



# Curvas ROC para Datos de Supervivencia. Aplicación a Datos Biomédicos

por

Anaderli Torres Ortiz

*Proyecto presentado en el cumplimiento de los requisitos para culminar el  
Master en Técnicas Estadísticas de la Universidad de Santiago de Compostela.*

Departamento de Estadística e Investigación Operativa  
Universidad de Santiago de Compostela  
Campus Universitario Sur, 15782, Santiago de Compostela.

Directora:

Carmen María Cadarso Suárez

Julio 2010

# Resumen

La curva ROC desempeña un papel fundamental en la evaluación de la precisión de las pruebas usadas para diagnosticar enfermedades. Cuando el objetivo de la prueba diagnóstica es la predicción de eventos en el tiempo, se involucran métodos de supervivencia y variables tiempo dependientes. En este contexto, la evaluación de la precisión de las pruebas diagnósticas se puede lograr haciendo uso de una extensión de los conceptos de sensibilidad y especificidad. En los últimos años se han propuesto varias definiciones y métodos de estimación de curvas ROC tiempo-dependientes. El propósito de este trabajo es realizar una revisión de estas definiciones, estudiar su capacidad para incorporar datos censurados, e ilustrar su uso a través de su aplicación a un conjunto de datos biomédicos.

# Agradecimientos

Porque el éxito nunca corresponde a una sola persona, y siempre es compartido en la medida en que se ha transferido el conocimiento, quiero expresar mi agradecimiento sincero

\* A la Dra. Carmen Cadarso, Directora de este trabajo, por sus apreciados y relevantes aportes para el desarrollo de esta investigación, además por la confianza depositada, por transmitir con esmero su pasión a la bioestadística y enseñarme la trascendencia del estudio de esta área.

\* A María Xosé Rodríguez, por su colaboración inestimable, y por su constante y paciente seguimiento y asistencia compartiendo su tiempo de manera generosa durante el desarrollo del presente trabajo.

\* Al Dr. Wenceslao Gonzales por sus consejos orientativos a la elección final de esta área de estudio.

\* Al Dr. Francisco Gude Sampedro por su asistencia en la interpretación médica de los resultados obtenidos, y al Servicio de Cardiología del Complejo Hospitalario Universitario de Santiago por prestar la base de datos real utilizada en esta memoria.

\* A mis colegas, Teresa Noguera e Isabel Martínez, por su constante espíritu servicial y desinteresada colaboración.

\* A Abelardo Monsalve, por compartir su discernimiento y los consejos brindados, por la motivación, el apoyo, y la disposición .

Por último también expreso mi más sincera gratitud y la hago extensiva a todos los extraordinarios profesionales que laboran como profesores en el Master de Técnicas Estadísticas de la USC, UDC, UVIGO, para mí ha sido un privilegio y un vivo ejemplo de excelencia profesional de los que pude aprender mucho.

# Índice general

<b>Índice general</b>	<b>IV</b>
<b>Índice de figuras</b>	<b>VI</b>
<b>Índice de tablas</b>	<b>VIII</b>
<b>1. Introducción</b>	<b>1</b>
<b>2. Base Teórica</b>	<b>4</b>
2.1. Estudio de Clasificación . . . . .	4
2.1.1. Test Diagnóstico . . . . .	5
2.1.2. Resultados de una Prueba Diagnóstica en Términos Probabilísticos . . . . .	6
2.1.3. Características Operativas . . . . .	7
2.1.4. Test Diagnóstico con Resultado Binario . . . . .	8
2.1.5. Test Diagnóstico con Resultado Continuo . . . . .	8
2.2. Relación entre Sensibilidad y Especificidad . . . . .	9
2.3. Curva ROC . . . . .	12
2.3.1. El Área Bajo la Curva ROC . . . . .	15
2.4. Análisis Supervivencia . . . . .	16
2.4.1. Datos Censurados . . . . .	17
2.4.2. El Modelo de Cox . . . . .	19
<b>3. Curva ROC para Datos de Supervivencia</b>	<b>22</b>
3.1. Incorporando la Dimensión del Tiempo . . . . .	22
3.2. Taxonomía, Exenciones de Sensibilidad y Especificidad . . . . .	23
3.2.1. Algunos Comentarios Importantes Sobre la Taxonomía . . . . .	24
3.3. Extensiones de la Definición de Curva ROC . . . . .	27
3.3.1. Acumulativo-Dinámico . . . . .	27
3.3.2. Incidente-Estático . . . . .	28
3.3.3. Incidente-Dinámico . . . . .	30
3.4. Censura y Eventos con Riesgos Competitivos . . . . .	31
3.5. Estimación de los Parámetros de Funcionamiento . . . . .	31
3.5.1. Enfoque de Leisenring et al. (1997) . . . . .	32
3.5.2. Enfoque de Etzioni et al. (1999) . . . . .	32
3.5.3. Enfoque de Heagerty et al. (2000) . . . . .	34
3.5.4. Enfoque de Cai et al. (2006) . . . . .	34

3.5.5. Enfoque de Heagerty and Zheng (2005) . . . . .	35
3.5.6. Enfoque de Song and Zhou (2008) . . . . .	36
3.6. Comparando las Ventajas . . . . .	37
<b>4. Aplicación</b>	<b>40</b>
4.1. Aplicación a Datos Simulados: AKI ( <i>Acute Kidney Injury</i> ) . . . . .	40
4.1.1. Biomarcador Baseline . . . . .	41
4.1.2. Biomarcador Longitudinal . . . . .	46
4.2. Aplicación a Datos Reales: ACS ( <i>Acute Coronary Syndrome</i> ) . . . . .	50
<b>5. Conclusión</b>	<b>59</b>
<b>A. Apéndice</b>	<b>60</b>
A.1. Metadatos AKI . . . . .	60
A.2. Generación de Datos Simulados AKI . . . . .	60
<b>B. Apéndice</b>	<b>63</b>
B.1. Metadatos ACS . . . . .	63
B.2. Selección Óptima del Modelo ACS . . . . .	64
<b>C. Apéndice</b>	<b>66</b>
C.1. Detalles del Software Utilizado . . . . .	66
<b>D. Apéndice</b>	<b>68</b>
D.1. Resumen de Notación . . . . .	68
<b>Bibliografía</b>	<b>69</b>

# Índice de figuras

2.1.	Distribuciones de sensibilidad y especificidad en un escenario supuesto de un test con resultado continuo. . . . .	10
2.2.	Cambios en el punto de corte afectan la sensibilidad y la especificidad.	11
2.3.	Ejemplo de una curva ROC. . . . .	13
2.4.	Ejemplo de funciones de densidad de la población sana y enferma que generan una curva ROC. . . . .	14
2.5.	Ejemplo de funciones de densidad de la población sana y enferma que generan una curva ROC para un test no informativo. . . . .	14
2.6.	Ejemplo de funciones de densidad de la población sana y enferma que generan una curva ROC para un test casi perfecto. . . . .	14
2.7.	Curva ROC para dos test, A y B, donde A es inequívocamente mejor.	16
2.8.	Ejemplo hipotético de datos censurados. . . . .	18
3.1.	Ilustración del funcionamiento de la variable $D_i(t)$ : siendo $t$ un tiempo cualquiera de interés en el estudio, para el individuo A: $D_A(t) = 1$ , pues $T_A \leq t$ , y para el individuo B: $D_B(t) = 0$ , pues $T_B > t$ . . . . .	23
3.2.	Ilustración hipotética de casos y controles en un estudio supervivencia con $TPF^A$ y $FPF^D$ . . . . .	28
3.3.	Ilustración hipotética de casos y controles en un estudio supervivencia con $TPF^I$ y $FPF^E$ . . . . .	29
3.4.	Ilustración hipotética de casos y controles en un estudio supervivencia con $TPF^I$ y $FPF^D$ . . . . .	30
4.1.	Distribución del biomarcador baselina AKI. Se muestra el ajuste con curvas lowess (detalles en el Apéndice C), para el biomarcador AKI medio y severo. . . . .	42
4.2.	Estimaciones de la curva ROC para el biomarcador baselina AKI, tomando como tiempo de interés $t$ un día (en color negro) y dos días (en color rojo) después de la cirugía. . . . .	43
4.3.	AUC usando marcador baselina y distintas taxonomía . . . . .	45
4.4.	Distribución del biomarcador longitudinal en el estudio de AKI medio y severo. Se muestra el ajuste con curvas lowess (detalles en el Apéndice C). . . . .	46

4.5.	Estimaciones de la curva ROC para el biomarcador longitudinal para AKI <b>Severo</b> , tomando como tiempo de interés $t + s$ un día, y tiempo de medición del marcador $s = 0,25$ (en color rojo), $s = 0,50$ (en color verde), y $s = 0,75$ (en color azul), y respectivamente tomando como tiempo de interés $t + s$ dos días para $s = 1,25$ (en color rojo), $s = 1,50$ (en color verde), y $s = 1,75$ (en color azul). . . . .	48
4.6.	Estimaciones de la curva ROC para el biomarcador longitudinal para AKI <b>Medio</b> , tomando como tiempo de interés $t + s$ un día, y tiempo de medición del marcador $s = 0,25$ (en color rojo), $s = 0,50$ (en color verde), y $s = 0,75$ (en color azul), y respectivamente tomando como tiempo de interés $t + s$ dos días para $s = 1,25$ (en color rojo), $s = 1,50$ (en color verde), y $s = 1,75$ (en color azul). . . . .	49
4.7.	Estimaciones de la curva ROC para el biomarcador baselina ACS, tomando como tiempo de predicción de la muerte $t$ un año (en color negro), dos años (en color rojo), tres años (en color verde), cuatro años (en color azul) después presentarse en el paciente el ACS. . . . .	52
4.8.	AUC para datos de Infarto Agudo del Miocardio, tiempo en días. . . . .	54

# Índice de tablas

2.1.	<i>Algunas variaciones de los elementos en el diseño de estudio de clasificación.</i>	5
2.2.	<i>Posibles resultados de una prueba diagnóstica.</i>	6
2.3.	<i>Algunas probabilidades en un estudio diagnóstico.</i>	7
2.4.	<i>Resultados de la prueba diagnóstica binaria en fracciones.</i>	8
2.5.	<i>Resultados de la prueba diagnóstica continua en fracciones, para un valor de corte <math>c</math>.</i>	9
3.1.	<i>Resumen de la taxonomía, para test con resultados B: binario, C:continuo, L: longitudinal (continuo y binario).</i>	25
4.1.	<i>AUC del marcador baselina AKI severo y medio, para tiempo de supervivencia uno y dos días.</i>	44
4.2.	<i>AUC para tiempo de supervivencia uno y dos días y para AKI severo y medio, con distintos tiempos de medición del marcador longitudinal <math>s</math>.</i>	50
4.3.	<i>AUC para tiempo de supervivencia: medio año, uno, dos, y tres años.</i>	53
4.4.	<i>Resumen TPF para tiempo de supervivencia un año, dos, tres y cuatro años, considerando distintas FPE.</i>	55
4.5.	<i>Característica de las variables utilizando <math>ROC_{t_1}^{A/D}</math> y <math>ROC_{t_1}^{I/D}</math> para el grupo de No Diabéticos, cuando se permite un 20 % de falsos positivos</i>	57
4.6.	<i>Característica de las variables utilizando <math>ROC_{t_1}^{A/D}</math> y <math>ROC_{t_1}^{I/D}</math> para el grupo de Diabéticos, cuando se permite un 20 % de falsos positivos</i>	58
A.1.	<i>Descripción y formato de las variables de los datos AKI</i>	61
B.1.	<i>Descripción y de las variables de los datos ACS utilizadas en el modelo</i>	63
B.2.	<i>AIC resultante para diferentes modelos probados.</i>	64

# 1

## Introducción

Las ciencias de la salud no son simplemente un conjunto de teorías y modelos que contestan las preguntas *¿cómo?* y *¿qué?*, su objetivo principal es lograr la “explicación” y contestar el *¿por qué?*. Entonces, son sobre todo, un proceso que permite explicar y explorar un cúmulo de enfermedades. El método utilizado es proponer una hipótesis que sea la posible solución a un problema médico, y comprobar o refutar mediante investigación dicha hipótesis. A tal efecto, la estadística es un instrumento indispensable para la realización de este proceso, y en muchos casos se convierte en la única herramienta que permite obtener resultados en el estudio de problemas médicos que no pueden ser abordados desde una perspectiva de las leyes deterministas.

Uno de los problemas médicos más comunes es el de clasificación o toma de decisiones clínicas. Clasificar a un paciente con la presencia o ausencia de determinado evento o enfermedad se logra a través de las “pruebas diagnóstica” de las cuales es “trascendente” conocer su exactitud o precisión. La curva ROC es una de las herramientas más usadas para evaluar la exactitud de estos tests, aún cuando la variable en la prueba que se desea estudiar sea el tiempo que tarda en ocurrir un suceso. Esta unión de la curvas ROC tradicional con el análisis de supervivencia da origen a las curvas ROC tiempo dependientes. Estas curvas ROC tiene un auge importante por el impacto que supone el permitir a un médico conocer o predecir el plazo de actuación que tiene ya sea ante una situación beneficiosa (curación, alta hospitalaria), ante una situación alternativa (cambio de tratamiento) o incluso hasta el punto relevante de permitir o agregar ese conocimiento en situaciones complejas y delicadas como son las perjudiciales para el paciente (muerte, aparición de efecto adverso, rechazo del trasplante, otros sucesos funestos o devastadores) donde actuar a tiempo hace la diferencia entre la vida y la muerte de un individuo.

Todo esto justifica y **motiva** el seguimiento de los últimos desarrollos en el contexto de un marco teórico compacto, que exponga la clasificación de esta curva

tiempo dependiente y los métodos de estimación propuestos en las últimas publicaciones especializadas a manera de resumen, que facilite el acceso conceptual y amplíe las posibilidades de aplicación y desarrollo a futuro de tan importante herramienta.

En este contexto se persiguen los siguientes **objetivos**:

- Resumir de forma sistemática e integrada el cuerpo de conocimiento que constituyen las ROCs en supervivencia en su estado más reciente.
- Mostrar la taxonomía de las curvas ROCs tiempo dependientes destacando cuándo el uso de alguna en particular es más adecuado.
- Mostrar las posibles relaciones entre los métodos de estimación propuestos clarificando lo que aporta cada método.
- Aplicar las definiciones dentro de la taxonomía conjuntamente con un método de estimación para analizar el comportamiento estadístico de un biomarcador.
- Mostrar los modos de operación ante tipos diferentes de biomarcadores (baselina, longitudinal, procedentes o no de un modelo multivariable)

Las **aportaciones** más significativas de este trabajo son:

- Se expone lo esencial de las curvas ROCs tiempo dependientes concretando taxonomías, métodos de estimación, y posibles biomarcadores.
- Se muestra la aplicabilidad de las curvas ROCs tiempo dependientes para evaluar la capacidad de un marcador a la hora de predecir tiempo de supervivencia de individuos en un caso real.

**Líneas futuras** de investigación:

- Exploración o extensión de los modelos propuestos a otras asunciones que no sea *proportional hazard*, o otros modelos que no sea el de Cox (Cox (1972)). Modelos de supervivencia como el *stratified proportional hazards* o el modelo de *accelerate failure time*, modelos donde se puedan obtener estimadores más consistente de la distribución de supervivencia.
- Proponer nuevos métodos de estimación de las curvas ROCs tiempo dependientes, en conjunto con un estudio de consistencia y convergencia de dichos métodos.
- Desarrollar metodologías para la inferencia a través de la AUC como resumen del funcionamiento de un marcador para predecir el desenlace de eventos temporales.

- Desarrollo de la comparación formal de biomarcadores basada en Áreas Bajo la Curva (AUC) tiempo dependientes.
- Implementación en software libre de las metodologías existentes y de las que se propongan.

**Estructura del trabajo:**

En el capítulo dos interesa dar un marco de referencia básico de corte teórico sobre los estudios de clasificación, las curvas ROCs (relación entre sensibilidad y especificidad), el análisis de supervivencia y finalmente, se hace referencia al modelo de Cox.

El capítulo tres, basándose en el marco conceptual establecido en el capítulo dos, expone y describe la clasificación de las curvas ROCs tiempo dependientes y los métodos de estimación propuestos recientemente, haciendo una comparación en cuanto a los aportes que cada método ofrece, por ejemplo, en cuanto a la conciliación con datos censurados o con los riesgos competitivos.

El cuarto capítulo, recoge las aplicaciones de este trabajo, dividiéndose en dos partes, la primera de ellas aplica las técnicas estadísticas mostradas en los capítulos anteriores a un conjunto de datos simulados sobre Lesión Aguda de Riñones donde destaca por ejemplo, el uso de un marcador longitudinal. En la segunda parte algunas curvas ROCs tiempo dependientes son calculadas para datos reales de Infarto Agudo del Miocardio. En este caso destaca el uso de un biomarcador resultante de un modelo multivariante. Ambas partes van acompañadas de sus respectivas conclusiones estadísticas.

Por último el capítulo 5 expone las conclusiones generales. Y finalmente los anexos incluyen una breve descripción de los datos utilizados, una forma alternativa de obtener por simulación los datos sobre Lesión Aguda de Riñones, un resumen de la notación, y un resumen de las funciones utilizadas del lenguaje R para obtener los resultados mostrados en el capítulo cuatro.

# 2

## Base Teórica

### 2.1. Estudio de Clasificación

Los seres humanos por racionalidad sentimos la necesidad eminente de ordenar, clasificar, y estructurar. En concreto, clasificar todas las cosas que nos rodean resulta importante porque nos permite agrupar y reaccionar ante un determinado grupo con un comportamiento específico o aplicando un principio de optimalidad.

Los primeros sistemas de clasificación se basaban en las semejanzas y diferencias que podían observarse en los organismos. Este enfoque fue aplicado en la medicina por muchos años para clasificar las enfermedades, pero yacía un problema en él, y era la subjetividad en la creación de los grupos.

Todas las clasificaciones que inciden de forma directa en la vida del hombre, como son las clasificaciones dentro del contexto de la medicina, han dejado en un segundo plano la subjetividad intentando agrupar con precisión. Hoy en día la clasificación correcta de las enfermedades, de los individuos como enfermos o sanos, o de una persona dentro de una escala que establece el grado, nivel, o etapa de una enfermedad, es vital. Por lo tanto, la comunidad científica, médica y bioestadística, emplea tiempo en el diseño de todo un *estudio de clasificación*, en donde se evalúan detallada y rigurosamente los elementos que intervienen en el estudio y las posibles formas de clasificación, y se desarrollan "pruebas" que los pacientes pasarán para ser clasificados.

Un estudio de clasificación tiene como principal objetivo evaluar la *precisión de las pruebas*. Básicamente un test será preciso si clasifica a los sujetos correctamente de acuerdo al status de enfermedad que posean. Una evaluación rigurosa de la precisión requiere procedimientos estadísticos para diseñar las pautas a seguir, y propiamente para aplicar la evaluación; comparando la prueba en cuestión con otras pruebas, verificando que cumple ciertos criterios de exploración útil (por ejemplo, pruebas no invasivas), etc.

Es necesario tener claro algunos *elementos claves* para diseñar el estudio de clasificación. La Tabla 2.1 resume algunas variaciones que se suelen encontrar en el diseño de un estudio de clasificación (Pepe (2003)).

Elementos	Posibles Variaciones
Tipo de Resultado del Test	Binario, Ordinal, Continuo, Otro
Selección de los Sujetos	Caso-Control, Corte, Corte con Selección
Comparativa del Diseño	Apareado, No Apareado
Integridad del Test y la Enfermedad	Inherente, Cegadora

**Tabla 2.1:** Algunas variaciones de los elementos en el diseño de estudio de clasificación.

### 2.1.1. Test Diagnóstico

Por lo amplio del contexto médico las pruebas para clasificar a pacientes se utilizan con muchos propósitos aunque estas sean comúnmente llamadas *tests diagnósticos* o *pruebas diagnósticas*. Se definen como procedimientos aplicados con la finalidad de detectar una “condición” médica determinada en un individuo. La definición se refiere a “condición” que bien puede ser una enfermedad, un síndrome, o un proceso patológico, ya que estas pruebas son utilizadas de forma rutinaria para realizar exploración médica, monitorizar la progresión de un individuo, y por supuesto diagnosticar. Solo por simplificación se utilizan indistintamente los términos: condición y enfermedad; enfermo y presencia de la condición; o sano y ausencia de la condición.

La complejidad de estos procedimientos radica en determinar la condición supuestamente patológica en el sujeto, ya que generalmente esta no es susceptible de ser observada directamente (con alguno de los cinco sentidos elementales).

Entonces, se comienza el estudio diagnóstico con la selección de la o las características cuantificables del individuo que pueden resultar de interés en un test determinando, y consecuentemente se procede a la medición de estas variables para poder clasificar al sujeto. Este clasificador, comúnmente llamado también marcador o biomarcador, proporciona un “resultado” o *Yield* en inglés, que se denota por  $Y$ .

Como se mencionó en la Tabla 2.1, el resultado de un test diagnóstico, o marcador, puede ser:

- *Binario*: Cuando solo existen dos posibles resultado, denotados por: Positivo  $P$ , y negativo  $N$  (por ejemplo: test de embarazo).
- *Ordinal*: Cuando el resultado puede tomar distintos valores ordenados siguiendo una escala establecida (por ejemplo: leve, moderado, grave).

- *Continuo*: Cuando el resultado puede ser cualquier valor dentro de un intervalo real especificado (por ejemplo: glicemia, colesterol, hemoglobina).

Por otra parte, la condición real del sujeto o status de la enfermedad (en inglés *Disease*) se denota por  $D$ . Cuando se desconoce la realidad del individuo se aplica un criterio de referencia (o *gold standard*) que se utiliza como criterio de la “verdad”.

Para valorar la validez de una prueba se establece una comparación entre el resultado  $Y$ , y el status real  $D$ . Considere entonces un escenario simple; un test con resultado binario. Esta comparación se logra de forma sencilla a través de una tabla en la que se cruzan los posibles resultados de la prueba (positivos o negativos), frente a la “verdad” valorada por el criterio de referencia (ausencia o presencia real de la enfermedad), como la Tabla 2.2.

		STATUS VERDADERO DE LA ENFERMEDAD (D)		
		Ausente	Presente	
RESULTADO DE LA PRUEBA DIAGNÓSTICA (Y)	Negativa	Verdadero negativo True Negative (TN)	Falso negativo False Negative (FN)	$TN + FN$
	Positiva	Falso positivo False Positive (FP)	Verdadero positivo True Positive (TP)	$FP + TP$
		$TN + FP$	$FN + TP$	

**Tabla 2.2:** Posibles resultados de una prueba diagnóstica.

La evaluación del desempeño de la prueba comienza por la cuantificación (estimación, más bien) de la magnitud de los aciertos, o su inverso, la magnitud de los errores que pueden cometerse. Dos errores son posibles; un resultado falso positivo  $FP$  (clasificar a un individuo sano como enfermo) y un resultado falso negativo  $FN$  (clasificar a un individuo enfermo como sano).

Al tratarse de seres vivos, resulta responsable hacer una estimación en términos de probabilidad, de la incertidumbre que existe en torno a la prueba, que permita responder por ejemplo a: ¿cuál es la probabilidad de que la prueba diagnóstica clasifique erróneamente?.

### 2.1.2. Resultados de una Prueba Diagnóstica en Términos Probabilísticos

Estadísticamente la Tabla 2.2 es una tabla de contingencia que permite calcular probabilidades. Por ejemplo ver la Tabla 2.3.

En especial estas cuatro probabilidades mostradas en la Tabla 2.3 son los cuatro índices de la eficacia de una prueba diagnóstica: sensibilidad, especificidad, valor predictivo positivo y valor predictivo negativo. Ahondemos en las dos primeras características operativas.

Probabilidades	Cálculo
Probabilidad de obtener un resultado negativo cuando el sujeto no tiene la enfermedad	$\frac{TN}{TN+FP}$
Probabilidad de obtener un resultado positivo cuando el sujeto tiene la enfermedad	$\frac{TP}{FN+TP}$
Proporción de resultados válidos entre los resultados negativos de la prueba	$\frac{TN}{TN+FN}$
Proporción de resultados válidos entre los resultados positivos de la prueba	$\frac{TP}{TP+FP}$

**Tabla 2.3:** Algunas probabilidades en un estudio diagnóstico.

### 2.1.3. Características Operativas

Yerushalmy (1947) introduce los términos de sensibilidad y especificidad como indicadores estadísticos que evalúan el grado de eficacia inherente a una prueba diagnóstica (resumen de la precisión diagnóstica).

#### Sensibilidad

La *sensibilidad* es un parámetro que se mide en el grupo de sujetos que verdaderamente están enfermos. Es el cociente entre verdaderos positivos y el total de las personas enfermas. Por tanto, es la probabilidad de obtener un resultado positivo cuando el individuo tiene la enfermedad, o la proporción de verdaderos positivos. Según Pepe (2003) el término proporción puede causar confusión en epidemiología resultando más apropiado sustituirlo por *fracción*. Se denota entonces la sensibilidad como *TPF* (*True Positive Fraction*).

La sensibilidad es especialmente importante cuando una enfermedad no debe pasar desapercibida y cuando el pronóstico mejora mucho con el tratamiento precoz.

#### Especificidad

La *especificidad* es un parámetro que se mide en el grupo de sujetos no enfermos. Es el cociente entre verdaderos negativos y el total de no enfermos. Por tanto, es la probabilidad de obtener un resultado negativo cuando el individuo no tiene la enfermedad. Se denota la especificidad como *TNF* (*True Negative Fraction*).

La especificidad es especialmente importante cuando la enfermedad cambia la vida del enfermo o tiene cierto estigma (no se quiere preocupar inútilmente o estigmatizar erróneamente a alguien con un falso positivo) y también cuando las consecuencias de un tratamiento suponen un riesgo para el enfermo (amputación, etc.).

### 2.1.4. Test Diagnóstico con Resultado Binario

Un test con dos posibles resultados da lugar a la siguiente notación:

Sea  $D$ , la variable binaria que denota el status verdadero de la enfermedad:

$$D = \begin{cases} 1 & \text{para presencia de la enfermedad;} \\ 0 & \text{para ausencia de la enfermedad.} \end{cases}$$

Sea  $Y$ , el resultado del test diagnóstico binario:

$$Y = \begin{cases} 1 & \text{positivo para la enfermedad;} \\ 0 & \text{negativo para la enfermedad.} \end{cases}$$

Ahora, los resultados en fracciones del test se resumen mediante la Tabla 2.4.

	$D = 0$	$D = 1$
$Y = 0$	<p>ESPECIFICIDAD</p> <p>True Negative Fraction (<math>TNF</math>)</p> $TNF = P[Y = 0 D = 0]$	<p>False Negative Fraction (<math>FNF</math>)</p> $FNF = P[Y = 0 D = 1]$
$Y = 1$	<p>False Positive Fraction (<math>FPF</math>)</p> $FPF = P[Y = 1 D = 0]$	<p>SENSIBILIDAD</p> <p>True Positive Fraction (<math>TPF</math>)</p> $TPF = P[Y = 1 D = 1]$

**Tabla 2.4:** Resultados de la prueba diagnóstica binaria en fracciones.

Los subíndices  $D$  y  $\bar{D}$  son usados ocasionalmente, y denotan el status de enfermo y sano respectivamente. Por ejemplo  $Y_D$  denota el resultado del test para un sujeto enfermo.

### 2.1.5. Test Diagnóstico con Resultado Continuo

La realidad suele ser más compleja que los modelos que el hombre busca para representarla. En algunas circunstancias, más de dos categorías pueden ser necesarias para enmarcar la condición de cada paciente, el resultado de una prueba, o de ambos. Ejemplo de estas circunstancias es cuando los resultados de una prueba son de naturaleza cuantitativa u ordinal, o sea, el resultado de realizar el test diagnóstico es un número, un rango, o un nivel (3,4 mmol/L, "ligero", 36 puntos, etc.).

Los valores continuos pueden ser transformados en binarios si se establece un "punto de corte" o umbral ( en inglés *Threshold*) a partir del cual se considerarán los resultados como positivos o negativos para la presencia de la condición (por ejemplo: glicemia mayor a 125 mg/dl) o como categóricos, si se establecen rangos.

Se deben realizar estudios para buscar cuál es el punto del espectro que permite separar a los enfermos de los no enfermos y escoger así el punto de corte

óptimo según las necesidades del estudio. La selección de un punto de corte óptimo es, con este tipo de pruebas, la tarea más importante. Sin embargo, la noción de punto de corte óptimo no es única ya que varía de acuerdo a las circunstancias, Pepe (2003).

Una vez seleccionado el, o los puntos de cortes necesarios, para cada *threshold* (denotado por  $c$ ) se puede obtener un resultado  $Y$  dicotómico:

$$\begin{array}{ll} \text{positivo} & \text{si } Y \geq c, \\ \text{negativo} & \text{si } Y < c, \end{array}$$

note que se asume que valores más grandes de  $Y$  implican mayor anormalidad o más enfermedad (Pepe (2003)).

Ahora, los resultados en fracciones del test, para un valor de corte  $c$ , se resumen mediante la Tabla 2.5.

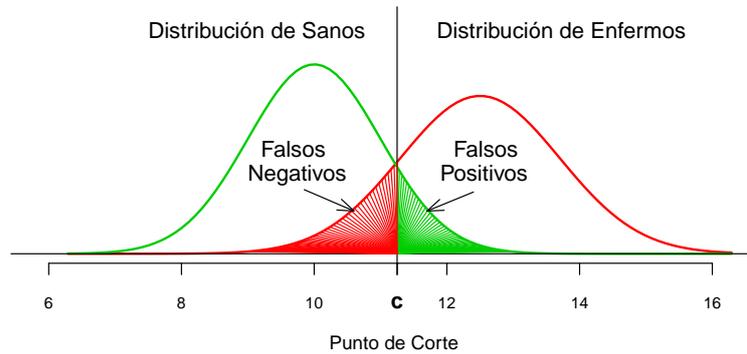
	$D = 0$	$D = 1$
$Y < c$	<p>ESPECIFICIDAD</p> $TNF(c) = P[Y < c   D = 0]$	$FNF(c) = P[Y < c   D = 1]$
$Y \geq c$	$FPF(c) = P[Y \geq c   D = 0]$	<p>SENSIBILIDAD</p> $TPF(c) = P[Y \geq c   D = 1]$

**Tabla 2.5:** Resultados de la prueba diagnóstica continua en fracciones, para un valor de corte  $c$ .

## 2.2. Relación entre Sensibilidad y Especificidad

En ocasiones se requieren pruebas que cumplan las dos exigencias de tener una sensibilidad y especificidad altas, y por esta razón es importante estudiar la relación entre estos dos conceptos.

Suponga un escenario en el que el resultado del test es continuo y ha sido dicotomizado mediante un punto de corte  $c$ . Las personas sanas y las enfermas siguen distribuciones diferentes, originando hipotéticamente las curvas que se muestra en la Figura 2.1. Según el diseño del test los valores obtenidos en general y la media de puntuaciones obtenidas por los sujetos enfermos originan la curva que se muestra a la derecha, y los valores en general conjuntamente con la media de las puntuaciones de los sujetos sanos originan la curva de la izquierda. Se puede ver también que las dos distribuciones se solapan, es decir, hay personas con puntuaciones por encima del punto de corte que no presentan la enfermedad, mientras por el contrario también hay personas con puntuaciones por debajo del mismo punto de corte que si padecen el trastorno. Los primeros serán falsos positivos y los segundos serán falsos negativos.



**Figura 2.1:** Distribuciones de sensibilidad y especificidad en un escenario supuesto de un test con resultado continuo.

Observe en la Figura 2.2 que cuando se desplaza al punto de corte hacia la derecha, es decir, se mueve hacia puntuaciones altas del test se disminuye la probabilidad de hallar un falso positivo y se aumenta así la especificidad de la prueba (a costa de disminuir la sensibilidad). Por el contrario cuando se desplaza el punto de corte hacia puntuaciones bajas, desciende la probabilidad de hallar falsos negativos y con ello se incrementa la sensibilidad de la prueba (pero se disminuye la especificidad). En conclusión, en cualquier prueba diagnóstica el aumento de sensibilidad siempre se hace en detrimento de la especificidad y viceversa.

Por otra parte, intentando describir la precisión del test mediante la probabilidad total de clasificación incorrecta:

$$P[Y \neq D] = \rho FNF + (1 - \rho) FPF,$$

vemos que depende de  $\rho = P[D = 1]$ , la prevalencia (proporción de personas en la población que presentan la enfermedad o el evento de interés), lo que la convierte en un resumen de la precisión diagnóstica inadecuado. Se puede ilustrar este problema si se considera por ejemplo, un test que no aporta información, es decir un test que simplemente clasifica a todos los individuos de una población como sanos con  $Y = 0$ , de modo que  $FPF = 1 - P[Y = 0|D = 0] = 1 - 1 = 0$  y  $FNF = 1 - P[Y = 1|D = 1] = 1 - 0 = 1$ , si la prevalencia es pequeña digamos de 0,001, entonces la probabilidad de mala clasificación también será baja a pesar de que claramente el test sea inútil.

En conclusión, como consecuencia del planteamiento anterior es conveniente reportar conjuntamente la  $FPF$  y  $FNF$  para describir la precisión de un test diagnóstico, y la razón es que ambos errores tienen un costo y unas consecuencias diferentes (Pepe (2003)):

- Cometer un error “Falso Negativo”, es decir, fallar no diagnosticando la enfermedad cuando esta está presente, conlleva a personas enfermas sin

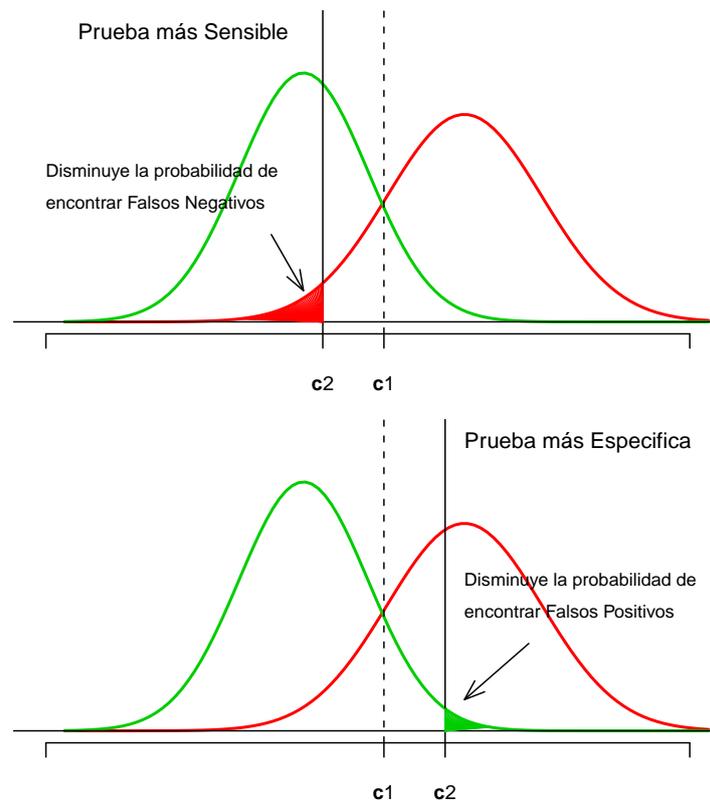


Figura 2.2: Cambios en el punto de corte afectan la sensibilidad y la especificidad.

tratamiento con muchas posibilidades de morir.

- Cometer un error "Falso Positivo", es decir, fallar diagnosticando la enfermedad cuando esta no está presente en el sujeto, conlleva a aplicar un tratamiento a personas sanas. En este caso, por su puesto que existirá un impacto negativo en el paciente pero son consecuencias consideradas menores en relación con la muerte.

Finalmente, Zweig and Campbell (1993) y Pepe (2003) proponen el uso de la  $FPF$  y  $TPF$  (que es igual a  $1 - Especificidad$  y la  $Sensibilidad$ ) para cuantificar la precisión del test.

Se plantea entonces un nuevo gráfico que ayude a identificar rápidamente cuán preciso es el test. Enfrentando la  $FPF$  (eje de las abscisas) contra la  $TPF$  (eje de las ordenadas). Este gráfico tendrá un solo punto  $(FPF, TPF)$  en el primer cuadrante unitario para el caso del test binario. De modo que un test con precisión perfecta será aquel donde  $FPF = 0$  y  $TPF = 1$ . Para el caso de test con resultado continuo se toma en cuenta  $\forall c$  los pares  $(FPF(c), TPF(c))$ . El gráfico obtenido se denomina Curva ROC detallado en la próxima sección.

### 2.3. Curva ROC

Para tests con resultados continuos u ordinales, la curva ROC (*Receiver Operating Characteristic*) es el índice de la exactitud diagnóstica y proporciona un criterio unificador en el proceso de evaluación de la prueba según Zweig (1992), debido a sus diversas aplicaciones:

- Evaluación visual de la exactitud
- Comparación visual entre pruebas
- Análisis estadístico ROC:
  - Evaluación cuantitativa de la exactitud mediante al área bajo la curva ROC
  - Comparación cuantitativa entre pruebas
- Selección de niveles de decisión o umbrales considerando la prevalencia y la relación costo beneficio de los resultados falsos
- Análisis de regresión logística y análisis discriminantes

La curva ROC es un gráfico en el que se observan todos los pares de Sensibilidad y complemento de la Especificidad, resultantes de la variación continua de todos los puntos de corte en todo el rango de resultados observados (Sackett et al. (1989)).

La curva ROC se desarrollo en los años cincuenta como herramienta para el estudio de detección e interpretación de señales de radar (Peterson et al. (1954)). A mediados de los años sesenta el uso de la metodología ROC se extendió a otros campos, como la Medicina y la Psicología (Green and Swets (1966), Swets (1973), Egan (1975)). Lusted (1971) inició su uso dentro del campo de los tests diagnósticos, y fue a partir del libro de Swets and Pickett (1982) cuando su uso se hizo más popular en este campo.

Denotando por  $c$  al punto de corte o umbral de decisión, las características operacionales ( $FPF, TPF$ ) serían:

$$FPF(c) = P[Y \geq c | D = 0],$$

$$TPF(c) = P[Y \geq c | D = 1],$$

y en consecuencia la curva ROC se define como:

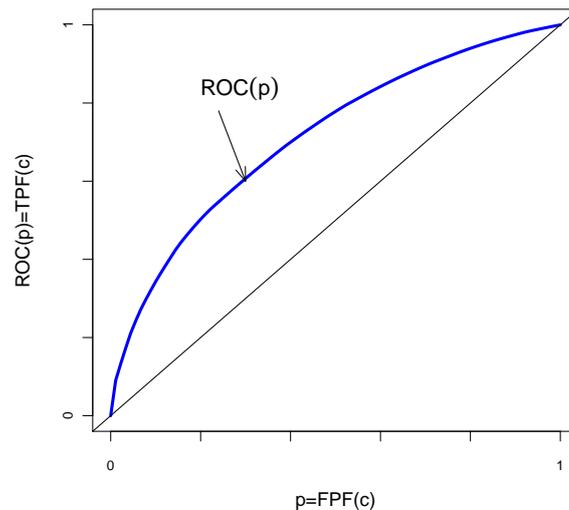
$$ROC(\cdot) = \{(FPF(c), TPF(c)), c \in (-\infty, \infty)\}.$$

Observe que, cuando el *threshold*  $c$  aumenta, tanto  $FPF(c)$  como  $TPF(c)$  disminuyen. Considere  $c$  en un extremo,  $c = \infty$ , tomando  $\lim_{c \rightarrow \infty} TPF(c) = 0$  y  $\lim_{c \rightarrow \infty} FPF(c) = 0$ . Considerando  $c$  en el otro extremo,  $c = -\infty$ , tomando  $\lim_{c \rightarrow -\infty} TPF(c) = 1$  y  $\lim_{c \rightarrow -\infty} FPF(c) = 1$ . Así, la curva ROC es una función

monótona creciente en el cuadrante positivo (ilustrada en la Figura 2.3). Entonces, también se puede escribir como:

$$ROC(\cdot) = \{(p, ROC(p)), p \in (0, 1)\},$$

donde la función  $ROC$  es aquella a la que cada valor de  $p$  le hace corresponder  $TPF(c)$ , siendo  $c$  un punto de corte particular, para el cual  $FPF(c) = p$ , Pepe (2003).

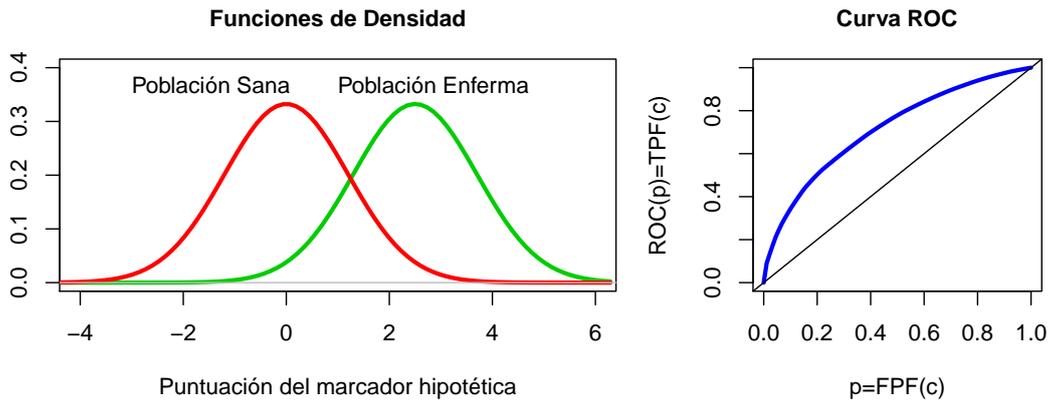


**Figura 2.3:** Ejemplo de una curva ROC.

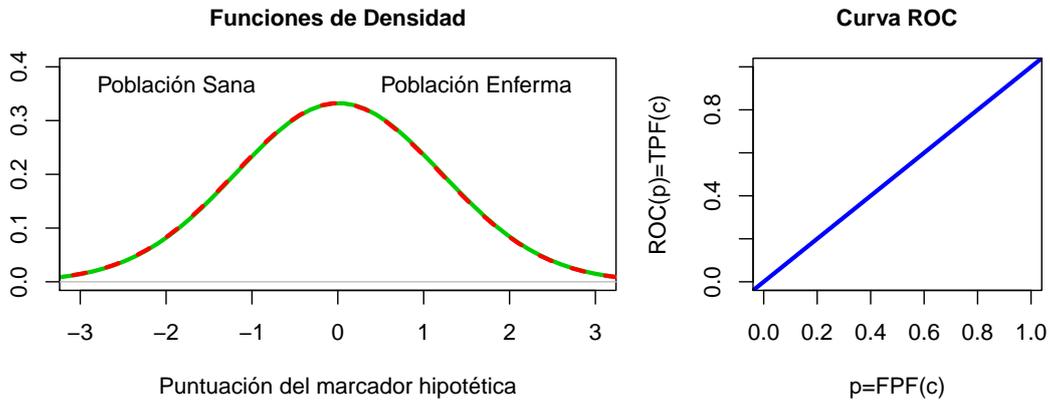
La Curva ROC permite describir que tan separadas están las distribuciones de la sensibilidad y la especificidad de un test, por ejemplo; si las funciones de densidad se cruzan la curva ROC resultante será similar a la mostrada en Figura 2.4, si las distribuciones se solapan casi por completo, entonces la curva ROC está representada por la línea diagonal como se muestra en la Figura 2.5 indicando que el test no es informativo. Si las funciones de densidad están simétricamente distantes como se muestra en la Figura 2.6 entonces la curva ROC será las líneas que forman la esquina superior izquierda, indicando un test perfecto.

En resumen, de acuerdo con Zweig and Campbell (1993), el uso de las curvas ROC en la evaluación de las pruebas diagnósticas presenta las siguientes ventajas: son curvas fáciles de interpretar, no requieren un nivel de decisión particular porque está incluido todo el espectro de puntos de corte, son independientes de la prevalencia, proporciona una comparación visual directa entre pruebas en una escala común.

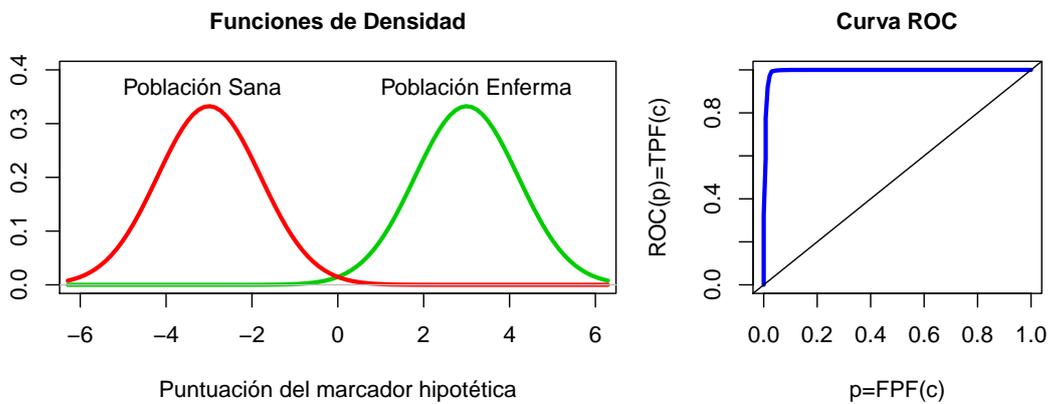
El nombre de curva “Característica Operativa del Receptor” proviene de la noción que podemos obtener de la curva, nosotros -los *receptores* de información-



**Figura 2.4:** Ejemplo de funciones de densidad de la población sana y enferma que generan una curva ROC.



**Figura 2.5:** Ejemplo de funciones de densidad de la población sana y enferma que generan una curva ROC para un test no informativo.



**Figura 2.6:** Ejemplo de funciones de densidad de la población sana y enferma que generan una curva ROC para un test casi perfecto.

que podemos usar (o *operar*) a cualquier punto de la curva usando el valor de corte apropiado. La aplicación clínica determina qué *características* de la prueba son necesarias (Zhou et al. (2002)).

### 2.3.1. El Área Bajo la Curva ROC

El Área Bajo la Curva ROC (*Area Under the Curve, AUC*) estima la capacidad de distinguir o de “discriminar” entre enfermos y no enfermos que tiene una prueba diagnóstica. Se define como:

$$AUC = \int_0^1 ROC(p) dp$$

Si el área bajo la curva valiese 1 (100 %, la curva ROC se dirige hacia la esquina superior izquierda como en la Figura 2.6) la prueba sería perfecta, ya que clasificaría al 100 % de los enfermos como enfermos y al 100 % de los exentos de enfermedad como sanos.

En cambio, si el área bajo la curva valiese 0,5 (50 %, área debajo de la línea diagonal representada en la Figura (2.5)) existiría la misma probabilidad de clasificar a un enfermo como sano que como enfermo. Un área de 0,5 bajo la curva equivale a no discriminar, se interpreta como una prueba “no informativa”. Esta curva ROC sería, de hecho, aquella que tendría valores de sensibilidad iguales a 1-especificidad en todos los posibles puntos de corte. En estas circunstancias se dice que la prueba es “inútil” ya que no reduce nada el grado de incertidumbre previo acerca de si el paciente tiene o no la enfermedad. La diagonal del gráfico que expresa lo que sería una prueba inútil se suele representar junto al área bajo la curva ROC. Al representar la línea diagonal se ve en cuánto supera la prueba que se valora a lo que sería una prueba no informativa.

Un área menor de 0,5 requiere invertir los criterios de diagnósticos: es decir, considerar los negativos como positivos y viceversa. Al invertirlos se comprueba que la prueba discrimina más del 50 %. La construcción de curvas ROC permite además la comparación de varias pruebas diagnósticas.

El área bajo la curva por ser una mediana global implica pérdida de información, por lo que no debe considerarse aisladamente sin examinar la curva. A veces, curvas ROCs de trazados muy distintos pueden tener áreas semejantes (Borgueño et al. (1995)).

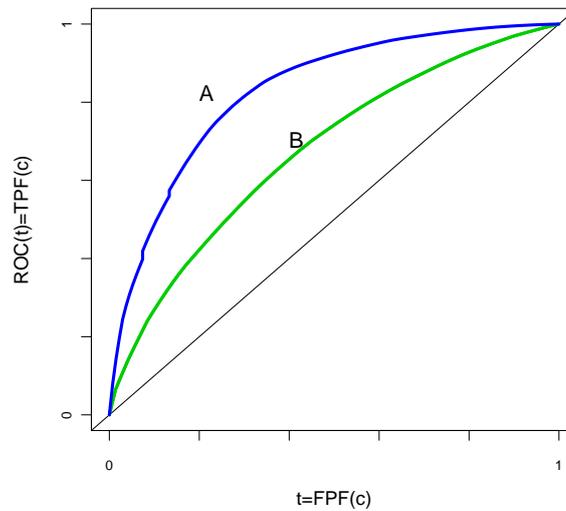
Pepe (2003) muestra que claramente, si dos test están ordenados tal que el test  $A$  es uniformemente mejor que el test  $B$  (ver Figura 2.7) se tiene que

$$ROC_A(p) \geq ROC_B(p), \quad \forall p \in (0, 1),$$

entonces las  $AUC$  estadísticamente también están ordenadas:

$$AUC_A \geq AUC_B$$

pero lo inverso no es necesariamente cierto, es decir, una curva ROC cualquiera con mayor  $AUC$  que otra puede no ser mejor en la situación clínica de interés.



**Figura 2.7:** Curva ROC para dos test, A y B, donde A es inequívocamente mejor.

Entonces solo es posible afirmar que un si una curva ROC es uniformemente mejor que otra tendrá mayor área bajo ella y el test correspondiente será la prueba de elección, debido a su *superior capacidad de discriminación*.

El área bajo la curva ROC tiene una interpretación interesante. Según Bamber (1975) y Hanley and McNeil (1982), se define como la probabilidad de clasificar correctamente un par de individuos sano y enfermo, seleccionados al azar de la población, mediante los resultados obtenidos al aplicar la prueba diagnóstica.

## 2.4. Análisis Supervivencia

El análisis de supervivencia tiene su origen a finales del siglo XVII. La primera referencia conocida es un trabajo de Halley (1693) denominado "*An estimate of the Degrees of the Mortality of Mankind*". A este trabajo debe su nombre pero esta disciplina se ha desarrollado y utilizado en diversas áreas: además de en la medicina, en control de calidad, en marketing, y en estudios que necesiten un largo tiempo de observación.

El análisis de supervivencia es un conjunto de técnicas que permiten estudiar la variable "*tiempo transcurrido hasta que ocurre un evento*", que se denota por  $T$ .

El evento de interés en el contexto médico puede ser: un fallo, la muerte, la curación, desaparición de un síntoma, cambio de un parámetro, etc..

Debido a que la variable tiempo es una variable continua podría ser, en principio, estudiada mediante las técnicas de análisis de la varianza o los modelos de regresión. Hay, sin embargo, dos dificultades importantes para este planteamiento. En primer lugar, en la mayor parte de los estudios citados la variable tiempo

no tiene una distribución normal, más bien suele tener una distribución asimétrica y aunque podrían intentarse transformaciones que la normalizaran. Sin embargo, existe una segunda dificultad que justifica un planteamiento específico para estas variables, y es que para observarlas se tiene que prolongar el estudio durante un período de tiempo suficientemente largo, en el cual suelen ocurrir “pérdidas”, que imposibilitan la observación del evento.

### 2.4.1. Datos Censurados

Existen tres motivos por los que pueden aparecer estas pérdidas, en primer lugar por fin del estudio. Supóngase, por ejemplo, que para evaluar una intervención quirúrgica arriesgada se sigue en el tiempo, durante un año, a dos grupos de pacientes. A los de un grupo se les practicó la intervención y a los de otro no, y se registró la duración del intervalo de tiempo entre la intervención (o la entrada en el estudio, para el grupo no intervenido) y la muerte. Al final del estudio puede haber individuos que no hayan muerto. Otra causa es la pérdida propiamente dicha, por ejemplo se quiere evaluar la eficacia de un tratamiento preventivo para el SIDA, y se sigue durante cinco años a individuos VIH+. Algunos de los individuos, y puede ser un número importante, desaparecerán del estudio en algún momento del mismo por diversos motivos: cambio de domicilio, falta de interés, etc.. Una última causa de pérdida es la ocurrencia de un evento competitivo, puede ser muerte por alguna otra causa ajena al estudio.

La Figura 2.8 muestra un esquema temporal de un estudio para observar tiempos hasta que ocurre un evento, suponga hipotéticamente que se trata de supervivencia en una intervención quirúrgica. Con el círculo se representan las pérdidas o censuras y con la x las muertes (ocurrencia del evento). El individuo *A* desaparece del estudio 3 meses después de la intervención (sería una censura en sentido estricto). El *B* fallece a los 2,5 meses. El *C* sigue vivo al acabar el estudio (sería una censura a los 12 meses por fin del estudio). El *D*, al que se le interviene en el mes 1, fallece en el 9, el tiempo de supervivencia sería 8 meses (hay 1 mes de censura por la izquierda). El *E*, al que se le interviene en el mes 2, se pierde en el 7 (sería una censura a los 5 meses, ya que hay censura en sentido estricto y censura por la izquierda). El *F*, al que se le interviene en el mes 6, sigue vivo al acabar el estudio, sería una censura a los 6 meses (existe censura por fin del estudio y censura por la izquierda).

Si se quisiera aplicar un modelo de regresión lineal a un estudio de este tipo, habría que eliminar del mismo las observaciones perdidas, ya que para ellas no se conoce el valor de la variable; sin embargo sí se tiene alguna información útil sobre la misma: se sabe que es mayor que el tiempo en el que se produjo la censura.

En el análisis de supervivencia, el análisis de los datos puede ser realizado utilizando técnicas paramétricas, semi paramétricas y no paramétricas:

a) *Métodos Paramétricos*: las distribuciones que se utilizan para modelar la supervivencia son; la Exponencial, Weibull y la Lognormal entre otras distribuciones de probabilidad.

b) *Métodos No Paramétricas o Semiparamétricas*: entre las técnicas no paramétricas y semi paramétricas se tienen las Tablas de vida, el método de Kaplan Meier

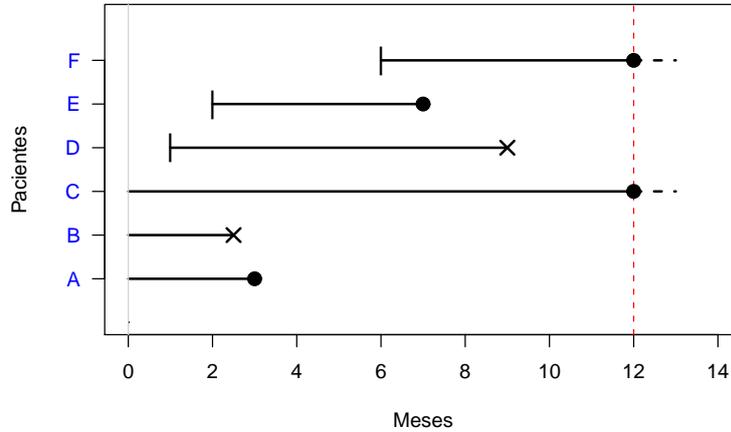


Figura 2.8: Ejemplo hipotético de datos censurados.

(Kaplan and Meier (1958)) y la Regresión de Cox (Cox (1972), Cox and Oakes (1984)).

Entonces, definimos  $T$  como una variable aleatoria no negativa que representa el tiempo hasta la ocurrencia de un evento de los individuos de alguna población.  $T$  puede ser discreto si se mide en días, años, meses, pero usualmente, se asume que  $T$  es continuo, en el intervalo  $[0, \infty)$ .

Sea  $f(t)$  que denota la función de densidad de probabilidad (f.d.p) de  $T$  y sea la función de distribución acumulada

$$F(t) = P(T \leq t) = \int_0^t f(u) du.$$

La probabilidad de que un individuo sobreviva por lo menos hasta el tiempo  $t$  es dado por la *función de supervivencia*.

$$S(t) = P(T \geq t) = \int_t^{\infty} f(u) du.$$

Esta función es continua monótona decreciente con  $S(0) = 1$  y  $S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$ .

Otra función que tiene una gran utilidad en el análisis de supervivencia es la *función de riesgo*. La función de riesgo especifica la tasa instantánea de muerte o falla en un intervalo de tiempo, dado que el individuo sobrevivió hasta  $t$ . En particular,  $\lambda(t)\Delta t$  es la probabilidad aproximada de muerte en  $[t, t + \Delta t)$ , dado que el sujeto ha sobrevivido hasta  $t$ .  $\lambda(t)$ , es definido como:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} = \frac{f(t)}{S(t)}.$$

La función de riesgo es denominada también, razón de riesgo, o razón de falla. Para algunos propósitos es conveniente definir la *función acumulada de riesgo*:

$$\Lambda(t) = \int_0^t \lambda(u) du = -\ln S(t), \quad t \geq 0.$$

El estudio de la censura en el análisis de supervivencia es importante porque si los datos censurados se consideran como no censurados se sesga la variable  $T$  hacia valores inferiores. Sin embargo, es importante tener en cuenta que: una de las condiciones indispensables del análisis tradicional de Kaplan-Meier es que las causas de censura sean *independientes* del evento principal (muerte, recidiva, rechazo, etc.) como por ejemplo, lo sería el paciente que se encuentra bien de la enfermedad de que se trata pero se marcha a vivir a Canadá y, por tanto, desaparece del seguimiento. V. Javier Escrig (2005), expone las siguientes reflexiones: ¿habría independencia de censura en un estudio sobre aparición de rechazo del órgano trasplantado en el caso del que cesa en el seguimiento por fallecer debido a una infección oportunista intercurrente?, ¿o en el caso de un enfermo operado de un tumor con sospecha de recidiva aún no confirmada que fallece por un accidente de tráfico?. Parece ser que en casos así y en otros de similares características *no hay tal independencia*, es más, parece ser que es común en medicina un análisis de supervivencia en el que aparezcan algunas de estas causas de censura que suponen un riesgo alternativo, o en cierta forma complementario, a lo que es el riesgo objeto del estudio. A estos riesgos que compiten con el principal en la limitación del tiempo de supervivencia de un sujeto se les denomina *riesgos competitivos*, como se mencionó anteriormente, y su consideración como una causa independiente de censura, tal como hace el método Kaplan-Meier tradicional, suele llevar a resultados sesgados.

Por otra parte, el tiempo hasta que ocurre un evento puede estar relacionado con múltiples factores o variables pronóstico. La significación y ponderación de todas estas variables puede estudiarse simultáneamente mediante el *método de riesgos proporcionales de Cox* (Cox (1972)).

#### 2.4.2. El Modelo de Cox

Se le llama modelo de riesgos proporcionales (en inglés: *proportional hazard model*), porque el modelo presupone, en primer lugar, una relación multiplicativa entre la función arbitraria de riesgo y la función log-lineal de las covariables. Este es el denominado supuesto de proporcionalidad según el cual el ratio de *hazard* es constante, para todo par de observaciones, en cualquier instante  $t$ . Asimismo, presupone un efecto log-lineal de las covariables sobre la función *hazard*, es decir que las covariables actúan sobre la función *hazard* de forma multiplicativa.

Una ventaja fundamental de la formulación de Cox radica en el hecho que puede realizarse una inferencia de los efectos de la información concomitante sin conocer la forma de la distribución de supervivencia.

Sea  $x_i$  las covariables, y  $\beta_i$  los coeficientes o parámetros desconocidos correspondientes a cada  $x_i$ , este método semiparamétrico (Cox (1972)) se define:

$$\ln(\lambda(t|x)) = a_t + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p,$$

por lo tanto se tiene que

$$\lambda(t|x) = \lambda_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p),$$

o resumiendo, si  $\beta$  es el vector de parámetros y  $x$  el vector de covariables:

$$\lambda(t|x) = \lambda_0(t) \exp(\beta x).$$

Se puede ver cómo el primer factor es una función de riesgo,  $\lambda_0$ , común a todos los individuos, y a la que no se le pone ninguna restricción. Esta es la parte no paramétrica del modelo, o sea,  $\lambda_0(t)$  es el riesgo cuando todas las variables  $x_i$  son 0, o *riesgo basal*, que es variable con el tiempo.

El modelo especifica el efecto de las covariables sobre el riesgo. El riesgo condicional se factoriza como producto de dos factores: uno que depende solo del tiempo, y otro que depende solo de las covariables. El modelo especifica para cada covariable un riesgo relativo constante en el tiempo:

$$\frac{\lambda(t|x)}{\lambda(t|x^*)} = \exp[\beta^t(x - x_*)],$$

si  $x = x^* + \mathbf{1}_j$ , entonces

$$\frac{\lambda(t|x)}{\lambda(t|x^*)} = \exp[\beta_j],$$

luego  $\exp[\beta_j]$  se interpreta como el efecto (constante) en el riesgo que tiene un incremento unitario de la  $j$ -ésima covariable.

Bajo el modelo de Cox, el riesgo acumulado se define:

$$\Lambda(t|x) = \int_0^t \lambda(u|x) du = \exp[\beta x] \Lambda_0(t),$$

donde  $\Lambda_0(t) = \int_0^t \lambda_0(u|x) du$ , es el riesgo acumulativo basal y la función de supervivencia se define como:

$$S(t|x) = \exp[-\Lambda(t|x)] = S_0^{\exp[\beta x]},$$

donde  $S_0(t) = \exp[-\Lambda(t)]$  es la supervivencia "baselina" o basal.

Para la estimación de un modelo de riesgos proporcionales se podría postular para la función de riesgo basal un modelo paramétrico (por ejemplo Weibull) o una estimación semiparamétrica donde se procede a estimar los parámetros maximizando la verosimilitud parcial.

Puede que las covariables en un modelo de Cox también dependan del tiempo, con lo cual la especificación del modelo sería:

$$\lambda(t|x(t)) = \lambda_0(t) \exp[\beta^t x(t)] = \lambda_0(t) \exp[\beta_1 x_1(t) + \dots + \beta_p x_p(t)],$$

de donde se tiene que:

$$\frac{\lambda(t|x(t))}{\lambda(t|\tilde{x}(t))} = \exp [\beta(x(t) - x^*(t))],$$

si  $x(t) = x^*(t) + \mathbf{1}_j$  entonces

$$\frac{\lambda(t|x(t))}{\lambda(t|x^*(t))} = \exp [\beta_j],$$

luego  $\exp [\beta_j]$  se interpreta como el efecto (no constante) en el riesgo que tiene un incremento unitario de la  $j$ -ésima covariable (en tiempo  $t$ ).

# 3

## Curva ROC para Datos de Supervivencia

Este capítulo pretende mostrar la forma como se adecúan los conceptos relacionados con la curva ROC al análisis de supervivencia, y para ello se comienza exponiendo como influye el tiempo en las variables hasta ahora definidas. En la segunda sección se describe como el reconocimientos de casos y controles en el tiempo da lugar a una taxonomía de las extensiones tiempo dependientes de la sensibilidad y especificidad, la combinación de esta taxonomía genera a su vez curvas ROCs para datos de supervivencia que son expuestas en la tercera sección, y por último se presenta un resumen de los enfoques para estimar los parámetros de funcionamiento de los biomarcadores, y se comenta a manera de revisión comparativa las aportaciones de esos métodos.

### 3.1. Incorporando la Dimensión del Tiempo

En el análisis de supervivencia se definió  $T_i$  como la variable tiempo hasta la ocurrencia de un evento para el sujeto  $i$ . El tiempo de supervivencia  $T_i$  puede también ser representado como un proceso continuo, como lo expone Heagerty and Zheng (2005), sin embargo a efectos de esta memoria resulta más didáctico el uso de la primera definición.

Previamente se definió también la variable status real de la enfermedad como:

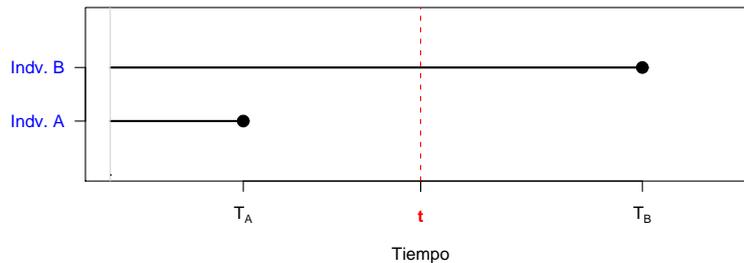
$$D = \begin{cases} 1 & \text{para presencia de la enfermedad;} \\ 0 & \text{para ausencia de la enfermedad.} \end{cases}$$

Heagerty et al. (2000) definió nuevamente esta variable para datos de supervivencia, de modo que extendió esta variable originalmente binaria a una variable binaria variante en el tiempo, renombrandola como status de la falla. Por ejemplo, si estamos interesados en evaluar el comportamiento de un marcador a la hora de discriminar entre aquellos individuos que presentan el fallo antes y después del

tiempo  $t$ , definiremos una nueva variable binaria que representa el “status de la falla” del individuo en el tiempo  $t$  de la siguiente manera:

$$D_i(t) = \begin{cases} 1 & \text{si } T_i \leq t, \\ 0 & \text{si } T_i > t. \end{cases}$$

Esta variable puede ser fácilmente ilustrada, como se muestra en la Figura 3.1.



**Figura 3.1:** Ilustración del funcionamiento de la variable  $D_i(t)$ : siendo  $t$  un tiempo cualquiera de interés en el estudio, para el individuo A:  $D_A(t) = 1$ , pues  $T_A \leq t$ , y para el individuo B:  $D_B(t) = 0$ , pues  $T_B > t$ .

Como se mencionó anteriormente en la Tabla 2.1, los sujetos en el estudio pueden ser casos o controles, que Pepe (2003) define como: sujetos que presentan la enfermedad y sujetos que no la presentan respectivamente. Para el contexto de curva ROC en el análisis de supervivencia la selección de casos y controles es sumamente importante porque definen las posibles extensiones de los conceptos de sensibilidad y especificidad hasta ahora vistos.

## 3.2. Taxonomía, Exenciones de Sensibilidad y Especificidad

Heagerty and Zheng (2005) propone las siguientes definiciones de casos y controles dependientes del tiempo y su clasificación, considerando  $t$  un tiempo cualquiera que resulte de interés en el estudio:

- **Casos:** Individuos que presentan el fallo. Utilizados para definir la sensibilidad.
  - *Casos Incidentes (I):* Individuos donde  $T_i = t$ , y por lo tanto se considera que presentan el fallo en el momento  $t$ .
  - *Casos Acumulativos (A):* Individuos donde  $T_i \leq t$ , y por lo tanto se considera que presentan el fallo antes o justo en el momento  $t$  de interés.

- **Controles:** Individuos que no presentan el fallo. Utilizados para definir la especificidad.
  - *Controles Estáticos (E):* Individuos donde  $T_i > t^*$ , con  $t^*$  es un valor fijo por lo general alto, y por lo tanto se considera que no presentan el fallo (supervivientes a largo plazo).
  - *Controles Dinámicos (D):* Individuos donde  $T_i > t$  y por lo tanto se considera que no presentan el fallo antes del tiempo  $t$ .

A continuación, se utilizarán los super índices  $\mathbb{A}$  (casos acumulativos) e  $\mathbb{I}$  (casos incidentes), para las definiciones de sensibilidad, y los super índices  $\mathbb{D}$  (controles dinámicos) y  $\mathbb{E}$  (controles estáticos) para las definiciones de especificidad. Además, en las definiciones de sensibilidad y especificidad, en vez de usar la variable  $D_i(t)$  se usa su definición expresada en tiempo de supervivencia, por ejemplo en vez de usar  $D_i(t) = 1$  se utilizará  $T_i \leq t$ . Por último, se utiliza  $Y_i$  para denotar el resultado de la prueba diagnóstica, que puede ser binario o continuo como se ha definido previamente, y además longitudinal. Un marcador longitudinal es aquel cuyas medidas son tomadas más de una vez para un mismo individuo durante un periodo de tiempo, y se denota cada medida como  $s$ , resultante por ejemplo de los estudios clínicos longitudinales.

La Tabla 3.1 resume la taxonomía propuesta por Heagerty and Zheng (2005), y extiende la notación para el caso de marcadores discretos y longitudinales

### 3.2.1. Algunos Comentarios Importantes Sobre la Taxonomía

#### En Cuanto a la TPF

Pepe et al. (2008) comenta que si en un estudio determinado existen factores de riesgos, es decir, que las personas que mueren antes probablemente tenían más riesgo que las personas que mueren en tiempos tardíos, entonces la fracción de verdaderos positivos  $TPF$  puede verse como una *función decreciente en el tiempo*.

Además, Pepe et al. (2008) hace la siguiente reflexión: si se supone que un marcador binario se mide en un tiempo  $s$ , como en los estudios longitudinales o cuando no hay un tiempo natural basal, entonces la  $TPF_t^{\mathbb{I}}$  se escribe:  $TPF_t^{\mathbb{I}} = P(Y(s) = 1|T = t + s)$  que representa la sensibilidad del marcador para eventos que ocurren  $t$  unidades después de que  $Y$  es medido, como se vio en la sección anterior. En algunas situaciones esta sensibilidad *podría depender no solo del tiempo,  $t$ , sino también del tiempo absoluto de medición  $s$* . Por lo tanto, se puede escribir como:

$$TPF_t^{\mathbb{I}}(s) = P(Y(s) = 1|T = t + s).$$

Por otra parte, Pepe et al. (2008) también comenta que la  $TPF_t^{\mathbb{I}}$ , es la versión adoptada por la mayoría de investigadores (Etzioni et al. (1999); Zheng and Heagerty (2004); Heagerty and Zheng (2005); Cai et al. (2006)). Aunque la versión alternativa, la  $TPF^{\mathbb{A}}$ , tiene la *ventaja* de evaluar la sensibilidad para eventos que ocurren en el intervalo de tiempo  $(s, s + t]$  en contraposición a eventos que ocurren  $t$  unidades de tiempo después de que  $Y$  es medido. Esta definición es

SENSIBILIDAD ( $TPF$ )	
<b>Incidente</b>	
$B$	sensibilidad <sup>I</sup> ( $t$ ) : $P(Y_i = 1 T_i = t)$
$C$	sensibilidad <sup>I</sup> ( $c, t$ ) : $P(Y_i > c T_i = t)$
$L_C$	sensibilidad <sup>I</sup> ( $c, t, s$ ) : $P(Y_i(s) > c T_i = t + s)$
$L_B$	sensibilidad <sup>I</sup> ( $t, s$ ) : $P(Y_i(s) = 1 T_i = t + s)$
<b>Acumulativa</b>	
$B$	sensibilidad <sup>A</sup> ( $t$ ) : $P(Y_i = 1 T_i \leq t)$
$C$	sensibilidad <sup>A</sup> ( $c, t$ ) : $P(Y_i > c T_i \leq t)$
$L_C$	sensibilidad <sup>A</sup> ( $c, t, s$ ) : $P(Y_i(s) > c T_i \leq t + s)$
$L_B$	sensibilidad <sup>A</sup> ( $t, s$ ) : $P(Y_i(s) = 1 T_i \leq t + s)$
ESPECIFICIDAD ( $TNF = 1 - FPF$ )	
<b>Estática</b>	
$B$	especificidad <sup>E</sup> ( $t^*$ ) : $P(Y_i = 0 T_i > t^*)$
$C$	especificidad <sup>E</sup> ( $c, t^*$ ) : $P(Y_i \leq c T_i > t^*)$
$L_C$	especificidad <sup>E</sup> ( $c, t^*, s$ ) : $P(Y_i(s) \leq c T_i > t^* + s)$
$L_B$	especificidad <sup>E</sup> ( $t^*, s$ ) : $P(Y_i(s) = 0 T_i > t^* + s)$
<b>Dinámica</b>	
$B$	especificidad <sup>D</sup> ( $t$ ) : $P(Y_i = 0 T_i > t)$
$C$	especificidad <sup>D</sup> ( $c, t$ ) : $P(Y_i \leq c T_i > t)$
$L_C$	especificidad <sup>D</sup> ( $c, t, s$ ) : $P(Y_i(s) \leq c T_i > t + s)$
$L_B$	especificidad <sup>D</sup> ( $t, s$ ) : $P(Y_i(s) = 0 T_i > t + s)$

**Tabla 3.1:** Resumen de la taxonomía, para test con resultados  $B$ : binario,  $C$ :continuo,  $L$ : longitudinal (continuo y binario).

muy utilizada en parte porque es *fácil de estimar* empíricamente. Especialmente en datos “no censurados” una estimación simple es la proporción de sujetos que han presentado el evento en el intervalo de tiempo y que tienen valores del marcador positivo. Al respecto de la estimación de la  $TPF$ , varios estimadores que consideran o “incluyen la censura” de forma adecuada han sido desarrollados (Heagerty et al. (2000); DeLong et al. (1985); Parker and DeLong (2003); Zheng and Heagerty (2007); Song and Zhou (2008)). Cai et al. (2006) se enfoca en la  $TPF^I$  pero señala

que la  $TPF^A$  puede ser de interés clínico si se calcula como

$$TPF_t^A(s) = \int (TPF_u^I(s) dF_T(u)) / F_T(t+s) - F_T(s)$$

donde  $F_T$  es la distribución acumulada del evento en el tiempo. Por otra parte, estimar la sensibilidad incidente en base a la  $TPF^A$  es una estimación más difícil por el hecho de que diferenciar es más difícil que integrar. Además la  $TPF^A$  *no distingue* entre la sensibilidad de los eventos que ocurren temprano versus los que ocurren tardíos porque agrupa todos los eventos juntos en  $(s, s+t]$ . Más aún, una serie de  $TPF_s^A$  indexadas por  $t$  muestra información redundante en el sentido de que  $TPF_{t_2}^A(s)$  incluye eventos en  $TPF_{t_1}^A(s)$  si  $t_1 < t_2$ , mientras que en una serie de  $TPF_s^I$  incidentes cada una mostrará esencialmente *información diferente*. En algunos estudios de exploración médica la ocurrencia exacta del evento en el tiempo es desconocido pero en su lugar son conocidas versiones de intervalos censurados donde la  $TPF^I$  es fácilmente calculada como una proporción (Pepe et al. (2008)).

### En Cuanto a la FPF

Otra observación interesante es realizada por Pepe et al. (2008) en cuanto a la  $FPF^E$ . En concreto, hay situaciones donde el grupo de control surge de una manera natural y este puede ser fácilmente identificado. Sin embargo, definir los controles en un estudio resulta problemático cuando por ejemplo, todos los sujetos eventualmente tendrán el evento de interés; un estudio donde el interés se centre en el tiempo hasta que un individuo muere no tiene un grupo de control natural puesto que en algún momento, sea por una enfermedad, accidente, etc., todos morimos. En tales situaciones, es posible identificar a los controles si se elige un punto de marcación en el tiempo lejano,  $t^*$ , y se define a los controles estáticos como los sujetos con  $T > t^*$ , como se indico en la selección anterior. La *elección optima* de  $t^*$  puede depender del contexto. Por ejemplo, si la intención es monitorizar a los individuos en intervalos de tiempo de longitud  $\delta$ , donde una determinada cirugía será adecuada si se realiza en el tiempo  $\delta$  antes de que ocurra el evento, entonces la elección de  $t^* = 2\delta$  sería suficiente. De modo que, sujetos con  $T > 2\delta$  conformarían el grupo de control. Por otra parte, limitaciones que puedan tener algunos datos podrían a su vez limitar la posibilidad de elegir  $t^*$ . Por ejemplo, si en otra situación se toma  $t^* = 2$  años, porque pocos sujetos fueron seguidos durante más de dos años y medio, entonces los controles resultante se describen mejor como un grupo de referencia para comparar sujetos con eventos de ocurrencia temprana, en lugar del verdadero grupo de control.

Sin embargo, existe una alternativa que permite que la  $FPF$  varíe con el tiempo transcurrido,  $t$ . La  $FPF^D$ , presenta también una desventaja en el sentido de que puede *sesgar la información del biomarcador* (Pepe et al. (2008)). Considérese aquellos sujetos cuyo evento ocurrió poco después del tiempo transcurrido  $t$ , ellos son tomados como controles dinámicos para el tiempo  $t$ . Si para tales sujetos el test ha dado un resultado positivo, y están siendo considerados como controles o individuos sanos, entonces la situación imputa al biomarcador contando el resultado como

un falso positivo. Tal vez sería mejor tener en cuenta la capacidad de la prueba de señalar los eventos futuros.

Otro problema práctico con las  $FPF_s^D$  es que debido a que los grupo de control varían con el tiempo, lo hace también el eje de las abscisas de la curva ROC dependiente del tiempo. Por lo tanto resulta más difícil *interpretar tendencias* en el tiempo en estas curvas ROCs (Pepe et al. (2008)). Con un grupo de control estático, las tendencias sobre el tiempo en la curva ROC están relacionadas a las tendencias en la detección de los eventos. Sin embargo, cuando el grupo de control es dinámico las tendencias pueden ser debido a una combinación entre los cambios en los grupos de controles y los cambios en las propiedades de detecciones eventos. De hecho, se considera que cuando se usan  $FPF_s^D$ , aún cuando la  $TPF$  está asociada con un *threshold* específico,  $I(Y(s) \geq c)$  que es constante en el tiempo, la curva ROC empezará a incrementarse para valores altos de  $t$  tanto como sujetos caigan el grupo de control (Pepe et al. (2008)).

Por otra parte, es importante destacar que el marcador en el momento  $s$ ,  $Y(s)$ , puede ser una función de marcadores históricos hasta el tiempo  $s$ , y no es necesariamente el valor de una sola medición puntual en  $s$ . Además, la distribución de  $Y(s)$  puede variar con  $s$ , y por lo tanto el umbral  $c(s)$  puede entonces depender de  $s$ . En algunas aplicaciones la discriminación lograda con el marcador depende de la escala del tiempo absoluto  $s$  así como el tiempo  $t$  (Pepe et al. (2008)).

### 3.3. Extensiones de la Definición de Curva ROC

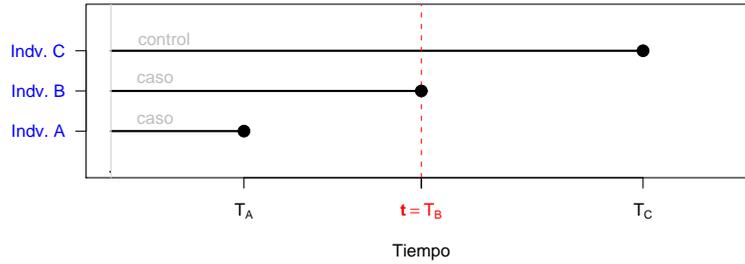
Las extensiones de la definición de la curva ROC para eventos dependientes del tiempo surgen por la elección de un tipo de casos (incidentes o acumulativo) y un tipo de controles (estáticos o dinámicos) a ser considerados en el estudio, veamos a continuación tres posibles exenciones de curva ROC: la Acumulativa Dinámica, la Incidente Estática, y la Incidente Dinámica.

#### 3.3.1. Acumulativo-Dinámico

*Esta definición la adopta* Heagerty et al. (2000) que habla de las particularidades que se presentan al combinar una sensibilidad acumulada y una especificidad dinámica.

Usando este enfoque, en cualquier tiempo fijo  $t$  considerado sobre la línea vital del tiempo, el conjunto de *individuos estudiados puede ser clasificado* o bien como un caso o como un control. Lo anterior es posible porque cada individuo desempeña el rol de un control para un tiempo  $t < T_i$  pero luego contribuye como un caso para tiempos  $t \geq T_i$ . Un ejemplo de la consideración de casos y controles en un tiempo específico  $t$  se muestra en la Figura 3.2, el individuo  $A$  es un caso pues su tiempo de supervivencia  $T_A$  es menor que  $t$ , también lo es el individuo  $B$  pues  $T_B = t$  y individuo  $C$  ya pasa a ser un control porque  $T_C > t$ .

*Definición matemática:* para un marcador continuo, sea  $c_p$  un punto de corte, y



**Figura 3.2:** Ilustración hipotética de casos y controles en un estudio supervivencia con  $TPF^A$  y  $FPP^D$ .

$p$  la fracción de falsos positivos dinámica;

$$\begin{aligned} FPP_t^D(c_p) &= 1 - \text{especificidad}^D(t, c_p) \\ &= P(Y_i > c_p | T_i > t) \\ &= p \end{aligned}$$

La curva  $ROC_t^{A/D}(p)$  es la fracción de verdaderos positivos acumulativa que se obtiene usando el punto de corte  $c_p$

$$\begin{aligned} ROC_t^{A/D}(p) &= TPF_t^A(c_p) \\ &= \text{sensibilidad}^A(t, c_p) \\ &= P(Y_i > c_p | T_i \leq t), \end{aligned}$$

que puede escribirse como la composición de  $TPF_t^A(c_p)$  y la función inversa  $[FPP_t^D]^{-1}(p) = c_p$ ;

$$ROC_t^{A/D}(p) = TPF_t^A\{[FPP_t^D]^{-1}(p)\}, \quad p \in [0, 1],$$

donde  $[FPP_t^D]^{-1}(p) = \inf\{c | FPP_t^D(c) \leq p\}$

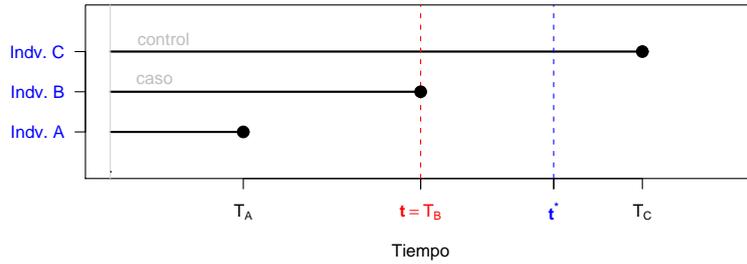
Estas definiciones de sensibilidad y especificidad como resumen de la “precisión acumulativa/dinámica”, son las *más adecuadas cuando* es importante un tiempo específico  $t'$  (o una pequeña colección de tiempos  $t'_1, t'_2, \dots, t'_m$ .) o cuando el interés científico radica en la discriminación entre los sujetos que mueren antes de un determinado tiempo  $t$  y aquellos que sobreviven más allá de un tiempo  $t$ .

### 3.3.2. Incidente-Estático

Esta definición la adopta Etzioni et al. (1999) y Slate and Turnbull (2000) como alternativa de la sensibilidad y especificidad tiempo-dependiente.

Usando las definiciones incidente estática, el conjunto de *individuos estudiados* puede ser clasificado como un caso o como un control, sin cambiar el status de su

enfermedad. Los casos son estratificados según el momento en el que el evento ocurre (incidente) y los controles son definidos como aquellos sujetos libres de evento en todo el periodo de seguimiento  $(0, t^*)$  (estático). Un ejemplo de la consideración de casos y controles se muestra en la Figura 3.3. El individuo  $B$  figura como un caso pues su tiempo de supervivencia  $T_B$  coincide con el tiempo de interés  $t$ , y el individuo  $C$  es considerado un control pues en todo el periodo estático  $(0, t^*)$  no presenta el evento, o lo que es igual  $T_C > t^*$ , el individuo  $A$  no es considerado como control obviamente, y será un caso cuando la incidencia se observe en  $T_A$ .



**Figura 3.3:** Ilustración hipotética de casos y controles en un estudio de supervivencia con  $TPF^I$  y  $FPF^E$ .

*Definición matemática:* para un marcador continuo, sea  $c_p$  un punto de corte, y  $p$  la fracción de falsos positivos estática;

$$\begin{aligned} FPF_{t^*}^E(c_p) &= 1 - \text{especificidad}^E(t, c_p) \\ &= P(Y_i > c_p | T_i > t^*) \\ &= p \end{aligned}$$

La curva  $ROC_t^{I/E}(p)$  es la fracción de verdaderos positivos incidente que se obtiene usando el punto de corte  $c_p$

$$\begin{aligned} ROC_t^{I/E}(p) &= TPF_t^I(c_p) \\ &= \text{sensibilidad}^I(t, c_p) \\ &= P(Y_i > c_p | T_i = t), \end{aligned}$$

que puede escribirse como la composición de  $TPF_t^I(c_p)$  y la función inversa  $[FPF_{t^*}^E]^{-1}(p) = c_p$ ;

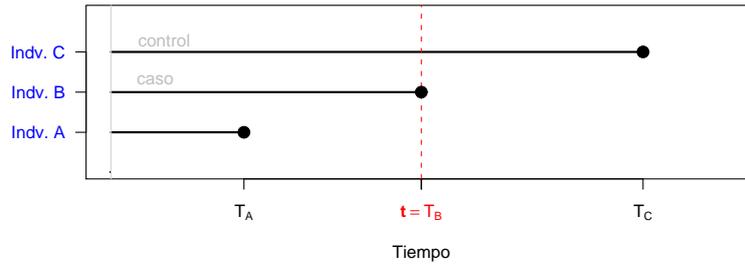
$$ROC_t^{I/E}(p) = TPF_t^I\{[FPF_{t^*}^E]^{-1}(p)\}, \quad p \in [0, 1],$$

donde  $[FPF_{t^*}^E]^{-1}(p) = \inf\{c | FPF_{t^*}^E(c) \leq p\}$

Estas definiciones facilitan el uso de aproximaciones de regresión estándar para caracterizar la sensibilidad y especificidad porque el evento en el tiempo,  $T_i$ , puede ser usado como una covariable.

### 3.3.3. Incidente-Dinámico

Usando este enfoque un sujeto puede desempeñar el rol de control para un tiempo cercano,  $t < T_i$ , pero luego desempeñar el papel de caso cuando  $t = T_i$ , como se muestra por ejemplo en la Figura 3.4 donde el individuo  $B$  es un caso y el individuo  $C$  es un control.



**Figura 3.4:** Ilustración hipotética de casos y controles en un estudio supervivencia con  $TPF^I$  y  $FPF^D$ .

La sensibilidad incidente y la especificidad dinámica son definidas por dicotomizar el riesgo determinado en el tiempo  $t$  entre aquellos en los que se observó la muerte (casos) y aquellos en los que se observó la supervivencia (controles). Estas definiciones incidentes-dinámicas tienen algunas características atractivas en relación con las definiciones alternativas.

Primero, la sensibilidad incidente y la especificidad dinámica son basadas en la clasificación del conjunto de riesgos en el tiempo  $t$  en casos y controles, siendo por tanto un acompañante natural del modelo *hazard*.

Segundo, las definiciones permiten fácilmente la extensión tiempo dependiente con marcadores longitudinales usando  $P(Y_i(s) > c | T_i = t + s)$  para definir la sensibilidad incidente y  $P(Y_i(s) \leq c | T_i \geq t + s)$  para definir la especificidad dinámica con un marcador longitudinal  $Y_i(s)$ .

Por otra parte, el uso de la sensibilidad incidente y la especificidad dinámica permiten tanto resúmenes de precisión en tiempos específicos, como también resúmenes en tiempos promedios (Heagerty and Zheng (2005)). En contraste, los enfoques incidente estático o acumulativo dinámico no proponen métodos para obtener resúmenes de precisión promedios significativos.

*Definición matemática:* para un marcador continuo, sea  $c_p$  un punto de corte, y  $p$  la fracción de falsos positivos dinámica;

$$\begin{aligned} FPF_t^D(c_p) &= 1 - \text{especificidad}^D(t, c_p) \\ &= P(Y_i > c_p | T_i > t) \\ &= p \end{aligned}$$

La curva  $ROC_t^{\mathbb{I}/\mathbb{D}}(p)$  es la fracción de verdaderos positivos incidente que se obtiene usando el punto de corte  $c_p$

$$\begin{aligned} ROC_t^{\mathbb{I}/\mathbb{D}}(p) &= TPF_t^{\mathbb{I}}(c_p) \\ &= \text{sensibilidad}^{\mathbb{I}}(t, c_p) \\ &= P(Y_i > c_p | T_i = t), \end{aligned}$$

que puede escribirse como la composición de  $TPF_t^{\mathbb{I}}(c_p)$  y la función inversa  $[FPF_t^{\mathbb{D}}]^{-1}(p) = c_p$ ;

$$ROC_t^{\mathbb{I}/\mathbb{D}}(p) = TPF_t^{\mathbb{I}}\{[FPF_t^{\mathbb{D}}]^{-1}(p)\}, \quad p \in [0, 1],$$

donde  $[FPF_t^{\mathbb{D}}]^{-1}(p) = \inf\{c | FPF_t^{\mathbb{D}}(c) \leq p\}$

### 3.4. Censura y Eventos con Riesgos Competitivos

La censura es a menudo, pero no siempre, un problema en estudios prospectivos cuyo resultado es un evento en el tiempo. La censura es un corte en los datos que no debería impactar en las definiciones de  $TPF$  y  $FPF$ . Pero, según Wang et al. (2006) el simple hecho de incluir todos los sujetos que no presenten el evento de interés, aun por censura, en el cálculo de la  $FPF$  es defectuoso porque esos sujetos censurados incluidos pueden estar contaminando el grupo de control. Entonces, ha de tenerse en cuenta la forma adecuada de incorporar la censura a los estudios de supervivencia.

Por otra parte, eventos que presentan riesgos competitivos son un fenómeno real que ocurre en la población y por lo tanto impacta sobre las definiciones de las  $TPF$  y las  $FPF$ . Una pregunta común al respecto es ¿los sujetos con eventos de riesgos competitivos deben ser considerado como casos o como controles?.

En algunos estudios es posible encontrar biomarcadores que sean predictivos de los eventos de riesgos competitivos, entonces se podría considerar a estos como un segundo grupo de casos y evaluar los marcadores en ellos por separado. Es decir, se podrían estimar dos curvas ROC: una de primordial interés que compare a los sujetos con el evento objetivo, con los controles, y una de interés secundario que compare los sujetos riesgos competitivos a los controles. Una alternativa es incluir a los sujetos con riesgo competitivo al grupo de control, lo que se justifica solo si estos son de algún modo señalados, para no caer en decisiones clínicamente erróneas (Pepe et al. (2008)). Incluso entonces, lo que podría ser de interés es comparar este grupo con los otros controles para que se pueda interpretar a toda la fracción de falsos positivos en termino de los componentes de cada grupo de control.

### 3.5. Estimación de los Parámetros de Funcionamiento

En la estimación de los parámetros de funcionamiento de los biomarcadores se han propuesto cinco métodos bajo la definición de curva ROC incidente estática

que se resumen a continuación (Leisenring et al. (1997), Etzioni et al. (1999), Cai et al. (2006), Heagerty and Zheng (2005), Song and Zhou (2008)). Estos enfoques pueden ser clasificados de forma general como prospectivos (tres primeros métodos) y retrospectivos (dos últimos métodos). Adicionalmente se ha propuesto un métodos para la estimación de curva ROC acumulativa dinámica que también se expone en esta memoria (Heagerty et al. (2000)).

Considere una muestra de datos no censurados  $\{Y_i; T_i; i = 1, \dots, n\}$ , y una muestra de datos censurados  $\{Y_i; T_i; C_i; \delta_i; i = 1, \dots, n\}$  donde  $C_i$  es el tiempo de censura y  $\delta_i = \mathbf{1}\{T_i \leq C_i\}$ .

### 3.5.1. Enfoque de Leisenring et al. (1997)

Un primer enfoque para estimar los parámetros de funcionamiento de un biomarcador es el de Leisenring et al. (1997).

Considérese un escenario simple donde  $Y$  es un marcador **binario** y **no hay censura**. Estos autores propone emplear métodos de regresión binaria para ajustar la  $TPF$  y la  $FPF$ .

Más específicamente estos autores se centran en la  $TPF^{\mathbb{I}}$  y la  $FPF^{\mathbb{E}}$  y proponen los siguientes modelos:

$$TPF_t^{\mathbb{I}} = P(Y = 1|T = t) = g(\alpha + \beta\eta(t)), \quad (3.1)$$

donde  $g^{-1}$  es una función link conocida y  $\eta(t)$  es una función conocida del tiempo. Por otra parte se tiene que

$$FPF_{t^*}^{\mathbb{E}} = P(Y = 1|T > t^*).$$

Para el ajuste de la expresión anterior se emplean todos los individuos que no presentan el fallo o evento de interés antes de tiempo  $t^*$ .

Aunque por simplicidad en la presentación anterior se ha asumido un marcador medido solo en el tiempo baselina, esta aproximación permite incorporar fácilmente marcadores longitudinales.

### 3.5.2. Enfoque de Etzioni et al. (1999)

Las curvas ROCs incidentes tiempo dependientes fueron inicialmente propuestas por Etzioni et al. (1999). Para un marcador **continuo** nuevamente en **ausencia de censura**, estos autores extiende el enfoque de regresión binaria anterior. Los controles en esta aproximación son nuevamente definidos como un grupo no dependiente del tiempo (supervivientes a largo plazo). Más específicamente se consideran como controles a aquellos individuos que no presentan el evento de interés antes de un tiempo  $t^*$ . Por otro lado los casos, son aquellos que experimentan el evento de interés en  $[0, t^*]$ .

Para un tiempo  $t$  y un valor de corte  $c_p$  la  $TPF$  y la  $FPF$ , se definen como

$$TPF_t^{\mathbb{I}}(c_p) = P(Y \geq c_p|T = t),$$

$$FPF_{t^*}^{\mathbb{E}}(c_p) = P(Y \geq c_p | T > t^*).$$

A partir de las expresiones anteriores, y tomando

$$p = FPF_{t^*}^{\mathbb{E}}(c_p) = P(Y \geq c_p | T > t^*) = S_D(c_p),$$

la curva ROC tiempo dependiente puede ser expresada como

$$\begin{aligned} ROC_t^{\mathbb{I}/\mathbb{E}}(p) &= TPF_t^{\mathbb{I}}(c_p) \\ &= P(Y \geq c_p | T = t) \\ &= P(Y \geq S_D^{-1}(p) | T = t). \end{aligned}$$

Estos autores asumen un modelo de regresión para curva ROC anterior. Más específicamente:

$$\begin{aligned} ROC_t^{\mathbb{I}/\mathbb{E}}(p) &= P(Y \geq S_D^{-1}(p) | T = t) \\ &= g(h(p) + \beta\eta(t)). \end{aligned} \quad (3.2)$$

donde  $g^{-1}$ , es la función link conocida,  $h$  es una función desconocida, y  $\eta$  es una función conocida. Etzioni et al. (1999) asumen una forma paramétrica para  $h$ , a saber  $h(p) = a_0 + a_1\Phi^{-1}(p)$ . Posteriormente Cai and Pepe (2002) dieron cabida a la función baselina  $h$  no paramétrica. Cabe señalar que,  $g(h(p))$  es la curva ROC baselina en  $t = 0$ .

El algoritmo propuesto para la estimación del modelo anterior se basa en los siguientes pasos:

- Paso 1: Elegir un conjunto de  $FPF$ :

$$P = \{p_1, \dots, p_k\} \subset (0, 1).$$

- Paso 2: Elegir un conjunto  $T$  de tiempos de interés,  $T = \{t_1, \dots, t_j\}$  y estimar la función de supervivencia en los controles  $S_D$ . Sea  $\hat{S}_D$  este estimador.
- Paso 3: Para cada  $p \in P$  y cada observación en los controles  $\{Y_i, i = 1, \dots, n_D\}$ , calcular las siguientes indicadoras binarias:

$$\hat{B}_{i,p} = \mathbf{I}[\hat{S}_D(Y_i) \leq p], \quad i = 1, \dots, n_D.$$

- Paso 4: Estimar el modelo de regresión para la curva ROC (3.2) mediante el ajuste de un modelo de regresión para datos binarios a partir de la siguiente muestra:

$$\{\{\hat{B}_{i,p}, (\Phi^{-1}(p), \eta(t))\}, t \in T\}.$$

### 3.5.3. Enfoque de Heagerty et al. (2000)

Heagerty et al. (2000) propone dos métodos de estimación para la curva ROC, en este caso acumulativa dinámica. Se define:

$$TPF_t^A = \text{sensibilidad}^A(t, c_p) = \frac{\{1 - S(t|Y > c)\}P(Y > c)}{(1 - S(t))},$$

$$1 - FPF_t^D = \text{especificidad}^D(t, c_p) = \frac{\{S(t|Y \leq c)\}P(Y \leq c)}{S(t)},$$

donde  $S(t)$  es la función de supervivencia,  $S(t) = P(T > t)$ , y  $S(t|Y > c)$  es la función de supervivencia condicional para el subconjunto definido por  $Y > c$ .

El primer método consiste en un estimador para la sensibilidad y la especificidad en un tiempo  $t$ , que se obtiene combinando el estimador de Kaplan and Meier (1958) para la función de supervivencia, y además la función de distribución empírica del marcador  $Y$  (en ausencia de censura).

Una ventaja de este estimador es su simplicidad pero por otra parte presenta problemas, por ejemplo, no garantiza que la sensibilidad o la especificidad sean monótonas. Heagerty et al. (2000) entonces propone un segundo método de estimación usando el estimador de la función de distribución bivariada propuesta por Akritas (1994) denominado el estimador del vecino más cercano (Nearest Neighbor Estimation, NNE), que además **acomoda la censura**.

### 3.5.4. Enfoque de Cai et al. (2006)

Luego, Cai et al. (2006), ofrece el más amplio de los enfoques retrospectivos, que abarca los métodos de Leisenring et al. (1997) y Etzioni et al. (1999) y los extiende a **datos censurados**. Con marcadores **binarios** las funciones que se estima son  $TPF_t^I$  y  $FPF_{t^*}^E$ .

$$TPF_t^I = P(Y = 1|T = t) = g(\alpha + \beta\eta(t)), \quad (3.3)$$

$$FPF_{t^*}^E = P(Y = 1|T > t^*) = p.$$

Los sujetos que no presentan censura son analizados como en el enfoque de Leisenring et al. (1997), es decir, como un caso si  $0 < T \leq t^*$  y como un control si  $T > t^*$ . Los sujetos censurados en  $X$  (siendo  $X = \min\{T, C\}$ ), serán considerados como controles si  $X > t^*$  o de lo contrario como promedios ponderados de los casos y los controles. En esta última situación obsérvese que  $X \leq t^*$ .

$$P(Y = 1|T > X) = FPF_{t^*}^E(T > t^*|T > X) + \int_X^{t^*} TPF_t^I dP(T \leq t|T > X).$$

Por lo tanto, la “probabilidad de contribución” (likelihood contribution) para  $Y$  es un promedio ponderado de  $FPF_{t^*}^E$  y  $TPF_t^I$  para el tiempo transcurrido  $t$  en

$(C, t^*)$ . Los pesos se determinan fácilmente estimando la distribución de  $t^*$  con un método de tiempo de fallo estándar. Note que si existen eventos con riesgos competitivos la distribución debe estimarse con los métodos de incidencia acumulativa (Kalbfleisch and Prentice (1980)) en lugar de tratarlos como eventos censurados.

Para biomarcadores **continuos**, Cai et al. (2006) adopta el modelo ROC-GLM (3.2) asumiendo una forma no paramétrica para la curva ROC baselina  $h$ .

Específicamente en este modelo la  $TPF_t^{\text{II}}$  y la  $FPPF_{t^*}^{\text{IE}}$  para un valor de corte  $c_k$  son:

$$FPPF_{t^*}^{\text{IE}}(c_k) = P(Y \geq c_k | T > t^*) = p_k.$$

$$\begin{aligned} TPF_t^{\text{II}}(c_k) &= g(h'(c_k) + \beta\eta(t)) \\ &= g(h(FPPF_{t^*}^{\text{IE}}(c_k)) + \beta\eta(t)) \\ &= g(h(p_k) + \beta\eta(t)) \\ &= g(\alpha_k + \beta\eta(t)). \end{aligned} \tag{3.4}$$

Este modelo puede ser implementado reemplazando cada registro del biomarcador,  $Y$ , con una serie de  $K$  registros de variables binarias de la forma  $I(Y > c_k)$ , correspondiente a los umbrales del biomarcador  $c_1, \dots, c_K$ . El algoritmo expuesto en la sección 3.5.2 puede entonces ser aplicado con una serie de  $FPPFs$ , es decir  $\{FPPF_{t^*}^{\text{IE}}(c_1), FPPF_{t^*}^{\text{IE}}(c_2), \dots, FPPF_{t^*}^{\text{IE}}(c_K)\}$ , correspondiente a los valores de corte escogidos. Además, se estima una serie de interceptos en (3.4),  $\{\alpha_1, \alpha_2, \dots, \alpha_K\}$  que corresponden a los  $K$  umbrales o puntos de corte seleccionados, son interpretado como  $\{h(FPPF_{t^*}^{\text{IE}}(c_1)), \dots, h(FPPF_{t^*}^{\text{IE}}(c_K))\}$ .

Por otra parte, las técnicas de regresión de riesgos, están establecidas para modelar datos de eventos en el tiempo y acomodan de forma natural la censura. Después de ajustar un modelo prospectivo se puede combinar este ajuste con con la distribución del predictor observado para calcular los parámetros de funcionamiento del biomarcador ( $TPF$  y  $FPPF$ ).

### 3.5.5. Enfoque de Heagerty and Zheng (2005)

Heagerty and Zheng (2005) emplean la técnica de regresión de riesgos proporcionales o modelo de Cox para un marcador baselina  $Y$ :

$$\lambda(t) = \lambda_0(t) \exp(\gamma(t)Y), \tag{3.5}$$

donde el parámetro de la regresión  $\gamma$  puede depender de  $t$ . Ellos notaron que al ajustar el modelo para un marcador **binario**, denotando al conjunto de riesgo en  $t$  como  $R(t)$  se tiene:

$$\widehat{TPF}_t^{\text{II}} = \frac{\sum_{i \in R(t)} Y_i \exp(\hat{\gamma}(t)Y_i)}{\sum_{i \in R(t)} \exp(\hat{\gamma}(t)Y_i)},$$

que es un estimador consistente de la  $TPF_t$ . Ello proviene de la observación de Xu and O'Quigley (2000); de que bajo el modelo de Cox la distribución de  $Y \exp(\gamma(t)Y)$  para sujetos en el conjunto de riesgo  $R(t)$  es igual a la distribución condicional de  $Y$  tomando a  $T = t$ . Para estimar  $FPF$ , estos autores emplean el método empírico en los controles del conjunto de riesgo de  $t^*$ :

$$\widehat{FPF}_{t^*}^{\mathbb{E}} = \sum_{i \in R(t^*)} Y_i / n(t^*)$$

donde  $n(t^*)$  es el tamaño del conjunto de riesgo. Con biomarcadores **continuos**, sea  $\widehat{F}_{t^*}$  la distribución empírica de  $Y$  para sujetos en el conjunto de riesgos en  $t^*$ , entonces

$$\widehat{ROC}_t^{\mathbb{I}/\mathbb{E}}(p) \equiv \frac{\sum_{i \in R(t)} I(Y_i \geq \widehat{F}_{t^*}^{-1}(1-p)) \exp\{\widehat{\gamma}(t)Y_i\}}{\sum_{i \in R(t)} \exp\{\widehat{\gamma}(t)Y_i\}},$$

### 3.5.6. Enfoque de Song and Zhou (2008)

Song and Zhou (2008) emplearon como Heagerty el modelo de riesgos proporcionales pero simplificando los parámetros para que no dependa del tiempo, es decir,  $\gamma(t) = \gamma$ , de forma que:

$$\lambda(t) = \lambda_0(t) \exp(\gamma Y). \quad (3.6)$$

Bajo (3.6) se puede escribir la función de supervivencia para  $T$  de forma análoga a como se vio en la sección 2.4.2, y también se puede escribir la función de densidad condicionada de supervivencia. Usando el teorema de Bayes para escribir la  $TPF_t^{\mathbb{I}}$  y la  $FPF^{\mathbb{E}}$  de un marcador binario:

$$\begin{aligned} TPF_t^{\mathbb{I}} &= P(Y = 1 | T = t) \\ &= P(Y = 1)P(T = t | Y = 1) / P(T = t) \\ &= P(Y = 1) \frac{\lambda_0(t) \exp(\gamma) \exp(-\Lambda_0(t) \exp(\gamma))}{\lambda_0(t) \exp(\gamma) \exp(-\Lambda_0(t) \exp(\gamma))P(Y = 1) + \lambda_0(t) \exp(-\Lambda_0(t))P(Y = 0)} \\ &= \text{logit}^{-1}\{\text{logit}P(Y = 1) + \gamma + \Lambda_0(t)(1 - \exp(\gamma))\}, \end{aligned} \quad (3.7)$$

donde  $\text{logit}^{-1}(x) = \exp(x)/(1 + \exp(x))$  y  $\Lambda_0$  es la función de riesgo (*hazard*) baselina acumulativa;

$$\begin{aligned} FPF^{\mathbb{E}} &= P(Y = 1 | T > t^*) \\ &= P(Y = 1)P(T > t^* | Y = 1) / P(T > t^*) \\ &= \text{logit}^{-1}\{\text{logit}P(Y = 1) + \Lambda_0(t)(1 - \exp(\gamma))\}. \end{aligned} \quad (3.8)$$

Obsérvese que si  $\gamma = 0$  entonces  $TPF_t^{\mathbb{I}} = FPF^{\mathbb{E}} = P(Y = 1)$ , el cual es un resultado intuitivo.

Con marcadores **continuos**, sobre las expresiones  $TPF_t^{\mathbb{I}}$  y  $FPF^{\mathbb{E}}$  (3.7) y (3.8) se incorporan integrales correspondientes al marcador dicotomizado  $I[Y > c]$

$$TPF_t^{\mathbb{I}}(c) = \frac{\int_c^{\infty} \exp(\gamma Y) \exp\{-\Lambda_0(t^*) \exp(\gamma Y)\} dF(Y)}{\int_{-\infty}^{\infty} \exp(\gamma Y) \exp\{-\Lambda_0(t^*) \exp(\gamma Y)\} dF(Y)}$$

$$FPF^{\mathbb{E}}(c) = \frac{\int_c^{\infty} \exp\{-\Lambda_0(t^*) \exp(\gamma Y)\} dF(Y)}{\int_{-\infty}^{\infty} \exp\{-\Lambda_0(t^*) \exp(\gamma Y)\} dF(Y)}$$

Song y Zhou sustituyen  $\hat{\gamma}$ ,  $\hat{\Lambda}_0$  y la distribución empírica de  $Y$ ,  $\hat{F}$ , en las expresiones anteriores para estimar la  $TPF_t^{\mathbb{I}}$  y la  $FPF^{\mathbb{E}}$ . El estimador de la curva ROC es calculada:

$$\widehat{ROC}_t(p) = TPF_t^{\mathbb{I}}\{[FPF^{\mathbb{E}}]^{-1}(p)\}.$$

El método de Song y Zhou tiene dos ventajas sobre el enfoque de Heagerty y Zheng. En primer lugar demostró ser más eficiente en estudios de simulación (Song and Zhou (2008)). Esto es debido probablemente a su empleo de un estimador de máxima verosimilitud parcial (mple) para  $\gamma$  y  $\Lambda_0$  y así los correspondientes estimadores de  $(TPF, FPF)$  son también mple. En contraste con el estimador de  $TPF_t$  empleado por Heagerty and Zheng (2005) que no es mple. Además su estimación empírica de  $FPF$  no utiliza la estructura conferida por el modelo de Cox, mientras que Song y Zhou si la utilizan. La segunda ventaja concierne a la censura. Mientras que el método de Heagerty and Zheng (2005) no permite que la censura dependa del marcador  $Y$ , el método de Song and Zhou (2008) sigue siendo válido si la censura depende del marcador.

### 3.6. Comparando las Ventajas

Entre los métodos retrospectivos, Cai et al. (2006) es el más amplio. Los otros métodos retrospectivos pueden ser vistos como casos especiales de Cai's, por lo tanto se comparará con los dos enfoques prospectivos; los métodos de Song and Zhou's y Heagerty y Zheng's, que de ahora en adelante se denotaran por Cai, S+Z y H+Z. La siguiente comparación es solo para la estimación de la curva  $ROC^{\mathbb{I}/\mathbb{E}}$

**Comparando el uso de métodos Prospectivos versus Retrospectivos.** La fracción de verdaderos y falsos positivos son definidas como cantidades retrospectivas porque ellas hacen referencia a la distribución de  $Y$  "condicionada" a el resultado "obtenido". Por lo tanto, un análisis retrospectivos es el enfoque que se adapta de forma natural para estimar estas fracciones. Sin embargo, los parámetros que dependen de  $t$  en un enfoque retrospectivo como por ejemplo  $\beta$  en la ecuación (3.1) cuantifican directamente cómo el funcionamiento del marcador va variando con  $t$ . De este modo, inferir sobre estos parámetros con el enfoque retrospectivo es una tarea sencilla, y en contraposición, bajo modelos prospectivos los parámetros no cuantifican directamente el efecto del tiempo en el funcionamiento del marcador.

**Comparación en cuanto a los supuestos que requiere cada modelización.** Los supuestos que requiere la modelización de Cai son de poca trascendencia. Esto es principalmente debido a que el método es no paramétrico con respecto a la distribución de  $Y$  en los controles y semiparamétrico en cuanto a la distribución de  $Y$  en los casos. Específicamente, semiparamétrica porque solo modela el efecto de  $T$  en la distribución de forma paramétrica. Por otra parte, los métodos prospectivos también tienen supuestos leves. En ellos se usan modelos semiparamétricos para la  $Prob(Y|T)$  pero dejan la distribución de  $Y$  en los grupos de individuos sin especificar.

**Comparando la consideración de la censura.** La censura que es independiente de  $Y$  es considerada o alojada por todos los métodos. Sin embargo, la censura que es dependiente de  $Y$  solo es considerada en el método de S+Z. Hay excepciones para los métodos de Cai y H+Z en escenarios generales de censura condicionalmente independiente, pero es posible que esta forma de modelizar no sea completamente trivial (Xu and O'Quigley (2000); Cai et al. (2006)).

**Comparación en cuanto a los eventos con riesgos competitivo.** Aunque no está especificado con palabras en alguno de los métodos propuestos, todos pueden ampliarse para acomodar los eventos con riesgos competitivos. Las funciones de riesgo (Hazard) son reemplazadas con funciones específicas de cada causa de riesgo y las funciones de supervivencia son reemplazadas con la probabilidad de no tener un evento (de cualquier tipo, ni el evento de interés ni eventos de riesgos competitivos).

**Comparación en cuanto a biomarcadores longitudinales.** Cai et al. (2006) desarrolló el método retrospectivo en el contexto general donde los datos del marcador son recolectados longitudinalmente sobre el tiempo. Los métodos prospectivos pueden ser generalizados para acomodar datos longitudinales usando un modelo de regresión marginal (Zheng and Heagerty (2007)). Específicamente cada medida del marcador  $Y(s)$  genera un registro del dato con el tiempo de origen  $T$  reajustado a  $s$ . Es decir, el tiempo de ocurrencia del evento o tiempo de censura asociado con  $Y(s)$  es  $T - s$  o  $C - s$ , respectivamente. Al permitir que el correspondiente riesgo baselino y los coeficientes de la regresión dependan de  $s$ , la  $TPF$  y la  $FPF$  pueden ser escritas como funciones de  $s$  y  $t$ .

**Comparando en cuanto a covariables.** Varios factores pueden afectar la distribución del marcador y/o el funcionamiento del marcador como predictor de eventos. Estos factores son llamados covariables. Por ejemplo, la severidad del evento AKI (*Acute Kidney Injury*, Lesión Aguda de Riñones, la severidad puede ser total o media) puede afectar la capacidad de el biomarcador de predecir este evento. Por ejemplo, PSA puede ser mejor predictor de un tipo de cancer de próstata que de otro. Covariables específicas de la enfermedad que son asociados sólo con los casos pueden ser modeladas como partes de la función  $TPF$  con el análisis de enfoque retrospectivo. Tales covariables no son generalmente acomodadas por méto-

dos prospectivos. Sin embargo, si la covariable es discreta, los eventos pueden ser clasificados y tratados como eventos de riesgos competitivos. Por ejemplo, eventos de AKI totalmente severos y eventos de AKI con severidad media pueden ser considerados tipos de eventos de riesgos competitivos y estimar  $TPF$  que puede ser calculada para cada uno de los dos tipos de eventos. Existen otras covariables que se aplican tanto a los casos como a los controles. Por ejemplo, lugar del estudio en un estudio multicéntrico o características del sujeto que realiza las pruebas, podrían influir en el marcador o su rendimiento. Cai et al. (2006) y Song and Zhou (2008) describen como incorporar tales covariables en el análisis. Referimos al lector a esos documentos para más detalles.

**Comparando en cuanto como se compara el comportamiento de biomarcadores**

Los métodos retrospectivos permiten comparar el comportamiento de diferentes predictores, sin más que incorporar, como covariable, el tipo de predictor en los modelos de regresión asumido para la  $FPF$  y la  $TPF$ . La forma de proceder es similar a la descrita en el capítulo 3 de Pepe (2003). Por el contrario, en la actualidad, los métodos prospectivos no permiten comparar el comportamiento de diferentes predictores.

**Comparando en cuanto a la combinación de marcadores.** Los métodos examinados en esta memoria no están centrados en cómo combinar, de forma óptima, diferentes predictores juntos. Sin embargo, una vez se defina esta combinación (por ejemplo empleando el modelo de Cox), el *score* de la combinación puede ser empleado como un predictor univariante.

# 4

## Aplicación

### 4.1. Aplicación a Datos Simulados: AKI (*Acute Kidney Injury*)

Pepe et al. (2008) utiliza un conjunto de datos de supervivencia denominados AKI, para estudiar sobre ellos las curvas ROCs.

**El contexto es el siguiente:** Los pacientes que han sido sometidos a una cirugía cardíaca mayor tienen un alto riesgo de sufrir una lesión en los riñones, debido a la interrupción del flujo sanguíneo que pasa a través de ellos durante la cirugía. Para detectar si un paciente sufre de una Lesión Aguda en los Riñones (en inglés *Acute Kidney Injury, AKI*) se monitoriza observando si hay aumento en los valores de creatinina sérica. Sin embargo, esta prueba diagnóstica tiene una desventaja importante y es que la creatinina sérica solo puede ser monitorizada a partir del tercer día después de la operación, porque antes la producción y equilibrio de estos valores se ven afectados por otros factores que no tienen que ver con los riñones. Entonces, lo que se busca es un biomarcador que sea capaz de detectar un AKI más temprano que la creatinina sérica (antes del tercer día después de la cirugía) de forma que un tratamiento apropiado pueda ser iniciado prontamente en aquellos pacientes afectados.

Un biomarcador es actualmente medido en la orina bajo investigación de un estudio multicéntrico de 1800 pacientes que han experimentado una cirugía cardíaca mayor. Las muestras de orina son tomadas a varios intervalos después de la operación, son congeladas y luego almacenadas. Al final del estudio los especímenes de orina almacenados serán ensayados por un laboratorio clínico.

Un AKI es un evento que ocurre cuando el nivel de creatinina sérica tres días después de la operación incrementa un 25% sobre el nivel preoperatorio y es sostenido así por 24 horas. Un AKI puede ser considerado como un evento

“Severo” o “Medio”, de modo que los pacientes son clasificados con un AKI severo si alcanzan un nivel  $> 200\%$  en relación con creatinina preoperativa, de lo contrario el individuo es clasificado como AKI Medio.

**Se conocen las siguientes estadísticas sobre el estudio:** aproximadamente el 80 % de los pacientes se recupera después de la cirugía sin presentar AKI u otro evento devastador y son dados de alta entre el tercer y quinto día después de la operación. Se espera que el 20 % de los pacientes experimenten un evento AKI, 12 % medio y 8 % severo. Un pequeño grupo adicional de pacientes  $< 1\%$  morirá por complicaciones asociadas con su enfermedad o con la cirugía sin reunir los criterios de un AKI. Llamaremos a estas muestras no-AKI.

**Algunas cuestiones de interés en este estudio son:** determinar el número de pacientes para quienes el diagnóstico de AKI puede ser adelantado con el nuevo marcador, el tiempo y el costo en términos de diagnósticos falsos. El análisis necesita dar cabida a eventos de riesgos competitivos debido a muertes por no-AKI, a diversos grados de severidad del AKI y a la naturaleza longitudinal del biomarcador.

Este estudio está en progreso y los datos no estarán disponibles durante algún tiempo. Tenemos datos simulados que se aproximan al diseño del estudio clínico, tal como se describe en el Apéndice A. Estos datos están disponibles en el sitio web: [www.fhcrc.org/science/labs/pepe/dabs/](http://www.fhcrc.org/science/labs/pepe/dabs/).

**Constitución de los datos simulados:** de acuerdo a las estadísticas conocidas, de los 1800 sujetos en estudio, 342 tienen un evento AKI, 136 graves AKI y 206 AKI medios. Además, 18 pacientes fallecen por causas aparentemente inconexas al daño renal. Para más detalle ver el Apéndice A.

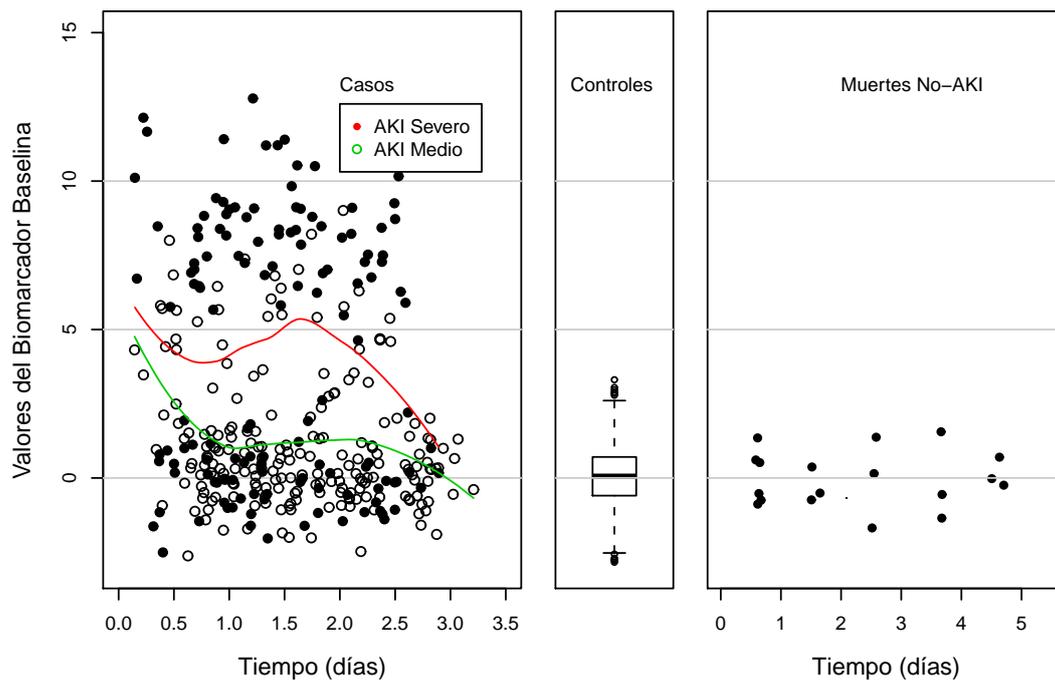
#### 4.1.1. Biomarcador Baseline

Considere el biomarcador medido desde la primera muestra de orina postoperativa que llamaremos marcador baseline.

Comencemos con un **análisis básico** del biomarcador, para ello la distribución de los cuatro grupos de pacientes se muestra en la Figura 4.1.

Vemos que en comparación con el grupo de sujetos controles, el resto tiene un valor de marcador baseline más alto, y particularmente para el grupo de sujetos que presentan un AKI severo los valores del marcador baseline están más alejados de los que presentan los controles que los valores del marcador correspondiente a sujetos con AKI medio. Los valores del biomarcador baseline en pacientes que mueren por no-AKI no parecen diferir de los valores del biomarcador baseline para el grupo de controles, y se consideraran dentro de este grupo (la comparación formal entre los grupos puede hacerse a través del test de rangos de Wilcoxon (Pepe et al. (2008))).

La Figura 4.2 muestra las curvas ROCs que han sido obtenidas para este biomarcador baseline. Esta Figura tiene la siguiente **estructura**: por filas se mues-

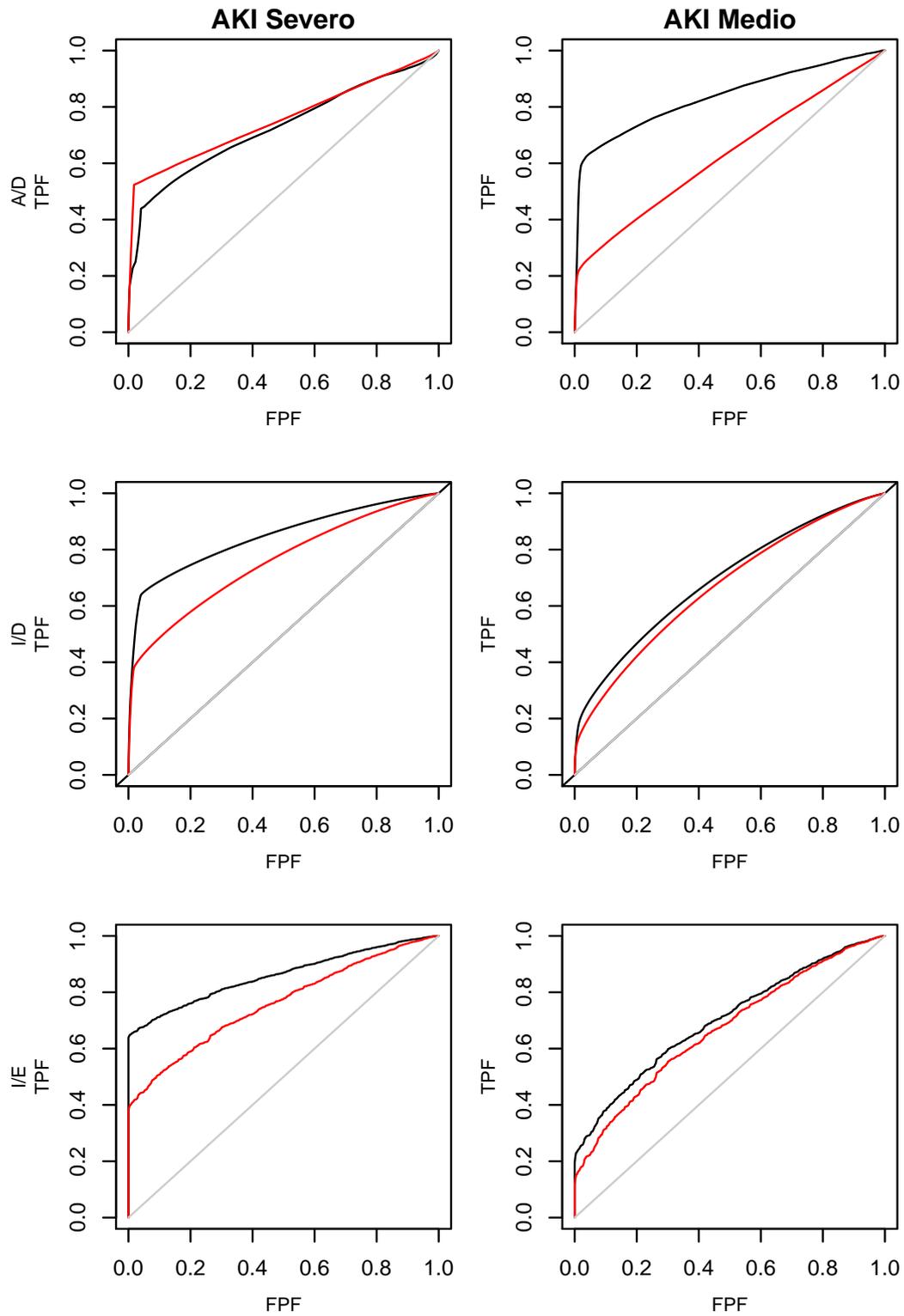


**Figura 4.1:** Distribución del biomarcador baselina AKI. Se muestra el ajuste con curvas lowess (detalles en el Apéndice C), para el biomarcador AKI medio y severo.

tran las curvas ROC obtenidas bajo la definición  $A/D$ ,  $II/D$ , e  $II/E$  que fueron explicadas en el capítulo anterior, y por columnas; a la izquierda las correspondiente para sujetos que presentaron un evento AKI severo y a la derecha para sujetos que presentaron un AKI medio.

Se tienen en cuenta las siguientes **consideraciones** en esta gráfica: se han incluido solo los sujetos con AKI severo junto con los controles para estimar las curvas ROCs de la columna izquierda y solo los sujetos con AKI medio junto con los controles para estimar las curvas ROCs de la derecha, queriendo decir con esto que, cada estimación y su correspondiente análisis se aplicaron separadamente para el estudio del AKI severo y del AKI medio. Dos curvas ROCs son obtenidas para cada caso; para  $t = 1$  día (en color negro), y para  $t = 2$  días (en color rojo). Se han elegido estos tiempos de interés porque se quiere estudiar la capacidad del biomarcador baselina para diagnosticar eventos AKI antes de su diagnóstico clínico con la creatinina serica que es posible solo al tercer día de la cirugía.

La **metodología** empleada para la construcción de estas curvas es la siguiente: estas curvas se obtuvieron mediante software R aplicando funciones de paquetes que han sido desarrollados especialmente para el análisis de supervivencia tiempo dependientes. A saber, para cada curva  $ROC^{A/D}$  se aplicó la función `survival-ROC` propuesta por Heagerty como complemento de su artículo del año 2002. Para cada curva  $ROC^{II/D}$  se aplicó la función `risksetROC` que también ha sido una contribución de Heagerty esta vez como complemento de su artículo del año 2005, esta



**Figura 4.2:** Estimaciones de la curva ROC para el biomarcador baselina AKI, tomando como tiempo de interés  $t$  un día (en color negro) y dos días (en color rojo) después de la cirugía.

función utiliza las estimaciones de los parámetros de funcionamiento mostrada en el capítulo tres como el método de Heagerty y Zhen (H+Z). Por último las curvas  $ROC^{I/E}$  son obtenidas como combinación de las  $TPF$  y las  $FPF$  producto de aplicar la función `risksetROC` para distintos tiempos; el tiempo considerado como incidente  $t$  y el considerado como estático  $t^*$  (para más detalle de las funciones ver el Apéndice C).

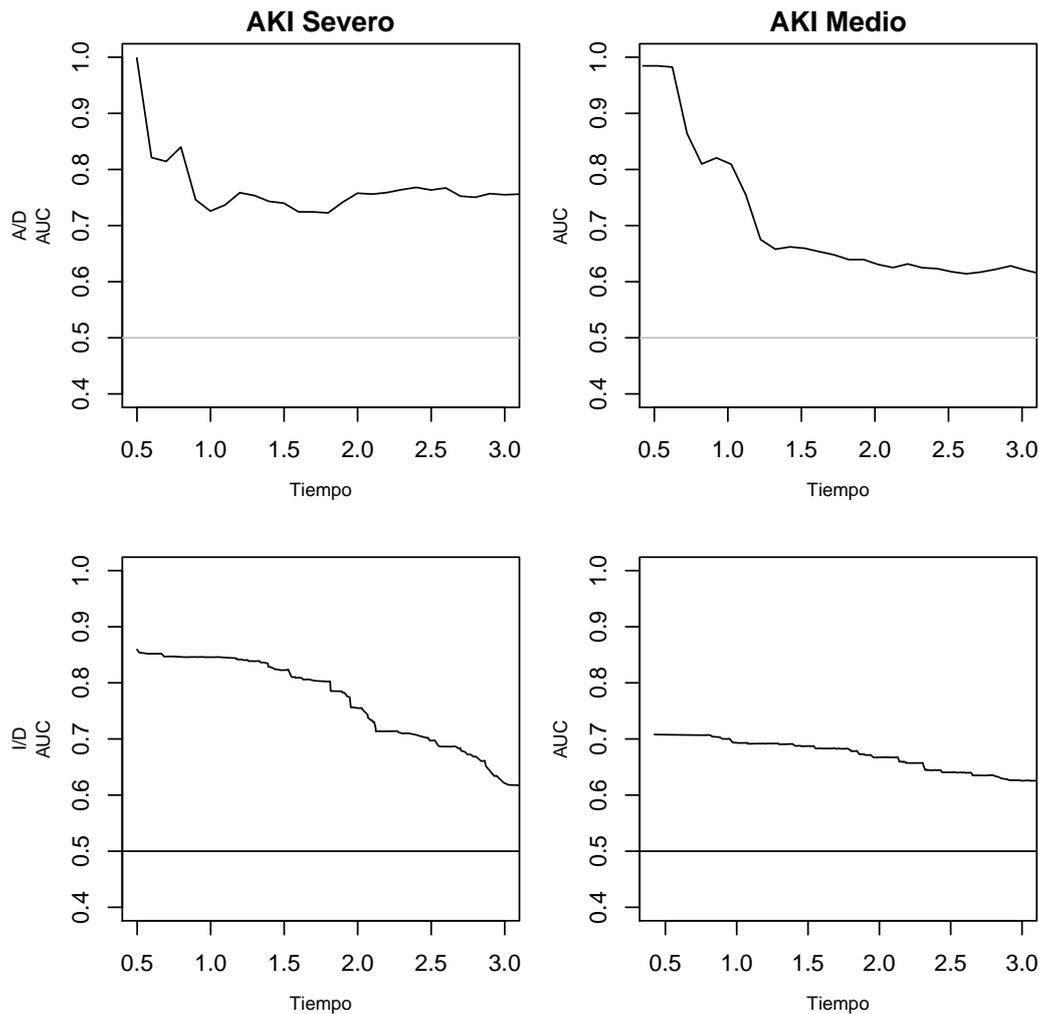
Sobre las ROCs obtenidas para el biomarcador baselina en el estudio de AKI severo se **aprecia**:

- La capacidad del biomarcador para predecir un evento AKI severo al primer día después de la cirugía es distinta a la capacidad para predecirlo al segundo día. vemos que las curvas  $ROC^{I/D}$  y  $ROC^{I/E}$  de la columna de la izquierda para  $t = 1$  (en color negro) están por encima de las curvas obtenidas para  $t = 2$  (en rojo). Específicamente la Tabla 4.1, por ejemplo para la  $ROC^{I/D}$ , indica que el área bajo esta curva para  $t = 1$  es el 84,57% del área total del cuadrado unitario mientras que para  $t = 2$  es de 75,49%, un notable 9,08% menos. Únicamente se evidencia para la  $ROC^{A/D}$  una situación opuesta, al parecer para algunas  $FPF$  la curva resultante considerando  $t = 2$  es superior a la de  $t = 1$  (Figura 4.2).

	AUC AKI Severo	AUC AKI Medio
$ROC_{t_1}^{A/D}$	0.72579	0.83601
$ROC_{t_2}^{A/D}$	0.75357	0.63017
$ROC_{t_1}^{I/D}$	0.84572	0.69340
$ROC_{t_2}^{I/D}$	0.75494	0.68347
$ROC_{t_1}^{I/E}$	0.85982	0.70198
$ROC_{t_2}^{I/E}$	0.75991	0.66964

**Tabla 4.1:** AUC del marcador baselina AKI severo y medio, para tiempo de supervivencia uno y dos días.

- En general las AUC para todos los posibles tiempos se encuentra por encima del 60% del área total (ver la Figura 4.3)
- En el futuro el tratamiento puede ser iniciado sobre la base del nuevo biomarcador pero los médicos tendrán que equilibrar los beneficios potenciales de aplicar un tratamiento a un individuo que ha sido diagnosticado con AKI,



**Figura 4.3:** AUC usando marcador baselina y distintas taxonomía

contrastando el hecho de que puede ser verdadero o simplemente una falsa alarma o un falso positivo. Para tratamientos potencialmente tóxicos o caros los falsos positivos deben ser bajos. Para tratamientos no tóxicos adicionales, como la hidratación, un gran número de falsos positivos sería aceptable. Una serie de opciones de tratamiento será examinada en el futuro, así que aquí son de interés una gama de fracciones de falsos positivos. Tomemos como ejemplo, un 20% de falsos positivos, todas las curvas ROCs presentadas en la Figura 4.2 para AKI “severo” demuestran que el biomarcador será capaz de predecir correctamente más del 50% de los individuos que presentarán un AKI severo.

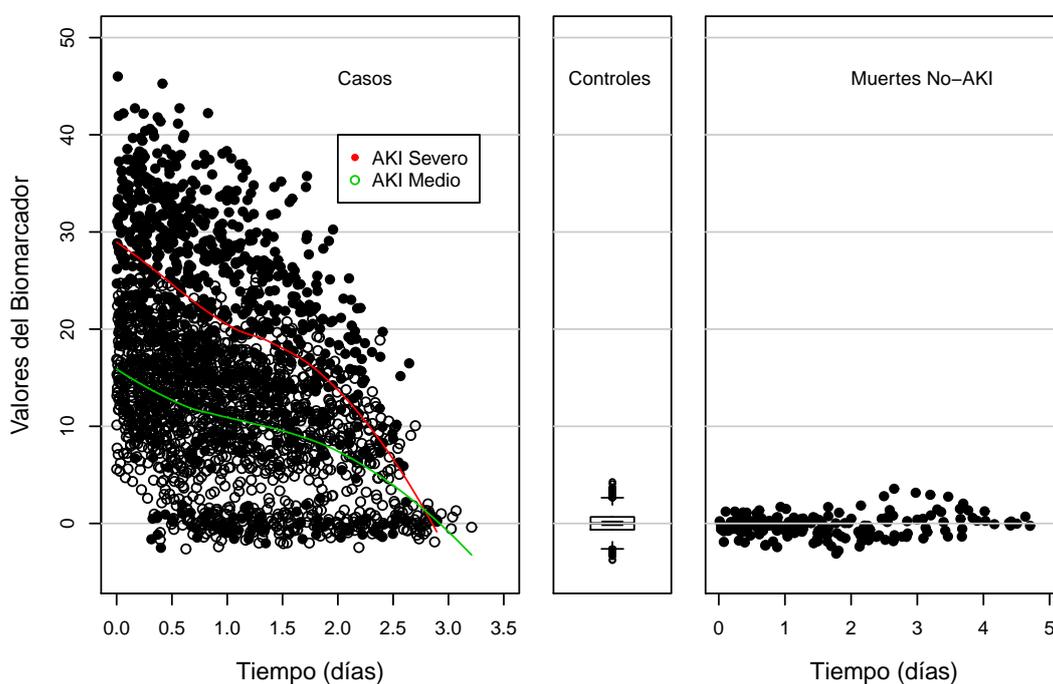
Para el biomarcador baselina en el estudio de AKI medio se pueden formular conclusiones análogas a las presentadas anteriormente para el AKI severo, sin

embargo lo que destaca más en la Figura 4.2 y se comprueba en la Tabla 4.1, es que las AUC para las ROCs cuando se estudia el evento AKI medio son inferiores a las AUC para las ROCs cuando se estudia en evento AKI severo, esto indica que el biomarcador parece predecir mejor los eventos severos que los eventos medios.

#### 4.1.2. Biomarcador Longitudinal

Considere el biomarcador medido longitudinalmente para cada individuo desde la primera muestra de orina postoperatoria hasta la última antes de darse por finalizado el estudio. Cada medición se realiza en un tiempo  $s$ .

Comencemos con un **análisis básico** del biomarcador longitudinal, para ello la distribución de los cuatro grupos de pacientes se muestra en la Figura 4.4.



**Figura 4.4:** Distribución del biomarcador longitudinal en el estudio de AKI medio y severo. Se muestra el ajuste con curvas lowess (detalles en el Apéndice C).

Vemos que en comparación con el grupos de controles, el resto tiene un valor de marcador más alto, y particularmente para el grupo de sujetos que presentan AKI severo los valores del marcador longitudinal están más alejados de los controles que los valores de AKI medio, igual que ocurría con el biomarcador baselina. Los valores del biomarcador longitudinal en pacientes que mueren por no-AKI no parecen diferir de los valores del biomarcador longitudinal en controles.

La Figura 4.5 muestra las curvas ROCs que han sido obtenidas para este biomarcador longitudinal en el estudio de AKI severo. Esta Figura tiene la sigu-

iente **estructura**: por filas se muestran las curvas ROC obtenidas bajo la definición  $A/D$ ,  $I/D$ , e  $I/E$ , y por columnas; a la izquierda las correspondiente a el tiempo de interés  $t + s = 1$  día y a la derecha pa el tiempo de interés  $t + s = 2$  días. Análogamente la Figura 4.6 presenta la misma estructura para AKI medio.

Se tienen en cuenta las siguientes **consideraciones** en estas figuras: se han incluido solo los sujetos con AKI severo junto con los controles para estimar las curvas ROCs de la Figura 4.5 y solo los sujetos con AKI medio junto con los controles para estimar las curvas ROCs de la Figura 4.6.

Tres curvas ROCs se obtienen para una de las gráficas que conforman estas figuras; cuando el tiempo en que interesa evaluar la capacidad predictiva del marcador es un día se han considerado tres tiempos de medición del marcador longitudinal,  $s = 0,25$  que equivale aproximadamente a las mediciones tomadas 6 horas después de la operación (en color negro), para  $s = 0,50$  equivalente a 12 horas (en color rojo), y para  $s = 0,75$  equivalente a 18 horas (en color azul). Análogamente, cuando el tiempo que interesa evaluar la capacidad predictiva del marcador es dos día se han considerado tres tiempos de medición del marcador longitudinal,  $s = 1,25$  que equivale aproximadamente a las mediciones tomadas 30 horas después de la operación (en color negro), para  $s = 1,50$  equivalente a 36 horas (en color rojo), y para  $s = 1,75$  equivalente a 42 horas (en color azul).

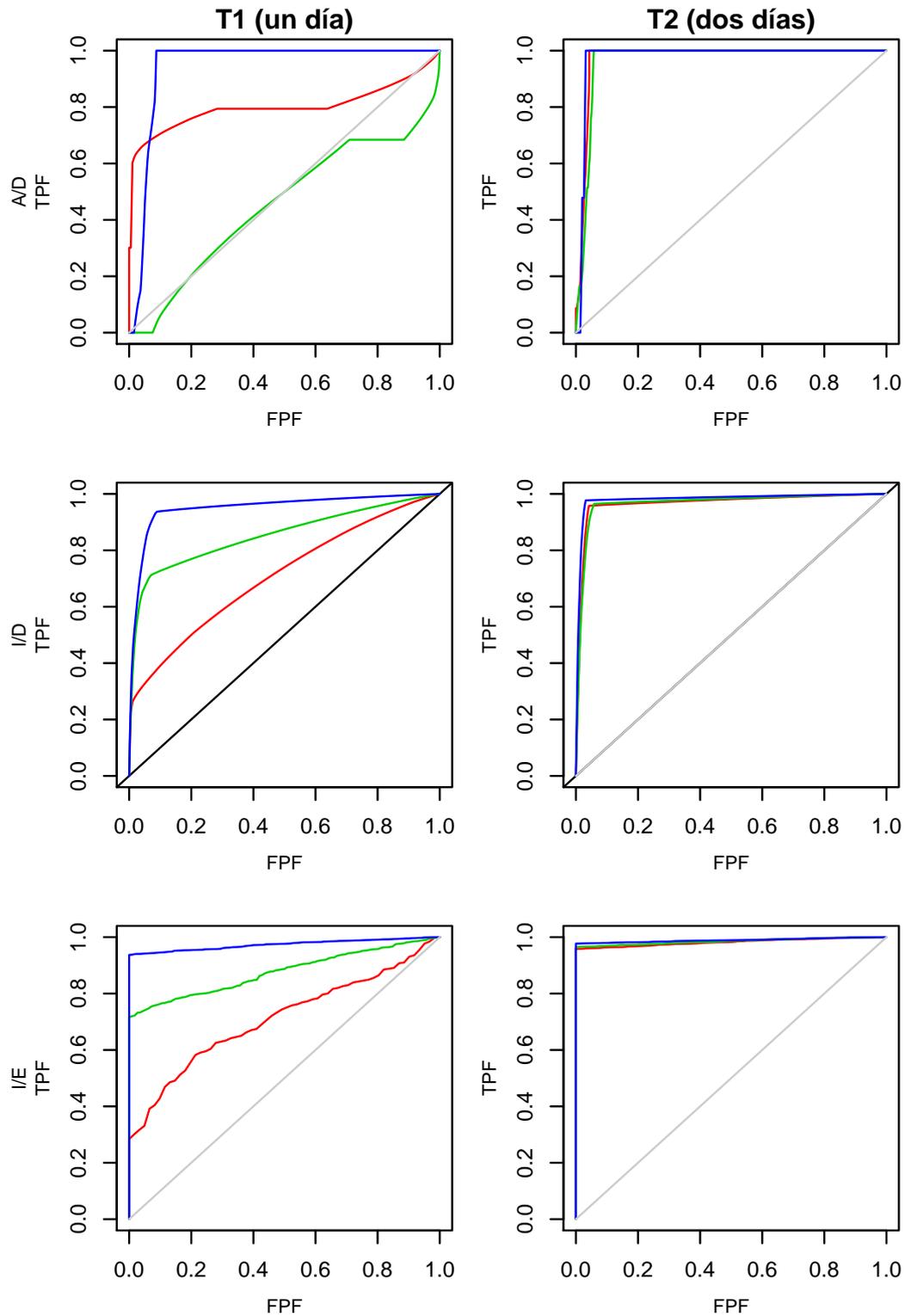
La **metodología** empleada para la construcción de estas curvas es similar a la aplicada para el marcador baselina, la única diferencia radica en que a las funciones utilizadas para generar las curvas ROCs se le suministra como tiempos de medición del marcador, los tiempos respectivo de medición longitudinal, en este caso:  $s = 0,25$ ,  $s = 0,50$ , ...,  $s = 1,75$  (tiempo en días), y no los tiempos baselina. Obteniendo sistemáticamente 18 curvas para AKI severo y 18 curvas para AKI medio.

Sobre las ROCs obtenidas para el biomarcador longitudinal en el estudio del AKI severo se **aprecia**:

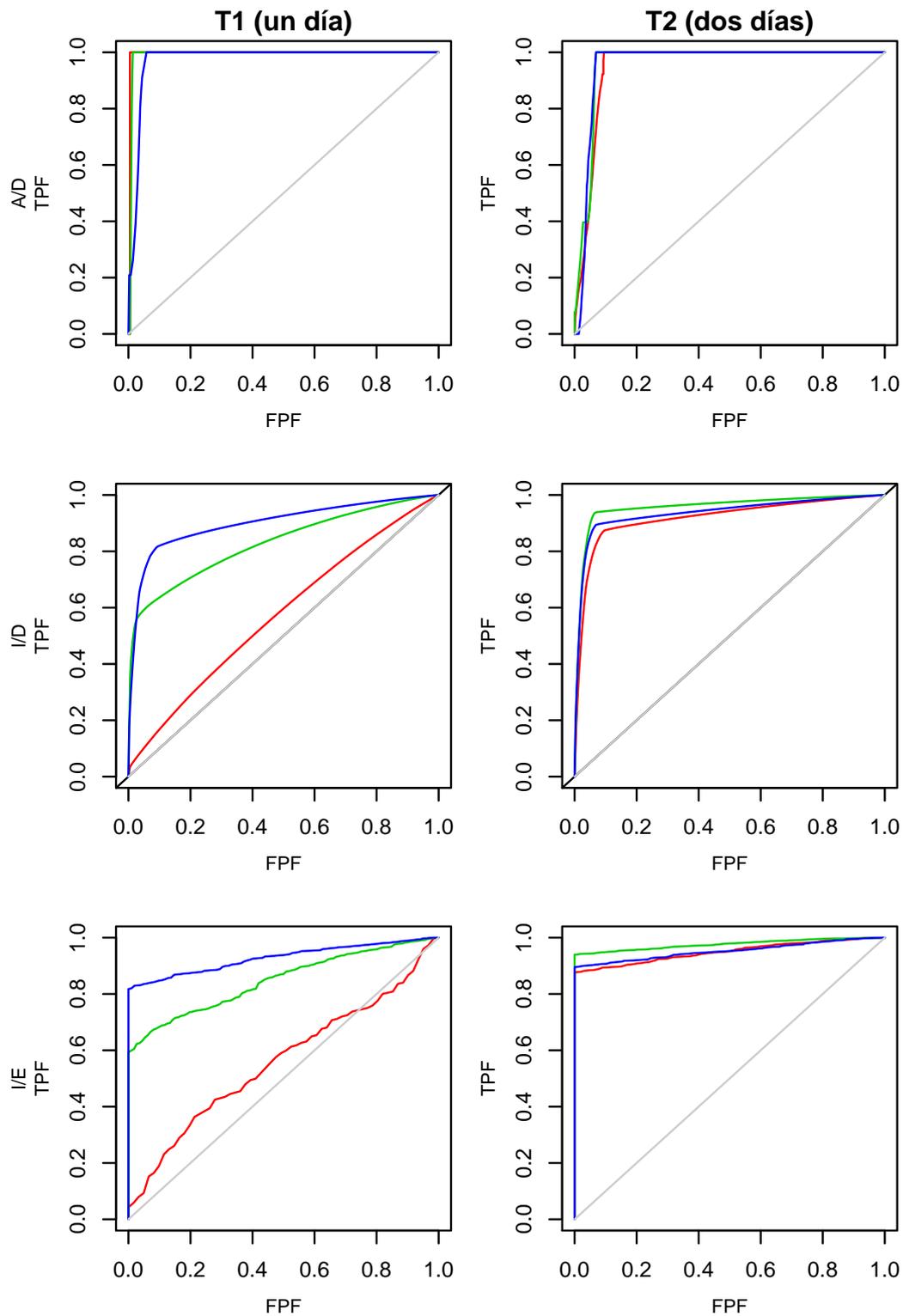
- La capacidad del biomarcador longitudinal para predecir la ocurrencia de un evento AKI justo 24 horas después de la cirugía (primer día) es mayor cuando el tiempo de medición del marcador  $s$  aumenta. La única excepción se observa en las curvas  $ROC^{A/D}$  obtenidas, ya que difieren de las restantes curvas ROCs, lo que incita a estudiar con detenimiento la cantidad de pacientes y el rol que juegan en cada tiempo de medición para esta taxonomía en particular.

Un resultado importante que se evidencia en la Figura 4.5 es que tomando la medición del marcador longitudinal después de las primeras 24 horas el biomarcador es capaz de predecir casi con exactitud la ocurrencia del evento AKI severo para las 48 horas. Específicamente la Tabla 4.2, por ejemplo para la  $ROC_t^{I/D}$  con  $t = 2$ , indica que el área bajo esta curva para  $s = 1,25$  es el 96,71 % del área total del cuadrado unitario, mientras que para  $s = 1,50$  es 96,81 % y para  $s = 1,75$  es 98,08 %.

Para el biomarcador longitudinal en el estudio del AKI medio se pueden formular conclusiones análogas a las presentadas anteriormente para el AKI severo,



**Figura 4.5:** Estimaciones de la curva ROC para el biomarcador longitudinal para AKI **Severo**, tomando como tiempo de interés  $t + s$  un día, y tiempo de medición del marcador  $s = 0,25$  (en color rojo),  $s = 0,50$  (en color verde), y  $s = 0,75$  (en color azul), y respectivamente tomando como tiempo de interés  $t + s$  dos días para  $s = 1,25$  (en color rojo),  $s = 1,50$  (en color verde), y  $s = 1,75$  (en color azul).



**Figura 4.6:** Estimaciones de la curva ROC para el biomarcador longitudinal para AKI **Medio**, tomando como tiempo de interés  $t + s$  un día, y tiempo de medición del marcador  $s = 0,25$  (en color rojo),  $s = 0,50$  (en color verde), y  $s = 0,75$  (en color azul), y respectivamente tomando como tiempo de interés  $t + s$  dos días para  $s = 1,25$  (en color rojo),  $s = 1,50$  (en color verde), y  $s = 1,75$  (en color azul).

		AUC AKI Severo			AUC AKI Medio		
	$s$	$ROC_t^{A/D}$	$ROC_t^{I/D}$	$ROC_t^{I/E}$	$ROC_t^{A/D}$	$ROC_t^{I/D}$	$ROC_t^{I/E}$
$t_1$	0,25	0.80009	0.71280	0.70944	0.99512	0.57505	0.55971
	0,50	0.45527	0.85326	0.87649	0.99127	0.83344	0.84694
	0,75	0.94422	0.94968	0.97351	0.97493	0.90135	0.92817
$t_2$	0,25	0.58104	0.71133	0.71307	0.57241	0.57337	0.55930
	0,50	0.80570	0.76829	0.77833	0.79658	0.69518	0.69783
	0,75	0.97666	0.94258	0.95227	0.96413	0.85995	0.87440
	1,25	0.97403	0.96711	0.98223	0.96246	0.91879	0.98223
	1,50	0.96585	0.96818	0.98519	0.96283	0.95687	0.97614
	1,75	0.97659	0.98089	0.98934	0.96760	0.93734	0.98934

**Tabla 4.2:** AUC para tiempo de supervivencia uno y dos días y para AKI severo y medio, con distintos tiempos de medición del marcador longitudinal  $s$

sin embargo lo que destaca más en la Figura 4.6 y se comprueba en la Tabla 4.2 es que las AUC para las ROCs del AKI medio son inferiores a las AUC para las ROCs del AKI severo, esto indica que el biomarcador parece predecir mejor los eventos severos que los eventos medios, aunque en ambos casos usando los tiempo de medición  $s = 1,25$ ,  $s = 1,50$ ,  $s = 1,75$  y prediciendo el evento AKI al día dos las AUC obtenidas son significativamente grandes.

## 4.2. Aplicación a Datos Reales: ACS (*Acute Coronary Syndrome*)

Cid-Álvarez et al. (2009) inicia el uso de este conjunto de datos bajo el siguiente contexto: En pacientes con Síndrome Coronario Agudo (en inglés *Acute Coronary Syndrome*, ACS) el incremento de los niveles de glucosa plasmática están asociados con peor pronóstico para los pacientes. El objetivo de Cid-Álvarez et al. (2009) era evaluar si los valores de la glucosa en el momento de admisión y de la glucosa en ayunas permitían predecir la muerte en pacientes con ACS y comparar su capacidad predictiva. El método que emplearon fue el siguiente: la relación de la mortalidad con los niveles de glucosa entre 811 pacientes que llegaron consecutivamente a ser hospitalizados con un ACS, es estimada usando un modelo de Cox. Las muestras de sangre son obtenidas en el momento de admisión o entrada al hospital y después de la primera noche, en ayunas. La capacidad predictiva de la

glucosa en ayunas y en admisión fueron comparadas usando curvas ROCs tiempo dependientes. Ellos obtuvieron como resultado que la glucosa en ayunas es mejor predictor.

Además de los niveles de glucosa para cada individuo, se recaudaron registros de otras variables, en total 43. De modo que también obtuvieron como resultado que existe una dependencia del ratio de riesgo de mortalidad sobre la glucosa en ayunas aun cuando se ajusta por covariables entre pacientes no diabéticos. Entonces, partiendo del modelo multivariable usado por Cid-Álvarez et al. (2009), se ha obtenido un nuevo modelo siguiendo un proceso de selección optima de variables (ver Apéndice B) del cual a su vez se obtienen los predictores lineales o score que será nuestro biomarcador. Estudiaremos el comportamiento de este biomarcador para una gama de fracciones de falsos positivos, y además seleccionando distintos tiempos de interés hipotéticos se evaluará la capacidad de clasificación que tiene el marcador cuando se responde a diferentes preguntas médicas, como por ejemplo, la capacidad de predecir la muerte después de padecer un ACS antes del primer año.

De 811 pacientes en el estudio, 151 murieron (18,6%). 558 no son diabéticos (68,8%) y 253 si lo son. Se empleó en el modelo de Cox multivariante entre otras las siguientes variables de los pacientes, consideradas como predictoras de la mortalidad: glucosa en ayunas, edad, cateterismo, fallo renal (mdrd), anemia, y otros factores como crujía previa, ser o no fumador (para más detalle de las variables y el modelo empleado ver el Apéndice B).

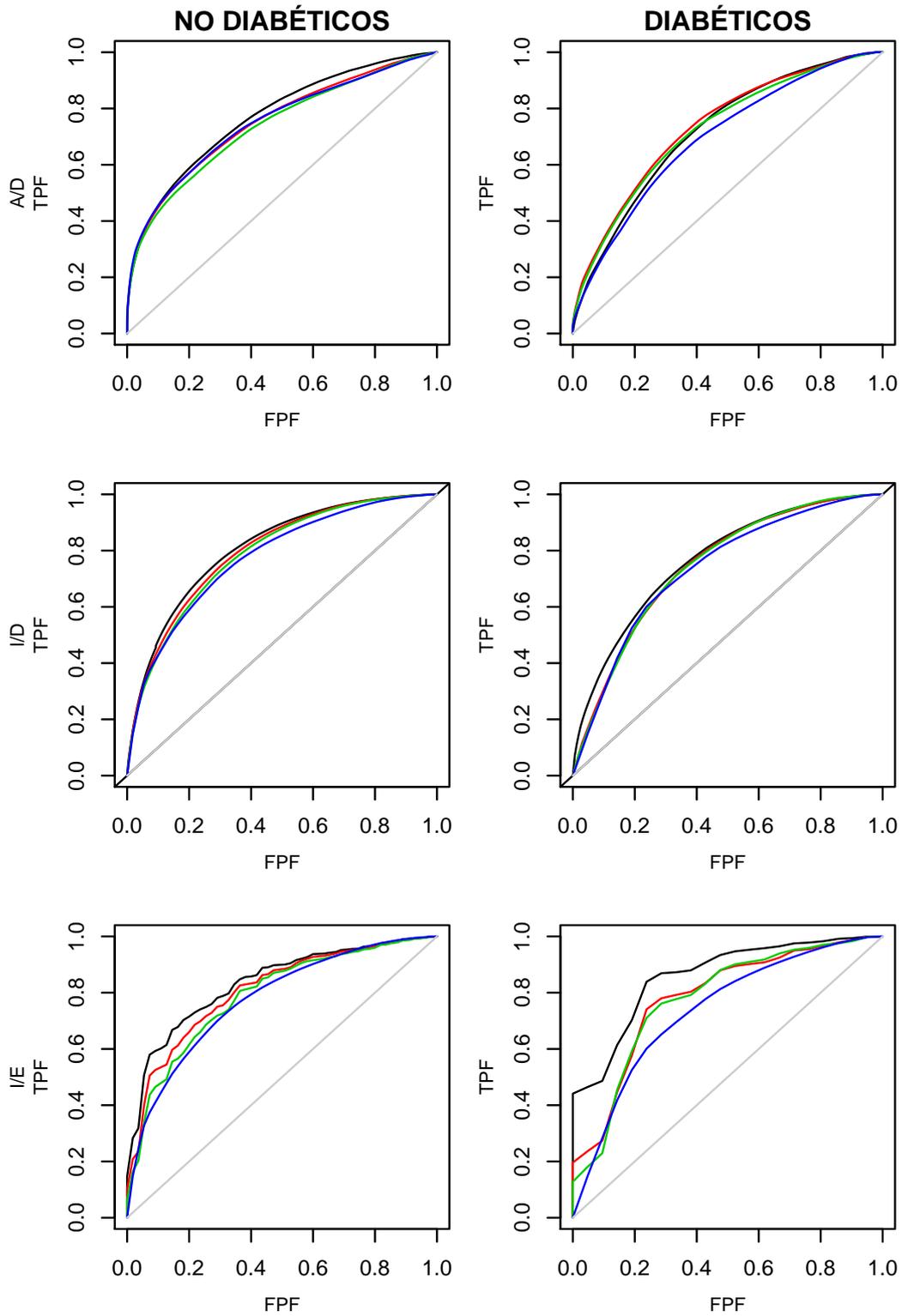
La Figura 4.7 muestra las curvas ROCs que han sido obtenidas para este biomarcador. Esta Figura tiene la siguiente **estructura**: por filas se muestran las curvas ROC obtenidas bajo la definición  $A/ID$ ,  $I/ID$ , e  $I/E$ , y por columnas; a la izquierda las correspondiente para sujetos no diabéticos y a la derecha para sujetos diabéticos.

Se tienen en cuenta las siguientes **consideraciones** en esta gráfica: se han incluido solo sujetos no diabéticos para estimar las curvas ROCs de la columna izquierda y solo los sujetos diabéticos para las curvas ROCs de la derecha, queriendo decir con esto que, cada estimación y su correspondiente análisis se aplicó separadamente para pacientes sin y con diabetes. Cuatro curvas ROCs se obtienen para cada caso; para  $t = 1$  año (en color negro), para  $t = 2$  años (en color rojo), para  $t = 3$  años (en color verde), y para  $t = 4$  años (en color azul). Se han elegido estos tiempos de interés por considerarse tiempos de revision habitual en pacientes con diabetes, y además porque es importante, si el marcador tiene capacidad para predecir la muerte del individuo, contar con un tiempo prudente donde se pueda monitorear con más frecuencia el estado de esos pacientes a los que se les pronostica próximo el evento muerte, de manera preventiva.

La **metodología** empleada para la construcción de estas curvas es análoga a la empleada anteriormente en la aplicación de AKI baselina en la sección 4.1.1.

Sobre las ROCs obtenidas para el biomarcador en no diabéticos se **aprecia**:

- La capacidad del biomarcador para predecir la muerte de un paciente no siendo diabético al primer año después del ACS es ligeramente distinta a la capacidad para predecirla al segundo, tercer y cuarto año, disminuyendo



**Figura 4.7:** Estimaciones de la curva ROC para el biomarcador baselina ACS, tomando como tiempo de predicción de la muerte  $t$  un año (en color negro), dos años (en color rojo), tres años (en color verde), cuatro años (en color azul) después presentarse en el paciente el ACS.

un poco con el paso de los años. En efecto, la Tabla 4.3 nos comprueba, por ejemplo para la  $ROC_t^{A/D}$ , que el área bajo esta curva para  $t = 1$  año es el 82,08 % del área total del cuadrado unitario, mientras que para  $t = 2$  años es de 80,92 %, para  $t = 3$  años es de 79,76 %, y para  $t = 4$  años es de 78,50 %. Una situación similar se presenta con la  $ROC_t^{A/D}$  con diferencias más pequeñas entre las áreas, y para la  $ROC_t^{I/E}$  con diferencia ligeramente más grande entre las áreas de cada tiempo.

$ROC_t$	$t$	$t$	AUC	
	(días)	(años)	No Diabéticos	Diabéticos
$ROC_t^{A/D}$	182	1/2	0.81562	0.70805
	365	1	0.77014	0.72007
	730	2	0.75142	0.73677
	1095	3	0.73736	0.72509
$ROC_t^{I/D}$	182	1/2	0.82080	0.77343
	365	1	0.80922	0.76257
	730	2	0.79761	0.74830
	1095	3	0.78505	0.76077
$ROC_t^{I/E}$	182	1/2	0.84514	0.87340
	365	1	0.83415	0.85874
	730	2	0.80831	0.78226
	1095	3	0.78831	0.77533

**Tabla 4.3:** AUC para tiempo de supervivencia: medio año, uno, dos, y tres años.

- En general las AUC para todos los posibles tiempos se encuentra aproximadamente por encima del 70 % del área total (ver la Figura 4.8). Para el caso de los diabéticos las áreas son ligeramente inferior a las áreas de los no diabéticos. Esto se puede comprobar visualmente también en las curvas ROCs obtenidas en la Figura 4.7, las curvas de la columna derecha parecen estar más cerca de la diagonal que las curvas de la columna izquierda correspondiente a los no diabéticos, de hecho, podemos comprobarlo numéricamente a través de la Tabla 4.3, por ejemplo, AUC de la  $ROC_t^{A/D}$  considerando como tiempo de interés  $t$  medio año después del ACS en el caso no diabético es 81,56 % mientras que la correspondiente AUC para los diabéticos es de 70,80 %. Un notable 10,76 % menos.

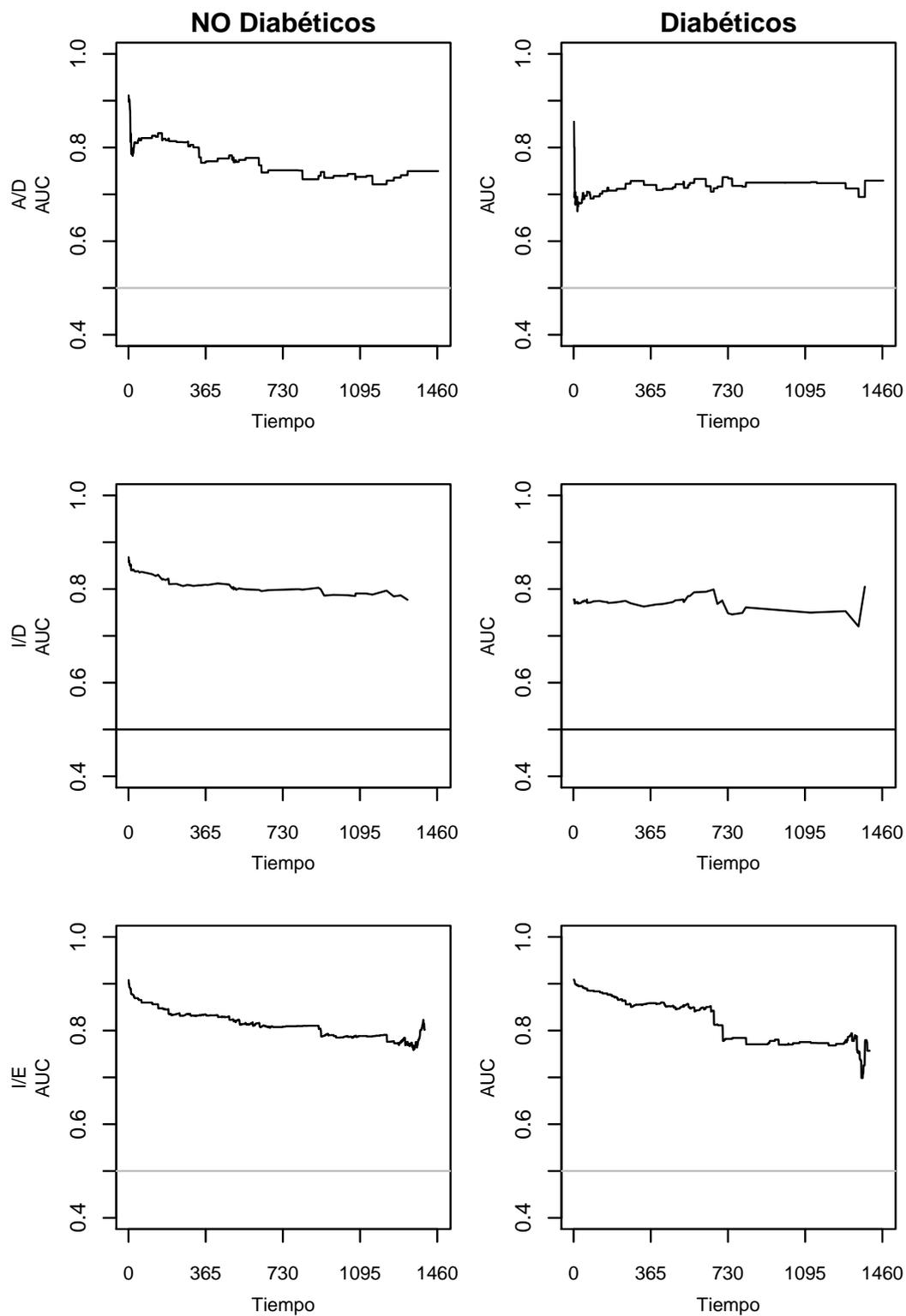


Figura 4.8: AUC para datos de Infarto Agudo del Miocardio, tiempo en días.

- La acción preventiva que el médico puede iniciar una vez que se ha pronosticado la muerte de un paciente que padeció un ACS con este biomarcador, es un seguimiento constante mediante consultas, y el chequeo de sus valores clínicos, intentando controlar estos con la aplicación de tratamientos disponibles en el mercado. Considerando que es una medida preventiva, podría aceptarse y evaluarse una gama de fracciones de falsos positivos.

Tomando como ejemplo la curva  $ROC^{A/D}$ , cuando se permite un 20 % de falsos positivos este biomarcador resultante de un modelo multivariable es capaz de predecir la muerte del 58,02 % de los individuos que mueren antes del primer año de forma correcta.

Si el interés está en predecir aquellos individuos que morirán justo al año de haber padecido el ACS, entonces este biomarcador es capaz de predecir un 64,45 % de los individuos correctamente.

En los dos casos anteriores la capacidad predictiva del biomarcador disminuye si el tiempo de interés varía en dos tres y cuatro años.

Es evidente que cuando se permite un porcentaje mayor de falsos positivos, como el 50 % o el 80 % mostrado en esta tabla la capacidad de predecir correctamente a individuos que si morirán del biomarcador aumenta, pudiendo llegar hasta un 89,41 % si es de interés aplicar la clasificación de casos controles I/D por ejemplo.

	$p = FPF$	$ROC(p)$ No Diabéticos				$ROC(p)$ Diabéticos			
		$t_1$	$t_2$	$t_3$	$t_4$	$t_1$	$t_2$	$t_3$	$t_4$
A/D	0.20	0.58025	0.56470	0.54228	0.56654	0.46123	0.50097	0.48989	0.43546
	0.50	0.83283	0.80354	0.78745	0.80167	0.80872	0.81865	0.79874	0.75948
	0.80	0.95859	0.93734	0.92945	0.92849	0.95541	0.95117	0.94612	0.93998
I/D	0.20	0.64457	0.61279	0.59723	0.58276	0.55538	0.50356	0.51554	0.54866
	0.50	0.89410	0.88393	0.87543	0.85303	0.85173	0.84605	0.84289	0.82433
	0.80	0.98165	0.98119	0.97904	0.96942	0.97188	0.97149	0.97285	0.95294
I/E	0.20	0.65471	0.62459	0.60443	0.58276	0.54531	0.51692	0.55221	0.54866
	0.50	0.89756	0.88775	0.87796	0.85303	0.84572	0.85142	0.85777	0.82433
	0.80	0.98271	0.98238	0.97981	0.96942	0.97034	0.97321	0.97717	0.95294

**Tabla 4.4:** Resumen TPF para tiempo de supervivencia un año, dos , tres y cuatro años, considerando distintas FPF.

Para el biomarcador en caso de diabéticos se pueden formular conclusiones análogas a las presentadas anteriormente para el caso de no diabéticos.

El uso en la práctica médica de un biomarcador como este, debe ser discutido ampliamente, sopesando los pro y contras que este biomarcador pueda aportar a los métodos clínicos empleados actualmente. Sería necesario además plantear adecuadamente la interpretación clínica que los médicos pueden dar a los resultados o medidas de resumen de este biomarcador, todo esto dentro del marco de un *estudio de clasificación formal*.

Sin embargo, un primer paso en el entendimiento de la trascendencia de este marcador que motiva un estudio a fondo, sería conciliar los valores característicos de las variables involucradas en la construcción del biomarcador, *apreciados en los grupos ya clasificados*, con aquellos valores característicos de las variables esperados en la práctica.

Considere los grupos ya clasificados resultante de aplicar este biomarcador mediante la definición de  $ROC_{t_1}^{A/D}$ , permitiendo un 20% de *FPF*. Un primer ejemplo, son los valores de la variable "edad": se espera en la práctica clínica que pacientes después de un ACS con mayor edad sea los que conforman en mayor medida el grupo de pacientes que morirán antes del primer año. En efecto, vemos para pacientes no diabéticos, que la edad media de los pacientes que se predice morirán antes del primer año es aproximadamente 78 años, mientras que en el grupo que se predice que no morirán tiene una edad media de 61 años aproximadamente (Ver Tabla 4.5).

Otra variable de interés es la "glucosa" en ayunas, siguiendo la misma consideración sobre la curva ROC y la *FPF* anterior, se observa nuevamente en la Tabla 4.5 para no diabéticos que es diferente el valor medio de la glucosa en los dos grupos ya clasificados, a saber, 112 mg/dL para el grupo con predicción de supervivencia contra 136 mg/dL para el grupo cuya predicción es la muerte (los valores normales de glucosa están en el intervalo de 90 mg/dL a 117 mg/dL). No obstante, estos valores pueden ser comparados con sus análogos en la Tabla 4.6 para pacientes diabéticos, observando 164mg/dL contra 176mg/dL, resultados también esperados en la práctica clínica.

Por último, también se muestra en las tablas los valores correspondientes a la variable "mdrd" (sobre el fallo renal) que se pueden analizar de forma análoga.

La caracterización de las variables anteriores solo pretende ser de carácter demostrativo, considerando la cantidad de variables involucradas en el modelo multivariante utilizado para obtener este biomarcador. Además, otras interpretaciones médicas podrían obtenerse de procedimientos como este, planificadas dentro de un estudio formal de clasificación.

	$ROC_{t_1}^{A/D}$ con $FPF = 0,20$		$ROC_{t_1}^{I/D}$ con $FPF = 0,20$	
	Punto de corte $c = 0,93093$		Punto de corte $c = 0,89228$	
	$Y_i < c$	$Y_i > c$	$Y_i < c$	$Y_i > c$
	GLUCOSA		GLUCOSA	
Min	69.0	66.00	69.0	66.0
Max	335.0	450.00	335.0	450.00
1st Qu	97.0	98.75	97.0	99.0
3rd Qu	120.0	157.80	112.4	157.0
Median	106.0	125.50	106.0	121.0
Mean	112.5	136.10	112.4	134.8
Var	630.5	3075.9	629.3935	2941.679
Desv	25.10976	55.4608	25.08772	54.23725
	MDRD		MDRD	
Min	9.618	11.97	9.618	11.97
Max	192.000	137.40	192.000	137.40
1st Qu	69.910	44.83	70.140	45.26
3rd Qu	100.700	76.86	101.300	76.81
Median	85.150	62.08	85.500	62.90
Mean	86.040	62.60	86.340	62.88
Var	618.6443	600.7965	622.8784	579.7548
Desv	24.87256	24.51115	24.95753	24.0781
	EDAD		EDAD	
Min	33.0	50.0	33.0	50.00
Max	88.0	95.0	88.0	95.00
1st Qu	51.0	74.0	51.0	74.00
3rd Qu	72.0	84.0	72.0	84.00
Median	62.0	80.0	62.0	80.00
Mean	61.7	78.8	61.4	78.79
Var	164.4194	53.06014	162.5056	52.52168
Desv	12.82261	7.28424	12.74777	7.247184

**Tabla 4.5:** Característica de las variables utilizando  $ROC_{t_1}^{A/D}$  y  $ROC_{t_1}^{I/D}$  para el grupo de No Diabéticos, cuando se permite un 20% de falsos positivos

	$ROC_{t_1}^{A/D}$ con $FPF = 0,20$		$ROC_{t_1}^{I/D}$ con $FPF = 0,20$	
	Punto de corte $c = 0,82368$		Punto de corte $c = 0,71035$	
	$Y_i < c$	$Y_i > c$	$Y_i < c$	$Y_i > c$
	GLUCOSA		GLUCOSA	
Min	60.0	68.0	70.0	60.0
Max	406.0	340.0	406.0	340.0
1st Qu	123.0	132.5	123.8	132.0
3rd Qu	192.0	218.0	192.5	218.0
Median	155.0	173.0	154.5	166.0
Mean	164.2	176.6	164.3	175.2
Var	3820.92	3784.397	3776.615	3953.760
Desv	61.8135	61.51745	61.45417	62.87893
	MDRD		MDRD	
Min	10.00	7.676	10.00	7.676
Max	184.20	85.040	184.20	85.040
1st Qu	56.99	30.270	58.38	29.690
3rd Qu	88.93	56.180	89.46	56.850
Median	72.28	41.280	74.84	41.320
Mean	73.30	42.330	75.27	43.330
Var	645.3467	344.3081	612.3348	344.4224
Desv	25.40368	18.55554	24.7454	18.55862
	EDAD		EDAD	
Min	44.00	67.00	44.00	57.00
Max	86.00	95.00	86.00	95.00
1st Qu	61.00	74.50	61.00	74.00
3rd Qu	76.00	85.00	75.00	84.00
Median	69.00	80.00	68.00	79.00
Mean	67.71	79.79	67.22	78.74
Var	90.44146	49.03738	90.41026	52.77375
Desv	9.510072	7.00267	9.508431	7.264555

**Tabla 4.6:** Característica de las variables utilizando  $ROC_{t_1}^{A/D}$  y  $ROC_{t_1}^{I/D}$  para el grupo de Diabéticos, cuando se permite un 20% de falsos positivos

# 5

## Conclusión

De acuerdo con los objetivos planteados en esta memoria nos enfocamos en la estimación de las curvas ROCs tiempo dependiente exponiendo un resumen sobre los métodos de estimación propuestos en las ultimas publicaciones especializadas. Entre estos es han seleccionado los de Heagerty para mostrar las diferencias que se obtienen al aplicar las definiciones de  $ROC^{A/D}$ ,  $ROC^{I/D}$ ,  $ROC^{I/E}$  sobre datos de supervivencia. Particularmente, se analizó el comportamiento estadístico de un biomarcador tanto baselina como longitudinal, y también para un biomarcador procedente de la aplicación de un modelo multivariante de Cox (biomarcador para el estudio de AKI, y para el estudio de ACS respectivamente). El modo de operar biomarcadores baselina con respecto a los longitudinales difiere en la consideración de tiempos de medición. Para la aplicación aquí mostrada se obtuvieron mejores predicciones en el marcador longitudinal posiblemente asociada a la información extra que este supone al consistir de mediciones en tiempos muy cercanos. Un marcador más complejo construido a partir de varias variables también puede arrojar resultados satisfactorios mediante las curvas ROC tiempo dependientes. La elección de la definición de curva ROC tiempo dependiente a aplicar dependerá de como se ha planteado el estudio de clasificación en el sentido de qué tiempos resultan de interés, y a cuáles preguntas desea el médico responder mediante la prueba diagnóstica. La curva ROC evaluará la capacidad del biomarcador correctamente si se ha elegido de forma adecuada el método de estimación que esté acorde con el estudio; si es prospectivo o retrospectivo, si se requiere o no que acomode la censura o si necesitamos considerar varios predictores. Finalmente el desarrollo de nuevas metodologías de estimación y de inferencia, son necesarias como líneas futuras de investigación requeridas para el avance de esta importante área de la estadística en la medicina.

# A

## Apéndice

### A.1. Metadatos AKI

**Origen de los datos:** El fichero `aki_sim.dta` que contiene los datos se puede obtener gratuitamente en

*<http://labs.fhcrc.org/pepe/dabs/datasets.html>*

**Descripción del contenido:** Los datos de Lesion Aguda de Riñones (en inglés Acute Kidney Injury, AKI) son datos simulados como se muestra en la sección A.2, sobre un biomarcador medido en la orina de pacientes después de ser sometidos a una cirugía cardíaca mayor y los correspondientes tiempos de medición.

Esta base de datos esta formada por 36,000 observaciones simuladas (20 por cada individuo en el estudio), de 6 variables de interés (ver Tabla A.1).

### A.2. Generación de Datos Simulados AKI

#### (i) Notación

$n$  = tamaño total de la muestra, 1800 pacientes

$i$  = índice del sujeto

$k$  = k-esimo espécimen de la muestra

$s_{ik}$  = tiempo en que el k-esimo espécimen fue tomado en el paciente  $i$ .

#### (ii) Muestreo de los tiempos $s_{ik}$

Se pueden tomar muestras de orina a los paciente aproximadamente hasta el quinto o sexto día después de la cirugía. Se genera de la siguiente forma:

Variable	Formato	Descripción
id	int	Identificador del paciente.
y	float	Valor del biomarcador medido en la orina.
s	float	Tiempo de medición del marcador longitudinal con respecto a la cirugía.
tevent	float	Tiempo faltante para ocurrir el evento, valido solo para <code>statusall=1,2,9</code> .
statusall	factor	Status (Controles: 0, AKI Severo:1, AKI Medio:2, Muerte No-AKI:9).
tmsfull	float	Tiempo en que se observó la ocurrencia del evento.

**Tabla A.1:** Descripción y formato de las variables de los datos AKI

$$s_{ik} = 0,25k + \varepsilon_{ik},$$

donde  $k = 1, \dots, 20$  y  $\varepsilon \sim \text{uniforme}(0, 0,25)$ . Estos tiempos son modificados luego dependiendo del status del paciente.

### (iii) Subgrupos de pacientes

#### Controles

Un conjunto aleatorio de 1,440 pacientes se le asignan el status de controles. Se simuló que ellos eran dados de alta en el día 3 un 30 %, en el día 4 un 40 %, y en el día 5 un 30 % mediante la medición de los tiempos  $s_{ik}$  que exceden el tercer y cuarto día para conjuntos aleatorios de controles de 30 % y 40 % respectivamente

#### Muertes por No AKI

A 18 pacientes se les asigna el status de muerte por No AKI. La medición de los tiempos son omitidas después del día 1 para 6 de los pacientes simulados en los que su evento por no AKI ocurre en el primer día. Similarmente se deja de medir los tiempos después de los días 2, 3, 4 y 5 para un conjunto de tres paciente en cada día.

#### Eventos AKI (T)

Los restantes 342 pacientes tiene un evento AKI. De estos, se le asignan a 206 el status de AKI medio y 135 AKI severo. Un evento AKI es observado en el tiempo  $E$  y se genera de la siguiente forma para pacientes con AKI severo:

$$E^{sev} \sim \text{uniforme}(0, 0,25) \text{ con probabilidad } 0,6$$

$$E^{sev} \sim \text{uniforme}(0,25, 1,25) \text{ con probabilidad } 0,4$$

Es decir,  $E^{sev}$ , el tiempo de ocurrencia del AKI, se distribuye uniformemente entre 0 y 0.25 en 60 % de los pacientes severos y también se distribuye uniformemente entre 0.25 y 1.25 en el 40 % de esos pacientes. Para pacientes que presentaron AKI medio se consideró

$$E^{med} \sim \text{uniforme}(0, 0,25) \text{ con probabilidad } 0,4$$

$$E^{med} \sim \text{uniforme}(0,25, 2,25) \text{ con probabilidad } 0,6$$

El tiempo  $T$  del diagnóstico clínico de AKI con la creatinina serica es generado como:

$$T = E + V \text{ donde } V \sim \text{uniforme}(0, 2,75)$$

#### (iv) Valores del Biomarcador

##### *Controles y muertes por no-AKI*

Los valores del biomarcador se distribuyen normalmente con media 0 y varianza 1 con ninguna tendencia en el tiempo. Una estructura auto-regresiva se usa para simularlos:

$$Y_{i,1} \sim N(0, 1)$$

$$Y_{i,k} = \alpha Y_{i,k-1} + \sqrt{1 - \alpha^2} \varepsilon_{ik}, \quad k \geq 2$$

donde  $\varepsilon_{ik}$  es un error independiente  $N(0, 1)$  y la correlación autorregresiva es determinado por  $\alpha$ . Se eligió  $\alpha = 0,8$

##### *Casos*

En los casos, el valor del biomarcador es generado como para los controles hasta el momento de sus evento temporal  $E$ . Sea  $s_{ik^*}$  el tiempo de la primera medición después de  $E$ . Se genera como:

$$Y_{i,k^*} = \Delta + \alpha Y_{i,k^*-1} + \sqrt{1 - \alpha^2} \varepsilon_{ik^*}$$

donde  $\Delta = u + \delta$ ,  $\delta$  independiente distribuida normalmente con media 0 y desviación estándar 2 y  $\mu$ , la media de  $Y_{i,k^*}$  depende de la severidad del AKI.

$$\mu = 8 \text{ para sujetos con AKI severo}$$

$$\mu = 4 \text{ para sujetos con AKI medio.}$$

Para tiempos medidos posteriormente,

$$Y_{i,k} = \alpha Y_{i,k-1} + \sqrt{1 - \alpha^2} \varepsilon_{ik}, \quad k > k^*$$

# B

## Apéndice

### B.1. Metadatos ACS

El banco de datos utilizado en la aplicación ACS, propiedad del servicio de cardiología de el Complejo Hospitalario Universitario de Santiago, está constituido por 811 observaciones procedentes del registro de casos incidentes en el periodo de Septiembre del 2003 a Marzo del 2007. Se disponen de 43 variable por observación de las cuales solo 11 se involucraron en el modelo utilizado en esta memoria (ver Tabla B.1)

Variable	Formato	Descripción
catereal	byte	Cateterismo.
tab	factor	No fuma, fuma, exfumador.
mdrd	float	Fallo renal.
edad	int	Edad del paciente en el momento del ACS.
sca	factor	Elevación o no del segmento ST.
sexo	float	Sexo del paciente.
ci_prev	byte	Cirugía previa.
hta	byte	Hipertensión arterial.
anemia	byte	Anemia.
gluc_a	float	Glucosa en ayunas.
tiempo	float	Tiempo de supervivencia después del ACS.
exitus	float	Status del paciente.
diab	byte	Diabético o no diabético.

**Tabla B.1:** Descripción y de las variables de los datos ACS utilizadas en el modelo

## B.2. Selección Óptima del Modelo ACS

El modelo de regresión empleado es el de Cox and Oakes (1984) porque permite una modelización multivariante de la supervivencia, con la que se puede cuantificar para cada individuo la evolución temporal del riesgo de fallecer desde la fecha de incidencia, teniendo en cuenta la influencia de las características individuales de cada paciente.

El modelo de partida en R usado por Cid-Álvarez et al. (2009) se muestra a continuación:

```
R> modelo.inicial<-coxph (Surv(tiempo,exitus)~
  catereal + tab + mdrd + edad + sca + sexo + ci\_prev + killip
+ hta + anemia + ns(gluc\_a), knots= c(74,105,126)),
data=belen)
```

$$\lambda(t) = \lambda_0 \exp(\beta_1 \text{catereal} + \beta_2 \text{tab} + \beta_3 \text{mdrd} + \beta_4 \text{edad} + \beta_5 \text{sca} + \beta_6 \text{sexo} + \beta_7 \text{ci\_prev} + \beta_8 \text{killip} + \beta_9 \text{hta} + \beta_{10} \text{anemia} + \beta_{11} \text{ns}(\text{gluc\_a}))$$

Se muestran en la Tabla B.1 algunos de los modelos probados y sus correspondientes AIC.

Modelo	df.	Criterio AIC	Variabes
Inicial	15.000	1716.514	catereal, tab, mdrd, edad, sca, sexo, ci_prev, killip, hta, anemia, gluc_a
Prueba 1	13.000	1715.681	catereal, tab, mdrd, edad, sca, ci_prev, killip, hta, anemia, gluc_a
Prueba 2	14.000	1714.574	catereal, tab, mdrd, edad, sca, sexo, ci_prev, killip, anemia, gluc_a
Prueba 3	13.000	1713.189	catereal, tab, mdrd, edad, sca, sexo, killip, anemia, gluc_a
Final	12.000	1712.237	catereal, mdrd, edad, sca, sexo, ci_prev, killip, anemia, gluc_a

**Tabla B.2:** AIC resultante para diferentes modelos probados.

El modelo óptimo seleccionado fue aquel que resultó con menor AIC, conjuntamente evaluando el cambio de los coeficientes para evitar la salida de variables de confusión, se aplicó a manera referencial el algoritmo Stepwise.

Se ajustó por igual este modelo a diabéticos y no diabéticos con la finalidad de que las diferencias apreciadas en esta memoria sean producto únicamente de las

curvas ROC distintas aplicadas y no a la utilización de diferentes modelos en estos grupos, con lo que se entiende que este es un modelo de referencia demostrativo.

Para el modelo final este es el resume:

```
R> summary(modelo.final)
```

```
Call: coxph(formula = Surv(tiempo, exitus) ~ catereal_f + mdrd
+ edad + sca + sexo_f + ci_prev_f + killip_f + anemia_f + ns(gluc_a,
knots = c(74, 105, 126)), data = belen, x = TRUE)
```

```
n= 811
```

	coef	exp(coef)	se(coef)	z	Pr(> z )
catereal_f1	-0.335320	0.715109	0.186454	-1.798	0.072112 .
mdrd	-0.012761	0.987320	0.003774	-3.381	0.000722 ***
edad	0.060212	1.062061	0.010236	5.882	4.05e-09 ***
scascaest	0.238159	1.268911	0.183107	1.301	0.193377
sexo_fMen	0.447458	1.564330	0.186018	2.405	0.016152 *
ci_prev_f1	0.131854	1.140942	0.176642	0.746	0.455398
killip_f2	0.770004	2.159775	0.179950	4.279	1.88e-05 ***
anemia_f1	0.347157	1.415038	0.181138	1.917	0.055298 .
ns(gluc_a)1	-1.045208	0.351619	0.937793	-1.115	0.265047
ns(gluc_a)2	1.038113	2.823883	0.665753	1.559	0.118924
ns(gluc_a)3	0.497103	1.643951	2.023021	0.246	0.805897
ns(gluc_a)4	0.268934	1.308568	1.064171	0.253	0.800487

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

	exp(coef)	exp(-coef)	lower.95	upper.95
catereal_f1	0.7151	1.3984	0.49621	1.0306
mdrd	0.9873	1.0128	0.98004	0.9947
edad	1.0621	0.9416	1.04097	1.0836
scascaest	1.2689	0.7881	0.88628	1.8167
sexo_fMen	1.5643	0.6393	1.08640	2.2525
ci_prev_f1	1.1409	0.8765	0.80706	1.6130
killip_f2	2.1598	0.4630	1.51787	3.0731
anemia_f1	1.4150	0.7067	0.99216	2.0181
ns(gluc_a)1	0.3516	2.8440	0.05595	2.2096
ns(gluc_a)2	2.8239	0.3541	0.76587	10.4121
ns(gluc_a)3	1.6440	0.6083	0.03118	86.6738
ns(gluc_a)4	1.3086	0.7642	0.16254	10.5347

```
Rsquare= 0.224 (max possible= 0.903 )
```

```
Likelihood ratio test=205.7 on 12 df, p=0
```

```
Wald test= 181.4 on 12 df, p=0
```

```
Score (logrank) test = 225.7 on 12 df, p=0
```

# C

## Apéndice

### C.1. Detalles del Software Utilizado

Todo el análisis estadístico se llevo a acabo utilizando el software R, que se encuentra disponible gratuitamente en:

*<http://www.r-project.org/>*

Se emplearon específicamente:

- **survivalROC**: Es un paquete de R para el cálculo de la curva ROC tiempo dependiente acumulativa dinámica, siguiendo las definiciones de sensibilidad, especificidad y curva  $ROC^{A/D}$  de Heagerty et al. (2000). Se empleó de este paquete la función `survivalROC()` para obtener los resultados mostrados en el Capítulo 4 correspondientes a las  $ROC_t^{A/D}$ , siendo necesario para la construcción de algunas tablas los valores de  $TPF$  y  $FPF$  puntuales rescatados de la función `survivalROC()` mediante `curva$FP` y `curva$TP` respectivamente, además para los valores y gráficas de  $AUC^{A/D}$  se rescato de la función `survivalROC()` el parámetro `curva$AUC`, finalmente con las áreas obtenidas para cada curva ROC se generaron las gráficas de curvas AUC, como se muestra por ejemplo en la Figura 4.8. Para más información sobre la función `survivalROC()` ver:

*<http://finzi.psych.upenn.edu/R/library/survivalROC/html/survivalROC.html>*

- **risksetROC**: Es un paquete de R para el cálculo de la curva ROC tiempo dependiente incidente dinámica, siguiendo las definiciones de Heagerty and Zheng (2005). Este paquete se basa en dividir el conjunto de riesgo en el tiempo  $t$  en casos y controles. Se emplearon de este paquete las funciones `risksetROC()` y `risksetAUC()` para obtener los resultados mostrados en el Capítulo 4 correspondientes con las  $ROC_t^{I/D}$ , siendo necesario para la construcción de algunas tablas valores de  $TPF$  y  $FPF$  puntuales rescatados de

la función `risksetROC()` mediante `curva$FP` y `curva$TP` respectivamente. Además, las gráficas de las curvas  $ROC_t^{I/D}$  también se obtuvieron a través de la función `risksetROC()` especificando como argumento `plot=TRUE`. Para más información sobre la función `risksetROC()` ver:

<http://finzi.psych.upenn.edu/R/library/risksetROC/html/risksetROC.html>

- `survival`: Es un paquete de R para el análisis de supervivencia, específicamente de este paquete se utilizó la función `Surv()` empleada para crear un objeto de tipo supervivencia (una variable), usada como variable respuesta del modelo multivariante para la aplicación de la sección 4.2. sobre Síndrome Coronario Agudo. Para más información sobre la función `Surv()` ver:

<http://finzi.psych.upenn.edu/R/library/survival/html/Surv.html>

- `coxph()`: Es una función del paquete `survival` de R, utilizada para ajustar un modelo de riesgos proporcionales, aquí se uso específicamente para obtener el modelo de Cox empleado en la sección 4.2. Otros detalles se pueden ver en el Apéndice B, y en:

<http://finzi.psych.upenn.edu/R/library/survival/html/coxph.html>

- `extractAIC()`: Es una función del paquete `stats` de R, utilizada para extraer el *Akaike An Information Criterio* (AIC), de un modelo ajustado, se aplica en la selección óptima del modelo (Apéndice B).

<http://finzi.psych.upenn.edu/R/library/stats/html/extractAIC.html>

- `lowess()`: Es una función del paquete `stats` de R, utilizada para realiza los cálculos del suavizado LOWESS con regresión polinómica localmente-ponderada, en esta memoria se aplica en las gráficas de dispersión del biomarcador AKI tanto baselina como longitudinal (sección 4.1.1, 4.1.2), siguiendo lo aplicado en Pepe et al. (2008), para dar idea de la diferencia entre los valores del biomarcador AKI severo y AKI medio. Para más información sobre la función `lowess()` ver:

<http://finzi.psych.upenn.edu/R/library/stats/html/lowess.html>

# D

## Apéndice

### D.1. Resumen de Notación

$D$	Status verdadero de la enfermedad.
$D, \bar{D}$	Subíndice para la población enferma y no enferma respectivamente.
$\rho$	Prevalencia de la enfermedad.
$Y$	Resultado del test o prueba diagnóstica.
$n_D, n_{\bar{D}}$	Número de sujetos enfermos y sanos en el estudio.
$c$	punto de corte o umbral.
$T_i$	Tiempo de supervivencia del sujeto $i$ .
$FPF$	Fracción de falsos positivos.
$TPF$	Fracción de verdaderos positivos.
$p$	Fracción variable de falsos positivos $p \in (0, 1)$ .
$\lambda(t)$	Función de riesgo.
$\Lambda(t)$	Función acumulada de riesgo.
$C$	Tiempo de censura.
$S(y)$	Función de supervivencia para $Y$ .
$ROC$	Curva Característica Operativa del Receptor.
$AUC$	Área bajo la curva ROC (Area Under the ROC Curve).
$\phi$	Función de distribución acumulativa normal.
$\Phi$	Función de densidad normal estándar.
$N(\mu, \sigma^2)$	Distribución normal con media $\mu$ y varianza $\sigma$ .
$\mathbb{A}$	Casos acumulativos.
$\mathbb{I}$	Casos Incidentes.
$\mathbb{E}$	Controles Estáticos.
$\mathbb{D}$	Controles Dinámicos.

# Bibliografía

- Akritis, M. (1994). Nearest neighbor estimation of a bivariate distribution under random censoring. *Annals of Statistics*, 22:1299–1327.
- Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*, 12:387–415.
- Borgueño, M., García-Bastos, J., and González-Buitrago, J. (1995). Las curvas roc en la evaluación de las pruebas diagnósticas. *Med Clin*, 104:661–670.
- Cai, T. and Pepe, M. (2002). Semi-parametric roc analysis to evaluate biomarkers for disease. *J Am Stat Assoc*, 97:1099–1107.
- Cai, T., Pepe, M. S., Zheng, Y., Lumley, T., and Jenny, N. S. (2006). The sensitivity and specificity of markers for event time. *Biostatistics*, 7:182–197.
- Cid-Álvarez, B., Gude, F., Cadarso-Suárez, C., Gonzalez-Babarro, E., Rodríguez-Álvarez, M., García-Acuna, J., and González-Juanatey, J. (2009). Admission and fasting plasma glucose for estimating risk of death of diabetic and nondiabetic patients with acute coronary syndrome: nonlinearity of hazard ratios and time-dependent comparison. *American Heart Journal.*, 158(6):989–997.
- Cox, D. and Oakes, D. (1984). *Analysis of Survival Data*. London.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220.
- DeLong, E., Vernon, W., and Bollinger, R. (1985). Sensitivity and especificity of a monitoring test. *Biometrics*, 41:947–958.
- Egan, J. (1975). *Signal Detection Theory and ROC analysis*. New York.
- Etzioni, R., Pepe, M. S., Longton, G., Hu, C., and Goodman, G. (1999). Incorporating the time dimension in receiver operating characteristic curve: a case study of prostate cancer. *Med Decis Making*, 19:242–251.
- Green, D. and Swets, J. (1966). *Signal Detection Theory and Psychophysics*. New York.
- Halley, E. (1693). An estimate of the degrees of the mortality of mankind. *Philosophical transactions*, 196.

- Hanley, J. and McNeil, B. (1982). The meaning and use of the area under a receiver operating characteristic(roc) curve. *Radiology*, 143:29–36.
- Heagerty, P. and Zheng, Y. (2005). Survival model predictive accuracy and roc curves. *Biometrics*, 61:92–105.
- Heagerty, P. J., Lumley, T., and Pepe, M. S. (2000). Time-dependent roc curves for censored survival data and a diagnostic marker. *Biometrics*, 56(2):337–344.
- Kalbfleisch, J. and Prentice, R. (1980). *The Statistical analysis of failure time data*. Wiley.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481.
- Leisenring, W., Pepe, M., and Longton, G. (1997). A marginal regression modelling framework for evaluating medical diagnostic test. *Stat Med*, 16:1263–1281.
- Lusted, L. (1971). Decision - making studies in patients management. *Journal of Medicine*, 284:416–424.
- Parker, C. and DeLong, E. (2003). Roc methodology within a monitoring framework. *Stat Med*, 22:3473–3488.
- Pepe, M. S. (2003). *The Statistical Evaluation of Medical test for Classification and Prediction*. Oxford.
- Pepe, M. S., Zheng, Y., and Jin, Y. (2008). Evaluating the roc performance of markers for future events. *Lifetime Data Analysis*, 14(1):86–113.
- Peterson, W., Birdsall, T., and Fox, W. (1954). The theory of signal detectability. transactions of the ire professional group on information theory. *PGIT*, 4:171–212.
- Sackett, D., Haynes, R., and Tugwell, P. (1989). *Epidemiología clínica, una ciencia básica para la medicina clínica*. Madrid.
- Slate, E. and Turnbull, B. (2000). Statistical models for longitudinal biomarkers of disease onset. *Statistics in Medicine*, 19:617–637.
- Song, X. and Zhou, X. (2008). A semiparametric approach for the covariate specific for curve with survival outcome. *Statistica Sinica*, 18:947–965.
- Swets, J. (1973). The relative operating characteristic in psychology. *Science*, 182:990–1000.
- Swets, J. and Pickett, R. (1982). *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. New York.
- Wang, T., Gona, P., Larson, M., Tofler, G., Levy, D., Newton-Cheh, C., Jacques, P., and Rifai, N. (2006). Multiple biomarkers for the prediction of first major cardiovascular events and death. *N Engl J Med*, 355:2631–2639.

- Xu, R. and O'Quigley, J. (2000). Proportional hazard estimate of the conditional survival function. *Journal of the Royal Statistical Society Series B*, 62:667–680.
- Yerushalmy (1947). Statistical problems in assessing methods of medical diagnosis, with special reference to x-ray techniques. *Pub Health*, 62:1432–49.
- Zheng and Heagerty (2007). Prospective accuracy for longitudinal markers. *Biometrics*, 63:332–341.
- Zheng, Y. and Heagerty, P. J. (2004). Semiparametric estimation of time-dependent roc curves for longitudinal marker data. *Biostatistics*, 5:615–632.
- Zhou, X., Obuchowski, N., and McClish, D. (2002). *Statistical Methods in Diagnostic Medicine*. Wiley, New York.
- Zweig, M. (1992). Inefficiency of diagnostic efficiency[carta]. *Clinical Chemistry*, 38:163–164.
- Zweig, M. and Campbell, G. (1993). Receiver operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39:561–577.