

Capítulo 2

DIAGNÓSTICO Y PRONÓSTICO

ASPECTOS METODOLÓGICOS Y CLÍNICOS

Carlos D. Tajer, Hernán Doval y Daniel Ferrante

INTRODUCCIÓN

La elaboración diagnóstico-pronóstica a través de la utilización de criterios semiológicos y métodos complementarios es un arte tan complejo como la terapéutica aplicado a la identificación de la enfermedad responsable del padecimiento o la estimación del riesgo de que ocurran complicaciones. El objetivo es configurar un modelo conceptual o hallar el criterio o método ideal, es decir, aquel que distingue totalmente a los pacientes enfermos de los sanos, o mejor, que nos dice si nuestro paciente está enfermo o sano, y cómo será su evolución. Los procesos para evaluar métodos diagnósticos y pronósticos son similares, aunque su validez se compara con un criterio de verdad reconocido (gold standard diagnóstico) o con el desarrollo de un evento en el seguimiento (gold standard pronóstico).

El diagnóstico puede tener un enfoque ontológico (precisar la "enfermedad". p. ej., apendicitis) u operativo (establecer un "síndrome" que orienta a una conducta: abdomen agudo quirúrgico). En el primer caso el acierto diagnóstico puede evaluarse respecto de un hallazgo específico, como el estudio anatomopatológico del apéndice, mientras que en el otro adquiere su validez cuando se compara con una conducta alternativa, por ejemplo no operar.

Las nuevas definiciones diagnósticas reflejan cambios en la mirada médica acordes a la necesidad de evaluar nuevas conductas, agrupando en forma diferente datos de la realidad clínica, e implica en casi todos los casos una estimación pronóstica que nos ayuda en la adopción de decisiones que en sí mismas conllevan riesgo (cirugía, anticoagulación, etc.).

A pesar de los avances y, en algunas patologías, justamente por la presencia de nuevas técnicas ultrasensibles que reconocen disfunciones en estadios asintomáticos cuya importancia clínica es difícil de evaluar, carecemos de métodos ideales y debemos adoptar una actitud que contemple la presunción clínica y la contribución relativa de cada paso, en el conocimiento de que todo criterio o estudio tiene cierto margen de error.

Cuando solicitamos una prueba, interpretamos su contribución de acuerdo con su resultado y lo conjugamos con el diagnóstico presuntivo previo, para juzgar una probabilidad diagnóstica nueva. Cuanto mayor sea la exactitud y confiabilidad de la prueba, mayor será la modificación de nuestro concepto previo según su resultado y así podemos aplicar pruebas sucesivas o escalonadas.

En esta nueva edición, el libro cuenta con un capítulo de evaluación pronóstica posinfarto y, a su vez, reorienta las estrategias de prevención primaria de acuerdo con estimaciones de riesgo y sistemas de puntajes. Esto ha motivado la inclusión de esta introducción, en la cual desarrollaremos aspectos metodológicos y conceptuales de la evaluación diagnóstica y pronóstica, para confluir luego en una interpretación de su uso en la vida real de la consulta.

INTERPRETACIÓN DE UN MÉTODO DIAGNÓSTICO.

Un paciente añoso nos consulta porque padeció un episodio sincopal breve. En el examen detectamos un soplo sistólico eyectivo que nos hace considerar la posibilidad de una estenosis aórtica grave. Pero nuestro enfoque tiene limitaciones: si bien es cierto que la estenosis aórtica raramente cursa sin soplo, por el contrario, el soplo es un hallazgo frecuente en el anciano y en la mayor parte de los casos no indica estrechez sino sólo esclerosis valvular. Este conocimiento clínico surge de comparar el soplo sistólico en el anciano con el diagnóstico establecido, por ejemplo, a través del eco-Doppler, que se toma como criterio de verdad o gold standard, con cifras que volcamos en la **tabla 1**. Evaluemos el criterio "soplo sistólico" para el diagnóstico de estenosis aórtica del anciano con la aplicación de las diferentes nomenclaturas. Éstas aún resultan "difíciles" de incorporar al lenguaje médico cotidiano y se resumirán en un breve glosario, para luego discutir algunos aspectos conceptuales vinculados con la medicina basada en la evidencia.

Tabla 1.
Tabulación y nomenclatura de estudios diagnósticos

	ENFERMOS Estenosis aórtica	SANOS Sin estenosis aórtica	
Prueba positiva <i>soplo sistólico</i>	a Verdaderos positivos 195	c Falsos positivos 700	a + c Total de positivos 895
Prueba negativa <i>sin soplo</i>	b Falsos negativos 5	d Verdaderos negativos 300	b + d Total de negativos 305
	a + b Total enfermos 200	c + d Total Sanos 1000	a + b + c + d Total casos 1200

Sensibilidad, especificidad, valor predictivo positivo y negativo.

Sensibilidad.

Es la capacidad de un método de reconocer a los enfermos como tales (estudio positivo frente a la enfermedad), de evitar diagnósticos falsos negativos. También puede denominarse tasa de verdaderos positivos.

$$\text{sensibilidad} = \frac{\text{verdaderos positivos}}{\text{falsos negativos} + \text{verdaderos positivos}} = \frac{a}{a + b}$$

En la estenosis aórtica del anciano, la presencia de soplo sistólico es casi constante: sensibilidad = $195/200 = 0,98$ o 98%, es decir, el criterio soplo es sensible; es rara la estenosis aórtica sin soplo, es decir, no tiene falsos negativos. Una sensibilidad perfecta indicaría que todos los pacientes enfermos lo tienen, es decir, que ningún paciente sin el criterio presente estará enfermo. Como ejemplo, dado que el centellograma de ventilación/perfusión tiene una sensibilidad elevada para detectar tromboembolia pulmonar, un resultado negativo nos orienta a descartar el diagnóstico. Cuanto más sensible es un método, mayor es su valor predictivo cuando es negativo.

Valor predictivo negativo. (VPN)

Porcentaje de pacientes que estarán sanos cuando la prueba es negativa. Depende en gran medida de la sensibilidad. $\text{VPN} = 300/305 = 0,98$ ó 98%, elevado.

$$\text{Valor predictivo negativo} = \frac{\text{verdaderos negativos}}{\text{verdaderos negativos} + \text{falsos negativos}} = \frac{d}{b + d}$$

Especificidad.

Es la capacidad de reconocer a los sanos como tales. Un método con especificidad ideal es aquel que no tiene falsos positivos, es decir, que siempre que el resultado fuera positivo el paciente estaría enfermo (estudio negativo frente a la salud). Es sinónimo de tasa de verdaderos negativos. El soplo es muy sensible, pero muchos ancianos sin estenosis aórtica (sanos) tienen soplo, es decir, tiene baja especificidad, muchos falsos positivos. $\text{Especificidad} = 300/1.000 = 0,3$ ó 30%. Un ejemplo de método muy específico es la troponina ante la sospecha de un episodio coronario agudo: no tiene falsos positivos o son excepcionales, por lo que su positividad confirma el diagnóstico (alto valor predictivo positivo).

$$\text{Especificidad} = \frac{\text{verdaderos negativos}}{\text{verdaderos negativos} + \text{falsos positivos}} = \frac{d}{c + d}$$

Valor predictivo positivo. (VPP)

Porcentaje de enfermos cuando la prueba es positiva. Está relacionado con la especificidad, dado que ambos dependen del porcentaje de falsos positivos.

En este caso, $\text{VPP} = 195/895 = 0,22$ ó 22%.

$$\text{Valor predictivo positivo} = \frac{\text{Verdaderos positivos}}{\text{verdaderos positivos} + \text{falsos positivos}} = \frac{a}{a + b}$$

Criterios para aplicar los métodos diagnósticos al paciente individual. El concepto bayesiano.

Prevalencia, likelihood ratio, probabilidad preprueba

y posprueba de enfermedad luego de un resultado negativo o positivo.

Para la interpretación del resultado de una prueba en el paciente se requerirá un enfoque "por etapas": el pensamiento clínico parte de una presunción de "prevalencia" del problema que es necesario diagnosticar y aplica estudios diagnósticos para multiplicar la posibilidad de confirmar el diagnóstico o de excluirlo; en términos metodológicos esto consiste en aplicar a una prevalencia previa el valor/resultado de una prueba (likelihood ratio positivo y negativo) para conocer la posibilidad diagnóstica posterior a la prueba.

Prevalencia.

En epidemiología, es el cociente entre el número de pacientes enfermos en un momento determinado y el número total de personas contemplado para la comparación. En el consultorio clínico, la "prevalencia" es la probabilidad que estimamos en un paciente de un diagnóstico determinado, en base a su presentación clínica y la semiología. Esto es lo que se denomina técnicamente probabili-

dad preprueba, es decir, previa a la aplicación de métodos diagnósticos más complejos. Cuando aplicamos un método de screening poblacional, nos basamos en conceptos epidemiológicos (p. ej., mamografías en grupos etarios con prevalencia elevada de tumores malignos). En el otro extremo tenemos la indicación frente a una fuerte presunción clínica (ECG ante dolor precordial prolongado).

Incidencia.

Es el número de casos nuevos de un problema que se desarrollan durante un período y, obviamente, es un parámetro clave en los estudios de seguimiento. La prevalencia de un problema poblacional es la suma de los casos viejos aún no curados y de los casos nuevos (incidencia).

Likelihood ratio (LR) positivo o negativo.

La traducción literal de este concepto es "tasa de probabilidad", que expresa poco de su extraordinario valor en el razonamiento clínico. Una traducción más conceptual es "multiplicador de chances", dado que el LR es el factor que multiplica la chance u odds de estar enfermo.

$$\text{Likelihood ratio positivo (LR+)} = \frac{\text{Sensibilidad}}{1 - \text{especificidad}}$$

$$\text{Likelihood ratio negativo (LR-)} = \frac{1 - \text{Sensibilidad}}{\text{Especificidad}}$$

En el ejemplo, $\text{LR+} = 0,98 / (1 - 0,3) = 0,98/0,7 = 1,39$; $\text{LR-} = (1-0,98) / 0,3 = 0,02/0,3 = 0,08$.

Ante la presencia de soplo, la chance de tener estenosis aórtica se multiplica por 1,39, aumenta el 39%; ante su ausencia se multiplica por 0,08, es decir, disminuye 12 veces. Así, una prueba que tenga un LR+ de 3 implica que la chance de estar enfermo sospechada en un paciente determinado se multiplica tres veces. Un LR- de 0,05 indica que la chance de estar enfermo se reduce 20 veces cuando la prueba es negativa. Tomado en forma individual, el LR positivo o negativo informa mucho más sobre una prueba que los parámetros clásicos y su aplicación clínica es más sencilla.

Probabilidad preprueba y posprueba.

La probabilidad preprueba es la prevalencia del problema de acuerdo con la epidemiología o la presentación clínica y la posibilidad posprueba es la prevalencia sospechada cuando el resultado es positivo o negativo.

Concepto bayesiano y estimación de la probabilidad preprueba y posprueba.

El concepto bayesiano ha sido de extraordinaria utilidad didáctica y conceptual para la selección de estudios y su interpretación.

El Reverendo Thomas Bayes dedicó parte de su vida a las matemáticas y enunció un teorema que siempre consideró incompleto e imperfecto y por ello decidió mantenerlo inédito, pero se publicó post mortem en 1763 (1). El teorema se aplicaba a tratar de corregir el cálculo de resultados ulteriores en base a los resultados de los primeros experimentos y dio lugar a la fundación del concepto de probabilidad condicional. En el ejemplo anterior, cuál es la probabilidad de tener estenosis aórtica en función de la presencia o la ausencia de la condición soplo.

Muy resumidamente, sin entrar en aspectos matemáticos del teorema que son absolutamente sub-

sidiarios¹ y aplicado al tema médico, afirma que cuando evaluamos la probabilidad de que algo ocurra o esté presente o se logre, esta probabilidad estará condicionada por el resultado de la prueba y por lo que sabíamos previamente sobre el tema. El pensamiento bayesiano comenzó a aplicarse a la denominada teoría de la decisión (que aplica cuantificación a la incertidumbre) y a fines de la década de los cincuenta médicos radiólogos plantearon la posibilidad de utilizarlo para el diagnóstico médico, considerando que también éste se daba en un marco de incertidumbre, con conocimientos limitados e imperfectos.

Conceptualmente, el tema es sencillo: Cuando enviamos a un paciente a efectuarse una prueba, cualquiera que ésta sea, interpretaremos los resultados de acuerdo con los hallazgos de la prueba y el diagnóstico presuntivo previo. En base a estos dos elementos, consideraremos una probabilidad diagnóstica nueva, diferente de la que teníamos antes del resultado de la prueba. Cuanto mayor sea la exactitud y confiabilidad de la prueba, mayor será la modificación de nuestro concepto previo. Si aún mantenemos una duda, podemos aplicar una segunda prueba diagnóstica, que interpretaremos tomando en cuenta la primera impresión diagnóstica y el resultado del primer estudio.

En forma muy simplificada podríamos enunciar el aporte del enfoque bayesiano al diagnóstico afirmando que la probabilidad de un diagnóstico determinado luego de conocer el resultado de una prueba dependerá del valor de ésta, el likelihood ratio positivo o negativo, derivado de su sensibilidad y especificidad, y de la probabilidad previa a la prueba de que ese diagnóstico esté presente (prevalencia u odds preprueba).

Odds o chance posprueba = odds preprueba x likelihood ratio.

Dada una prevalencia, por ejemplo, del 20% de enfermedad, expresamos el riesgo como el cociente de enfermos sobre el total de población = $20/100 = 0,20$ ó 20%. El odds o chance es el cociente entre enfermos y sanos = $20/80 = 0,25$.

Aunque es menos exacto, incluso podemos olvidar el término odds y expresar como concepto (no aritmético) que la probabilidad posprueba es función de la probabilidad preprueba y el valor de la prueba expresado en sus LR.

Probabilidad o prevalencia previa a la prueba.

La probabilidad preprueba parte siempre de un dato epidemiológico de prevalencia, sobre el cual podemos adicionar parámetros. Si en el caso anterior en el interrogatorio surge que el paciente tiene angina de pecho, multiplicaría aún más la posibilidad de que estemos frente a una estenosis aórtica.

En el ejemplo del diagnóstico de enfermedad coronaria, uno de los mejor estudiados, si conocemos que la prevalencia a los 65 años es del 25%, la probabilidad preprueba es de 0,25 y el odds 0,33. Si además analizamos si el paciente es sintomático o asintomático, su prevalencia se modifica. El likelihood ratio positivo de angina de esfuerzo típica es de 200, de manera que si la consulta es por angina, el odds posprueba es = $\text{odds preprueba} \times \text{LR}(+) = 0,33 \times 200 = 66$. Convirtiendo odds a probabilidad: $\text{probabilidad} = \text{odds} / (1 + \text{odds}) = 66/67 = 0,99$ ó 99%. Si pensamos indicar un estudio complementario, antes de efectuar la prueba sumamos para el cálculo todo el conocimiento previo, en forma del todo similar a la aproximación clínica.

Desde el punto de vista clínico, en ausencia de pruebas perfectas, es decir, sin falsos resultados negativos y/o positivos, la mayor utilidad de su aplicación se da cuando tenemos una duda diagnósti-

1. Se desarrolla en extenso en el artículo de Probabilidad Condicional contenido en el CD del libro.

ca importante: es poco útil cuando partimos de una prevalencia del 90% ó 2% y mucho más relevante en prevalencias del 30-70%.

Ejemplo clásico del valor de la ergometría.

Supongamos que la ergometría positiva por depresión del ST ≥ 2 mm tiene un LR positivo de 5, es decir, multiplica cinco veces la chance previa de tener obstrucciones en las arterias coronarias. Si hacemos un screening en jóvenes atletas asintomáticos menores de 20 años con una prevalencia de enfermedad coronaria muy baja, 1 en 100.000, la prueba positiva aumenta la posibilidad posprueba a 5 en 100.000, es decir, sigue siendo altamente improbable que exista coronariopatía. La inmensa mayoría de las pruebas positivas por ST en atletas jóvenes serán falsas. La prueba no tiene valor en ese grupo etario para detectar coronariopatía.

De esta manera, el concepto bayesiano nos ayuda a decidir si tiene sentido aplicar un método y cuál es el valor que le daremos al resultado, y esto es incuestionable desde la óptica clínica, conceptual y práctica. Si una persona tiene una probabilidad muy alta de estar sano, un resultado positivo en una prueba habitualmente refleja un falso positivo, y viceversa, si la probabilidad de estar enfermo es muy alta, un resultado negativo habitualmente es falso negativo.

Métodos diagnósticos no dicotómicos. Puntos de corte y curva ROC.

La mayoría de los métodos biológicos no tienen una interpretación dicotómica (positivos o negativos), sino valores graduales (niveles de troponina, glucemia, colesterol, magnitud y extensión de la depresión del segmento ST, etc.). Para interpretar su valor clínico, el análisis se asemeja al anterior: establece la utilidad diagnóstica de cada uno de los niveles.

Discutiremos un ejemplo a través de la relación entre la glucemia en ayunas y el diagnóstico de diabetes. Simulamos un relevamiento poblacional: el método para evaluar es la glucemia ambulatoria y el criterio de verdad "diabetes" se establecerá a través del resultado de la curva de tolerancia, nuestro "gold standard". Nos interesa saber si la glucemia nos ayuda a determinar la presencia de diabetes y cuál es el nivel (punto de corte) que discrimina mejor el diagnóstico.

Glucemia basal	Diabéticos (Curva +)		No diabéticos	Total
	n	% del total del grupo		
< 90 mg/dl	1	0,25	399	400
90-109 mg/dl	1	0,33	299	300
110-129 mg/dl	20	10	180	200
130-159 mg/dl	29	58	21	50
≥ 160 mg/dl	49	98	1	50
Total	100	10	900	1.000

La prevalencia de diabetes en nuestra población es del 10% y se incrementa progresivamente de acuerdo con el nivel de glucemia basal: sólo 1/400 (0,25%) cuando la glucemia es menor de 90 mg/dl y 49/50 (98%) cuando supera 160 mg/dl.

Nos formulamos dos preguntas:

1) **Utilidad del método en general:** ¿La glucemia en ayunas sirve para detectar pacientes diabéticos

que se diagnosticarán por curva de tolerancia?

2) **Punto de corte:** ¿Cuál es el mejor nivel de glucemia para utilizar en un screening?

Una forma sencilla y conceptual de analizar esta información es evaluar diferentes niveles, agrupando toda la población para cada corte en sólo dos grupos, por encima y por debajo del nivel establecido, como hemos volcado en la **tabla 3**.

	Sensibilidad	Especificidad	Índice de Young (sensibilidad + especificidad)	LR+	LR-
< ó > 90	99	44	143	1,78	0,02
< ó > 110	98	78	176	4,4	0,03
< ó > 130	78	98	176	32	0,23
< ó > 150	49	99	148	436	0,50
Método ideal	100	100	200		

LR+ ó -: likelihood ratio positivo o negativo

Observamos que a medida que incrementamos el punto de corte a niveles mayores de glucemia, perdemos en sensibilidad y ganamos en especificidad. Ésta es una regla general de los métodos diagnósticos de resultados graduales: en la medida en que somos menos exigentes con el criterio de diagnóstico positivo (niveles de corte más bajos) incluimos más diagnósticos falsos positivos y evitamos diagnósticos falsos negativos, de manera tal que ganamos sensibilidad y perdemos especificidad. Si, por el contrario, elevamos el punto de corte para el diagnóstico, perdemos sensibilidad (más falsos negativos) y ganamos especificidad (menos falsos positivos). La decisión del mejor punto de corte desde la óptica matemática es el punto que más se acerca al 100% de sensibilidad y especificidad. Sumando sensibilidad + especificidad (índice de Young) en el caso de 100% de ambas obtenemos 200. El nivel de glucemia que implique un resultado más elevado al sumar sensibilidad + especificidad indica el mejor punto de corte. Sobre la base de esta tabla podemos contestar ambas preguntas afirmativamente:

- 1) El dosaje de glucemia contribuye a la detección de diabetes: sí, dado que incluso cualquiera de los puntos de corte incrementa el valor diagnóstico con LR positivos cercanos a 2 en el primer caso y muy superiores al ir incrementando el punto de corte, así como LR negativos muy bajos con los niveles menores.
- 2) El mejor punto de corte puede ser 110 ó 130.

Aspectos matemáticos y clínicos de la elección del punto de corte.

La determinación del mejor punto de corte en un método o medición requiere considerar la sensibilidad y la especificidad para cada punto referido y luego cuál será el verdadero objetivo de su implementación. Cuando el objetivo es de screening, es decir, detectar posibles casos para continuar luego con otros estudios diagnósticos, nos interesa un método sensible, con pocos falsos negativos. Sinónimos: un mejor valor predictivo negativo, LR- elevado. De esa manera confiamos que los pacientes que excluimos para una evaluación posterior estén sanos. Si, en caso contrario, pretendemos identificar un diagnóstico con mucha certeza, para tomar decisiones críticas que pueden ser peligro-

sas, buscaremos un método específico, con muy pocos falsos positivos y algunos falsos negativos: sinónimos de mayor especificidad y valor predictivo positivo, LR+ elevado. Combinando el conocimiento de las propiedades del método a través de las evaluaciones que efectuamos con anterioridad, ajustamos el punto de corte a los fines de nuestras decisiones clínicas.

Curvas ROC.

Una forma didáctica de graficar el análisis de los diferentes niveles de puntos de corte de un método, y eventualmente compararlo con otro, es la curva ROC, del inglés, Receiver Operator Characteristics. Esta terminología surge de curvas que se utilizaban para evaluar la capacidad de los operadores de radares para distinguir entre señales con significado (avión enemigo al acecho) y el ruido de fondo. En el libro de Sackett (2), lo han traducido como COR: curva de la Característica con que Opera el Receptor. La curva se construye relacionando los niveles de sensibilidad (tasa de verdaderos positivos) y 1 - especificidad (tasa de falsos positivos).

De esta manera, el extremo superior izquierdo es el punto de nivel óptimo: sensibilidad = 1, especificidad = 1, 1 - especificidad = 0.

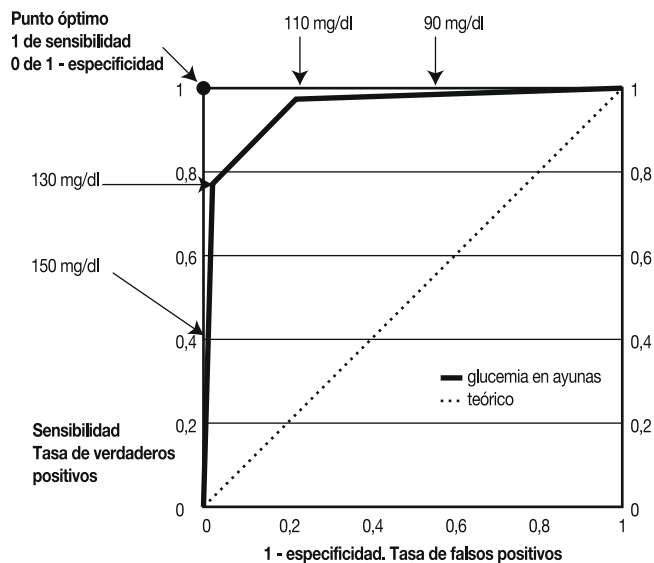


Figura 1.

En la figura vemos la curva ROC construida con los puntos de corte que hemos determinado en el análisis de la glucemia del ejemplo anterior.

La línea diagonal representa el azar absoluto. Considerando un valor de 1 para el área total del cuadrado, divide dos áreas iguales de 0,5. Un método aleatorio, por ejemplo, tirar la moneda, cara o ceca, para determinar si el paciente es diabético o no nos daría un área bajo la curva de 0,5. Expre-

sado en otros términos, la sensibilidad sería del 50%, la especificidad del 50%, el LR positivo y negativo 1, de manera que su contribución será nula (la chance previa a la prueba se multiplica por uno, es decir, no cambia). Con línea gruesa hemos unido los puntos de diferente nivel de sensibilidad y 1 - especificidad de la glucemia en ayunas. El área bajo la curva representa el valor diagnóstico de este método estimada de 0,92 (0,87-0,97). El método contribuye en forma significativa cuando el intervalo de confianza del área no cruza 0,5, nivel del azar. Existen diferentes programas que calculan el área exacta de la curva, aportan información interesante sobre el valor general del método y permiten su comparación con otras metodologías respecto del mismo diagnóstico: ¿cuál hubiera sido el área si en lugar de glucemia en ayunas se hubiera dosado hemoglobina glicosilada? En este ejemplo, el área bajo la curva es muy grande, abarca casi toda la superficie posible y los mejores puntos de corte son equidistantes del vértice de punto óptimo.

El área abarcada por la curva ROC, también denominada índice C (del inglés, C Index) puede categorizar la exactitud diagnóstica del método de acuerdo con la siguiente cuantificación (3):

- 0,5 (50%) = contribución nula (azar absoluto, tirar la moneda)
- 0,5-0,6 = fallido
- 0,6-0,7 = pobre
- 0,7-0,8 = aceptable
- 0,8-0,9 = bueno
- 0,9- 1 = excelente

Decisión epidemiológica en la aplicación de métodos diagnósticos.

Con la expansión de la ergometría en las décadas de los cincuenta y sesenta, se planteó efectuar ergometrías como "screening" masivo y adelantarse a los graves problemas coronarios de aparición súbita. La inmensa mayoría de los resultados positivos correspondían a diagnósticos falsos respecto de la presencia de enfermedad coronaria, y un dato aún más importante: el hallazgo de depresión del ST no pronosticaba eventos coronarios súbitos, que, como aprendimos luego, se relacionan con nuevos accidentes coronarios y rápida progresión. Cuando aplicamos un estudio en forma masiva, en realidad estamos tratando de adelantarnos al "pronóstico" a través del diagnóstico de formas silenciosas sobre las que no se consulta. Ésta es la intención de la mamografía, el Papanicolaou, el PSA, los chequeos cardiovasculares, etc. En la interpretación del resultado no sólo se debe comparar con el diagnóstico, sino con el impacto que la extensión de esa medida tiene sobre la salud, hechos no necesariamente paralelos. Llevando el concepto al extremo, si hacemos PSA en pacientes de 75 años asintomáticos y derivamos a los que tienen PSA positivo a una secuencia agresiva de biopsia y prostatectomía de acuerdo con el hallazgo, correremos el riesgo de hacer muchas biopsias en pacientes sin tumores (el 60-70% de los PSA elevados no tienen cáncer) y de detectar tumores con una chance menor de inducir muerte del 20% a los 10 años (4). Aplicamos además una cirugía que tiene su mortalidad y una tasa de morbilidad considerable (disfunción sexual, incontinencia) para un paciente asintomático. Vemos que el gold standard del diagnóstico de grandes grupos poblacionales no es la "enfermedad anatómica" sino el impacto que la secuencia de su aplicación tiene sobre la salud real. En tal sentido recientemente se ha cuestionado la mamografía, con el argumento de que su aplicación masiva no se ha asociado con una reducción de la mortalidad por cáncer de mama, lo cual despertó encendidas polémicas (5).

II. Interpretación área ROC: si tomamos todos los pares posibles formados por un sano y un enfermo, un área de 0,92 implicaría que el 92% de las veces el enfermo tendría un valor del test mayor que del sano. También indica el área de no superposición entre la distribución del test en sanos y enfermos.

La evaluación diagnóstico-pronóstica.

Utilidad y limitaciones del concepto bayesiano.

El pensamiento bayesiano en diagnóstico y pronóstico parte de varias premisas que se aplican a la práctica clínica:

- 1) Siempre que solicitamos una prueba, contamos con una idea previa sobre la posibilidad de que este paciente particular esté enfermo del problema que exploramos.
- 2) Todos los métodos diagnósticos tienen limitaciones en su precisión, de manera tal que pueden informar un resultado "falso" tanto cuando confirman como cuando descartan el problema.
- 3) Cuando se analiza el informe de una prueba, la conclusión que elaborará el médico tomará en cuenta el resultado de la prueba y la presunción diagnóstica previa. En base a esos dos elementos, el médico deberá decidir si considera confirmado o descartado el diagnóstico, si aplicará otra técnica o si comenzará un tratamiento. Este pensamiento se aplica a las decisiones médicas cotidianas: el hallazgo de QS en V1 y r minúscula en V2 y V3 en un adolescente inicialmente se considera normal, pero el mismo trazado en un anciano que padeció dolor precordial prolongado hace poco se interpreta como una secuela anteroseptal. En ambos casos combinamos una presunción previa derivada de la prevalencia (muy baja chance de enfermedad coronaria en la adolescencia, muy alta en un adulto que nos relata un dolor precordial prolongado) y del resultado del estudio (el hallazgo ECG es inespecífico, no del todo común, e incrementa la posibilidad de la presencia de enfermedad). Como en el primer caso la prevalencia es muy baja, ese pequeño incremento de posible enfermedad es insignificante, y en el segundo caso tiene mayor peso.

¿Este pensamiento puede ser equivocado? Por supuesto. El adolescente al poco tiempo puede desarrollar un episodio de síncope y descubrirse en un ecocardiograma una miocardiopatía hipertrófica de forma no obstructiva que le da explicación al hallazgo electrocardiográfico. Y nuestro adulto puede padecer una úlcera gastroduodenal y por su obesidad tener el corazón horizontalizado. Pero dentro de la imprecisión aceptada para las conductas médicas, la primera interpretación es aceptable. Vemos que en la práctica el pensamiento bayesiano funcionó adecuadamente: utilizamos la presunción previa de la enfermedad, el valor relativo del interrogatorio y del electrocardiograma y elaboramos una presunción posprueba.

Limitaciones del pensamiento bayesiano.

El enfoque tiene algunas limitaciones que cuestiona sus pretensiones cuantitativas y, en algunos casos, hasta cualitativas.

En esencia, el razonamiento bayesiano exige que el likelihood ratio positivo o negativo de la prueba, o dicho en términos prácticos, el resultado de la prueba no esté influido por la prevalencia de la patología. De manera tal que si el likelihood ratio (+) de la depresión del segmento ST de 2 mm es de 8, esto no se vea modificado por la edad o el sexo. Esto nos permite, conociendo la contribución del sexo y la edad, estimar la probabilidad posprueba.

En realidad, la sensibilidad y la especificidad de una prueba no son "fijas" o "invariables": surgen de estudios en determinadas poblaciones delimitadas por factores demográficos y clínicos sobre los que se efectúa un promedio, y cuando el método se evalúa nuevamente en poblaciones diferentes, cambian

parcialmente su sensibilidad y especificidad. Para poder aplicar el razonamiento matemático que utilizan las tablas para el diagnóstico de enfermedad coronaria, en el cual se combinan edad, sexo, interrogatorio y ergometría, la sensibilidad y la especificidad de la ergometría no deberían verse influidas por el sexo, la edad y el interrogatorio clínico. A su vez, la relación de cada criterio con el diagnóstico tampoco debería verse influida por los criterios restantes, algo casi imposible: el relato de angina atípica tiene un valor diferente en los ancianos que en los jóvenes, etcétera.

La fuente de error más frecuente en la interpretación del resultado de un estudio diagnóstico proviene de las diferencias entre la población en la cual se probó originariamente el método y la población que se está evaluando, denominado sesgo de espectro ("spectrum bias"):

En la primera etapa de evaluación de un método se eligen enfermos "muy enfermos" (casos severos) y sanos "muy sanos" (para que otros factores no confundan el resultado). En este escenario la sensibilidad será muy elevada, pero en estudios ulteriores al incluir pacientes menos graves la sensibilidad descenderá. La inclusión de pacientes "muy sanos" implica también que hay pocos falsos positivos por otras enfermedades que generan confusión en la interpretación, de manera que la especificidad de la prueba también será elevada. Al aplicarse luego a una población más representativa de la práctica clínica, los falsos positivos aumentan.

Un ejemplo clásico en este sentido es la caída de la fracción de eyección de esfuerzo en el ventriculograma radioisotópico de esfuerzo. En una primera etapa ese comportamiento se consideró altamente específico, es decir, siempre que caía la fracción de eyección existía coronariopatía, dado que las poblaciones estudiadas incluían pacientes coronarios o muy sanos. Cuando el método se extendió a otros problemas cardiológicos, se observó que en las miocardiopatías, las valvulopatías y aun en la hipertensión arterial la fracción de eyección puede caer sin que exista afectación coronaria.

Una forma pragmática de adaptar los resultados de un estudio a nuestros pacientes es ajustar (disminuir a mano) la sensibilidad y la especificidad de acuerdo con los siguientes criterios:

Sensibilidad: cuanto más severa es la población del estudio en relación con nuestra práctica clínica, más deberá reducirse la sensibilidad esperada en nuestros pacientes.

Especificidad: si el estudio incluye pacientes voluntarios sanos, deberá reducirse la especificidad esperada si evaluamos pacientes con otras enfermedades.

La segunda limitación del análisis bayesiano es que existen factores que influyen en el resultado de la prueba en forma no relacionada con su prevalencia, los denominados factores no bayesianos. Expresado en otros términos, la positividad de la prueba tiene una información diferente.

El valor diagnóstico del relato de angina típica de esfuerzo es diferente en hombres que en mujeres de edad mediana. Dado que en las mujeres la prevalencia de enfermedad coronaria es menor, inicialmente se supuso que esto podía interpretarse en el contexto lógico del enfoque bayesiano, es decir, la probabilidad posprueba (+) de enfermedad coronaria, en este caso de angina típica en el interrogatorio, era menor en las mujeres porque la prevalencia de enfermedad era menor. El sexo en ese caso sería un factor bayesiano, su interacción con el diagnóstico se explica por la prevalencia de enfermedad. Sin embargo, aun corrigiendo para la prevalencia, el valor encontrado de enfermedad coronaria en mujeres era muy lejano al esperado de acuerdo con el cálculo. Esto indicaba que existía algún motivo diferente de la prevalencia para explicar el menor valor diagnóstico de la angina en mujeres que en hombres, quizás atribuible a la mayor prevalencia de angina por mecanismos diferentes de la coronariopatía (síndrome X). Expresado en otros términos, ser mujer es un factor *no bayesiano*, es decir, condiciona la interpretación de la prueba en forma independiente de la prevalencia. Es muy probable que

este ejemplo en una patología muy estudiada como la enfermedad coronaria anatómica nos ayude a comprender lo imperfecto de una aproximación que presume condiciones constantes para resultados de determinadas pruebas diagnósticas. En la vida real esto no es así, aun cuando nos independicemos de la variabilidad de resultados entre instituciones, equipos, etc., y en el ejemplo anterior, de quien desarrolla el interrogatorio para establecer si la angina es típica o no.

EVALUACIÓN PRONÓSTICA Y PUNTAJES (SCORES) DE RIESGO

Decisiones médicas y riesgo individual

La categorización del riesgo individual es parte esencial del acto médico. El pronóstico es individual y absoluto (morir o no, infartarse o no) pero su multideterminación (causalidad múltiple) y la dificultad de asegurar lo que ocurrirá en cada persona nos lleva a intentar ubicar al individuo en "grupos" o "subgrupos" de riesgo de acuerdo con su "patrón clínico" y de esa manera estimar su riesgo en términos porcentuales y poblacionales. Así, al comentar la mortalidad del infarto, afirmamos que un primer episodio con localización inferior no supera el 4%, la de un segundo infarto el 14%, etc. Este pensamiento "probabilístico" es esencial en la adopción de conductas cotidianas en la práctica médica: muchas de las decisiones en cardiología se toman sobre la base del riesgo supuesto en el conocimiento de la "evolución natural" contrastada con el riesgo de las intervenciones o procedimientos y no para el alivio sintomático. Sin embargo, dentro de un mismo grupo de pacientes existen otros múltiples subgrupos, tantos como pacientes: siempre factores subjetivos irrepetibles condicionan, por ejemplo, el resultado de una intervención. Así, la incapacidad de afrontar adecuadamente las situaciones conflictivas de la vida (p. ej., evaluadas con el score de Rosebaum) multiplica el riesgo quirúrgico; intuimos muchas veces que por razones anímicas es conveniente postergar intervenciones por meses y psicoterapeutas entrenados pueden asesorarnos con mayor fundamento aun en los tiempos y eventuales estrategias. A su vez, hay instituciones y cirujanos con mejores resultados que otros, y otros etcéteras que hacen de la decisión final un dominio intransferible de la relación médico-paciente.

No sólo las "grandes" decisiones de intervención están condicionadas por el pensamiento probabilístico. La mayor parte de las conductas actuales del consultorio cardiológico en prevención, insuficiencia cardíaca, posinfarto, angina crónica, diabetes, etc., están basadas más en un concepto de riesgo poblacional que en aliviar síntomas o mejorar la calidad de vida actual del paciente. El tratamiento de la hipertensión arterial, la dislipemia, la aspirina preventiva, la espironolactona, entre otras, brinda un aporte sintomático actual nulo y un poderoso efecto de reducción de eventos graves en el futuro.

La caracterización del riesgo individual es parte crítica en la toma de decisiones terapéuticas de prevención primaria y secundaria.

Puntajes de riesgo y el problema del colesterol y la presión arterial

¿Tratar la hipercolesterolemia o la hipertensión, o bajar el colesterol y la presión?

Hemos aprendido en los últimos años que muchas de las variables fisiológicas en las que efectuábamos la dicotomía en normal o patológica de acuerdo con determinados niveles y así delimitábamos enfermedades como la hipertensión o la hipercolesterolemia, se comportan en realidad como varia-

bles continuas en su relación con el riesgo. Este tema se discute con más detalle en el capítulo de prevención primaria, pero introduciremos el concepto aquí para reflexionar sobre la utilidad de los índices y scores. En un metaanálisis sobre un millón de pacientes seguidos a 12 años para caracterizar su riesgo de acuerdo con la presión, los autores informaron una relación continua (6).

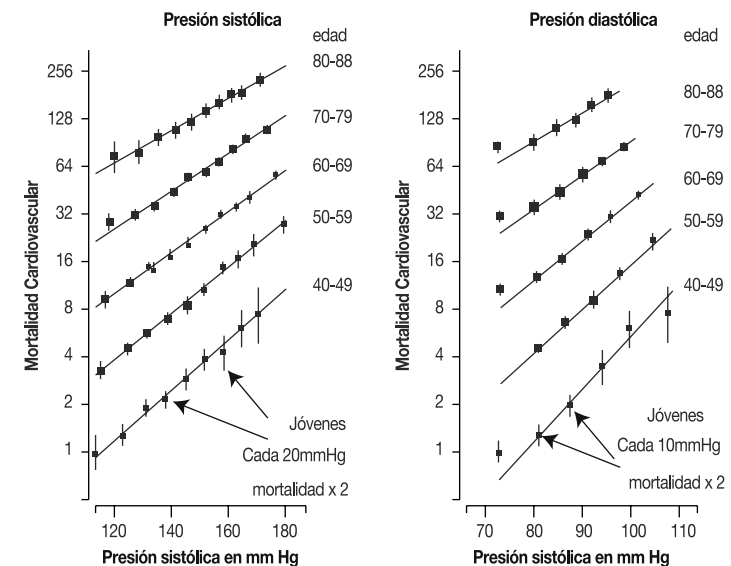


Figura 2. Relación entre presión y riesgo

Partiendo de valores de presión arterial sistólica de 115 mm Hg y diastólica de 75 mm Hg, la relación con el riesgo se incrementa en cada grupo etario sin un punto de inflexión que permita establecer el nivel normal; algo similar ocurre con el colesterol. La decisión de tratar o no dependerá de la posibilidad de hacer impacto sobre el pronóstico con la medicación balanceando sus costos y riesgos. En pacientes diabéticos sin antecedentes cardiovasculares, bajar la presión de 130 a 120, ambas en rango "normal", puede aportar un gran beneficio para evitar ataques cardíacos o cerebrovasculares. El estudio HPS (7) nos ha enseñado que cualquier colesterol es "alto" si el paciente ha padecido un infarto. De acuerdo con ese estudio, los pacientes con niveles de colesterol LDL de 90 mg/dl basal luego de un evento coronario o con otros factores de alto riesgo cardiovascular, si son tratados con simvastatina 40 mg y su LDL desciende a 55-60 mg/dl, disminuyen su riesgo en forma similar a una reducción de esa misma magnitud en pacientes con LDL de 150. Ante la pregunta del paciente: "¿para qué voy a tomar eso si yo no tengo colesterol?" nos vemos obligados a responder que ese "poco" es mucho para sus arterias y que bien vale la pena reducirlo.

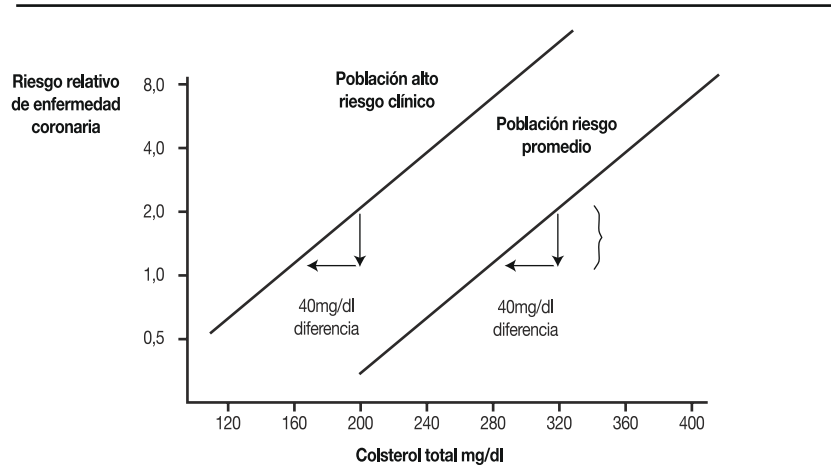


Figura 3. Hipótesis del estudio HPS y riesgo cardiovascular. Los autores plantearon (y confirmaron con los resultados del ensayo) que si se reducía el colesterol 40 mg/dl en pacientes con colesterol normal o bajo pero con riesgo cardiovascular elevado (antecedentes coronarios, diabetes), la reducción relativa de enfermedad coronaria sería similar a la obtenida con la reducción de 40 mg/dl de colesterol en pacientes hipercolesterolémicos con riesgo cardiovascular promedio.

También para la decisión del uso de aspirina el consenso norteamericano ha propuesto un lineamiento conceptual muy inteligente, ajustado al riesgo clínico del paciente (8).

Tabla 4.
Efectos esperados de beneficio o daño con la terapia con aspirina para pacientes con diferentes niveles de riesgo para eventos coronarios

Efectos de la aspirina sobre los eventos	Riesgo de eventos cardiovasculares a 5 años de acuerdo con scores (Framingham o similares)		
	1%	3%	5%
Mortalidad	No cambia	No cambia	No cambia
Eventos coronarios evitados	3 (1-4)	8 (4-12)	14 (6-20)
ACV isquémicos	No cambia	No cambia	No cambia
ACV hemorrágicos provocados	1 (0-2)	1 (0-2)	1 (0-2)
Hemorragias digestivas mayores provocadas	3 (2-4)	3 (2-4)	3 (2-4)

¿Como han razonado los autores?

Consumir aspirina incrementa el riesgo de accidente cerebrovascular hemorrágico y de hemorragias digestivas. Este incremento supera el beneficio en pacientes con bajo riesgo espontáneo de acuerdo con el score de Framingham (que discutimos más adelante) y es muy inferior al beneficio en pacientes con alto riesgo de eventos. Dado que el uso de aspirina en prevención primaria no ha demostrado que baje la mortalidad, la comparación entre riesgo de morbilidades es muy válida. *En la decisión de bajar o no el colesterol, la presión o indicar aspirina debemos caracterizar el riesgo individual del paciente.* La necesidad diaria de decidir de acuerdo con el riesgo y la cada vez mayor utilización de puntajes

pronósticos justifican una discusión conceptual y técnica de su interpretación y valoración.

¿Cómo podemos conocer el riesgo del paciente individual?

Una persona de 50 años nos consulta y plantea su inquietud sobre la necesidad de adoptar conductas de prevención o no. Lo primero que seguramente interrogaremos es la existencia de "factores o marcadores de riesgo": si tiene presión alta, si fuma, su nivel de colesterol, antecedentes familiares, actividad física, exceso de peso, etc. Se han publicado más de 250 variables que se asocian con mayor o menor riesgo de padecer un evento coronario: tomar vino con las comidas, dormir la siesta, número de relaciones sexuales semanales, en el caso del varón estar casado con mujeres con título universitario, consumo de vegetales y frutas, los niveles plasmáticos de homocisteína, proteína C reactiva, lipoproteína A, etc. Es importante distinguir un "factor de enfermedad" de un "marcador". En el caso del colesterol, su reducción inducida por fármacos retrasa o previene el desarrollo anatómico y clínico de la enfermedad aterosclerótica en la misma magnitud porcentual esperada de acuerdo con los cálculos epidemiológicos. Al contrario, la homocisteína elevada es un marcador de riesgo, pero aún no contamos con información que confirme que su reducción prevenga enfermedad vascular.

Procedimiento lógico-estadístico y configuración de un puntaje

La identificación de la asociación entre un parámetro y el riesgo se hace en varias etapas: primero se establece si está asociado en forma individual y luego se evalúa en forma conjunta con otras variables si esta asociación es "real" o si está mediada por otro factor "confundidor". Por ejemplo, los bebedores de café tienen mayor posibilidad de infarto. Sin embargo, el consumo de café se asocia con el hábito de fumar, y si corregimos para fumadores y no fumadores, el consumo de café no aporta al riesgo. En ese caso el tabaquismo funciona como un factor "confundidor". La asociación individual se efectúa a través de análisis univariados y luego en forma conjunta se aplican modelos en los que se incluyen varios factores en un análisis multivariado (9,19). El procedimiento más utilizado para establecer la relación cualitativa y cuantitativa entre un factor o marcador de riesgo para un evento determinado es la regresión logística múltiple. Este procedimiento es complejo y debe ser efectuado por profesionales muy entrenados; sin embargo, el clínico debe comprender su armazón conceptual para obtener un mejor provecho de la información: en todo caso, una mejor comprensión permitirá analizar en forma crítica las utilidades y las limitaciones de los puntajes en su aplicación clínica. La regresión logística tiene relación conceptual con la regresión lineal simple y múltiple, bastante más sencilla de entender.

Regresión lineal múltiple.

El procedimiento consiste en buscar una "función" que permita predecir un valor de una variable conociendo valores de otras. Supongamos que queremos relacionar el peso con la estatura. Peso sería la variable "dependiente" que queremos deducir y estatura la variable independiente, a partir de cuyo valor intentaremos calcular el peso. Para este ejemplo, se utilizó una base de datos de adolescentes mendocinos de 12 a 14 años provista por el Dr. Emilio Marigliano. Para explicar el concepto intentaremos estimar el peso partiendo de la estatura, para lo cual construiremos una fórmula conceptual **Peso = f (estatura)**, es decir, peso en función de la estatura.

Con el procedimiento de regresión lineal entre peso y altura, obtenemos los siguientes parámetros estadísticos:

r: coeficiente de correlación = 0,58, IC 0,54-0,62;

r²: coeficiente de determinación = 0,34;

La función se expresa como recta de regresión : $Y = a + b \cdot X$

peso = -85 kg + 0,865 kg/cm x altura en cm

Interpretación de los resultados.

De acuerdo con esta fórmula, el peso estimado para un adolescente de 150 cm de altura es = 44,9 kg. Cada cm más de altura implica 0,865 kg más de peso esperado. El coeficiente de regresión r nos dice que existe una buena correlación entre la talla y el peso (la ideal es 1) y el coeficiente de determinación r^2 nos enuncia que sólo el 0,34 (34%) de la variación del peso puede surgir de la valoración de la talla, es decir, que nuestro modelo tiene una capacidad muy limitada para predecir el peso individual. Esto coincide con la evidente percepción clínica de que personas de la misma estatura tienen pesos muy diferentes.

Si intentamos ser más precisos para predecir el peso no podemos restringirnos sólo a estimar la altura, sino explorar si existen otras variables que se asocian con el nivel de peso corporal. En este caso identificamos que el peso era mayor en mujeres que en varones, y en hijos únicos que en los que tenían hermanos. También la edad tenía relación con el peso, pero no fue significativa al considerarla en conjunto con la talla: la talla contenía toda la información que la edad aportaba o, lo que es igual, la edad no fue una variable independiente para estimar el peso. Efectuando el cálculo de regresión múltiple con inclusión de todas las variables que continuaron siendo significativas nos surge la fórmula:

$$Y = a + \beta_1 \times X_1 + \beta_2 \times X_2 + \beta_3 \times X_3$$

$$\text{Peso} = -95 + 0,87 \text{ kg/cm} \times \text{talla cm} + 2 \text{ kg} \times \text{sexo femenino} + 3,7 \text{ kg} \times \text{hijo único}$$

Esta fórmula indica que el peso aumentará en promedio 0,87 kg por cm de altura, que las mujeres tendrán 2 kg más que los hombres y que los hijos únicos agregan 3,7 kg a su peso. El r^2 del modelo con las tres variables alcanzaba 0,36, es decir, que mejoró el modelo que incluía sólo la talla (0,34). Este modelo sólo explica el 36% de las variaciones de peso entre los adolescentes, pero puede ser suficiente para tomar conciencia de que la condición de hijo único expone al joven a un riesgo de tener un promedio de 3,7 kg más en forma independiente de otros parámetros y podría guiar conductas educativas de prevención de obesidad.

Una vez configurado el modelo, debemos evaluar si éste cumple nuestra pretensión cuantitativa (con qué exactitud estamos pronosticando el peso en base a los parámetros considerados) y no sólo conceptual (el peso está relacionado con el sexo, la estatura y la condición de hijo único en la pubertad). Una forma sencilla de estimar la exactitud de nuestro modelo es calcular el peso de acuerdo con la fórmula en cada individuo y compararlo con el que realmente tiene. Si nuestra fórmula fuera "perfecta" podríamos predecir con exactitud el peso real, y a mayor "imperfección" mayor distancia entre el peso real y el calculado. Esto en gran medida puede adelantarse por el conocimiento del r^2 del modelo, en este caso 0,36. Nuestro modelo será sólo una aproximación, dado que no podemos presumir que estas tres variables contienen toda la información necesaria para estimar el peso que

tendrá un adolescente a esa edad, aun en esa población restringida. Y todavía nos quedaría validar que esta fórmula se mantiene igual en otras poblaciones culturalmente diferentes para que adquiera validez general y no restringida a la muestra de la cual se obtuvo.

Análisis de variables relacionadas con el pronóstico.

En el ejemplo anterior hemos construido una "fórmula" sumando factores con su contribución individual para estimar un parámetro cuantitativo. Para evaluar el riesgo se procede en forma similar, con la diferencia de que intentamos establecer si ocurrirá o no un evento dicotómico (sí o no) y no un valor cuantitativo. Para este cálculo asignaremos un valor de 1 ó 0 de según el evento ocurra o no^{III}.

Se repetirán los pasos anteriores: identificación de las variables que en forma individual se asocian con la mortalidad, construcción de una fórmula con asociación de variables y configuración de un modelo con las que mantienen su valor independiente. La evaluación del ajuste del modelo a la realidad es algo diferente: es posible comparar el valor de riesgo estimado para cada paciente (p ej., riesgo 0,3 ó 30% de que padezca un evento) con el valor del evento, que sólo puede ser 1 (100%) ó 0 (0%) si ocurrió o no. Habitualmente se prefiere construir grupos de riesgo ordenando a los pacientes de acuerdo con el pronóstico establecido y observar si el riesgo esperado según el cálculo se asemeja al observado en esa población. Este proceso se denomina calibración del modelo, es decir, el grado de adecuación entre el modelo y la evolución clínica de la población en la cual se elaboró. Si se demuestra validez interna, aplicaríamos este modelo conceptual a una nueva población para establecer la validez externa del modelo y de esa manera la posibilidad de generalizarlo como un concepto para aplicar en pacientes similares. Cuando se trabaja con grupos grandes, el modelo puede desarrollarse en la mitad de la población y validarse "prospectivamente" en la segunda mitad, pero adquiere mayor confiabilidad cuando se confirma en un grupo independiente de pacientes.

Ejemplo con el uso de un análisis estratificado de riesgo.

En una base de datos de 7.004 pacientes seguidos por seis meses luego de un infarto (11), la mortalidad fue del 3,8%. Tomando en cuenta tres variables, sexo, edad mayor de 65 años y diabetes, en el primer análisis univariado se observó que las tres se asociaron con mayor mortalidad en el seguimiento.

Efectos de la aspirina	Prevalencia (%)	Mortalidad (%)	Odds ratio
Población (n = 7.004 p)		3,8	
Diabetes	10	7,1	2,17 (1,6-3) 0,001
No diabetes	90	3,4	
Sexo femenino	22,4	5,3	1,59 (1,2-2,1) 0,001
Sexo masculino	77,6	3,4	
Edad ≥ 65 años	40	6,2	3 (2,2-3,9) 0,001
Edad < 65 años	60	2,2	

El odds ratio nos da una idea de la chance de morir: se duplica en presencia de diabetes, aumenta el 59% en mujeres y se triplica por encima de los 65 años. (Tabla 5)

III. Los estadígrafos efectúan este análisis utilizando la regresión logística múltiple. En el modelo de regresión lineal los parámetros suman su contribución. **Ecuación de la regresión lineal** $Y = a + b_1 \times X_1 + b_2 \times X_2 \dots b_n \times X_n$.

¿Cada una de estas variables contribuye en forma independiente a estimar la mortalidad?

Los pacientes más añosos tenían una prevalencia mayor de diabetes que los jóvenes (11,5 vs. 9,2%) y también las mujeres eran más añosas: tenían una diferencia de 6 años (mujeres 65 vs. varones 59 años), y expresado por grupos, el 33% de los mayores de 65 años eran mujeres contra el 16% de los menores. Esta asociación entre las variables hace posible que el mayor riesgo observado en el análisis univariado esté en realidad condicionado por las otras. Para poder establecer el valor independiente de cada criterio necesitamos encarar un análisis estratificado (multivariado) del valor de cada uno de los tres parámetros en forma independiente de los otros. (Tabla 6)

Tabla 6.

Mortalidad posinfarto y cuatro variables clínicas en el estudio GISSI II internacional. Análisis multivariado.

Variable	Coefficiente B multivariado	Odds	IC 95%	Puntaje relativo
Diabetes	0,62066	1,9	(1,3 - 2,6)	2
Edad ≥ 65 años	0,96411	2,6	(2- 3,4)	3
Insuficiencia cardíaca	1,02392	2,8	(2-3,8)	3
Infarto previo	0,70811	2	(1,6- 2,6)	2

Se observa que los cuatro parámetros contribuyen en forma significativa, pero la condición de sexo femenino se ha excluido del modelo porque no mantuvo su valor independiente al incorporar los otros datos.

Transformación de un análisis logístico en un score clínico.

El análisis multivariado tiene el objetivo de establecer si cada una de las variables contribuye o no en forma independiente a caracterizar la posibilidad de un evento determinado, lo cual tiene importancia fisiopatológica y conceptual. Para estimar el riesgo individual en la práctica clínica es sumamente útil traducir estos coeficientes y odds en un puntaje que pueda aplicarse individualmente. La construcción de un puntaje o score requiere un análisis pormenorizado de la relación entre cada una de las variables contempladas y el pronóstico. En el ejemplo anterior se requiere establecer un peso relativo para cada una de ellas, que surge en forma sencilla de sus odds: por ejemplo, hemos redondeado el valor del odds y podemos atribuir un valor de 2 si el paciente es diabético o tiene infarto previo, y de 3 si tiene insuficiencia cardíaca o pertenece a un grupo etario mayor. En forma conceptual, esto implica que la chance de muerte se duplica en los diabéticos y se triplica en los pacientes añosos, pero a los fines del puntaje los valores tienen sólo una estimación relativa, es decir, si el antecedente de infarto aporta 2, proporcionalmente el padecer insuficiencia cardíaca contribuye con 3. Si aplicamos este sistema de puntajes en forma individual, un paciente sin ningún parámetro tendrá un valor total de 0 y con los cuatro presentes sumará 10. Adjudicando a cada paciente el puntaje de acuerdo con la presencia o la ausencia del criterio construimos la tabla 7:

Tabla 7.

Mortalidad posinfarto de acuerdo con el puntaje propuesto según el análisis multivariado

Puntaje	Prevalencia		Mortalidad (%)
	n	%	
0	3.089	45	1,4
2-3	2.661	38	3,6
4-5	838	12	9,2
6-10	375	5	12,8

Sobre la base del análisis anterior podemos construir un puntaje de riesgo que en forma clara se asocia con mayor mortalidad en nuestra población.

Este primer cálculo nos confirma que hemos construido un modelo que marca efectivamente grupos con diferente riesgo. La validez externa del modelo, es decir, su posibilidad de extenderlo a la práctica clínica debe surgir de su aplicación a otras poblaciones posinfarto y confirmar si mantiene o no su valor predictivo^{IV}.

Valor del modelo y curvas ROC para métodos pronósticos.

De la misma manera que hemos explicado en el análisis de métodos diagnósticos, podemos evaluar la sensibilidad y la especificidad de cada nivel de puntaje para pronosticar la mortalidad. En los estudios pronósticos, el gold standard o criterio de verdad es la aparición o no del evento pronosticado en la evolución. Ingresamos los resultados anteriores, es decir, el puntaje individual obtenido para cada paciente y el desarrollo o muerte, en el programa Medcalc.be®, lo cual nos permitió establecer una curva ROC y seleccionar el mejor punto de corte. El área bajo la curva ROC hallado fue 0,713 (IC 95% 0,702-0,724) y el mejor punto de corte individual fue > ó ≤ al puntaje 2.

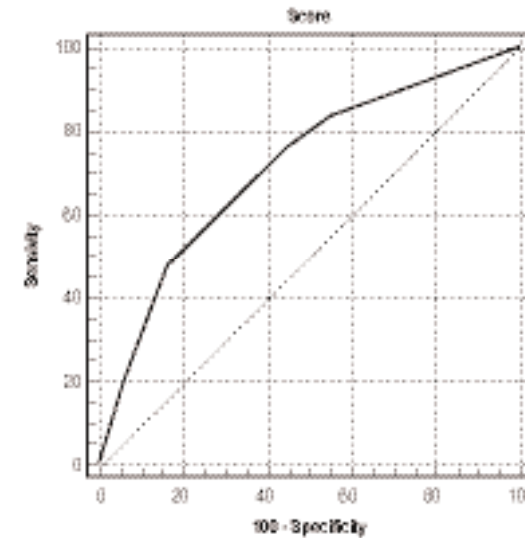


Figura 4. Evaluación del modelo pronóstico de puntaje posinfarto. Curva ROC

El uso de puntajes de riesgo en la práctica cardiológica.

La cardiología ha incorporado a la práctica clínica varios sistemas de puntajes en los últimos años. En la angina inestable e infarto no Q se ha propuesto el score de TIMI, que facilita la decisión de la indicación farmacológica y la derivación eventual a procedimientos invasivos. Para adoptar estrategias de prevención, la conjugación de puntajes de riesgo a través del score de Framingham o el re-

IV. A estos datos se aplica el método de bondad de ajuste de Hosmer y Lemeshow, que compara el porcentaje de mortalidad calculado para cada grupo de acuerdo al modelo con el efectivamente observado. En este caso lo ideal es que el porcentaje de

ciente score europeo tienen un gran valor de orientación para el médico y poder didáctico para el paciente. Discutiremos cada uno de estos puntajes. En el CD que acompaña al libro se proveen planillas de Excel @ diseñadas para su uso en el consultorio donde al ingresar los datos del paciente orienta sobre su puntaje de riesgo y su variación al modificar cada factor.

Puntaje de TIMI para riesgo evolutivo en la angina inestable e infarto no Q.

El grupo TIMI construyó un puntaje de acuerdo con el resultado obtenido en el estudio randomizado TIMI 11 B (12). Delimitaron un grupo inicial para elaborar la prueba a través del análisis de regresión logística, luego lo validaron en forma interna con el resto de los pacientes y en forma externa con los pacientes incluidos en el estudio ESSENCE, con un diseño muy similar. En la tabla 8 se resumen las 11 variables que se analizaron y las 7 que finalmente entraron en el modelo. Dado que los odds eran bastante similares (el más bajo fue 1,51 y el más alto 1,75), los autores adjudicaron para un valor equivalente a cada uno de los siete parámetros, es decir, cada uno aporta un punto al puntaje. De esta manera, un paciente sin criterios de riesgo tendrá un puntaje 0 y aquel con todos los criterios presentes alcanzará 7.

Tabla 8.
Características basales analizadas para el desarrollo del puntaje de riesgo de TIMI para el síndrome coronario sin elevación persistente del segmento ST

Variable	Análisis univariado			Análisis multivariado			puntos
	Coef b	p	OR (95% CI)	Coef b	p	OR (95% CI)	
Edad ≥ 65 años	0,4681	0,001	1,60 (1,25-2,04)	0,5575	0,001	1,75 (1,35-2,25)	1
≥ 3 factores de riesgo	0,3717	0,009	1,45 (1,10-1,91)	0,4336	0,003	1,54 (1,16-2,06)	1
Estenosis coronaria conocida > 50%	0,5473	0,001	1,73 (1,34-2,23)	0,5284	0,001	1,70 (1,30-2,21)	1
Infarto previo	0,2386	0,06	1,27 (0,99-1,63)				
Cirugía coronaria previa	0,3004	0,07	1,35 (0,97-1,88)				
Angioplastia previa	0,4828	0,004	1,62 (1,16-2,26)				
Alteraciones del ST	0,3356	0,02	1,40 (1,06-1,85)	0,4125	0,005	1,51 (1,13-2,02)	1
≥ 2 episodios < 24 h	0,4521	0,001	1,57 (1,24-2,00)	0,4279	0,001	1,53 (1,20-1,96)	1
Aspirina previa (7 días)	0,6179	0,002	1,86 (1,26-2,73)	0,5534	0,006	1,74 (1,17-2,59)	1
Marcadores séricos elevados	0,3486	0,004	1,42 (1,12-1,80)	0,4420	0,001	1,56 (1,21-1,99)	1
Historia de insuficiencia cardíaca	-0,1058	0,70	0,90 (0,53-1,53)				

Existió una relación progresiva entre el puntaje y el riesgo de padecer un evento global (necesidad de cirugía de emergencia, infarto o muerte) o evento mayor (infarto o muerte). (Tabla 9).

El área ROC (o C index) para el puntaje es de sólo 0,65 e incluso menor en el grupo tratado con enoxaparina, 0,63.

Tabla 9.
Relación entre el puntaje de TIMI y el riesgo de evento combinado o muerte.

Puntaje	Prevalencia(%)	Evento combinado (%)	Infarto o muerte (%)
0-1	4,3	4,7	2,9
2	17,3	8,3	2,9
3	32	13,2	4,7
4	29,3	19,9	6,7
5	13	26,2	11,5
6/7	3,4	40,9	19,4

Comentario crítico: en el estudio ENAI (13) y en el relevamiento de angina inestable ECLA III (14), observamos que la angina recurrente tenía un odds ratio de 5 a 8 para el desarrollo de infarto o muerte, comparado con sólo 1,5 de cada uno de los criterios incluidos en el modelo. Esto explica probablemente la baja funcionalidad del puntaje de TIMI al ingreso: no todo lo que va a ocurrir en la angina inestable está contenido en la información de ingreso. La evolución inicial y la respuesta al tratamiento delimitan riesgos muy diferentes. Un puntaje de riesgo inicial es necesario si al ingreso se adoptarán conductas en forma invasiva casi sistemática, como en la práctica norteamericana, y quizá de mucho menos valor si se adopta la práctica de enfriar con 48 h de heparinas de bajo peso molecular, como es la práctica europea [FRISC II (15) y estudio RITA 3 (16)] o en nuestro medio.

Puntajes de riesgo de desarrollar enfermedad coronaria.

Score de Framingham.

El estudio de cohorte Framingham (17) incluyó 2.489 hombres y 2.856 mujeres de 30 a 74 años, que fueron seguidos desde 1949 hasta la fecha. Ha contribuido como ningún otro ensayo a la comprensión del papel de los factores de riesgo en la enfermedad cardiovascular. Sobre la base de su población han desarrollado un puntaje de riesgo que se recomienda como criterio básico para decidir la indicación de aspirina o hipolipemiantes en los últimos consensos.

Tabla 10.
Planilla para el cálculo del puntaje de Framingham en hombres

Presión arterial sistólica/diastólica	Colesterol total mg/dl	Colesterol HDL mg/dl	Edad
< 129 y < 85	< 160	< 35	30-34
≥ 130 y/ó ≥ 85	160-199	35-44	35-39
≥ 140 y/ó ≥ 90	200-239	45-49	40-44
≥ 160 y/ó ≥ 100	240-279	50-59	45-49
	≥ 280	≥ 60	50-54
			55-59
Diabetes sí	Fumador		60-64
Diabetes no	No fumador		65-69
			70-74

Una vez obtenido el puntaje total del paciente se establece el riesgo de eventos coronarios de acuerdo con la tabla 10. El estudio Framingham estableció como evento coronario la aparición de angina de pecho en cualquier forma clínica, infarto o muerte atribuible a patología coronaria. Otro cálculo propuesto permite la predicción de eventos coronarios "duros": infarto o muerte.

Tabla 11.
Riesgo de padecer un evento coronario a los 10 años de acuerdo con el puntaje obtenido con el score de Framingham

Score	<-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Riesgo %	2	3	3	4	5	7	8	10	13	16	20	25	31	37	45	> 53

El índice C, área bajo la curva ROC, es de 0,69 en hombres y 0,72 en mujeres al usar el score y aumenta a 0,74 y 0,77 si se utilizan las variables como dato continuo en un cálculo matemático.

Se sugiere comparar el score con el riesgo esperado para cada grupo etario y de esa manera establecer si el paciente está en un grupo de mayor riesgo que el esperado para la edad.

Tabla 12.

Riesgo de desarrollar enfermedad coronaria a los 10 años en pacientes promedio de cada grupo etario y en el paciente con criterios de bajo riesgo								
Score	30-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74
Riesgo %								
Promedio	< 1	2	5	8	12	12	13	14
Bajo	< 1	2	3	5	7	8	8	8

Interpretación, uso práctico y limitaciones del score de Framingham. Esta propuesta ha tenido amplia aceptación con la ventaja de que sus datos se han recogido en un periodo en el cual los factores de riesgo no se corregían o trataban. Existen programas sencillos para computadoras personales y sistemas PDA (palm o pocket-PC) que permiten un cálculo rápido. Aunque es un criterio opinable, se ha tomado como criterio de alto riesgo de eventos 1% por año, moderado 0,5% y bajo 0,2%. Cabe destacar que los pacientes con antecedentes coronarios cursan con el 2% de eventos por año y, por lo tanto, siempre se consideran de "alto riesgo" respecto de la indicación de medidas preventivas.

Sin embargo, se han señalado limitaciones prácticas y conceptuales:

- 1) Los resultados predictivos del puntaje no son exactos al aplicarse a poblaciones con riesgo diferente. En grupos europeos sobrestima la posibilidad de enfermedad coronaria, incluso para países con alta prevalencia (véase más adelante).
- 2) Los eventos pronosticados no son fáciles de comparar en otras series: contempla todo tipo de angina de pecho, aun sin comprobación clínica, infartos silenciosos como hallazgo electrocardiográfico, etcétera.
- 3) Por su número limitado de pacientes, en el modelo no pueden incluirse otros parámetros como obesidad, sedentarismo, antecedentes familiares, etc. Asimismo el número de pacientes diabéticos es reducido y la predicción estimada es imprecisa.

Aún tomando en cuenta sus limitaciones, el puntaje es de gran utilidad para la exploración de riesgo en forma general y facilita el diálogo médico-paciente acerca del papel y el peso relativo de los factores.

El proyecto SCORE europeo.

El nuevo puntaje europeo de riesgo (18) se ha construido sobre la base de información prospectiva en 205.178 pacientes seguidos por más de 10 años, como producto de un proyecto establecido a comienzos de la década de los noventa. El elevado número de pacientes y eventos ha permitido utilizar como criterio de riesgo la mortalidad cardiovascular, que a los fines de análisis se dividió en coronaria y vascular no coronaria. Las áreas bajo la curva ROC de los modelos oscilaron de 0,71 a 0,84. Utiliza sólo como criterios la edad, el tabaquismo, el sexo, el nivel de presión sistólica y de colesterol total o alternativamente la relación colesterol total/HDL.

Aunque no contempla por separado la diabetes (no existió una definición común prospectiva inicial que permitiera unificar los criterios). en ese caso sugiere multiplicar el riesgo obtenido por 2 en hombres y por 4 en mujeres.

Algunas observaciones epidemiológicas han sido de gran interés para la interpretación de los puntajes:

- 1) Fue necesario dividir la población europea en países de alto y bajo riesgo. Aunque la contribución relativa (coeficiente u odds ratio) de los factores de riesgo es similar (p. ej., el tabaquismo duplica el riesgo en forma homogénea en todos los países), el riesgo absoluto espontáneo es muy diferente: en España la mortalidad por enfermedad cardiovascular a los 65 años se estimó en 3% mientras que en Finlandia (12,8%), Rusia (12%) o Inglaterra (7,9%) es mucho mayor. La Argentina tiene tasas de mortalidad por enfermedad cardiovascular similares a las de países de alto riesgo.
- 2) El porcentaje de mortalidad coronaria como componente de la mortalidad cardiovascular es mayor en países de alto riesgo, mientras que en los de bajo riesgo adquieren mayor peso las causas vasculares no coronarias de muerte.

Hemos desarrollado una planilla de Excel (véase el CD) que permite el cálculo sobre la base de las fórmulas aportadas por el artículo. Resumimos un ejemplo en la tabla 13.

Ser fumador a esta edad intermedia duplica el riesgo de morir por causa cardiovascular a los 10 años.

Tabla 13.

Cálculo del riesgo de mortalidad vascular y por causa en un paciente de sexo masculino (1), edad 55 años, presión sistólica 140 y colesterol total 300				
	a		b	
Sexo	1		1	
Edad	55		55	
Presión sistólica	140		140	
Fumador	1		0	
Colesterol total	300		300	
	Mortalidad 10 años		Mortalidad 10 años	
	Coronaria	No coronaria	Coronaria	No coronaria
	10,1%	1,7%	5,1%	0,9%
Vascular total	11,8%		6%	

a: fumador; b: no fumador.

Interpretación del criterio de alto riesgo en el paciente individual.

La *Task Force Europea* (19) consideró el 20% de eventos cardiovasculares a los 10 años (2% por año) como criterio de alto riesgo como umbral para una intervención activa en la corrección de los factores de riesgo. El Consenso Americano sobre aspirina propuso el 1% por año, pero el *NCEP III* (panel sobre colesterol) el 2% por año de eventos duros (infarto o muerte), es decir, un umbral mucho más alto (29), todos ellos referidos al score de Framingham. Algunos autores han criticado este enfoque, dado que en mayores de 65 años prácticamente todas las personas tienen "alto riesgo". Para evitar la tentación de entrar en una discusión en la que la mezcla de complejos argumentos epidemiológicos, farmacoeconómicos y subjetivos pueden hacerla incomprensible o, lo que es peor, imposible de proyectar al paciente individual, puede proponerse una lectura alternativa.

Dado que nuestro concepto de riesgo no se aplica al de establecer el costo de una póliza de seguro sino a la posibilidad de contribuir a reducirlo a través de intervenciones, lo que buscamos por medio de los puntajes son niveles a partir de los cuales la colaboración de un eventual tratamiento pueda hacer impacto con riesgos bajos de complicaciones. Así como todo paciente luego de un evento coronario es medicado con aspirina, hipolipemiantes, betabloqueantes y en muchos casos inhibidores de la enzima convertidora, nuestro enfoque práctico partirá de balancear el riesgo con el costo y la factibilidad del enfoque-medicación propuesta para su reducción. Queda claro que en personas

menores de 40 años la indicación de medidas preventivas se hará sólo ante marcadores muy elevados, mientras que en personas mayores de 65 años, con diabetes o enfermedades cardiovasculares conocidas la indicación será casi universal. En el grupo etario de 40-65 años es donde los puntajes pueden ayudarnos a tomar decisiones, aunque siempre con un margen de subjetividad y preferencias para discutir con los pacientes, para lo cual pueden ser muy útiles los scores. En forma similar a lo observado al analizar métodos diagnósticos, la utilidad del puntaje será mayor en grupos de riesgo intermedio y contribuirá poco en personas con riesgo muy alto o muy bajo.

Desde una óptica más extrema pero con argumentos muy sólidos, Wald y Law han propuesto no evaluar factores de riesgo luego de los 55 años, sino indicar en todas las personas una "polypill", una única pastilla que contenga dosis bajas de aspirina, estatinas, betabloqueantes, diuréticos tiazídicos, inhibidores de la enzima convertidora y ácido fólico. Con esta intervención estiman que el riesgo de muerte se reducirá un 80% y la vida podría prolongarse en promedio 10 años (21).

El enfoque bayesiano y la decisión médica computarizada.

Durante la década de los setenta y principios de los ochenta, la difusión del enfoque bayesiano y la disponibilidad en los grandes centros de computadoras para efectuar cálculos rápidos de fórmulas complejas llevó a pensar que una buena parte de las decisiones médicas podrían formularse como enunciados matemáticos. Autores brillantes como un oncólogo con formación psicoanalítica, Lucien Israel (22), en un bello libro sobre la estructura de la decisión médica, llegó a proponer la necesidad de sistemas de cálculos que contemplen todos los factores e independicen al médico de la fuerte subjetividad con que toma todas sus decisiones. Los modelos de teoría de la decisión aplicados a la medicina tuvieron su fruto en los modelos de inteligencia artificial, como por ejemplo los programas de apoyo al diagnóstico que se utilizan en algunos centros en guardia o en cuidados intensivos. Sin embargo, su importancia es mucho menor de lo que se esperaba, dado el gran avance de la computación y de las tecnologías de diagnóstico médico en las últimas décadas.

Algunos autores, como Alvan Feinstein (23), han criticado la pretensión cuantitativa de la aplicación del análisis bayesiano como una tabla de *Ouija* moderna (*del francés, Oui y del alemán, ja*, y consiste en una tabla de adivinación de datos o del futuro a semejanza del juego de las copas con letras). Adjudicar valores matemáticos al interrogatorio de la angina de pecho implica una tecnología determinada que es el interrogatorio adecuado y el entrenamiento médico, así como su presunción diagnóstica que lo lleva a interrogar de una u otra manera. También ocurre con muchos de los métodos clínicos y complementarios. Pretender un diagnóstico o una predicción precisa que elimine la incertidumbre de la decisión médica era visto por este autor como una ilusión poco aplicable en la medicina real o, expresado en otros términos, como intentar contener al azar en una fórmula mágica que evite el error. De allí su analogía con los juegos de adivinación.

Scores de riesgo surgidos del análisis de regresión logística.

Una alternativa diferente, aunque en el mismo marco conceptual.

Recientemente se publicaron una serie de ensayos bajo la denominación de "Medicina Científica", y el en tercero de ellos el Dr. Miettinen (24) se refirió a la pretensión de cuantificar el diagnóstico. Primero cuestiona el análisis bayesiano:

"... ¿No varía el likelihood entre subtipos de la enfermedad, por ejemplo con la severidad? ¿No es demasiado am-

plio el número de posibles perfiles para obtener sobre este concepto un aprendizaje-investigación significativo?" Y propone un enfoque alternativo "... El asunto obvio es reconocer, para un perfil diagnóstico dominante definido (digamos en forma amplia para un rango de edad y género del síntoma principal), la prevalencia como una función conjunta del conjunto de indicadores diagnósticos involucrados (edad, particularidades del síntoma, etc.) y evaluar a éste en función del perfil diagnóstico. Con esta mirada, en conjunción con la moderna teoría estadística, *el marco teórico adecuado es el de los modelos de regresión logística para dominios definidos*. En resumen, un sistema de puntajes como el que comentamos anteriormente para riesgo".

Es difícil a nuestro juicio sostener en forma general la validez de esta conclusión desde el punto de vista cuantitativo. Los modelos de regresión logística tienen la ventaja respecto del de Bayes de ser aún más difíciles de comprender por el médico práctico, de manera que paradójicamente pueden generar una confianza mayor en la presunción de que su complejidad va de la mano con su exactitud. Lamentablemente, los cálculos de regresión logística, como todos los multivariados, están fuertemente influidos por el "modelo" utilizado para el análisis. En el ejemplo de score de riesgo posinfarto observamos que el sexo femenino era un factor pronóstico, pero su contribución desaparecía si incluíamos en el modelo la edad y diabetes. En este caso la modificación es conceptual (aumenta el riesgo o no lo hace). En muchos otros casos la influencia es cuantitativa: la obesidad aislada es un factor de riesgo importante, pero a su vez se asocia con sedentarismo, hipertensión, diabetes y dislipemia. Si aislamos estos factores en un análisis multivariado, su contribución individual al riesgo es muy pequeña. Incluso en los scores de Framingham tabaquismo se considera sí o no, cuando sabemos que la cantidad de cigarrillos es una variable fuertemente tóxica, etcétera.

El enfoque práctico ante la consulta.

La aproximación a la estimación diagnóstico-pronóstica siempre ocurre en un marco de incertidumbre:

- 1) No conocemos con exactitud la prevalencia/probabilidad del problema en el paciente que nos consulta ni tampoco el valor diagnóstico exacto de las diferentes pruebas en ese contexto.
- 2) La aplicación cuantitativa del análisis bayesiano tiene grandes límites desde el punto de vista conceptual y la alternativa de la regresión logística está derivada de modelos en continua modificación.
- 3) Si nos equivocamos, podemos ocasionar graves daños a nuestros pacientes.

Desconocer el aspecto multicausal o multideterminado de las enfermedades y la existencia de múltiples factores que ignoramos o no "dominamos" (p. ej., el carácter del paciente y su historia de vida, sin duda relacionados con los eventos coronarios agudos) puede llevarnos a errores profundos. Por otro lado, es ilusorio y probablemente innecesario pretender una aproximación cuantitativa exacta de la probabilidad diagnóstica o pronóstica.

Volvamos a la consulta clínica, terreno fértil para el arte médico, y analicemos en realidad cómo actuamos y cuáles son nuestras dificultades reales:

- 1) Un paciente nos consulta por un síntoma. Queremos saber si su relato nos indica la presencia de una "enfermedad" (en términos de problemas identificables en los cuales su reconocimiento nos permite conocer la implicación pronóstica, riesgos y eventuales beneficios de nuestra intervención). Diferenciar entre "malestares de la vida" (achaque) y enfermedades no siempre es fácil y es nuestro punto de partida. Siempre elegimos entre los síntomas del paciente aquellos que pre-

sumimos de mayor riesgo o importancia médica (a nuestro juicio) y con frecuencia luego somos corregidos por la evolución y los deseos del paciente.

- 2) Todas nuestras decisiones estarán condicionadas por la primera impresión sobre la consulta, por nuestra predisposición a escuchar, nuestro carácter, las experiencias previas en estos cuadros y, en particular, las experiencias recientes.
- 3) El encuadre inapropiado (tiempos breves, contextos inadecuados, mal estado de ánimo o mala predisposición a escuchar, dificultades en el cobro de ese sistema de cobertura) lleva a errores a veces graves.
- 4) La selección de estudios complementarios estará condicionada por el diagnóstico presuntivo (que surge de la prevalencia poblacional y de los signos clínicos reconocidos). Elegiremos el método que mejor información nos dará desde el punto de vista diagnóstico y pronóstico, aun sin precisar su valor cuantitativo. Recordemos que el diagnóstico del paciente siempre es "cualitativo" o discreto: tiene o no un problema determinado que debemos asistir médicamente.
- 5) En la elección de la secuencia de estudio tomaremos en cuenta los resultados de la literatura y nuestra experiencia con los aplicadores de los métodos. No tiene sentido indicar métodos con sensibilidad o especificidad inadecuadas^v. También, por supuesto, evitamos enviar al paciente a operadores que tienen malos resultados. Hay ergometrías de alta calidad y otras donde apenas se ve el trazado en un canal, absolutamente "formales" y sin valor diagnóstico.
- 6) Desde el primer momento, siempre adoptamos conductas de protección frente al problema que eventualmente podría ser grave y luego corregimos la mira al obtener el resultado de los diferentes estudios.

Pero, se dirá, estas afirmaciones son todas perogrulladas: nadie hace otra cosa en el consultorio. Hemos descubierto entonces que en la práctica médica pensamos en términos bayesianos o multivariados, que aplicamos probabilidades condicionales, pero que esto es sólo una parte del complejo accionar médico.

Aplicar la pretensión cuantitativa del pensamiento bayesiano o de regresión logística (scores de riesgo) puede ser muy impreciso, tanto como lo son sus premisas. Con todo, más riesgosa parece la actitud de no contemplar el aspecto conceptual del enfoque bayesiano o logístico: la interpretación diagnóstica y pronóstica de los diferentes signos clínicos o resultados debe hacerse en un contexto abarcador que tome en cuenta todos los elementos disponibles. La confianza exclusiva en el resultado de un análisis o un aspecto clínico, no considerado en el contexto, es más probable que induzca a error que contemplar el "todo".

COMENTARIOS FINALES.

El arte del diagnóstico y la evaluación pronóstica se enriquece por la comprensión profunda de las virtudes y las limitaciones del conocimiento epidemiológico y de la interpretación de los estudios diagnósticos y pronósticos. Estamos (afortunadamente) todavía lejos de fórmulas que nos predigan todo en el paciente, y tampoco ése será a nuestro juicio el aporte genético, pero sí contamos con elementos que sumados a nuestra intuición y empatía nos permitan establecer la aproximación suficiente que facilite la toma de decisiones. También estamos lejos del diagnóstico por inspiración

^v En este aspecto el valor del análisis metodológico es absoluto: si no queremos perder ningún caso utilizaremos métodos de gran sensibilidad y si no queremos tratar ningún caso dudoso seleccionaremos métodos de gran especificidad. Es lógico que en los bancos de sangre se apliquen métodos de detección de HIV muy sensibles; de hecho nuestra aspiración es que ningún paciente

o intuición, aunque ambas son bienvenidas a la práctica médica. Un autor analizó cualitativamente el mecanismo de razonamiento aplicado para acercarse al diagnóstico en los ateneos del *New England Journal of Medicine* (25): "El desafío del diagnóstico diferencial es seleccionar las causas más probables de una condición del paciente, aun cuando la magnitud del problema, la naturaleza de la información médica y la notoria inhabilidad de los seres humanos para manipular probabilidades en sus cabezas conspiran todas contra el diagnosticador de manera tal que virtualmente hacen imposible el empleo del teorema de Bayes en el diagnóstico de rutina. Imposibilitados de estimar las probabilidades deseadas explícitamente, los médicos redimensionan el problema de una manera que permita utilizar una de las más efectivas habilidades mentales, la de comparar patrones. Pasos:

1) **asociar los diferentes hallazgos dentro de patrones.**

2) **selección de un 'pivote' o hallazgo clave.**

3) **generación de una lista de causas.**

4) **podar la lista de causas.**

5) **selección de un diagnóstico.**

6) **validación del diagnóstico.**

Aún cuando las conferencias clinicopatológicas difieren en varios caminos importantes de los problemas del diagnóstico en la vida real, creemos que los principios que se describen aquí reflejan ajustadamente los que se usan en la práctica".

Respecto del pronóstico, vale la pena relatar una anécdota.

Hace años atendimos a un paciente que había padecido un infarto anterior, con compromiso moderado de la función ventricular. Fue ingresado prospectivamente en un protocolo de múltiple evaluación diagnóstico-pronóstica. Se le efectuó una cinecoronariografía que evidenció una lesión total de la arteria descendente anterior, un infarto correspondiente de extensión moderada, fracción de eyección del 45%, ecocardiograma respetando el septum desde la base, sin incremento de la distancia mitroseptal, ausencia de arritmia compleja del Holter y perfusión miocárdica sin isquemia residual. Pudo desarrollar una ergometría hasta alcanzar 600 kgm sin evidencias de angor ni cambios del ST y buena progresión de la presión arterial. Se trataba de un paciente que sumaba factores de buen pronóstico, con excepción de la depresión de la fracción de eyección que no era crítica. En ese momento estábamos trabajando en colaboración con un grupo de psiquiatras que aplicaban una prueba de Rorschach cuantitativa, que evaluaba la capacidad simbólica del paciente con un puntaje. La hipótesis era que la incapacidad de elaborar los conflictos emocionales con riqueza simbólica debía traducirse en incremento del riesgo de enfermedades "somáticas". La operadora salió horrorizada de la evaluación del paciente: nunca había detectado un score tan bajo y nos transmitió que la gravedad del caso era extrema. Con todas las evaluaciones en la mano, incluida la coronariografía, tomamos el tema con algo de humor y condescendencia. Se incluyó al paciente en un grupo de rehabilitación y también psicoterapéutico, dentro de la rutina de aquel momento. Fue el único de su grupo que incumplió el tratamiento, faltaba a las sesiones y en 8 meses abandonó tanto la rehabilitación como el grupo con múltiples quejas. Dos meses más tarde fue internado con un infarto inferior extenso con más de 24 horas de evolución, en franco shock cardiogénico. Tenía en la angiografía una oclusión total de una coronaria derecha dominante, sana en el primer estudio, que lo llevó a la muerte. El riesgo de padecer un nuevo evento, la "accidentalidad", probablemente estaba contenida en la complejidad de su carácter y en alguna nueva situación de vida y no en los parámetros que pudimos evaluar con las

pruebas invasivas y no invasivas específicamente cardiológicas. La habilidad para pronosticar un nuevo infarto, el evento sin tiempo descrito por Carlos Bertolasi, en general es limitada, y en diferentes series no han colaborado la depresión del segmento ST ni la coronariografía, pero sí el hecho de continuar con el hábito tabáquico.

Este relato nos resalta que, aun con la genética y la alta tecnología, en la evaluación pronóstica deban sumarse aspectos que hacen al carácter y a la situación de vida del paciente y es muy probable que la intuición médica o, lo que es mejor, una preparación adecuada en psicología pueda colaborar en la evaluación pronóstica del paciente individual tanto como los métodos complementarios más "técnicos".

Bibliografía

1. Bayes T. An essay towards solving a problem in the doctrine of chances. *Philosophical transactions of the Royal Society of London* 1763;53:370-418.
2. Sacket D, Haynes B, Guyatt G, Tugwell P. *Epidemiología clínica. Ciencia básica para la medicina clínica*. Editorial Médica Panamericana: 1998.
3. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29-36.
4. Perron L, Moore L, Bairati I y col. PSA screening and prostate cancer mortality. *CMAJ* 2002;166:586-91.
5. Olsen O, Gotzsche P. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2001;(4):CD001877.
6. Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* 2002; 360:1903-13.
7. Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20 536 high-risk individuals: a randomised placebo-controlled trial. *Lancet* 2002;360:7-22.
8. U.S. Preventive Services Task Force. Aspirin for the Primary Prevention of Cardiovascular Events: Recommendation and Rationale. *Ann Intern Med* 2002;136:157-60.
9. Hosmer D, Lemeshow S. *Applied Logistic Regression*. 2nd ed. John Wiley & Sons: 2000.
10. Silva Aycaguer L. *Excursión a la regresión logística en ciencias de la salud*. Ediciones Diaz de Santos SA: 1995.
11. GISSI 2. A factorial randomised trial of alteplase versus streptokinase and heparin versus no heparin among 12,490 patients with acute myocardial infarction. *Lancet* 1990;336:65-71.
12. Antman E, Cohen M, Bernink M y col. The TIMI risk score for unstable angina/non-ST elevation MI. A method for prognostication and therapeutic decision making. *JAMA* 2000;284:835-42.
13. Tajer C, Grancelli H, Hirschson Prado A y col. Enalapril in unstable angina: a randomized double blind multicentre trial. *Eur Heart J* 1995;16:259.
14. Bazzino O, Diaz R, Tajer C y col. Clinical predictors of in-hospital prognosis in unstable angina: ECLA 3. The ECLA Collaborative Group. *Am Heart J* 1999;137:322-31.
15. FRISC II investigators. Invasive compared with non-invasive treatment in unstable coronary artery disease: prospective randomised multicentre study. Fragmin and fast revascularisation during instability in coronary artery disease. *Lancet* 1999;354:708-15.
16. Fox K, Poole-Wilson P, Henderson R y col. Interventional versus conservative treatment for patients with unstable angina or non-ST-elevation myocardial infarction: the British Heart Foundation RITA 3 randomised trial. *Lancet* 2002; 360: 743-751.
17. Wilson P, D'Agostino R, Lewy D y col. Prediction of coronary heart disease using risk factors categories. *Circulation* 1998;97:1837-47.
18. Conroy R, Pyörälä K, Fitzgerald A y col, on behalf of the SCORE project group. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J* 2003;24:987-1003.
19. Wood D, De Backer G, Faergeman O y col. Prevention of coronary heart disease in clinical practice. Recommendations of the Second Joint Task Force of European and other Societies on coronary prevention. *Eur Heart J* 1998;19:1434-503.
20. Third report of the National Cholesterol Education Program (NCEP). Expert Panel on Detection, Evaluation, and Treatment of High blood cholesterol in adults. Panel III. *Circulation* 2002;106:3113-21.
21. Wald N, Law M. A strategy to reduce cardiovascular disease by more than 80%. *BMJ* 2003;326:1-6.
22. Israel L. La decisión médica. Editorial Emecé: 1983.
23. Feinstein AR. *Clinical Biostatistics. The haze of Bayes, the aerial palaces of decision analysis, and the computerized Ouija*