar los coeficientes, es el denominado modelo de riesgo proporcional o modelo de Cox, que es un modelo que plantea el logaritmo del riesgo relativo (RR) como una función lineal de las variables independientes. Se supone, por lo tanto, que el RR, a diferencia del riesgo propiamente dicho, no depende del tiempo o, dicho de otra manera, que es constante a lo largo del tiempo (de ahí el nombre de modelo de riesgo proporcional). Esta es la única asunción del modelo y conviene destacar que no siempre es razonable y se debe evaluar en cada caso.

Con este modelo, el coeficiente de cada variable independiente es el logaritmo del RR cuando la variable aumenta

una unidad manteniéndose constantes las demás variables y, por tanto, su exponencial es el RR. Los paquetes estadísticos ofrecen tanto los coeficientes como sus exponenciales.

La tabla 1 muestra los coeficientes estimados por el modelo, los RR, sus IC del 95% y la significación estadística para las variables edad en el momento del trasplante (años), sexo y tratamiento, para los pacientes trasplantados del Hospital Ramón y Cajal. Se muestran las estimaciones realizadas con los modelos univariados (solo una variable en cada modelo) y con el modelo multivariable (un solo modelo con todas las variables), a efectos de comentar las diferencias. Para las variables continuas (edad en nuestro caso), el RR que estima el modelo es por aumento de unidad (1 año), es decir, por cada año adicional a la edad del trasplante, el riesgo de muerte en todo tiempo se multiplica por 1,04 o, lo que es lo mismo, es un 4% mayor, y este RR es significativamente distinto de 1 (ya que $p = 0,000$, o equivalentemente el IC del 95% [1,02-1,05] no incluye el valor 1). En el caso del sexo, codificado como 0 para varones y 1 para mujeres, el RR (en el modelo multivariable) de 0,79 quiere decir que el riesgo de las mujeres es el 79% (se multiplica por 0,79) del de los varones, aunque este RR no es significativamente distinto de 1 ($p = 0,083$, o el IC del 95% [0,60-1,03] sí incluye el valor 1). Se llama la atención sobre que si se hubiera codificado al revés, es decir 1 para varones y 0 para mujeres, se estaría considerando el riesgo de las mujeres como referencia, y en consecuencia, el RR obtenido sería el inverso, $RR = 1/0,79 = 1,27$ (RR de los varones con respecto a las mujeres).

Es importante detenerse sobre algunos aspectos relevantes de los modelos de regresión, en particular de los de Cox. En la tabla 1 se observa que en el análisis realizado con los modelos univariantes se encontró un efecto significativo del sexo (RR = 0,74; $p = 0,031$) y del tratamiento ($p = 0,010$), que en el análisis multivariable, es decir cuando se ha tenido en cuenta otras variables, deja de ser significativo (RR = 0,79; $p = 0,083$ para el sexo y $p = 0,083$ para el tratamiento). La interpretación de esta diferencia es que la distinta supervivencia para los sexos es un resultado que está confundido por el hecho de que los pacientes de ambos sexos también difieren en alguna de las otras variables del modelo y que cuando se controla por ellas, el efecto desaparece. Y la misma interpretación para el tratamiento.

Hay que señalar que el modelo de Cox univariante, para variables categóricas, es equivalente a la prueba del *log-rank* y se obtiene un valor p tanto más parecido cuanto mayor sea el tamaño muestral, aunque el modelo de Cox produce también una estimación del tamaño del efecto, el RR, que no se obtiene con el *log-rank*.

En cuanto al tratamiento, al ser una variable categórica con 3 categorías, el modelo ofrece el valor p que corresponde al contraste de hipótesis en el que se comparan globalmente los 3 grupos ($p = 0,010$ en el modelo univariante) y además los valores p y los RR, con sus IC, de las comparaciones de cada grupo con el que se haya tomado como referencia. En el ejemplo se tomó como referencia el grupo con tacrolimus y se compararon con él los grupos con azatioprina (RR = 2,50; $p = 0,008$, en el análisis univariante) y ciclosporina A (RR = 1,97; $p = 0,043$). Es decir, la supervivencia es diferente según el tratamiento ($p = 0,010$), aunque esta diferencia disminuye y deja de ser significativa al controlar por las otras variables.

UTILIZACIÓN DE LOS MODELOS DE COX

Los modelos de Cox se pueden usar tanto para identificar factores específicos asociados con el pronóstico (objetivo estimativo), como para la predicción del riesgo individual de un futuro evento (objetivo predictivo). Con el objetivo estimativo, el interés del investigador se centra en estimar la relación de una variable con la supervivencia. Este es

el objetivo más frecuente en la literatura clínica⁶ y ejemplos en el campo del trasplante son el de Costa et al² y Opelz y Döhler¹. En estos estudios, el resultado es la estimación del efecto de la variable de interés, es decir el RR de esa variable y las demás variables están en el modelo para evitar su posible efecto perturbador^{8,16}. En cambio, en un modelo predictivo, el resultado es el modelo completo en forma de fórmula, o gráfica (nomograma), que permite calcular la probabilidad futura del evento para un sujeto en función de los valores observados de los variables, o bien como una puntuación obtenida de la fórmula, con sus puntos de corte para generar grupos de riesgo. El uso de los modelos pronósticos para hacer predicciones individuales es más preciso y en general preferible a la creación de grupos de riesgo, aunque los grupos de riesgo pueden informar para opciones de tratamiento y permitir la estratificación por riesgo de gravedad en los ensayos clínicos¹⁷. Un ejemplo de estudio con objetivo predictivo en el campo del trasplante renal es el de Akl et al¹⁸.

En un estudio estimativo existe confusión cuando la relación entre la variable en estudio y la supervivencia difiere según que se considere o no otra variable; a esta última variable se le denomina variable de confusión para la asociación. Por ejemplo, si el análisis mostrado en la tabla 1 tuviera como objetivo estimar el efecto del tratamiento se encontraría que las variables edad y sexo son variables de confusión, porque los RR estimados en el modelo univariante (2,50 para el grupo con azatioprina y 1,97 para el de la ciclosporina) son distintos que los del modelo multivariable (2,13 y 1,85, respectivamente). Lo habitual es considerar que existe confusión cuando el RR cambia en más del 10%.

Otra dificultad que se puede presentar en un estudio estimativo es la existencia de interacción o modificación del efecto. Existe interacción cuando la asociación entre 2 variables, en este caso factor y evento, varía según los diferentes niveles de otra u otras variables. El análisis más sencillo de la interacción se hace mediante el análisis de subgrupos creados según los niveles de la variable modificadora de efecto¹⁹, para ello hay que evaluar el efecto del factor en las distintas categorías de la variable modificadora del efecto; por ejemplo, el efecto de la muerte por causa cardíaca del donante (MCC) sobre los resultados de

un trasplante de riñón puede ser distinto según los valores del índice de riesgo del donante (KDRI), si fuera así, el efecto de la MCC se debería mostrar, no como un RR sino como varios: uno para cada valor del KDRI. Se diría que hay interacción entre la MCC y el KDRI o que el KDRI modifica el efecto de la MCC. Esta manera de analizarlo, muy utilizada en ensayos clínicos, presenta 2 dificultades: está restringida a variables modificadoras categóricas u obliga a categorizarlas y no compara estadísticamente si los efectos son distintos. Para evaluar estadísticamente la interacción entre un factor y una variable modificadora del efecto se suelen usar modelos de regresión en los que se incluyen como variables independientes el factor, la variable modificadora y el producto de las 2 anteriores; la significación obtenida para este producto contrasta estadísticamente la interacción entre el factor y la variable. Si hay interacción, el efecto del factor no es único sino que será distinto según los valores de la variable considerada, análogo a la consideración realizada anteriormente en el caso de análisis de subgrupos. Si la interacción se realiza con una variable continua, el efecto será distinto según cada valor de la variable; una alternativa de presentación podría ser categorizar la variable²⁰, o mostrar el efecto del factor mediante un gráfico en el que se represente en el eje X el RR y en el eje de ordenadas la variable modificadora.

EVALUACIÓN DEL MODELO

Los modelos de regresión son una herramienta muy poderosa para analizar factores pronósticos y construir modelos para predecir el riesgo individual de un futuro evento, pero se deben evaluar. Aunque sigue siendo una asignatura pendiente en la investigación clínica⁶ hay 3 aspectos que hay que evaluar de los modelos, y son: evaluación de las asunciones, bondad del ajuste y evaluación de la validez del modelo.

Evaluación de las asunciones

La única asunción del modelo de Cox es la del riesgo proporcional, que se puede evaluar gráfica o estadísticamente. Una primera manera de evaluarla gráficamente, válida solo para variables categóricas, consiste en comparar visualmen-

te las funciones de supervivencia predichas por el modelo con las observadas (estimadas por el método de Kaplan-Meier). Si se cumpliera la asunción serían muy parecidas. En la figura 3A se muestran ambas gráficas para los grupos de tratamiento en los pacientes trasplantados en el Ramón y Cajal, en la que se observa una notable coincidencia. Nunca coincidirán del todo, porque la estimación realizada por Kaplan-Meier solo se hace en los tiempos en los que ocurren eventos, mientras que la estimación con el modelo se puede hacer en todo tiempo. El Stata, paquete estadístico con el que se ha realizado la figura 3A, presenta, para cada grupo, las estimaciones en los tiempos en que ocurren eventos o censuras. Por ejemplo, a los 10 años en el grupo de tacrolimus hay una censura, en ese tiempo no se observa cambio en la estimación de Kaplan-Meier y sí en la del modelo.

Otra forma usada más frecuentemente es la representación, en función del tiempo, de los logaritmos de las funciones de riesgo. Si los riesgos son proporcionales, estas gráficas son paralelas. Con frecuencia, las gráficas del logaritmo del riesgo suelen ser de tipo exponencial y no es fácil comprobar visualmente el paralelismo, por lo que se suele presentar el logaritmo del riesgo en función del logaritmo del tiempo, gráficas que también deben ser paralelas, si se cumple la asunción de riesgo proporcional y en las que el paralelismo es más fácil de observar. En la figura 3B se muestran estas gráficas para los grupos de tratamiento. Un problema con estas evaluaciones gráficas es decidir cuándo las líneas dejan de ser paralelas, decisión que siempre tiene un componente subjetivo. Hay una recomendación general¹⁰, debido a la robustez del método, de usar una estrategia conservadora y asumir la proporcionalidad, a menos que haya una fuerte impresión de falta de paralelismo. Usando esta recomendación se aceptaría el paralelismo en esta gráfica.

Otras pruebas gráficas son las incluidas en el denominado análisis de residuos. Los residuos son diferencias entre «cantidades» observadas y predichas por el modelo. Hay varios residuos propuestos, entre ellos los más habituales son los de *Schoenfeld*, los de *Cox-Snell* y los *martingala*, aunque algunos de ellos son difíciles de interpretar debido a las censuras¹⁶. En general, si se cumple la asunción, las gráficas de los residuos presentan una distribución dentro de una banda horizontal.

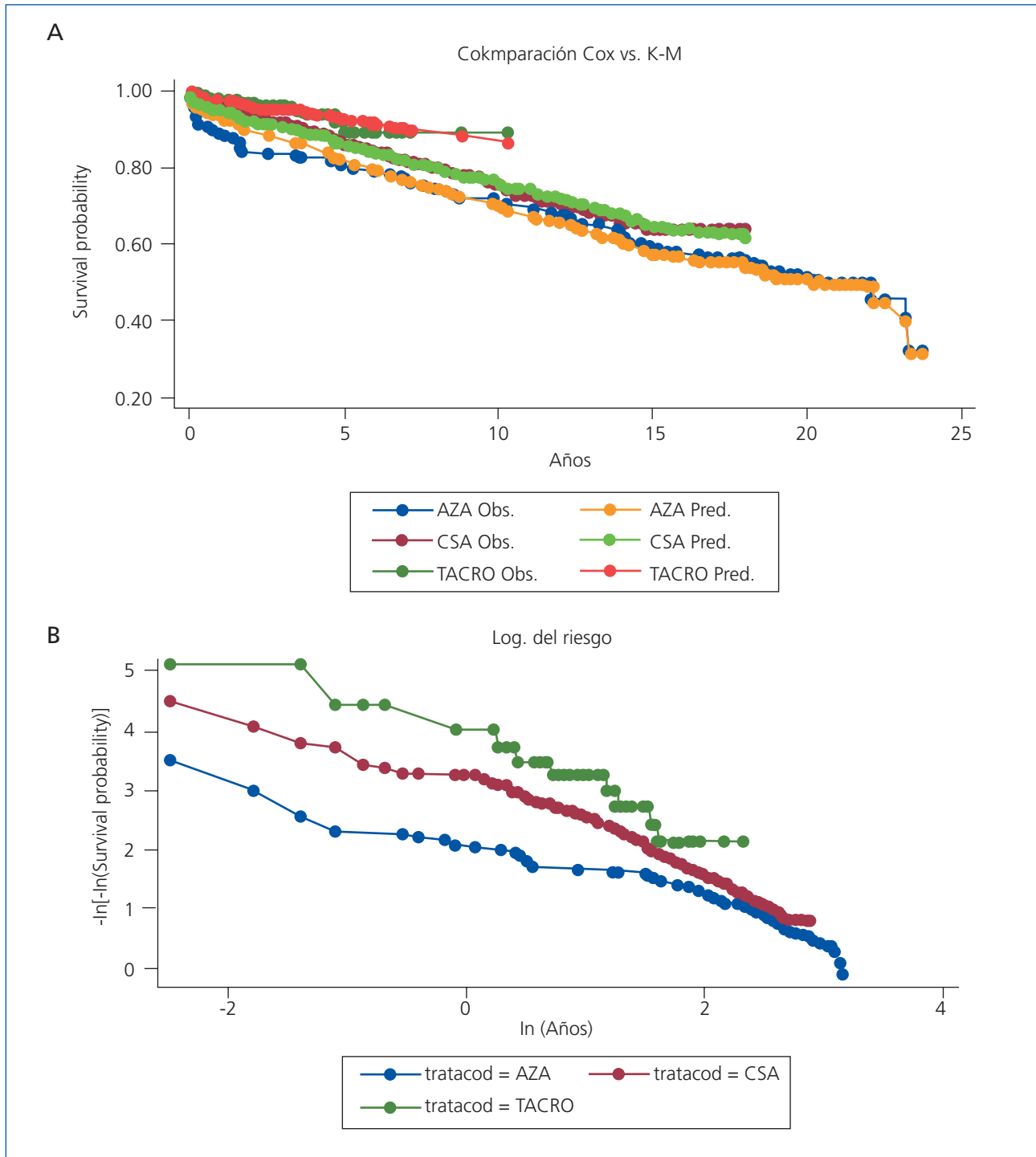


Figura 3. Métodos gráficos de comprobación de la asunción de riesgo proporcional del modelo Cox. A) Superposición de las curvas estimadas por Kaplan-Meier (observados —obs—) y por el modelo (predichas —pred—). Si se cumple la asunción son muy parecidas. B) Logaritmo de la función de riesgo en función del logaritmo del tiempo. Aunque el eje Y está etiquetado como $-\ln[-\ln(S(t))]$ hay que tener en cuenta que $-\ln(S(t))$ es la función de riesgo⁸. Si se cumple la asunción, las gráficas de los grupos deben de ser paralelas. AZA: azatioprina; CsA: ciclosporina; TACRO: tacrolimus.

Una evaluación estadística, que puede usarse tanto para variables categóricas como continuas, consiste en incluir en el modelo una variable de interacción con el tiempo, si esta variable fuera significativa indicaría que el riesgo varía con el tiempo y, por lo tanto, no se cumple la asunción. Si así fuera podría dejarse esta variable en el modelo, aunque se complica su interpretación: los RR son distintos para cada tiempo.

Bondad del ajuste

Hay varios métodos que evalúan globalmente lo bien que ajustan los datos al modelo. Uno muy general de todos los modelos de regresión que, como el de Cox, estiman los coeficientes por el método de máxima verosimilitud⁸ consiste en calcular un estadístico, el logaritmo del cociente de las verosimilitudes del modelo estimado y de los datos sin modelo, y que en la hipótesis nula de que los datos no ajusten al modelo se distribuye como una χ^2 con k grados de libertad (siendo k el número de variables en el modelo).

Validez del modelo

El último aspecto que hay que evaluar en un modelo es su validez o fiabilidad, es decir, ¿se comporta igual en otras muestras extraídas de la misma población (*reproducibilidad*) y/o de otras similares (*transportabilidad*)?²¹.

El modo más completo de evaluarlo será repetir el estudio con otra muestra y comparar los resultados (validación externa). Otra aproximación similar es partir la muestra inicial en 2 grupos, preferiblemente de forma no aleatoria²², por ejemplo por centro, un grupo de trabajo con el que se desarrolla el modelo y un grupo de validación, con el que se hace la validación. Una última aproximación consiste en usar técnicas de remuestreo como «*jackknife*» y «*bootstrap*»²³.

La evaluación se hace de modo distinto en los modelos estimativos que en los predictivos. En modelos estimativos, simplemente se ajusta el modelo en el grupo de trabajo, luego se hace la estimación del efecto en el de validación y se comparan.

En los modelos predictivos, la validación es un poco más compleja^{21,24}. Tiene 2 componentes: *calibración*, que es el grado en que la supervivencia predicha coincide con la observada, y puede evaluarse comparando la supervivencia predicha por el modelo en el grupo de validación con la estimada en el mismo grupo por Kaplan-Meier, y *discriminación*, grado en que el modelo distingue entre sujetos de alto y bajo riesgo.

EXTENSIONES DE LOS MODELOS DE COX

El modelo de Cox comentado hasta aquí exige que las variables independientes sean constantes a lo largo del tiempo (p. ej., sexo) o medidas en el momento inicial del seguimiento (p. ej., edad en el trasplante), pero a veces interesa evaluar el efecto de variables cambiantes en el tiempo, sobre todo en seguimientos largos, como ocurre en el trasplante. Hay una extensión del modelo de Cox que permite incluir estas variables (variables dependientes del tiempo) y contempla 2 esquemas: una variable que puede aparecer, o no, en algún momento del seguimiento, por ejemplo la aparición de diabetes mellitus después del trasplante y su efecto en la supervivencia²⁵; la otra situación es la de una variable que se mide en distintos momentos del seguimiento, por ejemplo la creatinina²⁶. Con esta extensión se puede estimar el efecto de la variable, pero no se pueden hacer predicciones para un paciente concreto.

Aunque raramente se usa, también conviene mencionar la extensión del modelo para contemplar riesgos competitivos²⁷. Se empieza a usar para el estudio de muerte por episodios agudos²⁸, con el alta hospitalaria como episodio competitivo, o la duración de la estancia hospitalaria, con la muerte como evento competitivo²⁹. En el caso del estudio del fallo del injerto, como ya se ha mencionado, la muerte con el injerto funcionando podría ser más un evento competitivo que una censura.

ELEMENTOS CLAVE PARA HACER UNA LECTURA CRÍTICA DE UN ARTÍCULO QUE UTILICE ANÁLISIS DE SUPERVIVENCIA

Como ya se ha comentado varias veces, el análisis de supervivencia y sus resultados no se describen en la litera-

tura médica con la calidad que sería de desear^{5,6}. Por ello es necesario que los lectores de estos artículos hagan una lectura crítica de los elementos clave de esta metodología, que se pueden resumir como⁶:

- Definición clara y precisa del evento, o eventos, de interés (p. ej., si fuera muerte, especificar si es por todas las causas o por alguna causa específica, o si fuera fallo del órgano, cómo se determina, si clínica o histológicamente) y descripción de las causas de censura. El lector crítico debe considerar la pertinencia y relevancia de estas definiciones, así como tener presente la necesidad de que las censuras sean no informativas. La importancia de esta asunción es tanto mayor cuanto mayor sea el número de censuras.
- Los números de pacientes y de eventos deben presentarse y evaluar si cumplen los requisitos del tamaño necesario para hacer los contrastes de hipótesis y/o uso de los modelos. Estos requisitos deben describirse en el apartado de tamaño muestral. Si se presentan análisis para varios eventos debe presentarse para todos ellos.
- Deben quedar claros los tiempos inicial y final del estudio, así como información sobre el tiempo promedio de seguimiento. Para ello, y aunque cuando se presenta suele ser la media o la mediana, el estadístico recomendado es la mediana de los tiempos censurados.
- Debe cuantificarse la incertidumbre de las estimaciones, idealmente mediante la presentación de IC o, al menos, presentar, junto a las curvas de supervivencia o riesgo, los sujetos en riesgo en diferentes tiempos.
- Cuando se usan modelos de regresión debe informarse si se han comprobado, y cómo, las asunciones del modelo (p. ej., en modelos de Cox, el riesgo proporcional).
- En modelos de regresión debe informarse sobre el número de variables incluidas en el modelo máximo, el criterio para su selección y la estrategia de construcción del modelo.
- En los modelos de regresión, especialmente los que tienen objetivo predictivo, se debe evaluar y presentar

el resultado, la bondad del ajuste e idealmente validar el modelo.

Conflicto de intereses

El autor declara que no tiene conflicto de intereses potenciales relacionados con los contenidos de este artículo.

REFERENCIAS BIBLIOGRÁFICAS

1. Opelz G, Döhler B. Cardiovascular death in kidney recipients treated with renin-angiotensin system blockers. *Transplantation*. 2014;97:310-5.
2. Costa NA, Kshirsagar AV, Wang L, Detwiler RK, Brookhart MA. Pretransplantation erythropoiesis-stimulating agent hyporesponsiveness is associated with increased kidney allograft failure and mortality. *Transplantation*. 2013;96:807-13.
3. Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part I: basic concepts and first analyses. *Br J Cancer*. 2003;89:232-8.
4. Curran SP, Famure O, Li Y, Kim SJ. Increased recipient body mass index is associated with acute rejection and other adverse outcomes after kidney transplantation. *Transplantation*. 2014;97:64-70.
5. Altman DG, De Stavola BL, Love SB, Stepniowska KA. Review of survival analyses published in cancer journals. *Br J Cancer*. 1995;72:511-518.
6. Abraira V, Muriel A, Emparanza JI, Pijoán JI, Royuela A, Plana MN, et al. Reporting quality of survival analyses in medical journals still needs improvement. A minimal requirements proposal. *J Clin Epidemiol*. 2013;66:1340-6.
7. Ciecka JE. Edmond Halley's Life Table and its uses. *Journal of Legal Economics*. 2008;15:65-74.
8. Abraira Santos V, Pérez de Vargas A. *Métodos multivariantes en Bioestadística*. Madrid: Centro de Estudios Ramón Areces; 1996.
9. Abraira V, Quereda C. Curvas de vida y supervivencia en el tratamiento sustitutivo de la insuficiencia renal: diálisis y trasplante renal. En: Arias M, editor. *Métodos estadísticos en el trasplante renal*. Madrid: Drug Farma; 2004. p. 67-133.
10. Kleinbaum DG. *Survival Analysis: a self-learning text*. New York: Springer-Verlag; 1996.
11. Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Stat Med*. 1999;18:695-706.

12. Pocock SJ, Clayton TC, Altman DG. Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. *Lancet*. 2002;359:1686-9.
13. Casariego E, Abraira V, Corredoira JC, Alonso García MP, Rodríguez Feijoo A, López Álvarez MJ, et al. A predictive model for mortality of bloodstream infections. Bedside analysis with the Weibull function. *J Clin Epidemiol*. 2002;55:563-72.
14. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. *BMJ*. 2013;346:e5595.
15. Abraira V. Contrastes de hipótesis: el valor p. *SEMERGEN*. 2002;28:374-5.
16. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis part III: multivariate data analysis -- choosing a model and assessing its adequacy and fit. *Br J Cancer*. 2003;89:605-11.
17. Steyerberg EW, Moons KG, Van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: Prognostic model research. *PLoS Med*. 2013;10:e1001381.
18. Akl A, Mostafa A, Ghoneim MA. Nomogram that predicts graft survival probability following living-donor kidney transplant. *Exp Clin Transplant*. 2008;6:30-6.
19. Wagner M, Balk EM, Kent DM, Kasiske BL, Ekberg H. Subgroup analyses in randomized controlled trials: the need for risk stratification in kidney transplantation. *Am J Transplant*. 2009;9:2217-22.
20. Singh SK, Kim SJ. Does expanded criteria donor status modify the outcomes of kidney transplantation from donors after cardiac death? *Am J Transplant*. 2013;13:329-336.
21. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med*. 1999;130:515-24.
22. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:b605.
23. Harrell FE Jr. Regression modeling strategies with applications to linear models, logistics regression, and survival analysis. New-York: Springer; 2001.
24. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000;19:453-73.
25. Kasiske BL, Snyder JJ, Gilbertson D, Matas AJ. Diabetes mellitus after kidney transplantation in the United States. *Am J Transplant*. 2003;3:178-85.
26. De Bruijne MH, Sijkens YW, Paul LC, Westendorp RG, Van Houwelingen HC, Zwinderman AH. Predicting kidney graft failure using time-dependent renal function covariates. *J Clin Epidemiol*. 2003;56:448-55.
27. Fine JP, Gray RJ. A Proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94:496-509.
28. Payen D, Lukaszewicz AC, Legrand M, Gayat E, Faivre V, Megarbane B, et al. A multicentre study of acute kidney injury in severe sepsis and septic shock: association with inflammatory phenotype and HLA genotype. *PLoS One*. 2012;7:e35838.
29. Brock GN, Barnes C, Ramírez JA, Myers J. How to handle mortality when investigating length of hospital stay and time to clinical stability. *BMC Med Res Methodol*. 2011;11:144.