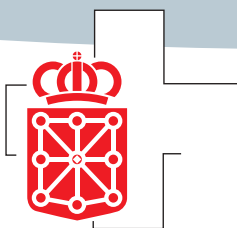


Epidemiología clínica aplicada a la toma de decisiones en medicina

Begoña Bermejo Fraile



ANALES

DEL SISTEMA SANITARIO DE NAVARRA

**EPIDEMIOLOGÍA CLÍNICA APLICADA
A LA TOMA DE DECISIONES
EN MEDICINA**

MONOGRAFÍA Nº 1: CIENCIAS BÁSICAS

EPIDEMIOLOGÍA CLÍNICA APLICADA A LA TOMA DE DECISIONES EN MEDICINA

Begoña Bermejo Fraile

*Especialista en Medicina Preventiva y Salud Pública
Máster en Salud Pública y Metodología de la Investigación Biomédica.
Universidad Autónoma de Barcelona.
Máster en Economía de la Salud y Gestión de Servicios Sanitarios.
Universidad de Barcelona.
Doctora en Medicina y Cirugía. Universidad Autónoma de Barcelona*



Gobierno de Navarra
Departamento de Salud

Titulo: Epidemiología clínica aplicada a la toma de decisiones en medicina
Autora: Begoña Bermejo Fraile
Edita: GOBIERNO DE NAVARRA
Departamento de Salud

© Gobierno de Navarra
Composición: Página, S.L.
Impresión: Ona Industria Gráfica, S.A.
Diseño de cubierta: Alberto Navarro

ISBN: 84-235-2133-8
Dep. Legal: NA. 2.061/2001

Promociona y distribuye: Fondo de Publicaciones del Gobierno de Navarra
(Departamento de Presidencia, Justicia e Interior)
C/ Navas de Tolosa, 21
Teléfono: 948 427 121
Fax: 948 427 123
Correo electrónico: fpubli01@cfnavarra.es
www.cfnavarra.es/publicaciones
31002 PAMPLONA

*A Laia, la más bonita
decisión de mi vida*

La naturaleza es probabilística
La información, incompleta
Los resultados, esenciales
Los recursos, limitados
Las decisiones, inevitables

H.S. Frazier
Universidad de Harvard, 1980

ÍNDICE

PRÓLOGO	15
INTRODUCCIÓN AL PROCESO DIAGNÓSTICO	17
PROBABILIDAD	23
Conceptos básicos de probabilidad	23
Definición.....	23
Propiedades.....	23
Leyes de la probabilidad	24
Odds	26
Teorema de Bayes	26
Resumen	29
Problemas	30
ÁRBOLES DE DECISIÓN	31
Planteamiento del problema	31
Elaboración de un árbol de decisión	31
Cálculo del resultado esperado	35
Fuentes de datos	39
Análisis de sensibilidad	40
Resumen	41
Problemas	42
PROBABILIDAD PREVIA	43
Experiencia personal	44
Representatividad.....	44
Disponibilidad	46
Anclaje y ajuste.....	46
Otras fuentes de datos	47
Incidencia, prevalencia, odds previa	48
Resumen	51
Problemas	52
FIABILIDAD	53
Calidad de la información	53
Fuentes de variación	55
Modos de aumentar la fiabilidad	56

Medición de la fiabilidad	57
Concordancia global.....	57
Índice kappa	58
Coeficiente de variación	63
Coeficiente de correlación intraclases	64
Resumen	65
Problemas	66
VALIDEZ DE LAS PRUEBAS DIAGNÓSTICAS	69
Medidas de la validez de las pruebas diagnósticas	69
Errores en la estimación de la sensibilidad y la especificidad ..	74
Ajuste de los datos	76
Ajuste mediante la regresión logística	77
Resultados de variables continuas. Curvas ROC	78
Elección de un punto de corte	80
Comparación de la validez de las pruebas diagnósticas	83
Comparación de áreas bajo la curva ROC	87
Resumen	89
Problemas	91
PROBABILIDAD POSTERIOR	93
Estimación de la probabilidad posterior	93
Tablas de 2x2.....	94
Teorema de Bayes	95
Razón de verosimilitud	97
Regresión logística	100
Probabilidad previa, sensibilidad, especificidad y probabilidad posterior	102
Resumen	107
Problemas	109
CUÁNDO APLICAR UNA PRUEBA DIAGNÓSTICA	111
Introducción	111
Umbral terapéutico	114
Umbral terapéutico y ratio beneficio/coste.....	116
Umbral de diagnóstico y umbral de tratamiento	117
Utilidad de la prueba.....	121
Diagnóstico precoz	123
Resumen	128
Problemas	130
RESULTADOS SOBRE LA SALUD	131
Introducción	131

Vidas salvadas, casos evitados	132
Incidencia y prevalencia	132
Medidas de los resultados sobre la salud	133
Esperanza de vida	139
DEALE	139
Modelos de Markov	141
Kaplan Meier	145
Años de vida ajustados por calidad (AVAC)	150
Técnicas de obtención de los AVAC	150
Escala de categorías (EC)	150
Estimación de la magnitud (EM)	151
Equivalencia temporal (ET)	152
Equivalencia de personas (EP)	153
Juego estándar (JE)	153
Comparación de los métodos de medida de la calidad de vida	155
Instrumentos de medida de la calidad de vida	157
Resultados multidimensionales sobre la salud	158
Curva de utilidad de la esperanza de vida	161
Resumen	162
Problemas	164
EVALUACIÓN ECONÓMICA	169
Introducción	169
Técnicas de evaluación económica	170
Medida de los costes	170
Análisis coste-beneficio (ACB)	172
Análisis coste-efectividad (ACE)	175
Análisis coste-utilidad (ACU)	176
Análisis de minimización de costes (AMC)	177
Valor monetario de la vida humana	178
Actualización o descuento	179
Análisis incremental	181
Análisis de sensibilidad	183
Presentación de resultados	184
Asignación de recursos	185
Resumen	187
Problemas	188
SOLUCIONES A LOS PROBLEMAS	191
BIBLIOGRAFÍA	223

PRÓLOGO

El progreso de la Medicina va registrándose en los libros de texto más utilizados. La comparación de ediciones sucesivas permite constatar cómo se incorporan nuevos conocimientos sobre etiología de enfermedades y eficacia de tratamientos pero también cómo se aplican métodos tradicionales a problemas nuevos. En efecto, el avance del conocimiento requiere, primero, curiosidad y duda para efectuar preguntas y, segundo, método para contestarlas.

La decisión clínica no se consideraba un problema en sí. Fue la creciente conciencia por parte de los profesionales de variaciones aparentemente arbitrarias en la práctica médica junto a la constatación de que los recursos sanitarios son asignados, mayoritariamente, a través de las decisiones clínicas, la que convirtió a la decisión clínica en “problema nuevo”, en objeto de análisis. La decisión clínica puede ser abordada desde diversas perspectivas disciplinarias: Medicina, Estadística, Ética, Economía y fue efectivamente abordada desde esas múltiples disciplinas en uno de los menos frecuentes y más fructíferos casos de colaboración interdisciplinaria.

Constituyen hitos de esta colaboración entre disciplinas que toman la decisión clínica como objeto de estudio dos libros de *Cambridge University Press* publicados en 1988: 1/ Bell D, Raiffa H, Tversky A (eds): *Decision making. Descriptive, normative, and prescriptive interactions...* con los Premios Nobel de Economía K Arrow y H Simon (éste fallecido recientemente), el clásico M Weinstein, R Hogarth actualmente en el Departamento de Economía y Empresa de la Universidad Pompeu Fabra, Kahneman, Thaler...2/ Dowie J, Elstein A (eds). *Professional judgment. A reader in clinical decision making...* con David Eddy (quien publicaría después en *JAMA* la ampliamente difundida serie de 'Decisión clínica: De la teoría a la práctica', J Kassirer (quien sería nombrado director del *New Engl J Med* a principios de 1992)...

En España también se han realizado aportaciones fundamentalmente desde la Epidemiología (sensibilidad pionera de la revista *Medicina Clínica*) y desde la Economía (papel seminal de la Asociación de Economía de la Salud, www.aes.es) así como –lo que es más importante– algunos cambios tanto en la forma de practicar la medicina como de organizar los

servicios sanitarios, cambios que suelen subsumirse bajo la denominación Gestión Clínica.

La autora del texto al que estas líneas sirven de prólogo utiliza ejemplos de investigaciones propias para ilustrar unas explicaciones que desgrana con una contundente claridad pedagógica. “Epidemiología clínica aplicada a la toma de decisiones en medicina” lleva al lector con pasmosa suavidad a cuestiones que sólo saben abordar con facilidad personas como la Dra Begoña Bermejo, muy acostumbradas a investigar, con fuerte vocación y rodaje docente, y curiosidad suficiente como para ponerse a desbrozar otros campos disciplinarios. De hecho la conocí como destacada integrante de la primera promoción del Master de Economía de la Salud y Gestión Sanitaria de la Universidad de Barcelona, la primera especialista en Salud Pública y Medicina Preventiva que lo cursó. He ido siguiendo sus trabajos y tras la lectura de la obra que nos ocupa tengo la seguridad de poderlo recomendar con tranquilidad, y muy altas garantías de satisfacción, tanto a profesionales sanitarios como a estudiantes de ciencias de la salud.

Vicente Ortún Rubio
Departamento de Economía y Empresa
Centro de Investigación en Economía y Salud
Universidad Pompeu Fabra, Barcelona

INTRODUCCIÓN AL PROCESO DIAGNÓSTICO

El mayor reto y atractivo de la medicina es tratar con individuos, diferentes unos de otros, distintos en el desarrollo de las enfermedades, en la manifestación de las mismas y en la respuesta a los tratamientos. A ello se une el hecho de que los médicos también son individuos diferentes, con distinta experiencia y diferentes conocimientos, que interpretan de distinto modo un mismo signo. De ahí que el proceso diagnóstico sea uno de los temas más problemáticos de la medicina, y la capacidad de hacerlo correctamente todo un arte (y mucha ciencia).

Si el problema fuese tan sencillo como que los enfermos siempre manifestasen determinados signos que nunca están presentes en los pacientes no afectados de la enfermedad, no haría falta todo un proceso para llegar al diagnóstico, sólo sería necesario saber si están presentes o no dichos signos, pero por desgracia esta situación es pura ficción. El verdadero estado del paciente permanece oculto dentro de su cuerpo y no puede observarse directamente, por lo que se deben usar señales externas, indirectas e imperfectas para inferir el verdadero estado.

El objetivo último del diagnóstico correcto es aplicar un tratamiento adecuado, pero muchas veces se elige el tratamiento sin tener la seguridad absoluta sobre su eficacia, ni tan siquiera sobre el diagnóstico, casi siempre hay cierto grado de incertidumbre. De hecho, la anamnesis, la exploración física y la aplicación de pruebas más o menos invasivas o el estudio de la respuesta a determinadas manipulaciones se hacen con el fin de disminuir esta incertidumbre.

La incertidumbre que rodea al proceso diagnóstico se refleja también en el lenguaje de los médicos. Muchas veces utilizan distintas palabras para expresar el mismo juicio acerca de la probabilidad de un hecho, o por el contrario las mismas palabras quieren decir cosas muy distintas en boca de uno y otro clínico. La terminología suele ser muy vaga en este sentido: pocas veces, raro, bastante probable, común, frecuente. ¿Significan lo mismo para todos estas palabras? ¿Cuándo es más probable una enfermedad, si se le dice al paciente "es probable que usted tenga esta enfermedad" o si le dice "no podemos descartar que usted tenga esta enfermedad"? Con el fin de homogeneizar el lenguaje de la incertidumbre, a las probabilidades se les asignan números, y así todos los médicos estarán de acuerdo en que una hipótesis diagnóstica es menos incierta en un

paciente cuya probabilidad de presentar la enfermedad es del 95% o del 5% que si dichas probabilidades son del 60 y 40% por ejemplo.

Desde la entrada en contacto con el paciente ya se generan una serie de hipótesis diagnósticas, dependiendo de los síntomas que refiera, su edad, sexo, aspecto, etc. Estas hipótesis pueden ser generales (fiebre) o específicas (infarto agudo de miocardio), un síndrome (nefrótico) o una enfermedad (talasemia) y en número no suelen exceder de las siete, dada la escasa capacidad de memoria que tenemos a corto plazo.

Si se parte de un elevado número de hipótesis es recomendable contrastar primero aquellas más probables o más importantes, entendiendo por importantes aquellas con un tratamiento más efectivo, o que el paciente puede morir o quedar incapacitado si no se trata.

Posteriormente, las preguntas (y quien dice preguntas dice pruebas) que realiza un clínico experimentado no son innumerables, ni mucho menos al azar, sino aquellas que tratan de contrastar la verosimilitud de un pequeño conjunto, manejable y eficiente de hipótesis diagnósticas. Lo que se trata de hacer con estas preguntas es rechazar o corroborar determinados diagnósticos. Los procesos cognitivos que se utilizan en este proceso de entrada y salida de nuestra mente de las diferentes hipótesis pueden ser de varios tipos y básicamente siguen tres modelos:

- *Modelo lineal*: en un modelo lineal, a los signos o síntomas que tienden a confirmar un diagnóstico se les asigna un peso positivo y a aquellos que tienden a rechazarlo un peso negativo. La decisión de confirmar o no un diagnóstico se basa en el signo dominante (positivo o negativo) de la suma de todos estos pesos.
- *Modelo bayesiano*: basado en probabilidades condicionadas. El médico cambia su estimación de la probabilidad de la enfermedad según cada nueva información adquirida, paso a paso.
- *Modelo algorítmico*: el clínico, conforme va comprobando hipótesis, va siguiendo las ramas de un proceso lógico.

A su vez, y para la construcción de estos modelos, se manejan asociaciones entre las diferentes variables, y estas asociaciones pueden ser de tres tipos:

- *Probabilística*: es la asociación estadística hallada entre dos variables. Por ejemplo, cuál es la probabilidad de hiperplasia prostática asintomática en los varones de 55 años o la probabilidad de que se detecte en estos individuos el antígeno específico de próstata.
- *Causal*: son asociaciones basadas en las relaciones causa-efecto. Definen o describen el mecanismo (anatómico, fisiológico o bioquímico) que puede simular el comportamiento normal del cuer-

po humano. Por ejemplo, la probabilidad de anemia ferropénica en una mujer de 30 años con sintomatología inespecífica pero con una menstruación abundante.

- *Determinista*: es aquella asociación que está basada en una serie de reglas bien definidas o algoritmos. Por ejemplo, si el bicarbonato plasmático es menor de 8 mmol/l y el pH es bajo, el diagnóstico es de acidosis metabólica.

De estas asociaciones, la que más fuerza da a la hora de validar, verificar, comprobar o rechazar una hipótesis es la asociación causal. Las asociaciones probabilísticas pueden a veces estar sesgadas o confundidas por un tercer factor y ser causalmente inconsistentes. En cualquier caso, y aunque se expliquen de forma separada, todos estos tipos de asociaciones se utilizan de forma superpuesta a lo largo del proceso diagnóstico.

El proceso diagnóstico termina en el momento en el que el clínico ha identificado un diagnóstico lo suficientemente aceptable como para indicar un pronóstico o instaurar ya un tratamiento. Para que un diagnóstico se considere efectivamente aceptable se tiene que haber comprobado su coherencia (no hay signos o síntomas contradictorios con el diagnóstico realizado) y adecuación (este diagnóstico explica los síntomas de este paciente).

La probabilidad, el nivel de certidumbre que se le exige a una hipótesis diagnóstica depende de las consecuencias que pueda conllevar una equivocación, lo que se traduce en un tratamiento correcto o no, o incluso la ausencia de tratamiento porque la verdadera enfermedad del paciente ha sido incorrectamente descartada. Lógicamente, no se precisa tanta exactitud para diagnosticar la gripe en una persona joven previamente sana que en el caso de una neoplasia de colon en un varón de 65 años.

Como decíamos previamente, las pruebas diagnósticas se aplican para aumentar el conocimiento sobre la probabilidad de una enfermedad en el paciente, pero es imprescindible conocer cómo varía esta probabilidad a medida que se adquiere mayor información y ello va ligado a un principio básico: la interpretación de cualquier resultado de unas pruebas depende de la probabilidad de la enfermedad antes de realizar la prueba. Cada vez que uno comienza a preguntarse ¿de qué informa el resultado de esta prueba? uno debe preguntarse primero ¿qué pensaba antes de indicar esta prueba? No es lo mismo una cardiomegalia en la radiografía de tórax en un paciente joven deportista que se realiza la placa en un chequeo rutinario, que en un paciente mayor hipertenso con signos de insuficiencia cardiaca.

Las pruebas (o preguntas o exploración física) no van orientadas hacia enfermedades muy poco probables de entrada, porque aunque se aplicase una prueba diagnóstica y ésta resultase positiva seguiría siendo

tan poco probable la enfermedad que no merecería la pena administrar un tratamiento que muy posiblemente tendrá algunos efectos adversos.

Cuando la probabilidad de presentar la enfermedad es muy elevada tampoco se realizan más pruebas, se administra directamente el tratamiento que se considera más adecuado, porque si aún por confirmar la presencia de la enfermedad, se hiciese la prueba y ésta resultase negativa, todavía la enfermedad sería lo suficientemente probable como para que mereciera la pena administrar un tratamiento.

Como resumen, y dependiendo del rango de probabilidades en que se encuentra la enfermedad, la actitud que se adopta ante el paciente es diferente (Fig. 1):

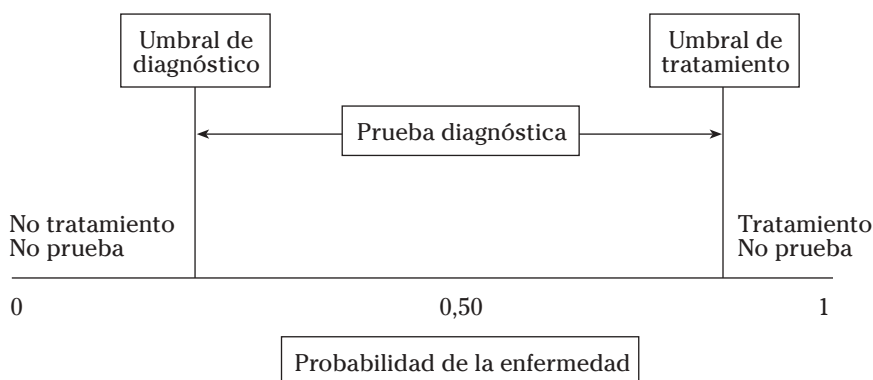


Figura 1. Proceso diagnóstico.

1. Cuando la probabilidad de una enfermedad es muy baja, en general no se hace nada respecto a ella, no se aplica ninguna prueba diagnóstica esperando un resultado positivo (ya que fácilmente sería un falso positivo), ni mucho menos se da su tratamiento específico.
2. Si la probabilidad de tener la enfermedad es intermedia entonces se busca más información (preguntando, explorando o aplicando otra prueba diagnóstica) con el fin de decantarnos por uno u otro lado (confirmar o rechazar una hipótesis diagnóstica).
3. Cuando la probabilidad de presentar la enfermedad es muy alta, en general se pasa directamente al tratamiento, sin obtener más información.

En cualquier caso, probabilidad alta o baja son términos relativos, que deben establecerse para cada enfermedad y en función fundamentalmente de las consecuencias que tenga una equivocación en el diagnóstico y de las características de las pruebas de que disponemos para diagnosticarla.

En términos generales se denomina probabilidad previa a la probabilidad de presentar la enfermedad antes de aplicar una prueba y probabilidad posterior a la probabilidad de presentar la enfermedad una vez conocido el resultado de la prueba. Pero previo y posterior también son términos relativos, porque por ejemplo, la probabilidad posterior de infarto de miocardio tras conocer el ECG es también la probabilidad previa antes de la determinación enzimática.

En la literatura a veces se utilizan indistintamente como sinónimos probabilidad previa y prevalencia y se debe al hecho del primer desarrollo del análisis de la decisión médica en el campo de la salud pública, donde muchas veces se aplican pruebas de cribaje con el fin de diagnosticar enfermedades asintomáticas, y la primera probabilidad de la que se parte es precisamente la prevalencia de la enfermedad en el grupo en estudio.

Tras el conocimiento del resultado de la prueba, el cambio de la probabilidad previa a la posterior depende no sólo de la probabilidad previa sino también de la validez de la prueba diagnóstica, concepto que se resume en dos términos muy utilizados y no tan bien conocidos: sensibilidad y especificidad.

Este monográfico está dedicado a todos los aspectos relevantes en el proceso diagnóstico, desde la primera toma de contacto con el paciente hasta que la enfermedad se ha resuelto (ya sea la curación, incapacidad o muerte).

Existen dos capítulos generales dedicados al estudio y manejo de las probabilidades y los árboles de decisión, herramientas básicas en la toma de decisiones.

La probabilidad de una enfermedad una vez conocido el resultado de una prueba diagnóstica depende de la probabilidad de la enfermedad antes de aplicar la prueba. Otro capítulo se ha dedicado a la estimación de esta probabilidad previa, basada en la propia experiencia o en la literatura.

Las pruebas diagnósticas no son infalibles, y de hecho se cometen muchos errores por creer que un resultado positivo de una prueba quiere decir presencia de enfermedad y un resultado negativo ausencia. Hay dos requisitos indispensables para que las pruebas diagnósticas aporten una información de calidad: fiabilidad y validez.

Una vez que se ha aplicado la prueba válida y fiable, dependiendo del resultado de las mismas variará la probabilidad de tener la enfermedad, y del estudio de esta probabilidad posterior se encarga otro capítulo.

A veces las probabilidades de presentar la enfermedad son tan extremas (bajas o altas) que apenas experimentan un cambio tras la aplicación de pruebas. La actitud del clínico seguirá siendo la misma que antes de conocer el resultado de las pruebas, por lo que no merecía la pena apli-

carlas. En el cálculo de estas probabilidades umbral se profundiza en el capítulo denominado “Cuándo aplicar una prueba diagnóstica”.

El objetivo último de la toma de decisiones es la mejora de la salud de los individuos, un concepto realmente difícil de medir, pero se explican diversas aproximaciones en el capítulo de “Resultados sobre salud”.

Por último, se trata el tema de la evaluación económica, de innegable utilidad, ya que el cada vez mayor desarrollo de nuevas tecnologías y la limitación de los recursos obliga a utilizar éstos de la manera más eficiente.

PROBABILIDAD

CONCEPTOS BÁSICOS DE PROBABILIDAD

Definición

La medicina no es una ciencia exacta y por ello los médicos rara vez pueden predecir un resultado con absoluta certeza. La incertidumbre acompaña tanto al proceso diagnóstico como terapéutico y el modo de cuantificar esta incertidumbre es mediante las probabilidades. No todos los pacientes con una glucemia elevada son diabéticos; una paciente cuya mamografía es normal no quiere decir con toda seguridad que no tenga un cáncer de mama; todos los pacientes a los que se les ha realizado un bypass coronario no muestran la misma supervivencia.

Lo mejor es definir la probabilidad en términos de frecuencia relativa (proporción). La probabilidad p de que un hecho A ocurra, escrito como $p(A)$, se estima del siguiente modo:

$$p(A) = \frac{\text{número de veces que ocurre } A}{\text{número de veces que puede ocurrir } A}$$

Por ejemplo, si de cada 1000 mujeres de 50 años, 100 desarrollan cáncer de mama, la probabilidad de dicho cáncer en esta población es = $100/1000 = 0,1 = 10\%$

Propiedades

La probabilidad siempre es un número que está entre 0 y 1 (o entre 0 y 100 si se expresa en porcentajes).

- a) La probabilidad de un suceso cierto es 1 (quiere decir que siempre ocurre).
- b) La probabilidad de un suceso incierto es 0 (nunca ocurre).
- c) Un valor de 0,5 significa que la probabilidad de que un suceso ocurra es la misma que de que no ocurra.
- d) La suma de las probabilidades o frecuencias relativas de todos los sucesos que pueden ocurrir debe ser 1. De aquí se deduce que la probabilidad de que ocurra un suceso más la probabili-

dad de que no ocurra es 1 (la negación se representa con un guión encima del suceso).

$$p(A) + p(\bar{A}) = 1 \qquad p(B) + p(\bar{B}) = 1$$

En la figura 2 se representa una hipotética población de 100 individuos y su distribución según dos características, A y B.

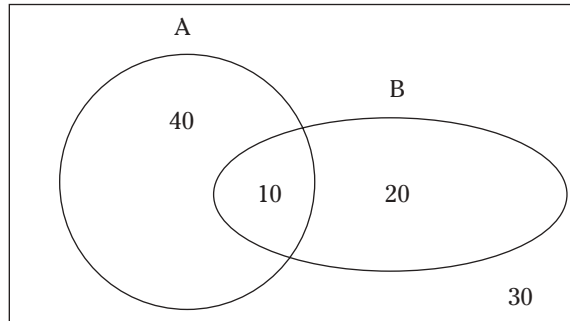


Figura 2. Cálculo de probabilidades.

$$p(A) = (40+10)/100 = 0,5$$

$$p(\bar{A}) = (20+30)/100 = 0,5$$

$$p(A) + p(\bar{A}) = 0,5+0,5 = 1$$

$$p(B) = (10+20)/100 = 0,3$$

$$p(\bar{B}) = (30+40)/100 = 0,7$$

$$p(B) + p(\bar{B}) = 0,3+0,7 = 1$$

Leyes de la probabilidad

La probabilidad de la intersección entre A y B es la probabilidad de que ocurran A y B al mismo tiempo. Se representa del siguiente modo:

$$p(A \cap B) \text{ en el ejemplo} = 10/100 = 0,1$$

Dos sucesos son independientes cuando la probabilidad de que ocurra un suceso es independiente de la presencia o ausencia del otro (por ejemplo, la probabilidad de desarrollar una diabetes mellitus tipo II es independiente del hecho de ser fumador). En estos casos, la probabilidad de la intersección es el producto de las probabilidades (ley multiplicativa):

$$p(A \cap B) = p(A) * p(B)$$

Dos sucesos son dependientes cuando la probabilidad de que ocurra un suceso depende de la presencia o ausencia del otro suceso (por

ejemplo, la probabilidad de desarrollar una hemocromatosis depende del sexo). En este caso, la probabilidad de la intersección ya no es igual al producto de las probabilidades.

$$p(A \cap B) \neq p(A) * p(B)$$

En el ejemplo, A y B no son independientes.

$$p(A \cap B) = 10/100 = 0,1 \neq 0,5 * 0,3 = 0,15$$

La probabilidad de la unión de A y B (ley aditiva) es la probabilidad de que ocurra cualquiera de los dos sucesos, A o B o la intersección. Se representa del siguiente modo:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

La probabilidad de la unión es la suma de las probabilidades menos la probabilidad de la intersección (porque si no esta área sería contada dos veces, cuando se calcula la probabilidad de A y cuando se calcula la probabilidad de B).

$$p(A \cup B) = 0,5 + 0,3 - 0,1 = 0,7$$

Cuando los sucesos A y B son mutuamente excluyentes (no pueden ocurrir al mismo tiempo), la probabilidad de la intersección es 0, por lo tanto la probabilidad de la unión es la suma de las probabilidades.

$$p(A \cup B) = p(A) + p(B)$$

La probabilidad de un suceso cualquiera es:

$$p(A) = p(A \cap B) + p(A \cap \bar{B})$$

$$p(A) = 10/100 + 40/100 = 0,5$$

donde $p(A \cap B)$ y $p(A \cap \bar{B})$ son mutuamente excluyentes.

La probabilidad condicional es la probabilidad de que ocurra A dado que ocurre B. Por ejemplo, probabilidad de tener un tumor cerebral dado que se tiene cefalea. Se representa con una barra vertical. Es el cociente entre la probabilidad de que ocurran A y B (intersección) y la probabilidad de que ocurra B:

$$p(A/B) = \frac{p(A \cap B)}{p(B)} = 0,1/0,3 = 0,33$$

En la práctica, la probabilidad de la ocurrencia simultánea de los sucesos A y B (probabilidad de la intersección) se acostumbra a calcular a partir de la probabilidad condicionada, que es más fácil de evaluar. De la

definición de probabilidad condicionada se deriva la fórmula de la ley multiplicativa de la probabilidad:

$$p(A \cap B) = p(B) * p(A/B)$$

Si dos sucesos A y B son independientes, como la misma palabra indica, la probabilidad de que ocurra A es independiente de que ocurra B, por tanto $p(A/B) = p(A)$

También se puede explicar porque si dos sucesos son independientes $p(A \cap B) = p(A) * p(B)$

$$p(A/B) = p(A) * p(B) / p(B) = p(A)$$

Odds

Odds es el cociente entre la probabilidad de que un hecho ocurra y la probabilidad de que no ocurra.

$$\text{Odds} = p / (1-p)$$

Por ejemplo, al tirar un dado la odds de que salga un as es 1 a 5 (0,20) que resulta de dividir la probabilidad de que salga un as (1/6) por la probabilidad de que no salga (5/6).

Los anglosajones utilizan más las odds que las probabilidades. Sobre todo son muy útiles cuando se habla de hechos muy raros (1 a 99 odds más que una probabilidad de 0,01) o muy frecuentes (99 a 1 odds más que una probabilidad de 0,99).

La transformación de odds en probabilidades es sencilla:

$$p = \text{odds} / (1 + \text{odds})$$

$$p(B) = 0,3 \quad \text{Odds (B)} = 0,3 / 0,7 = 0,428$$

(también se expresa como 3:7 ó 1:2,3)

Este resultado quiere decir que por cada 3 veces que ocurre el suceso B, 7 veces no ocurre, o poniendo los datos en referencia al 1 el resultado es 1:2,3 queriendo decir que por cada vez que ocurre el suceso B, 2,3 veces no ocurre.

TEOREMA DE BAYES

El teorema de Bayes permite estudiar las probabilidades condicionales en función de la condición inversa, o lo que es lo mismo, permite cambiar las probabilidades condicionales.

Si se parte de las fórmulas para el cálculo de la probabilidad condicional

$$p(A/B) = \frac{p(A \cap B)}{p(B)} \quad p(B/A) = \frac{p(A \cap B)}{p(A)}$$

De aquí se deduce que:

$$p(A \cap B) = p(A/B) * p(B) = p(B/A) * p(A)$$

Por otro lado, la probabilidad de un suceso cualquiera es:

$$p(B) = p(B \cap A) + p(B \cap \bar{A})$$

$$p(A/B) = \frac{p(A \cap B)}{p(B)} = \frac{p(B/A) * p(A)}{p(B \cap A) + p(B \cap \bar{A})} = \frac{p(B/A) * p(A)}{p(B/A) * p(A) + p(B/\bar{A}) * p(\bar{A})}$$

$$p(A/B) = \frac{p(B/A) * p(A)}{p(B/A) * p(A) + p(B/\bar{A}) * p(\bar{A})}$$

Como se puede observar, la probabilidad condicional $p(A/B)$ se ha puesto en función de la probabilidad condicional $p(B/A)$ o $p(B/\bar{A})$. Este teorema tiene su aplicación práctica más importante en el cálculo de la probabilidad de un diagnóstico correcto. Las características de la validez de una prueba diagnóstica son la sensibilidad (probabilidad de que en un individuo enfermo de un resultado positivo) y la especificidad (probabilidad de que en un individuo sin la enfermedad la prueba resulte negativa). Cuando un paciente acude a la consulta se le realiza la historia clínica y una serie de pruebas diagnósticas, y habitualmente no se tiene la certeza absoluta de si el paciente en estudio tiene o no la enfermedad, sólo se sabe el resultado de la prueba diagnóstica y se trata de calcular la probabilidad de que el paciente tenga la enfermedad si el resultado es positivo (valor predictivo positivo) o la probabilidad de que no la tenga si el resultado de la prueba ha sido negativo (valor predictivo negativo). Se han cambiado las probabilidades condicionales.

EJEMPLO: se ha evaluado la determinación de la ferritina sérica (tomando como punto de corte 65 mmol/l) para el diagnóstico de anemia ferropénica en un grupo de 2.579 mujeres. Los resultados son los de la tabla 1. (Tomado de: *Sackett, Richardson, Rosenberg and Haynes: Evidence-Based Medicine; How to practice and teach MBE. London: Churchill Livingstone, 1997*).

- La probabilidad de que una mujer presente anemia ferropénica $p(E+)$ es $809/2579 = 0,3137$
- La probabilidad de que la ferritina sérica resulte baja (<65 mmol/l) es $p(T+) = 1001/2579 = 0,3881$
- La probabilidad de que una mujer presente anemia y ferritina sérica baja es $p(E+ \cap T+) = 731/2579 = 0,2834$.

Tabla 1. Validez de la determinación de la ferritina sérica para el diagnóstico de anemia.

	Anemia	No anemia	
Ferritina <65 mmol/l	731	270	1.001
Ferritina ≥65 mmol/l	78	1.500	1.578
	809	1.770	2.579

Si los sucesos son independientes la probabilidad de la intersección es el producto de probabilidades y en este caso $0,2834 \neq 0,3137 \cdot 0,3881 = 0,1217$. No son sucesos independientes porque el que resulte la ferritina sérica baja o alta depende de que se sea o no anémica.

- La probabilidad de que en una mujer anémica, la ferritina resulte baja es $p(T+/E+) = 731/809 = 0,9036$. Este resultado es la sensibilidad de la prueba.
- La probabilidad de que en una mujer no anémica la ferritina resulte por encima de 65 mmol/l es $p(T-/E-) = 1500/1770 = 0,8475$. Este resultado es la especificidad de la prueba.
- La probabilidad de que presente anemia una mujer en la que la ferritina ha resultado baja es $p(E+/T+) = 731/1001 = 0,7303$. Si se aplica el teorema de Bayes el resultado es el mismo.

$$p(E+/T+) = \frac{p(T+/E+) \cdot p(E+)}{p(T+/E+) \cdot p(E+) + p(T+/E-) \cdot p(E-)} =$$

$$= \frac{0,9036 \cdot 0,3137}{0,9036 \cdot 0,3137 + (1-0,8475) \cdot (1-0,3137)} = 0,7303$$

- La probabilidad de que no presente anemia una mujer en la que la ferritina ha resultado por encima de 65 mmol/l es: $p(E-/T-) = 1500/1578 = 0,9506$.

Si se aplica el teorema de Bayes el resultado es el mismo:

$$p(E-/T-) = \frac{p(T-/E-) \cdot p(E-)}{p(T-/E-) \cdot p(E-) + p(T-/E+) \cdot p(E+)} =$$

$$= \frac{0,8475 \cdot (1-0,3137)}{0,8475 \cdot (1-0,3137) + (1-0,9036) \cdot 0,3137} = 0,9506$$

RESUMEN**Probabilidad de un suceso:**

$$p(A) = \frac{\text{número de veces que ocurre A}}{\text{número de veces que puede ocurrir A}}$$

$$p(A) + p(\bar{A}) = 1$$

$$p(A) = p(A \cap B) + p(A \cap \bar{B})$$

Probabilidad de la intersección (ley multiplicativa):

1. Sucesos no independientes:

$$p(A \cap B) = p(A/B) * p(B) = p(B/A) * p(A)$$

2. Sucesos independientes:

$$p(A \cap B) = p(A) * p(B)$$

Probabilidad de la unión (ley aditiva):

1. Sucesos excluyentes:

$$p(A \cup B) = p(A) + p(B)$$

2. Sucesos no excluyentes:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

Probabilidad condicional:

$$p(A/B) = \frac{p(A \cap B)}{p(B)}$$

$$\text{Odds} = p/(1-p)$$

Teorema de Bayes:

$$p(A/B) = \frac{p(B/A) * p(A)}{p(B/A) * p(A) + p(B/\bar{A}) * p(\bar{A})}$$

PROBLEMAS

1. Si la probabilidad de que nazca una niña es 0,5, ¿cuál es la probabilidad de que los siguientes cuatro nacimientos sean niñas? ¿Cuál es la probabilidad de que el cuarto hijo sea un niño si las tres primeras han sido niñas?
2. La probabilidad de fallo de la alarma de monitorización cardiaca en una unidad de cuidados intensivos es 0,001. Para aumentar la seguridad se coloca una segunda alarma. ¿Cuál es la probabilidad de que en caso de fallo cardiaco no suene ninguna de las dos alarmas?
3. Un grupo de pacientes acude a un programa de cribaje de cáncer de mama mediante mamografía. Las mamografías son vistas por dos médicos, el Dr. A y el Dr. B. El Dr. A refiere que en el 10% de las pacientes la mamografía es sugestiva de cáncer de mama. El Dr. B dice que son positivas el 17%. Los dos médicos se ponen de acuerdo (respecto a la positividad) en el 8% de los casos.
 - a) ¿Son independientes las respuestas de uno y otro médico?
 - b) ¿Cuál es la probabilidad de que el Dr. B confirme el diagnóstico positivo dado por el Dr. A?
 - c) En qué porcentaje de casos el Dr. B contradice el diagnóstico negativo del Dr. A.
 - d) Todas las pacientes cuya mamografía es sugestiva de cáncer, independientemente del médico, son remitidas para realizar más pruebas, ¿en qué porcentaje se realizarán pruebas complementarias para confirmar el diagnóstico?
4. La odds de un suceso es 1:9 ¿Cuál es su probabilidad?
5. La probabilidad de un suceso es 0,8 ¿Cuál es la odds?
6. La enfermedad E se presenta en el 10% de la población. Para su diagnóstico se aplica una prueba T, que resulta positiva en el 80% de los casos con la enfermedad y en el 5% de los individuos sin la enfermedad. Acude un individuo a la consulta y la prueba resulta positiva ¿qué probabilidad tiene de presentar la enfermedad?

ÁRBOLES DE DECISIÓN

PLANTEAMIENTO DEL PROBLEMA

A veces resulta difícil tomar una decisión rápidamente porque es mucha la información que es necesario mantener a un tiempo en la cabeza, ya que las consecuencias de cada una de las alternativas pueden ser muchas y muy diversas.

Una forma de calcular los efectos de cualquier decisión sería a partir de los datos que aportase un experimento real, un ensayo clínico por ejemplo. De hecho, la mayoría de los expertos consideran que ésta es la opción preferible. Sin embargo, bien porque son costosos o irrealizables por motivos éticos, porque el problema es muy complejo o se precisan los resultados con cierta urgencia, los ensayos clínicos no son siempre la mejor opción. Además, puede resultar más interesante evaluar los efectos de los diagnósticos y tratamientos en las condiciones reales de la práctica asistencial, no en las condiciones ideales en que se desarrollan los experimentos. En este caso, la solución es elaborar un modelo teórico que represente con la mayor fidelidad posible el comportamiento del mundo real y genere resultados con un grado de validez y fiabilidad adecuados al problema planteado. Lógicamente, dichos modelos siempre son una versión simplificada de la realidad que pretenden simular, y la complejidad dependerá de la precisión que se exija a los datos (hecho que a su vez depende de la importancia de las consecuencias de la decisión), del tiempo y recursos disponibles para el análisis y del conocimiento sobre el proceso en estudio.

Uno de los métodos de modelización son los árboles de decisión: un método gráfico en el que se representan y comparan los resultados esperados de cada una de las posibles alternativas de elección.

ELABORACIÓN DE UN ÁRBOL DE DECISIÓN

La mejor manera de entender cómo se elabora un árbol de decisión es mediante un ejemplo.

En los exámenes periódicos de salud se realiza un hemograma a toda la población, prueba mediante la cual se puede averiguar si el

paciente tiene anemia. Sin embargo, en las mujeres en edad fértil especialmente, la prevalencia de ferropenia preanémica es una patología frecuente, que puede provocar una sintomatología inespecífica (irritabilidad, cansancio, ansiedad, intolerancia al frío, caída del pelo, fragilidad de las uñas, etc.), es fácilmente diagnosticable (mediante la determinación de ferritina sérica) y tiene un tratamiento relativamente barato y efectivo. En la figura 3 se presenta el árbol de decisión en el que se ha basado la evaluación económica de la determinación de la ferritina sérica de cribaje.

La elaboración de un árbol de decisión, en términos generales, consta de los siguientes pasos:

- 1) Definir correctamente el problema de decisión. Son los objetivos del estudio, y suele ser conveniente escribirlos para repasar cada una de las palabras y comprobar que no existe ningún término ambiguo. En el ejemplo de la ferropenia el objetivo concreto era averiguar si la determinación de la ferritina sérica de cribaje en las mujeres en edad fértil era coste-efectiva, lo que implica que es preciso conocer los costes extras de la determinación de la ferritina sérica respecto al hemograma sólo (alternativa en uso) y el número de estados con ferropenia preanémica que se logran evitar.
- 2) Identificar todas y cada una de las alternativas dentro del rango de acción del investigador. Las alternativas en este caso concreto son la realización sólo del hemograma (y determinación de ferritina sérica en los casos en los que se detecta anemia para comprobar si el origen es ferropénico) o hemograma y ferritina sérica de entrada. No se cuestionaron otras alternativas (sideremia, transferrina, porcentaje de saturación de la transferrina, punción de médula ósea) porque ya revisando la literatura se observó que estas pruebas o no eran aplicables, o eran muy caras o no eran realmente efectivas (baja sensibilidad y especificidad).
- 3) Listar los posibles resultados de cada una de las alternativas. Al realizar un hemograma, la paciente puede resultar anémica o no anémica, definida la anemia como hemoglobina inferior a 12 mg/dl. A las pacientes anémicas se les determina la ferritina sérica y ésta puede resultar baja (anemia ferropénica) o normal (anemia no ferropénica). A las pacientes con anemia ferropénica se les aconseja un tratamiento, y dependiendo de la efectividad del mismo y del grado de cumplimiento, la paciente curará totalmente, curará parcialmente de modo que tendrá una ferropenia preanémica o no curará en absoluto y permanecerá con una anemia ferropénica. Las pacientes con anemia no ferropénica, a efectos

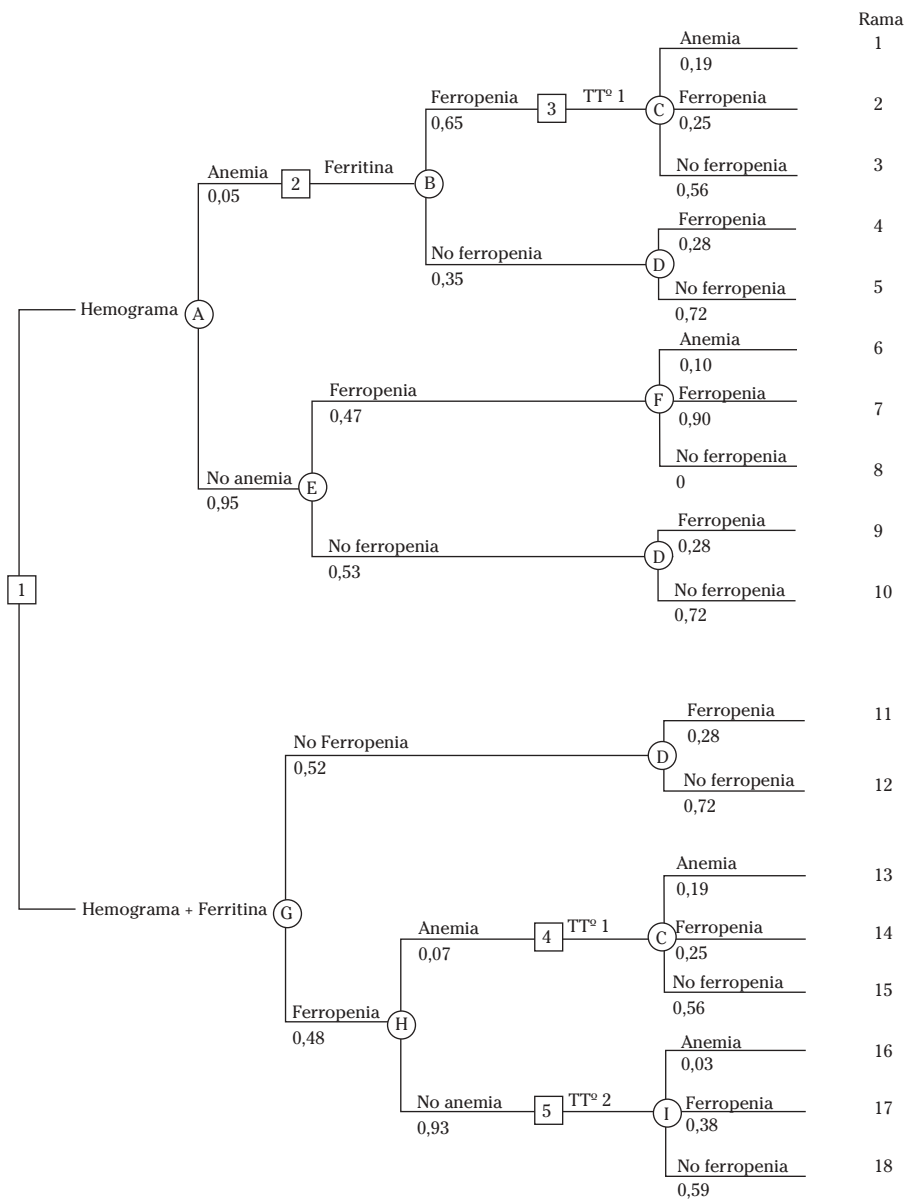


Figura 3. Árbol de decisión.

Tomado de: Bermejo B, Rovira J, Olona M, Serra M, Soriano B, Vaqué J. Análisis coste-efectividad de la determinación de ferritina sérica de cribaje en los exámenes periódicos de salud en las mujeres en edad fértil. Med Clin 1996;106:445-450.

de este estudio se consideran sanas, pero dependiendo de la incidencia de ferropenia anual, al cabo de un año pueden o no haber desarrollado la ferropenia. Mediante la segunda alternativa, de entrada se realiza un hemograma y la determinación de ferritina sérica, luego desde el primer momento se sabe si las pacientes tienen una anemia ferropénica, una anemia no ferropénica, están sanas (en todos estos casos se procede igual que en la primera alternativa) o una ferropenia preanémica, donde estriba la gran diferencia, ya que a estas pacientes se les aplica un tratamiento y dependiendo de la efectividad del mismo, el grado de cumplimiento y la incidencia de ferropenia anual, al cabo de una año estas pacientes pueden estar curadas, seguir con la ferropenia preanémica o incluso desarrollar la anemia ferropénica.

- 4) Representar la secuencia de sucesos por una serie de nudos de azar y nudos de decisión.

Los “nudos de decisión” son el punto del árbol en el que diversas opciones son posibles. Se representan por un cuadrado del que emergen una serie de líneas que representan las alternativas de decisión. En el ejemplo, los nudos de decisión se presentan cuando se trata de decidir qué pruebas diagnósticas aplicar y cuando se decide instaurar un tratamiento.

Los “nudos de azar” son el punto del árbol de decisión en el que las probabilidades determinan qué suceso ocurrirá. Se simbolizan por un círculo de donde emergen unas líneas con el nombre del resultado. Los diferentes sucesos de un nudo de azar deben ser exhaustivos (todos los posibles resultados han de estar representados, sus probabilidades suman 1) y ser mutuamente excluyentes. En el ejemplo, tras la realización de un hemograma la paciente puede o no tener anemia (nudo A), no existe otra posibilidad, esta anemia puede ser ferropénica o no ferropénica (nudo B), tras la instauración del tratamiento en la anemia ferropénica la paciente puede curar, mejorar (ferropenia preanémica) o mantenerse anémica (nudo C), de modo que todas las posibilidades están cubiertas.

- 5) Elegir un horizonte temporal para el problema según el alcance que pretende obtenerse. Lo más conveniente es elegir un horizonte comprensible pero a la vez exhaustivo. Si se trata de una enfermedad aguda es sencillo, el problema de decisión acaba cuando se resuelve la enfermedad, pero si se trata de una enfermedad crónica se puede representar todo lo que es la vida del paciente, cuidando de no hacer el árbol innecesariamente complicado. En el ejemplo que nos ocupa, se simularon los resultados tras la aplicación del programa anualmente desde los 20 años de la paciente hasta los 50 años (edad media estimada de la meno-

pausia), finalizando por tanto el estudio cuando se estimó que debido al cese de la menstruación disminuía en gran medida la prevalencia de ferropenia.

- 6) Determinar la probabilidad de cada resultado. Para ello se pueden utilizar datos de la bibliografía existente al respecto o se realiza de un modo intuitivo basado en la propia experiencia. A menudo la incertidumbre que acompaña a un hecho es elevada, pero se pueden conocer la probabilidad más alta y más baja de que ocurra, y comprobar los resultados variando estas probabilidades (análisis de sensibilidad). En el estudio de la ferropenia, las probabilidades de la validez de las pruebas se estimaron a partir de un estudio observacional transversal en la población laboral femenina que acude a la revisión periódica, en las que se realizaron las dos pruebas, hemograma y ferritina sérica; las probabilidades de incidencia de ferropenia y anemia se estimaron a partir de los datos de prevalencia por grupos de edad (derivados del mismo estudio) y las probabilidades de cumplimiento y curación tras el tratamiento se obtuvieron de una encuesta realizada a un grupo de expertos a nivel nacional. En el gráfico se han representado las probabilidades en el caso de tratarse de una cohorte de mujeres concretamente de 20 años (ya que la incidencia de ferropenia, anemia y curación espontánea cambian según la edad). Llamo la atención del lector sobre el hecho de que las probabilidades de todas las ramas que salen de un nudo de azar (círculo) siempre suman 1.
- 7) Asignar un valor a cada uno de los resultados. Si se trata de diferentes estados de salud, estos resultados pueden ser vivo o muerto, esperanza de vida, o en términos más generales utilidad. La utilidad se puede definir como la preferencia de los individuos por un determinado estado de salud. Por ejemplo, se asigna una utilidad mayor a vivir 8 años en perfecto estado de salud que a vivirlos encamado. En el capítulo 9 se desarrolla más el tema de la medida de los resultados sobre la salud. En el ejemplo de la ferropenia, los resultados sobre la salud fueron tres: sana (desde el punto de vista del hierro), ferropenia preanémica y anemia ferropénica.

CÁLCULO DEL RESULTADO ESPERADO

El valor esperado de una alternativa de decisión se calcula por un proceso conocido como “promediando resultados”, y es la suma de cada uno de los resultados por la probabilidad de que ocurran, teniendo en cuenta que esta probabilidad final es el producto de las probabilidades a lo largo de las ramas del árbol de decisión.

Por la ley multiplicativa de la probabilidad sabemos que:

$$p(A \cap B) = p(A) * p(A/B)$$

Esta ley se puede generalizar a la intersección de más de dos sucesos:

$$p(A \cap B \cap C \cap D) = p(A) * p(B/A) * p(C/A \cap B) * p(D/A \cap B \cap C)$$

Por ejemplo, ¿cuántas mujeres anémicas, ferropénicas y sanas desde el punto del hierro se espera que haya al cabo de un año de la aplicación de cada uno de los programas de cribaje?

1) Tras la aplicación del hemograma:

Mujeres anémicas serán las anémicas de entrada con la ferritina sérica baja y que no curaron tras el tratamiento (rama final 1) más las que tenían una ferropenia preanémica que desarrollaron la anemia al cabo de un año (rama 6):

$$0,19 * 0,65 * 0,05 = 0,006 \text{ (rama 1)}$$

Aunque las probabilidades se han colocado en el orden inverso al de la fórmula, 0,006 es el producto de la probabilidad de presentar una anemia (0,05), por la probabilidad de presentar una ferropenia dado que se tiene una anemia (0,65), por la probabilidad de no curar dado que se tiene una anemia ferropénica (0,19). El orden de factores no altera el producto y si lo que se busca es la probabilidad de un resultado concreto, resulta más fácil ir desde el final de la rama (resultado) hacia el tronco principal.

$$0,10 * 0,47 * 0,95 = 0,045 \text{ (rama 6)}$$

El producto de la probabilidad de no tener una anemia (0,95), por la probabilidad de presentar ferropenia dado que no se tiene anemia (0,47) y por la probabilidad de desarrollar una anemia dado que se tiene una ferropenia preanémica (0,10) da como resultado 0,045.

La suma de ambas probabilidades da el total de mujeres con anemia ferropénica al cabo de un año de la aplicación de la primera alternativa.

$$0,006 + 0,045 = 0,051$$

Mujeres con ferropenia preanémica serán aquellas con anemia ferropénica que no curaron totalmente (rama 2) más aquellas con anemia no ferropénica que desarrollaron la ferropenia al cabo de un año (rama 4) más aquellas con ferropenia preanémica que permanecieron ferropénicas (rama 7) más las sanas de entrada que desarrollaron la ferropenia al cabo de un año (rama 9):

$$0,25 * 0,65 * 0,05 = 0,008$$

$$0,28 * 0,35 * 0,05 = 0,005$$

$$0,90 * 0,47 * 0,95 = 0,402$$

$$0,28 * 0,53 * 0,95 = 0,141$$

$$0,008 + 0,005 + 0,402 + 0,141 = 0,556$$

Mujeres sanas desde el punto de vista del hierro serán las que presentaban una anemia ferropénica y que curaron tras el tratamiento (rama 3) más aquellas con anemia no ferropénica que permanecieron igual respecto al hierro (rama 5) más aquellas con ferropenia preanémica que curaron de forma espontánea (rama 8) más aquellas sanas desde el punto de vista del hierro que permanecieron igual (rama 10):

$$0,56 * 0,65 * 0,05 = 0,018$$

$$0,72 * 0,35 * 0,05 = 0,013$$

$$0,00 * 0,47 * 0,95 = 0,000$$

$$0,72 * 0,53 * 0,95 = 0,363$$

$$0,018 + 0,013 + 0,000 + 0,363 = 0,394$$

Como se puede observar, las probabilidades de presentar anemia ferropénica, ferropenia preanémica o ser normoférrica suman 1, ya que son todos los resultados posibles:

$$0,051 + 0,556 + 0,394 = 1$$

Los resultados se suelen dar en forma de probabilidades o en números absolutos: si suponemos que se parte de una cohorte de 1000 mujeres a las que se les aplica el programa de cribaje, se puede decir que al cabo de un año 51 presentarán anemia ferropénica, 556 ferropenia preanémica y 394 estarán sanas respecto al hierro.

2) Tras la aplicación del programa de cribaje con determinación del hemograma y la ferritina sérica se procede del mismo modo:

Mujeres anémicas serán aquellas con anemia ferropénica que no curaron tras el tratamiento (rama 13) más aquellas con ferropenia preanémica que no sólo no curaron sino desarrollaron la anemia al cabo de un año (rama 16):

$$0,19 * 0,07 * 0,48 = 0,006$$

$$0,03 * 0,93 * 0,48 = 0,013$$

$$0,006 + 0,013 = 0,019$$

Mujeres con ferropenia preanémica serán aquellas sanas de entrada que desarrollaron la ferropenia al cabo de un año (rama 11) más aquellas con anemia ferropénica que sólo curaron parcialmente (rama 14) más aquellas con ferropenia preanémica que ni curaron ni empeoraron (rama 17):

$$0,28 * 0,52 = 0,146$$

$$0,25 * 0,07 * 0,48 = 0,008$$

$$0,38 * 0,93 * 0,48 = 0,170$$

$$0,146 + 0,008 + 0,170 = 0,324$$

Mujeres sanas desde el punto de vista del hierro serán aquellas sanas de entrada que permanecieron igual (rama 12) más las que presentaban anemia ferropénica que curaron (rama 15) más aquellas con ferropenia preanémica que curaron (rama 18):

$$0,72 * 0,52 = 0,374$$

$$0,56 * 0,07 * 0,48 = 0,019$$

$$0,59 * 0,93 * 0,48 = 0,263$$

$$0,374 + 0,019 + 0,263 = 0,656$$

Tras la aplicación del programa de cribaje con determinación de ferritina sérica se reduce en gran medida el número de pacientes que sufrirán una anemia ferropénica o una ferropenia preanémica. Pero los resultados sobre la salud en diferentes estados no son directamente sumables, por lo que es preciso convertirlos en una unidad común, que puede ser la utilidad: imaginemos que en una escala de 0 a 1, la utilidad del estado sano es 1, la del estado ferropénico es 0,9 y la del estado anémico es 0,8.

El resultado sobre la salud de la primera alternativa sería:

$$(0,051 * 0,8) + (0,556 * 0,9) + (0,394 * 1) = 0,9352$$

El resultado sobre la salud de la segunda alternativa sería:

$$(0,019 * 0,8) + (0,324 * 0,9) + (0,656 * 1) = 0,9628$$

Como decía previamente, si el perfecto estado de salud se valora como 1, el estado de salud medio de una mujer que se somete al primer programa sería de 0,9352, mientras que con la determinación de la ferritina sérica de entrada la utilidad media sería mayor, 0,9628.

Del mismo modo que se promedian los resultados sobre la salud, también se pueden calcular los costes medios de cada alternativa, y el proceso es el mismo, se suman cada uno de los costes por la probabilidad de que se incurra en ellos. Los costes de la determinación del hemograma se estimaron en 275 pts., de la ferritina sérica en 1.140 pts., el coste del tratamiento de la anemia ferropénica, que incluye además del tratamiento médico propiamente dicho un control analítico con determinación del hemograma y ferritina sérica, se estimó en 2.975 pts., y el coste del tratamiento de la ferropenia preanémica, que incluye además del tratamiento médico un control analítico con determinación de ferritina sérica, se estimó en 2.180 pts.

1) Así, tras la aplicación del hemograma de cribaje, los costes son los del hemograma a toda la cohorte (nudo de decisión 1), más la determinación de ferritina sérica sólo en aquellas en las que el hemograma refleja anemia (nudo de decisión 2) más los costes del tratamiento de las anemias ferropénicas (nudo de decisión 3):

$$\begin{aligned} 275 * 1 &= 275 \\ 1.140 * 0,05 &= 57 \\ 2.975 * 0,65 * 0,05 &= 96,688 \\ 275 + 57 + 96,688 &= 428,688 \text{ pts.} \end{aligned}$$

Este es el coste medio por individuo al aplicar este programa de cribaje. Para estimar el coste total de la cohorte de 1.000 mujeres, basta con multiplicar esta cifra por 1.000.

2) En el segundo programa, los costes son los del hemograma y ferritina sérica a toda la cohorte (nudo de decisión 1), más los costes del tratamiento de la anemia ferropénica (nudo de decisión 4), más los costes del tratamiento de la ferropenia preanémica (nudo de decisión 5):

$$\begin{aligned} (275 + 1.140) * 1 &= 1.415 \\ 2.975 * 0,07 * 0,48 &= 99,96 \\ 2.180 * 0,93 * 0,48 &= 973,15 \\ 1.415 + 99,96 + 973,15 &= 2488,11 \text{ pts.} \end{aligned}$$

FUENTES DE DATOS

Cuando se trata de decidir entre varias alternativas terapéuticas, los analistas están de acuerdo en que los datos se deben recoger de estudios experimentales, que son los que confieren una mayor fiabilidad a los resultados. Sin embargo, lo importante para los verdaderos resultados sobre la salud que tendrá una opción, es calcularlos en condiciones reales. De aquí que sea preciso diferenciar dos conceptos que a veces, de forma errónea, se utilizan indistintamente; son la eficacia y la efectividad. La eficacia es la consecución de unos objetivos en condiciones ideales (las del experimento) y efectividad se refiere a dicha consecución en condiciones reales. Por ejemplo, un tratamiento antihipertensivo puede ser muy eficaz, pero debido a que se debe tomar un comprimido cuatro veces al día, el paciente olvida varias veces el tratamiento y al final su tensión arterial no baja lo suficiente. De todo ello se concluye que aunque se recojan datos de ensayos clínicos, éstos deben ajustarse según la efectividad, con el fin de que reflejen mejor los resultados que presumiblemente se darán en la realidad.

A veces hay varios estudios similares sobre el mismo tema que podrían constituir la fuente de información, pero pueden ser de pequeño tamaño y no dar todos precisamente los mismos resultados. En estos casos

se puede realizar un meta-análisis: es el proceso de utilizar métodos estadísticos para combinar los resultados de diferentes estudios, la selección de los cuales se basa en su validez interna (ausencia de sesgos), y en su validez externa (las condiciones de aplicación del estudio deben ser representativas de la situación a la que se pretenden generalizar, extrapolar los datos).

Si no se dispone de fuentes documentadas de datos se puede recurrir a la opinión de los expertos, ya que aunque no dejan de ser suposiciones, una suposición informada siempre es mejor que la ausencia total de información objetiva. De todos modos, no son métodos sencillos, requieren métodos especiales. En la técnica del Grupo Nominal se reúne a un grupo de expertos (entre cinco y nueve) a los que se permite discutir durante fases específicas del proceso, de ahí que en la mayoría de las fases, de grupo sólo tiene el nombre. En la técnica Delphi, más elaborada, no se reúne a los participantes, se les van realizando cuestionarios sucesivos que se envían sólo después de analizar los resultados del anterior.

En el peor de los casos no se dispone de otra fuente de información que la propia experiencia, y si se trata de un problema de decisión complejo esta experiencia suele ser muy pobre, por no decir nula.

En la mayoría de los casos, las fuentes de datos para un estudio suelen ser múltiples, rara vez un ensayo coincide plenamente con los objetivos de nuestro estudio, y si coincide ¿para qué realizar otro estudio igual?. La fiabilidad de las diferentes fuentes de información suelen ser muy distintas y por ello es conveniente reflejarlas con la máxima transparencia.

ANÁLISIS DE SENSIBILIDAD

Como hemos podido observar, la mayoría de los procesos ocurren de un modo aleatorio, los datos recogidos de la literatura pueden reflejar hechos a veces muy diferentes o incluso contradictorios, o no hay otra información disponible que la experiencia personal, y ésta puede ser pobre en el problema planteado (de ahí que se haya convertido en un problema), por lo que la incertidumbre rodea todo el proceso de decisión. Por ello es preciso realizar un análisis de sensibilidad para comprobar la robustez de los resultados. Se trata de realizar el análisis según diferentes supuestos para los parámetros que pueden presentar un margen de variabilidad (probabilidad de morir en quirófano, probabilidad de curar tras un tratamiento, de dar un resultado positivo a una prueba, coste de las pruebas diagnósticas, etc.), que desafortunadamente pueden ser casi todos. Cuando los resultados no cambian mucho según los distintos valores de las variables se dice que es un análisis robusto, de mayor fiabilidad que cuando son sensibles a los cambios.

RESUMEN

Los árboles de decisión son un método gráfico en el que se representan y comparan los resultados esperados de cada una de las alternativas de elección. Su elaboración requiere los siguientes pasos:

1. Definir el problema de decisión (son los objetivos del estudio).
2. Identificar las alternativas de decisión.
3. Listar todos los posibles resultados de cada una de las alternativas de elección.
4. Representar la secuencia de sucesos tras cada alternativa mediante nudos de decisión y nudos de azar. Los nudos de decisión, que se representan mediante cuadrados, son el punto del árbol de donde emergen las ramas que representan decisiones a tomar, y los nudos de azar, que se representan mediante círculos, son el punto del árbol del que emergen ramas que representan lo que puede suceder.
5. Elegir un horizonte temporal consistente con el problema.
6. Determinar la probabilidad de cada resultado, teniendo en cuenta que las probabilidades de todas las ramas que salen de un nudo de azar deben sumar 1.
7. Asignar un valor a cada uno de los resultados: si se trata de efectos sobre la salud, éstos pueden ser vivo o muerto, esperanza de vida o más generalmente utilidad.
8. Calcular el valor esperado de cada alternativa de decisión, sumando cada uno de los resultados por la probabilidad de que ocurra, teniendo en cuenta que esta probabilidad final es el producto de las probabilidades a lo largo de las ramas del árbol de decisión.
9. Elegir la alternativa con el valor esperado sobre la salud más elevado (si sólo se tienen en cuenta los resultados sobre la salud), con los costes más bajos (si lo que se valoran son los costes) o con la relación costes/resultados sobre la salud menor (si lo que se valora es la eficiencia).
10. Realizar un análisis de sensibilidad para comprobar la robustez de las conclusiones del estudio.

PROBLEMAS

1. Tras aplicar el tratamiento A, el 15% de los pacientes mueren durante el primer año, el 35% durante el segundo año, el 35% durante el tercer año, el 10% durante el cuarto año y el 5% durante el quinto año. Tras aplicar el tratamiento B, el 5% mueren durante el primer año, el 15% durante el segundo año, el 25% durante el tercer año, el 35% durante el cuarto año, el 15% durante el quinto año, y el 5% durante el sexto año. ¿Cuál de los dos tratamientos es más efectivo?
2. Acude a su consulta un paciente y usted sospecha que tiene la enfermedad de Sosika. El diagnóstico se realiza en base a la exploración física, porque existe un signo, las manchas de Pinbian (zonas eritematosas rodeadas de un borde azulado en las palmas de las manos), que siempre están presentes en estos enfermos, pero también pueden presentarlo individuos sin la enfermedad. Si no se trata, la enfermedad de Sosika resulta letal en el 15% de los casos. El tratamiento quirúrgico cura la enfermedad en el caso de que el paciente no muera en quirófano, hecho que ocurre en el 5% de las veces. ¿Con qué probabilidad de tener la enfermedad, dado que presenta las manchas de Pinbian, sería conveniente intervenir quirúrgicamente al paciente?
3. La enfermedad E está producida por un virus que permanece latente por un tiempo de 6 meses a partir del momento del contagio. La incidencia anual de la infección en la población es del 10%. Existe un tratamiento preventivo, con un coste de 100 pts. por individuo, que evita la aparición de la enfermedad en el 100% de los casos y no tiene ningún efecto secundario. Alternativamente, la enfermedad puede ser tratada eficazmente cuando aparecen los primeros síntomas, sin que llegue a producirse ningún efecto apreciable sobre la salud del paciente, con un coste global de 1.100 pts./caso.
 - a) ¿Es más eficiente prevenir o curar?
 - b) Si la eficacia del tratamiento preventivo fuese sólo del 80% ¿Sería preferible prevenir o curar?

PROBABILIDAD PREVIA

Como veremos en el capítulo siete, la probabilidad de presentar una enfermedad tras conocer el resultado de una prueba diagnóstica depende del resultado de la prueba y de la probabilidad de la enfermedad antes de aplicar dicha prueba.

La probabilidad de presentar la enfermedad antes de aplicar una prueba se denomina probabilidad previa (o preexamen), que es, claro está, un concepto relativo. Una misma probabilidad estimada de enfermedad puede ser previa o posterior dependiendo del momento en el proceso diagnóstico en que nos encontremos. La probabilidad posterior de un cálculo nefrítico tras la exploración física es la probabilidad previa a la realización de la ecografía, por ejemplo. Sin embargo, se tiende a pensar que probabilidad previa es sólo aquella primera probabilidad estimada por el médico durante la anamnesis y la exploración física. Quizás el error provenga del hecho de que no tenemos en mente que cada uno de los apartados de la historia y el examen físico constituye un examen diagnóstico que aumenta o disminuye la probabilidad de un trastorno objetivo. Imaginemos un paciente varón de 28 años, que acude a urgencias por disnea de inicio agudo y un vago dolor torácico que comenzaron tras finalizar un viaje de 10 horas en automóvil. El paciente había experimentado episodios similares en el pasado, aunque ninguno de esta gravedad y es muy aprensivo. La exploración física es normal. El clínico, con cada uno de estos datos va recalculando la probabilidad de la enfermedad, supongamos embolismo pulmonar, y así la probabilidad aumenta por el hecho de que haya permanecido inmóvil durante 10 horas, pero la edad del paciente, la ausencia de enfermedad previa y el examen físico la disminuyen.

Otro error frecuente es confundir probabilidad previa con prevalencia de enfermedad en la población general. Este hecho muy probablemente deriva del primer desarrollo de la teoría de la decisión en salud pública, donde se aplican pruebas de cribaje a grupos de la población independientemente de la presencia de síntomas, y la primera probabilidad estimada es, por supuesto, la prevalencia de la enfermedad en la población a la que va dirigido el cribaje. Si hablamos de cribaje de cáncer de mama a las mujeres entre 50 y 65 años, por ejemplo, la probabilidad previa será la prevalencia de esta patología en este grupo de la población. Pero esta probabilidad no es la misma en el caso de una mujer de 55 años que

acude a la consulta por presentar desde hace 2 semanas una tumoración en la mama.

De todos modos, este capítulo se dedica a la estimación de la probabilidad que realiza el médico en la primera visita, y por todo lo aclarado hasta ahora, para ser más correctos, debería titularse algo así como "Estimación de la probabilidad de la enfermedad durante la anamnesis y la exploración física".

La situación ideal consiste en basarse en datos documentados y de hecho existen revisiones sobre la validez y fiabilidad de la historia clínica y el examen físico. Incluso para algunos trastornos objetivos como la isquemia miocárdica, los análisis multivariantes pueden proporcionarnos sistemas de combinar la información para generar probabilidades preexamen muy precisas. Pero en general no se dispone de mucha información sobre las propiedades de los componentes de la anamnesis y la exploración, y ello conlleva que a menudo los clínicos realicen las estimaciones de estas probabilidades de forma intuitiva, basándose en la propia experiencia, con una amplia variación entre ellos. En estos casos, una opinión de consenso puede ser más precisa. Y por último, cuando aún seguimos teniendo dudas acerca de la probabilidad previa (preexamen), podemos suponer la más alta y baja plausibles, y comprobar si esto modifica nuestro curso de acción.

EXPERIENCIA PERSONAL

La propia experiencia con pacientes de características similares es lo que más influye a la hora de estimar la probabilidad de la enfermedad. Estos "limitados" datos los tenemos almacenados en la memoria, y cada vez que se estima una probabilidad se deben traer a la mente, mediante unos procesos cognitivos que se denominan heurísticos. A continuación se exponen los principales y los errores más importantes a los que se pueden ver sujetos (al fin y al cabo heurística es el arte de inventar).

Representatividad

Según esta regla heurística, la probabilidad de la enfermedad se calcula por la semejanza entre el paciente de la consulta y un grupo típico de pacientes, aquellos que presentan las características fundamentales que clásicamente han definido la enfermedad en cuestión. Las características de este grupo típico las hemos podido aprender por el estudio, pero generalmente se matizan por la propia experiencia. Por ejemplo, acude a urgencias un paciente de 52 años que al acostarse comenzó con un dolor precordial opresivo, que irradiaba hacia la mandíbula y el hombro izquierdo, y la duración fue de 4 ó 5 minutos. Son los síntomas típicos de la angina de pecho y por ello se le asigna una probabilidad alta a esta enfermedad.

Este tipo de proceso mental puede estar sujeto a errores por los siguientes motivos:

- Por comparar al paciente con una experiencia pobre y poco representativa de la enfermedad, porque en muchos casos la presentación de la enfermedad puede ser atípica. Por ejemplo, si sólo se han visto cuatro casos de hipotiroidismo, no se puede esperar una gran concordancia entre el paciente actual y esta experiencia previa, ya que además esta enfermedad se puede manifestar de muy diferentes modos (presenta síntomas tan inespecíficos como fatiga, piel seca, sensación de frío, estreñimiento, caída del cabello, menorragia, etc.).
- Cuando la enfermedad es rara, se sobrevalora la probabilidad de la enfermedad por la coincidencia de los síntomas, pero se olvida el hecho de que de entrada es una enfermedad muy poco frecuente. Por ejemplo, diagnosticar una leishmaniasis visceral ante un paciente con fiebre, adenopatías y hepatoesplenomegalia porque los síntomas concuerdan perfectamente con el cuadro típico, olvidando que en nuestro país es una enfermedad poco frecuente.
- Se utilizan, para valorar la representatividad, factores poco correlacionados con la enfermedad. Por ejemplo, no se realiza el diagnóstico de asma por el hecho de que el paciente no presenta sibilancias, cuando precisamente el asma grave no se correlaciona con la presencia de sibilancias.
- Presencia de predictores redundantes. Si un conjunto de síntomas casi siempre aparecen juntos, no se debe asignar más probabilidad a un diagnóstico por cada uno de estos signos o síntomas que aparezcan. Por ejemplo, se sobrestima la probabilidad de úlcera de duodeno en un paciente que refiere dolor abdominal que calma con antiácidos, leche y la ingesta en general. Se trata de signos redundantes, reflejando todos ellos el mismo hecho fisiológico de calmar el dolor por la neutralización del ácido.
- Por utilizar la regresión a la media erróneamente como evidencia diagnóstica. La regresión a la media es el hecho de que las determinaciones extremas (muy altas o muy bajas) en un principio, tienden a no ser tan extremas (más parecidas a la media) en una segunda determinación, por ello a veces suele ser conveniente, antes de emitir un diagnóstico, repetir las pruebas. Si en parte el diagnóstico de la hipertensión se realiza porque disminuye la tensión arterial tras la administración de antihipertensivos, esta respuesta podría deberse a la regresión a la media más que a una verdadera respuesta al tratamiento.

Disponibilidad

Este heurístico está basado en el hecho de que la probabilidad de un suceso se estima por la facilidad con que se recuerdan sucesos similares (disponibilidad de datos). Presenta la ventaja de que los hechos más frecuentes (con una probabilidad más alta) se recuerdan mucho más fácilmente que los infrecuentes. Sin embargo, en este proceso cognitivo influyen en gran medida factores emocionales que pueden llevarnos, por diferentes motivos, a estimar erróneamente la probabilidad de la enfermedad:

- La rareza del caso también favorece el que se recuerde con facilidad. El haber diagnosticado una porfiria en un paciente con dolor abdominal puede sobrestimar la probabilidad de porfiria en otros pacientes que acudan con esta misma sintomatología.
- La intensidad con la que se vivió determinado caso, por raro que fuera, también influye en su recuerdo. Si el diagnóstico de su primer paciente fue una arteritis de Horton, se puede sobrestimar la probabilidad de esta enfermedad.
- La capacidad de recordar un diagnóstico también depende de las consecuencias que tuvo para el paciente: se tiene mucho más presente el diagnóstico de aneurisma de aorta abdominal ante un dolor lumbar cuando el paciente, además de ser un ser muy querido, ha muerto por ello.
- Las consecuencias que tuvo para el médico también pueden influir en su recuerdo. Ante una paciente que presenta cansancio, irritabilidad y somnolencia se puede sobrestimar el diagnóstico de ferropenia si este fue el motivo de su tesis doctoral.
- Por último, la cercanía en el tiempo también favorece el recuerdo.

Anclaje y ajuste

Según este heurístico, se parte de una probabilidad inicial de enfermedad (probabilidad de anclaje) y ésta se va ajustando según se van conociendo características especiales del paciente. Por ejemplo, acude a la consulta una paciente de 35 años con un bulto en la mama poco sugestivo de carcinoma (móvil, no adherido, bien delimitado), por lo que estima que la probabilidad de esta enfermedad es del 15%. Sin embargo, durante la anamnesis, preguntando por los antecedentes familiares descubre que su madre y dos hermanas murieron de adenocarcinoma de mama, por lo que la probabilidad de que la paciente padezca esta enfermedad asciende mucho.

También hay razones para equivocarse por este método de estimación de las probabilidades y son las siguientes:

- Se establece una probabilidad de anclaje incorrectamente, porque generalmente, las probabilidades estimadas en un principio suelen ser muy extremas, o muy bajas (cerca de cero) o muy altas (cerca de 1). Si se presentan varias características conjuntamente se sobrestima la probabilidad de la enfermedad, y en cambio, cuando los síntomas se presentan de forma aislada no se da mucha fuerza a estas hipótesis diagnósticas. De hecho, al acabar la carrera, cuando la experiencia es escasa (no así el nivel de conocimientos en cuanto a la cantidad de datos) se tiende a buscar síndromes donde encajar todos los síntomas y signos que presenta el paciente, cuando éstos pueden corresponder a patologías diferentes.
- Generalmente no se ajustan bien las probabilidades conforme se adquiere más información, se tiende a ser muy conservador con el juicio inicial, lo que suele conllevar el realizar más pruebas diagnósticas de las necesarias.

Después de la lectura de este capítulo, no se trata de realizar el diagnóstico pensando en cada momento “ahora aplico el heurístico de la representatividad, ahora estimo la probabilidad de anclaje y ahora la ajusto”. Pero la comprensión del proceso cognitivo que nos lleva al diagnóstico y sobre todo el conocimiento de las posibles fuentes de error ayudan mucho a realizar una estimación más correcta de la probabilidad de la enfermedad.

OTRAS FUENTES DE DATOS

Los datos de la bibliografía son el segundo punto de referencia, de gran utilidad, ya que aportan información sobre la incidencia y la prevalencia de las distintas enfermedades, y éstos son unos puntos de partida (probabilidad de anclaje) muy útiles. Además, si se trata de enfermedades raras, los estudios pueden aportar mucha más experiencia de la que podría llegar a tener el más experimentado clínico.

Pero los estudios también pueden tener errores, entre los que el más importante es el sesgo de selección. Este sesgo se produce, por ejemplo, cuando los criterios de selección de los individuos del estudio son diferentes de las condiciones que generalmente llevan al paciente a buscar atención médica. La mayoría de los estudios están basados en los pacientes vistos en centros de referencia (donde más investigación se realiza) y por tanto no son representativos de los pacientes vistos en atención primaria, por ejemplo, donde la prevalencia de la enfermedad lógicamente será más baja. Por otro lado, los pacientes estudiados por el médico especialista son generalmente pacientes que presentan ya algún síntoma, o incluso el resultado de algunas pruebas diagnósticas y en estos casos la probabilidad de presentar la enfermedad es mayor.

A veces deben tenerse en cuenta características especiales de los pacientes al calcular la probabilidad de una enfermedad, porque pueden ser muy importantes, pero por su rareza no se encuentran como factores discriminantes en la literatura. Por ejemplo, un hombre de 30 años tiene las características de lo que podría ser una angina de pecho. Su padre y dos hermanos tuvieron un infarto de miocardio fatal antes de los 40 años. Es tan raro presentar tantos familiares tan jóvenes con IAM que estas variables no eran factores independientes predictores en los análisis multivariados descritos en la bibliografía. Pero el sentido común nos indica que es una historia alarmante que debe modificar mucho nuestro juicio sobre la probabilidad de la enfermedad.

INCIDENCIA, PREVALENCIA, ODDS PREVIA

Las medidas de frecuencia de una enfermedad más comúnmente utilizadas son la incidencia y la prevalencia.

La prevalencia es la proporción de individuos de una población que tiene la enfermedad en un momento determinado. Esta medida estima la probabilidad de que un individuo esté enfermo en aquel momento.

$$\text{Prevalencia} = \frac{\text{Número de personas con la enfermedad}}{\text{Población total}}$$

Por simplicidad y a efectos didácticos se habla de población total, pero se puede estimar la prevalencia de una enfermedad en grupos específicos de la población (por ejemplo cáncer de mama en mujeres mayores de 50 años) e incluso en presencia de determinados síntomas (prevalencia de cáncer de mama en mujeres mayores de 50 años que acuden a la consulta por notar una tumoración en la mama).

El momento en el que se estima la prevalencia puede corresponder a un punto concreto del calendario (por ejemplo, 1 de enero de 2001) o, por el contrario, puede representar un instante fijo para el acontecimiento que se mide, pero varía de una persona a otra en cuanto a tiempo real se refiere (por ejemplo, el momento del diagnóstico o el décimo día de tratamiento).

Ya que se trata de una proporción (el numerador está contenido en el denominador), su valor siempre está entre 0 y 1 (aunque generalmente se expresa en porcentajes, o en tantos por 1.000 o por 100.000).

La incidencia (sólo comentaré la incidencia acumulada) es la proporción de individuos que a lo largo de un período de tiempo determinado desarrolla la enfermedad.

$$\text{Incidencia} = \frac{\text{Número de casos nuevos en un período de tiempo}}{\text{Total de la población en riesgo}}$$

El numerador son los casos nuevos que aparecen a lo largo de un período de tiempo, y el denominador la población en riesgo de desarrollar la enfermedad.

Por ejemplo, si queremos estudiar la incidencia de diabetes a lo largo de un año, el numerador estará constituido por los pacientes que a lo largo del año desarrollen la diabetes (casos nuevos) y del denominador no deben formar parte aquéllos que al inicio del estudio ya eran diabéticos, puesto que no están en riesgo.

La incidencia estima la probabilidad o riesgo de que un individuo presente la enfermedad en un período determinado de tiempo.

Al igual que con la prevalencia, también la incidencia de una enfermedad se puede estimar en grupos específicos de la población.

Para entendernos, la incidencia es la película a lo largo de un período de tiempo y la prevalencia la foto de la situación en un momento concreto. Pero prevalencia e incidencia están estrechamente relacionadas:

$$\text{Prevalencia} = \text{Incidencia} * \text{Duración}$$

A mayor incidencia mayor prevalencia, cuantos más casos nuevos aparezcan más fácilmente es detectarlos en el momento de la foto. Pero a su vez depende de la duración de la enfermedad: si se trata de una patología crónica dará lugar a una prevalencia más alta que si la enfermedad cura o mata rápidamente.

Una odds es el cociente entre la probabilidad de que un hecho ocurra y la probabilidad de que no ocurra. Es otra forma de expresar la frecuencia con la que puede estar presente una enfermedad, ampliamente utilizada en los países de habla inglesa.

$$\text{Odds} = \frac{p}{1-p}$$

Si la probabilidad previa de una enfermedad es 0,25, la odds previa es 0,33.

$$\text{Odds previa} = \frac{0,25}{0,75} = 0,33 \text{ (1:3 ó 1 a 3)}$$

Otra forma de expresar esta odds es 1 a 3, queriendo decir que por cada individuo enfermo hay 3 individuos sin la enfermedad.

La transformación de odds en probabilidades también es sencilla:

$$p = \frac{\text{odds}}{1+\text{odds}}$$

Si la odds de una enfermedad es 0,25 (1 a 4), la probabilidad de esta enfermedad es 0,20.

$$p = \frac{0,25}{1,25} = 0,20$$

RESUMEN

La estimación de la probabilidad de una enfermedad se basa fundamentalmente en la experiencia personal y en los datos de la literatura. Sin embargo, ambas fuentes pueden estar sujetas a errores.

La experiencia puede inducir a errores especialmente cuando no se tiene mucha y sobre todo si se trata de enfermedades raras, cuando se presentan signos y síntomas redundantes o los casos vistos se han vivido de una forma especial. Además, se tiende a ser extremista en la primera estimación de la probabilidad de la enfermedad y conservador a la hora de ajustar dicha probabilidad por la nueva información adquirida.

Los datos de la bibliografía pueden inducir a errores fundamentalmente por el sesgo de selección: los pacientes sobre los que se ha realizado el estudio difieren en características relevantes de los pacientes que habitualmente un clínico puede ver en su consulta.

Las medidas de probabilidad, de frecuencia de una enfermedad, más ampliamente utilizadas son la incidencia (casos nuevos que aparecen a lo largo de un período de tiempo) y la prevalencia (número de casos con la enfermedad en un momento determinado). Esta probabilidad también se puede expresar en forma de odds.

PROBLEMAS

1. Un colega os informa de datos sobre un determinado paciente. Uno de los datos es erróneo. ¿Cuándo le inducirá a usted más a error, si el dato erróneo se lo dice al principio o si se lo dice al final?
2. Ante un lactante varón que presenta vómitos alimentarios en escopetazo se estima que la probabilidad de una estenosis hipertrófica del píloro es del 50%. ¿De qué manera se puede cometer un error basado en la representatividad?
3. Se presenta una paciente con disnea de esfuerzo, ortopnea, taquicardia, claudicación y vértigo. Usted estima que la probabilidad de anemia es del 60%. ¿De qué manera se puede cometer un error basado en la representatividad?
4. Usted es especialista en glomerulonefritis. Ve a un niño con edema palpebral y estima que la probabilidad de padecer un síndrome nefrótico es del 15%. ¿Qué tipo de proceso mental está implicado en la estimación de la probabilidad? ¿Qué errores se pueden estar cometiendo?
5. Un artículo presenta una serie de 128 pacientes con enfermedad de Crohn atendidos consecutivamente en un hospital provincial. Once de estos pacientes están afectados de psoriasis. ¿Qué medida de frecuencia de la enfermedad se ha utilizado, prevalencia o incidencia?
6. Un neumólogo publica una revisión de todas las personas ingresadas durante 1986 en el hospital donde trabaja. De 23.147 pacientes ingresados, 294 tenían tuberculosis pulmonar. El autor concluye que la incidencia de tuberculosis pulmonar en el área de cobertura del hospital es 12,7 casos por 1000 habitantes y año $((294/23.147) * 1000)$. ¿Correcto o falso?

FIABILIDAD

CALIDAD DE LA INFORMACIÓN

Antes de entrar en materia de fiabilidad es importante definir una serie de conceptos que suelen llevar consigo bastante confusión y todos ellos afectan a la calidad de la información en la práctica clínica.

Los individuos somos y reaccionamos de un modo diferente, y esta variabilidad explica que la medicina no sea una ciencia exacta. Por ello, para describir con un sólo dato qué ocurre en un grupo de individuos, continuamente se calculan medias de colesterol, porcentajes de individuos que desarrollan una diabetes, media de mg de hierro diarios que resultan eficaces para el tratamiento de la anemia, porcentaje de individuos que cumplen el tratamiento y curan, milímetros de Hg que ha disminuido la tensión arterial, etc.

Para la estimación de estas medias y porcentajes se selecciona una muestra de la población global, lo más representativa posible, pero nada garantiza que esta muestra sea efectivamente representativa, siempre existe una posibilidad de error debido al azar y es el error aleatorio (error derivado de trabajar con muestras y no con poblaciones completas). Esta es la razón de que en las publicaciones científicas se expresen las estimaciones acompañadas del intervalo de confianza (lo que equivaldría a dar un margen de error). Cuanto mayor es el tamaño de la muestra estudiada, menor es el error aleatorio, más estrecho es el intervalo de confianza, mayor es la precisión en la estimación del dato, hasta el punto de que si se estudiase a toda la población el error aleatorio sería nulo.

Pero con estudiar grandes muestras no se soluciona el problema, también existe la posibilidad de los errores sistemáticos o sesgos. Son desviaciones de los resultados de la verdad que ocurren de un modo sistemático, como seleccionar para el estudio de la especificidad de una prueba voluntarios sanos. La presencia de estos sesgos afecta a la representatividad de la muestra, a la validez interna de los resultados.

Estos dos tipos de errores (aleatorio y sistemático) no sólo afectan al muestreo, sino también a la medición.

La validez o exactitud de una prueba es el grado en el que ésta mide lo que tiene que medir. Es la ausencia de errores sistemáticos o sesgos, como el caso de la balanza que siempre mide un kg de más, por ejemplo. (este tema se desarrolla más ampliamente en el siguiente capítulo).

La fiabilidad es el grado de estabilidad conseguido cuando se repite una medición en condiciones similares. A pesar de tomar a un paciente la tensión arterial todos los días, a la misma hora y en la misma postura, pueden resultar cifras diferentes. Pero aquí surge el problema, si por ejemplo, nuestro verdadero nivel de colesterol es de 190 mg/dl y un día concreto ha resultado 210 mg/dl (lo que podría inducir a cambiar los hábitos dietéticos), esta desviación del verdadero valor ¿se debe a un error sistemático del aparato de medida, a que los resultados dados varían según quién analice los datos, a nuestra propia variabilidad o al azar? En cualquier caso, el resultado no es el correcto y tan importantes son la validez como la fiabilidad de una prueba, la ausencia de cualquiera de ellas conlleva un error que hay que tratar de minimizar. En la figura 4 se describen las cuatro combinaciones posibles: de izquierda a derecha y de arriba a abajo, la primera figura es una medida válida (da en la diana) y fiable (no varían mucho los resultados de una medición a otra); la segunda medida no es válida pero sí fiable, la tercera es válida (en media el resultado es el correcto, es el centro de la diana) pero no fiable (hay una gran variabilidad) y la cuarta medida no es ni válida ni fiable.

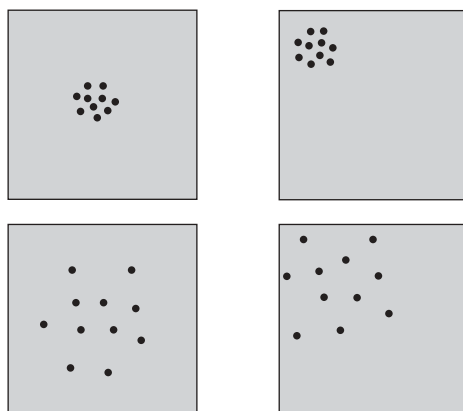


Figura 4. Validez y fiabilidad. De izquierda a derecha y de arriba abajo, se trata de cuatro pruebas o medidas con las siguientes características: válida y fiable, medida no válida pero fiable, válida no fiable y medida ni válida ni fiable.

Tan importante es la validez como la fiabilidad, porque (Fig. 5) ¿qué es preferible, el resultado extremo de una medida válida pero no fiable (con lo que no se parece en nada a la realidad) o el resultado extremo de

una medida que ni tan siquiera es válida pero por azar en esta determinación ha dado un resultado correcto? El día que hagamos la determinación en el paciente no sabemos si se trata de un valor extremo o no, con lo que no sabemos si es el correcto o no.

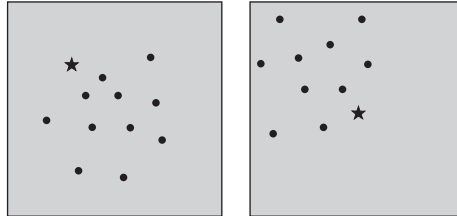


Figura 5. Elección entre validez y fiabilidad.

En la literatura muchas veces se utilizan indistintamente los conceptos de fiabilidad, reproducibilidad o repetibilidad, cuando estos últimos sólo se refieren a la acción de realizar algo más de una vez, no al hecho de que siempre den los mismos resultados. A un individuo se le puede tomar la temperatura axilar varias veces (es una prueba repetible, reproducible), lo cual no quiere decir que siempre vaya a dar el mismo resultado (fiabilidad).

Por último, la validez externa es la capacidad de extrapolar, de generalizar unos resultados. Para poder generalizar unas conclusiones hace falta primero, que estas conclusiones sean válidas y fiables, y en segundo lugar que la población a la que se quieren extrapolar no difiera en las características relevantes de la población en la que se ha realizado el estudio. Por ejemplo, por razones obvias sería ilógico extrapolar la prevalencia de anemia ferropénica en las mujeres en edad fértil a todo el conjunto de mujeres, pre y postmenopáusicas, ya que difieren en uno de los principales factores de riesgo de la anemia, la menstruación.

FUENTES DE VARIACION

La variación de una medida puede deberse a la variabilidad intra-observador, variabilidad interobservador (por falta de objetividad), debida al procedimiento o a la variabilidad de los individuos observados. Por ejemplo, dos enfermeras miden la tensión arterial de un sujeto y registran resultados diferentes. Las posibles explicaciones a estas diferencias son:

- el individuo se puso nervioso porque dos personas lo examinaban y aumentó su tensión arterial.

- si usaron el mismo esfigmomanómetro éste no da siempre los mismos resultados, o utilizaron diferente aparato con la consiguiente variabilidad.
- la sensibilidad auditiva de las dos enfermeras es distinta.

La característica medida puede variar, en el mismo individuo, por un sinnfín de factores. Siguiendo con el ejemplo de la tensión arterial, ésta varía según la postura en la que se toma, la hora del día, el estrés del paciente. Muchas determinaciones hormonales tienen un ritmo circadiano, el colesterol sérico puede variar en función del contenido en grasas de las comidas previas al análisis, la respuesta a un cuestionario puede variar según la hora del día, el grado de aburrimiento, las motivaciones, si el paciente está en casa o en el hospital, si hay otra persona delante mientras contesta. Una persona puede variar su conducta o su respuesta al saberse estudiada, variando su dieta o lo que dice que come si sabe que la ingesta calórica es lo que se está midiendo. En los ensayos clínicos, la respuesta al tratamiento puede variar si el paciente sabe que está tomando el tratamiento o es un control (efecto placebo). Los valores extremos (muy elevados o muy bajos) en una segunda determinación tienden a no ser tan extremos, es lo que se denomina regresión a la media.

Del mismo modo que los instrumentos diferentes pueden dar resultados diferentes, un mismo instrumento puede también variar en su respuesta en función de factores como el frío, el calor, la luz, el ruido, etc.

La variación entre diferentes observadores viene dada por la diferente formación (no ausculta igual un soplo un experimentado cardiólogo que un médico residente), su sensibilidad (la diferente agudeza visual en el examen del fondo de ojo), su propensión a cometer errores (en un ensayo clínico se puede juzgar de manera diferente a los pacientes si se sabe si son sujetos experimentales o controles), la diferente motivación (si se busca un determinado efecto se pregunta con más insistencia sobre su presencia), la diferente costumbre (un médico de la unidad de coronarias considera normal a un paciente con unos niveles de hemoglobina tales que de ser vistos por un médico de asistencia primaria resultaría con el diagnóstico de anemia), etc.

MODOS DE AUMENTAR LA FIABILIDAD

Es prácticamente imposible anular la variabilidad de las mediciones (siempre queda una parte que se la asignamos al azar), pero se debe intentar reducir al mínimo. La clave es la siguiente: la prueba diagnóstica a utilizar (preguntas de la anamnesis, exploración física, análisis clínico) ha de ser estándar y aplicada en condiciones estándar. Quiere decir que la determinación de colesterol se debe realizar por el mismo método, y en ayunas siempre, por ejemplo; la tensión arterial se debe tomar en la mis-

ma postura; para conocer la prevalencia del hábito tabáquico, se debe hacer la pregunta siempre del mismo modo, no insistiendo más en los pacientes con bronquitis crónica. Todas aquellas pruebas que dependen especialmente de la subjetividad deben poseer definiciones operativas claras, lo que las convierte casi en pruebas objetivas. Si la prueba es difícil de realizar, el o los investigadores deben estar formados y todos del mismo modo. Si a pesar de todo la variabilidad interobservador es elevada, es conveniente que los datos sean recogidos por un único investigador.

Y en general, el aumentar el número de mediciones, dando como resultado la media, aumenta en gran medida la fiabilidad. De hecho, el diagnóstico de hipertensión arterial está basado en tres tomas en condiciones basales.

MEDICIÓN DE LA FIABILIDAD

La fiabilidad de una medida es el grado de estabilidad en la misma. Si se realizan dos análisis del ácido úrico en el mismo individuo, ¿varía mucho de una a otra vez?; si un fondo de ojo lo examinan dos expertos oftalmólogos ¿llegan a las mismas conclusiones? Para el estudio de la fiabilidad, por lo tanto, es preciso medir la variable en cuestión dos o más veces y valorar la variación que se produce para un mismo individuo. Para el estudio de la fiabilidad del método, se deben mantener constantes todos los demás factores que puedan influir en la variabilidad de los resultados, como puede ser el observador. Si, por el contrario, lo que se pretende estudiar es la fiabilidad del observador, el método debe ser el mismo.

Las pruebas estadísticas que se emplean varían según la naturaleza de la variable en estudio: si se trata de una variable cualitativa (enfermo o sano, resultado positivo o negativo, grado leve, moderado o severo), la fiabilidad se estudia generalmente mediante el índice de concordancia global y el índice kappa; si se trata de una variable cuantitativa (presión de oxígeno, número de colonias), los estadísticos más utilizados son el coeficiente de variación y el coeficiente de correlación intraclases.

Concordancia global

Con un ejemplo resultará más sencilla la comprensión de este análisis. Con el fin de estudiar la concordancia intraobservador para el diagnóstico rápido de la infección por *Chlamydia trachomatis* mediante inmunofluorescencia directa, se presentaron las mismas muestras dos veces al mismo observador. Los resultados son los de la tabla 2.

Tabla 2. Concordancia en el diagnóstico de infección por *Chlamydia trachomatis*. Tomado de: Hernández Aguado I, Ruiz-Atienza L, Fernández García I, Fernández García E, Alvarez Dardet C. Evaluación de la variabilidad del diagnóstico rápido de infección por *Chlamydia trachomatis* mediante inmunofluorescencia directa. Med Clin (Barc) 1990; 94: 9-11

		2ª vez		
		+	-	
1ª vez	+	22 (a)	4 (b)	26
	-	6 (c)	160 (d)	166
		28	164	192

La primera vez, resultaron positivas para *Chlamydia trachomatis* 26 muestras y la segunda vez, en cambio, 28. En cuatro muestras, el diagnóstico la primera vez fue positivo y la segunda vez negativo, y en seis muestras la primera vez el diagnóstico fue negativo y la segunda vez positivo. Éstos son lo que se denominan pares discordantes, existe un desacuerdo en la interpretación entre la primera y la segunda vez.

El resultado fue el mismo una y otra vez en $22+160 = 182$ casos, que expresado en porcentajes es el 94,8% ($182/192$) de concordancia global.

Índice kappa

La concordancia global del 94,8% vista en el ejemplo anterior, en parte podría ser debida al azar, ya que si los analistas decidiesen el diagnóstico lanzando una moneda al aire también coincidirían en algunos casos.

El índice kappa es una medida de concordancia que corrige por este efecto del azar y la fórmula es la siguiente:

$$\text{Índice Kappa} = \frac{p_o - p_e}{1 - p_e}$$

donde:

p_o = concordancia observada

p_e = concordancia esperada por azar

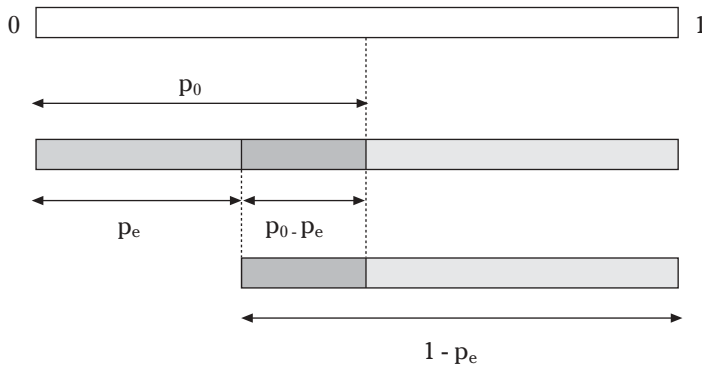


Figura 6. Índice Kappa.

El razonamiento detrás de esta fórmula (Fig. 6) es que de todo el acuerdo posible no debido al azar ($1-p_e$), qué cantidad de acuerdo real ha habido (p_0-p_e).

Si el 13,5% (26/192) de las muestras resultan positivas la primera vez y el 14,6% (28/192) resultan positivas la segunda, por puro azar, si las respuestas en ambos casos fuesen independientes, el 1,97% (13,5% * 14,6%) del total de las muestras (192*1,97% = 3,78) resultarían positivas las dos veces.

Del mismo modo, si en el 86,5% (166/192) de las muestras el diagnóstico es negativo la primera vez, y en el 85,4% (164/192) resulta también negativo en la segunda, si ambas respuestas fuesen independientes, sólo por azar el 73,8% (86,5% * 85,4%) del total de las muestras (192*73,8% = 141,8) resultarían negativas las dos veces.

Como norma general, los valores esperados por azar se calculan multiplicando los marginales (totales de filas y columnas) y dividiendo por el total (Tabla 3). El esperado de la celda 'a' es $28*26/192 = 3,8$. El esperado de la celda 'd' es $164*166/192 = 141,8$. En realidad, en una tabla de 2x2 sólo es preciso calcular uno de los valores esperados, el resto salen por la diferencia con los totales. El número de esperados a calcular es (filas-1)*(columnas-1). En las tablas de 2x2 hay dos filas y dos columnas, luego el número de valores a calcular es $(2-1)*(2-1) = 1$.

Tabla 3. Valores esperados por azar.

		2ª vez		
		+	-	
1ª vez	+	3,8 (a)	22,2 (b)	26
	-	24,2 (c)	141,8 (d)	166
		28	164	192

Entonces, la concordancia esperada por azar es:

$$p_e = (3,8+141,8)/192 = 75,8\%$$

Si la concordancia global observada era del 94,8%, el índice kappa es:

$$\text{Índice Kappa} = \frac{0,948-0,758}{1-0,758} = 0,785$$

Los valores del índice kappa oscilan entre -1 y +1, de modo que un índice kappa negativo indica que el acuerdo es menor que el que se podría haber obtenido por azar, si es igual a cero todo el acuerdo se explica por el azar, y un índice kappa igual a 1 indica un acuerdo perfecto. Como norma general, cuando el índice es menor de 0,4 indica una escasa fiabilidad, entre 0,4 y 0,75 la fiabilidad es buena y si es superior a 0,75 la fiabilidad es excelente.

No se debe utilizar la prueba de la Ji al cuadrado para el estudio de la concordancia, pues éste es un estadístico que mide el nivel de asociación solamente. Si por ejemplo, en todas las muestras en las que se da un diagnóstico positivo la primera vez se da un diagnóstico negativo la segunda y al contrario, en todas las muestras en las que se da un diagnóstico negativo la primera vez se da un diagnóstico positivo la segunda, la prueba de la Ji al cuadrado indicará una asociación estadísticamente significativa, y sin embargo la concordancia sería nula.

Cuando la variable cualitativa tiene más de dos categorías ordenadas, se calcula el índice kappa ponderado, porque no es lo mismo el desacuerdo de la categoría I a la II que de la categoría I a la III, por ejemplo. Este índice, básicamente se calcula igual que el índice kappa simple, pero se multiplica el valor de cada casilla por un peso que pondera la magnitud del desacuerdo con sus casillas vecinas.

En la tabla 4 se muestran los resultados de la interpretación, en dos veces sucesivas, de 237 imágenes colposcópicas, con el fin de diagnosticar la presencia y el grado de la neoplasia cervical intraepitelial (CIN).

Tabla 4. Concordancia en el diagnóstico del grado de la neoplasia cervical intraepitelial (CIN).
Tomado de: *Hopman EH, Voorhorst FJ, Kenemans P, Meyer CJ, Helmerhorst TJ. Observer agreement on interpreting colposcopic images of CIN. Gynecol Oncol 1995; 58: 206-269.*

		Primera sesión				
		NO CIN	CIN I	CIN II	CIN III	
Segunda sesión	No CIN	56	12	3	3	74
	CIN I	10	23	5	6	44
	CIN II	3	9	31	12	55
	CIN III	0	3	13	48	64
		69	47	52	69	237

En la tabla 5 se presentan las frecuencias que se hubiesen hallado si ambas medidas se hubiesen basado en el azar, si ambas medidas fuesen independientes. Si en la primera sesión en el 29,1% (69/237) de las imágenes no hay evidencia de CIN y en la segunda sesión no hay lesiones sugestivas en el 31,2% (74/237) de los casos, si ambas medidas fuesen independientes, sólo por azar en el 9,1% (29,1%*31,2%) del total (9,1%*237 = 21,5) no habría evidencia de CIN en ninguna de las dos ocasiones. Si en la primera ocasión la imagen es sugestiva de CIN I en el 19,8% de los casos (47/237) y esto ocurre en el 18,6% (44/237) de las imágenes en la segunda ocasión, si ambas medidas fuesen independientes resultaría CIN I en las dos ocasiones en el 3,68% (19,8%*18,6%) de los casos, lo que supondría en números absolutos serían 8,7 (237*3,68%). Del mismo modo se calculan las frecuencias esperadas por azar de todas las casillas. En esta tabla sólo es preciso calcular (4-1)*(4-1) = 9 valores esperados y el resto vienen dados por la diferencia con los totales.

Tabla 5. Valores esperados por azar.

		Primera sesión				
		NO CIN	CIN I	CIN II	CIN III	
Segunda sesión	No CIN	21,5	14,7	16,2	21,5	74
	CIN I	12,8	8,7	9,7	12,8	44
	CIN II	16,0	10,9	12,1	16,0	55
	CIN III	18,6	12,7	14,0	18,6	64
		69	47	52	69	237

El índice kappa ponderado se calcula a partir de la proporción de desacuerdos, pero asignando a cada frecuencia un factor de ponderación (w) en función de la magnitud del desacuerdo. La fórmula para el cálculo de estos pesos más ampliamente utilizada es la de elevar al cuadrado el nivel de desacuerdo. Así, a las diferencias en una categoría se les asigna el peso 1 (1²), a las diferencias en dos categorías el peso 4 (2²) y así sucesivamente.

La fórmula del Índice Kappa ponderado es:

$$Kp = 1 - \frac{q_o}{q_e}$$

donde q_o = proporción de desacuerdos real.

q_e = proporción de desacuerdos esperados por azar.

La explicación detrás de esta fórmula es que la proporción de acuerdos es el complementario de la proporción de desacuerdos, pero no debidos al azar (Fig. 7). Se calcula en base a los desacuerdos porque son éstos los que se tienen que ponderar.

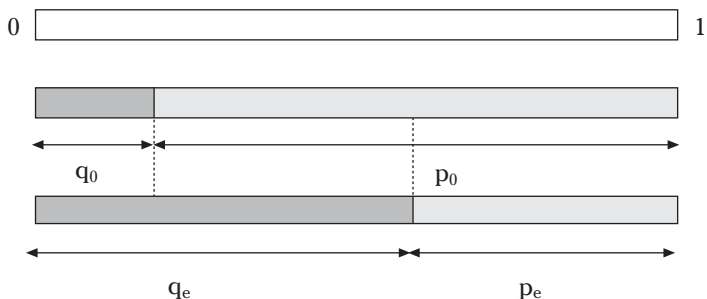


Figura 7. Índice Kappa ponderado.

$$q_o = 10*1 + 3*4 + 0*9 + 12*1 + 9*1 + 3*4 + \\ 3*4 + 5*1 + 13*1 + 3*9 + 6*4 + 12*1 = 148$$

$$q_e = 12,8*1 + 16*4 + 18,6*9 + 14,7*1 + 10,9*1 + 12,7*4 + \\ 16,2*4 + 9,7*1 + 14*1 + 21,5*9 + 12,8*4 + 16*1 = 669,8$$

$$Kp = 1 - \frac{148}{669,8} = 0,78$$

Con este esquema de ponderación (el cuadrado de las diferencias) el resultado del índice kappa es el mismo que el coeficiente de correlación intraclases.

Una de las principales limitaciones del índice kappa es que varía según la prevalencia de la enfermedad (o el signo) en estudio, aunque la variabilidad interobservador sea escasa. En general, si la probabilidad es muy baja o muy alta, el índice kappa es menor (porque la concordancia debida al azar ya es muy alta), y por ello no se pueden comparar índices basados en prevalencias diferentes.

Coefficiente de variación

Quando se trata de una variable cuantitativa, una de las medidas de fiabilidad es el coeficiente de variación medio de todos los individuos estudiados. Si las diferentes medidas en un mismo individuo difieren mucho, una desviación estándar elevada dará cuenta de dicha variabilidad. Pero no se pueden comparar desviaciones estándar directamente, ya que dependen de la variable en estudio. Si los valores de la variable son elevados (media elevada) la desviación estándar también será elevada. Por ejemplo, se mide el peso de un grupo de personas en kilogramos y la media es de 55 kg y la desviación estándar de 5 kg; en otro grupo se mide el peso en gramos y la media es de 55.000 gr y la desviación estándar de 5000 gr ¿en cuál de las dos muestras varía más el peso? En las dos la variabilidad es igual, pero la desviación estándar es diferente porque son distintas las unidades de medida. El coeficiente de variación es precisamente la desviación estándar dividido por la media y multiplicado por 100 (para expresarlo en porcentajes).

Se ha medido el estradiol en sangre de 5 mujeres postmenopáusicas mediante dos métodos diferentes A y B. Los resultados son los de la tabla 6.

Tabla 6. Medición del estradiol en sangre mediante dos métodos.

	A	B
1	20,5	18,7
2	18,4	19,8
3	8,0	9,4
4	16,2	14,8
5	10,4	12,0

La diferencia absoluta entre las dos determinaciones, la media del resultado de los dos métodos, la desviación estándar (DE) y el coeficiente de variación (CV) se presentan en la tabla 7.

Tabla 7. Coeficiente de variación.

	Diferencia	Media	DE	CV%
1	1,8	19,6	1,27	6,48
2	-1,4	19,1	0,99	5,18
3	-1,4	8,7	0,99	11,38
4	1,4	15,5	0,99	6,39
5	-1,6	11,2	1,13	10,09
				7,90

CV= coeficiente de variación; DE= desviación estándar.

El coeficiente de variación medio de las 5 mujeres es del 7,90%, lo que indica una elevada fiabilidad. En general, coeficientes de variación superiores al 30% indican baja fiabilidad y por debajo del 20% la fiabilidad se puede considerar buena.

Coeficiente de correlación intraclases

El coeficiente de correlación intraclases es una medida de fiabilidad cuando la variable que se mide dos o más veces es cuantitativa.

A un grupo de individuos se les realiza una extracción de sangre y ésta se separa en dos alícuotas para determinar la hemoglobina mediante dos métodos diferentes. La diferencia en los niveles de hemoglobina medios de los individuos se debe en parte a que los individuos son diferentes (varianza entre-individuos) y en parte a la diferencia de la medición con uno y otro método (varianza intra-individuos).

El coeficiente de correlación intraclases valora cuánto representa la varianza entre-individuos de la varianza total entre las medias de las estimaciones individuales (suma de la varianza intra-individuos y entre-indi-

viduos). Si los individuos dan siempre resultados muy similares, la varianza intra-individuos será muy pequeña, el método es fiable, el resultado del cociente será cercano a uno. En cambio, si para un mismo individuo los resultados en una y otra determinación son muy diferentes, el coeficiente será muy cercano a 0.

El desarrollo matemático es quizás algo complicado y se escapa a los objetivos de este libro, pero hoy en día la mayoría de los programas estadísticos calculan el coeficiente de correlación intraclases.

Este coeficiente puede tomar valores entre 0 y 1 y como norma general y al igual que el índice kappa, si es menor de 0,4 indica una escasa fiabilidad, entre 0,4 y 0,75 la fiabilidad es buena y si es superior a 0,75 la fiabilidad es excelente.

Un error muy frecuente es utilizar el coeficiente de correlación de Pearson como medida de fiabilidad y es incorrecto porque este coeficiente sólo mide la intensidad de la asociación, no tiene en cuenta las diferencias entre una y otra medida. Por ejemplo, si mediante un método siempre resulta la hemoglobina 1 g/dl superior que con el otro método, la correlación es perfecta, el coeficiente de correlación de Pearson es igual a 1, y este dato, sin embargo, no está informando de la diferencia entre los dos métodos.

RESUMEN

La calidad de la información clínica depende de varios factores:

- Validez o exactitud de la información. Depende de la ausencia de sesgos o desviaciones sistemáticas del valor real.
- Fiabilidad, o grado de estabilidad de una medición cuando se realiza en condiciones similares.

La variabilidad en los resultados se debe, entre otras causas, a la variabilidad del observador, a la variabilidad del método y a la variabilidad intraindividual. Por ello, para aumentar la fiabilidad, se deben aplicar pruebas estándar en condiciones estándar.

Para medir la fiabilidad se debe realizar la prueba dos o más veces en un mismo individuo y cuantificar la variación. El estadístico utilizado varía según la naturaleza de la variable en estudio: si es cualitativa se utilizan sobre todo el índice de concordancia global y el índice kappa; si se trata de una variable cuantitativa se utilizan el coeficiente de variación y el coeficiente de correlación intraclases.

Como norma general, si tanto el índice kappa como el coeficiente de correlación intraclases son menores de 0,4 indican una escasa fiabilidad, entre 0,4 y 0,75 la fiabilidad es buena y si es superior a 0,75 la fiabilidad es excelente.

PROBLEMAS

1. Dos expertos radiólogos observan 110 radiografías con el fin de valorar si existen o no signos de neumoonosis. Los resultados se presentan en la tabla 8. ¿Cuál es la fiabilidad de los radiólogos?

Tabla 8. Concordancia entre radiólogos.

		Radiólogo 1		
		Sí	No	
Radiólogo 2	Sí	14	7	21
	No	8	81	89
		22	88	110

2. Un observador valora en dos veces diferentes el estadiaje del cáncer de mama en 100 pacientes. El resultado se presenta en la tabla 9. ¿Cuál es la fiabilidad del observador? ¿Cuál es el coeficiente de correlación intraclases?

Tabla 9. Concordancia en el estadiaje del cáncer de mama.

		1. ^a vez				
		I	II	III	IV	
2. ^a vez	I	25	7	2	1	35
	II	4	14	5	0	23
	III	3	6	17	3	29
	IV	0	1	2	10	13
		32	28	26	14	100

3. A un grupo de individuos se les realiza una extracción de sangre y ésta se separa en dos alícuotas para determinar la ferritina sérica mediante dos métodos diferentes. Los resultados son los de la tabla 10. ¿Cuál es el coeficiente de variación medio?

Tabla 10. Determinación de ferritina sérica mediante dos métodos.

	A	B
1	22	20
2	12	13
3	4	4
4	41	37
5	18	19

4. Se ha determinado el nivel de calidad de vida según dos cuestionarios y los resultados han sido los que se muestran en la tabla 11. ¿Cómo cree que será el coeficiente de correlación de Pearson respecto al coeficiente de correlación intraclases?

Tabla 11. Calidad de vida según dos cuestionarios.

Cuestionario 1	Cuestionario 2
12	22
15	25
20	30
25	35
28	38
30	40
36	46
40	50
45	55
48	58

VALIDEZ DE LAS PRUEBAS DIAGNÓSTICAS

MEDIDAS DE LA VALIDEZ DE LAS PRUEBAS DIAGNÓSTICAS

Las pruebas diagnósticas, término que incluye la anamnesis, exploración física, técnicas de diagnóstico por la imagen (TAC, resonancia, radiografías), otros procedimientos (ECG), además de las pruebas de laboratorio, son el instrumento que utilizamos en medicina para diferenciar, de la forma más exacta y fiable posible, a los individuos enfermos de los no enfermos. Y digo de la forma más exacta y fiable porque no siempre un resultado positivo tras la aplicación de una prueba diagnóstica quiere decir presencia de enfermedad y un resultado negativo ausencia.

Básicamente hay dos tipos de pruebas diagnósticas:

- *Cualitativas*: clasifican a los pacientes como enfermos o no enfermos de acuerdo a la presencia o ausencia de un signo o síntoma. Por ejemplo, una radiografía puede confirmar o descartar la presencia de una fractura. Parece fácil el diagnóstico, sin embargo no todos los expertos radiólogos se ponen de acuerdo sobre la benignidad o malignidad de una imagen en la mamografía.
- *Cuantitativas*: clasifica a los pacientes como enfermos o no según estén por encima o debajo de unos límites (puntos de corte) de positividad. Uno de los criterios más utilizados para establecer este punto de corte es el estadístico (también llamados normal o Gausiano). La glucemia, por ejemplo, presenta una distribución normal con una media y una desviación estándar en los individuos sanos y una distribución normal con otra media y desviación estándar en los individuos diabéticos. Generalmente se asumen como normales los individuos con valores de la variable que caen en el intervalo de la media más y menos dos desviaciones estándar, lo que en una distribución normal supone el 95% de los individuos. Para el caso de la glucemia, esto significa que en el 2,5% de las personas normales el resultado se considerará positivo (no normal): son los llamados falsos positivos (sería el caso de los individuos no diabéticos con un nivel de glucemia elevado). Del mismo modo, dentro de los individuos enfermos, los ha-

brá con valores “normales”, fuera de los límites característicos de los enfermos, son los denominados falsos negativos (Fig. 8).

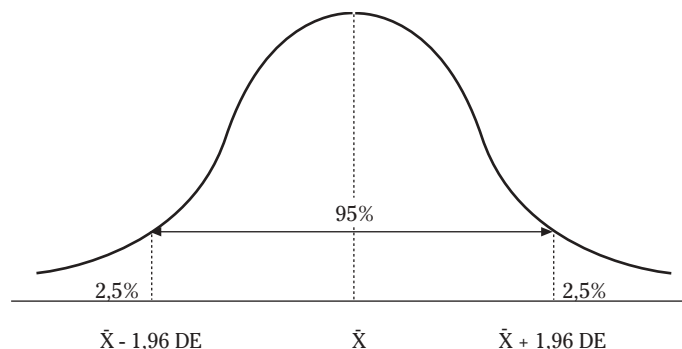


Figura 8. Distribución normal.

La eficacia de las pruebas viene dada por su validez o exactitud, y los índices claves de dicha validez son la sensibilidad y la especificidad. Los individuos pueden o no presentar la enfermedad en estudio (prefiero utilizar la terminología enfermo y no enfermo, en lugar de enfermo y sano, porque el hecho de que el paciente no presente la enfermedad en estudio no quiere decir que no tenga ninguna enfermedad) y la prueba puede resultar positiva o negativa. De la combinación de estas dos variables resultan cuatro posibilidades:

- Individuos enfermos en quienes la prueba resulta positiva. Son los verdaderos positivos (VP).
- Individuos enfermos en quienes la prueba resulta negativa. Son los falsos negativos (FN).
- Individuos no enfermos (al menos sin la enfermedad en estudio) en quienes la prueba resulta negativa. Son los verdaderos negativos (VN).
- Individuos no enfermos en quienes la prueba resulta positiva. Son los falsos positivos (FP).

La sensibilidad (S) o tasa de verdaderos positivos (TVP) es la capacidad de detectar a los enfermos. De todos los enfermos ¿en cuántos resulta la prueba positiva?

$$\text{Sensibilidad} = \frac{\text{VP}}{\text{VP} + \text{FN}}$$

El complementario de la sensibilidad es la tasa de falsos negativos (TFN):

$$TFN = \frac{FN}{VP + FN}$$

La especificidad (E) o tasa de verdaderos negativos (TVN) es la capacidad de detectar a los individuos sin la enfermedad. De todos éstos, ¿en cuántos resulta la prueba negativa?

$$\text{Especificidad} = \frac{VN}{VN + FP}$$

El complementario de la especificidad es la tasa de falsos positivos (TFP):

$$TFP = \frac{FP}{VN + FP}$$

Para el cálculo de estas probabilidades se puede construir una tabla 2x2 de modo que en las columnas se representa el verdadero estado del paciente (enfermo o no enfermo) y en las filas el resultado de la prueba diagnóstica (Tabla 12).

Tabla 12. Tabla 2x2.

	Enfermedad		
	+	-	
Prueba +	VP	FP	VP + FP
Prueba -	FN	VN	FN + VN
	VP + FN	VN + FP	

Por ejemplo, se han estudiado 150 individuos diabéticos y 200 no diabéticos, y se les ha medido el nivel de glucemia una hora después de la comida. La prueba se considera positiva si la glucemia es superior a 120 mg/100 ml, y así ha resultado en 133 individuos diabéticos y en 33 no diabéticos. ¿Cuál es la sensibilidad y especificidad de la prueba?

Si construimos la tabla 2x2 (Tabla 13), 150 son diabéticos y 200 no lo son. De los 150 diabéticos, en 133 la prueba ha resultado positiva y por lo tanto en 17 negativa. De los 200 no diabéticos, en 33 la prueba ha resultado positiva y en el resto, en 167, negativa. En total la prueba ha resultado positiva en 166 individuos y negativa en 184.

La sensibilidad (o tasa de verdaderos positivos) es la probabilidad de que en los individuos diabéticos la prueba resulte positiva, la glucemia sea superior a 120 mg/100 ml.

$$\text{Sensibilidad} = \frac{133}{133 + 17} = 88,7\%$$

Tabla 13. Validez de la glucemia para el diagnóstico de diabetes.

		Diabetes		
		+	-	
Glucemia (mg/100ml)	> 120	133	33	166
	≤ 120	17	167	184
		150	200	

La tasa de falsos negativos es por tanto el complementario de la sensibilidad; de los individuos diabéticos, ¿en cuántos es la glucemia igual o inferior a 120 mg/100 ml?

$$\text{TFN} = \frac{17}{133 + 17} = 11,3\%$$

La especificidad (o tasa de verdaderos negativos) es la probabilidad de que en los individuos no diabéticos la prueba resulte negativa (glucemia inferior o igual a 120 mg/100 ml):

$$\text{Especificidad} = \frac{167}{167 + 33} = 83,5\%$$

La tasa de falsos positivos es el complementario de la especificidad; ¿en cuántos individuos no diabéticos la glucemia es superior a 120 mg/100 ml?

$$\text{TFP} = \frac{33}{167 + 33} = 16,5\%$$

Una buena prueba diagnóstica es aquella que tiene una sensibilidad y especificidad elevadas. La validez también puede expresarse mediante una sola cifra, la razón de verosimilitud (o likelihood ratio: LR). La razón de verosimilitud de un resultado positivo es la probabilidad de un resultado positivo en los enfermos dividido por la probabilidad de un resultado positivo en los no enfermos:

$$\text{LR}_+ = \frac{\text{Probabilidad de un resultado positivo en los enfermos}}{\text{Probabilidad de un resultado positivo en los no enfermos}} =$$

$$\text{LR}_+ = \frac{\text{TVP}}{\text{TFP}} = \frac{\text{Sensibilidad}}{1 - \text{Especificidad}}$$

Cuanto mayor sea esta cifra (LR+) mejor discrimina la prueba. Nos indica cuánto más probable es un resultado positivo en un individuo enfermo que en uno no enfermo.

Del mismo modo, la razón de verosimilitud de un resultado negativo es el cociente entre la probabilidad de un resultado negativo en los enfermos y la probabilidad de un resultado negativo en los no enfermos:

$$LR^- = \frac{\text{Probabilidad de un resultado negativo en los enfermos}}{\text{Probabilidad de un resultado negativo en los no enfermos}} =$$

$$LR^- = \frac{TFN}{TVN} = \frac{1 - \text{Sensibilidad}}{\text{Especificidad}}$$

Cuanto menor sea esta cifra (LR-) mejor discrimina la prueba. Nos indica cuánto menos probable es un resultado negativo en un individuo enfermo que en uno no enfermo, o lo que es lo mismo, el inverso de este cociente expresa cuánto más probable es un resultado negativo en un individuo no enfermo comparado con uno enfermo.

Siguiendo con el ejemplo de la diabetes, la razón de verosimilitud de un resultado positivo es:

$$LR^+ = \frac{133/150}{33/200} = \frac{0,887}{0,165} = 5,38$$

Este resultado quiere decir que un resultado positivo es 5,38 veces más probable en un individuo diabético que en uno no diabético.

Y la razón de verosimilitud de un resultado negativo es:

$$LR^- = \frac{17/150}{167/200} = \frac{0,113}{0,835} = 0,135 \quad (1 \text{ a } 7,4)$$

Un resultado negativo es 7,4 (1/0,135) veces más probable en un individuo no diabético que en uno diabético.

En general, si el cociente de verosimilitud es superior a 10 o inferior a 0,1, la prueba discrimina muy bien, genera cambios importantes y a menudo concluyentes desde una probabilidad previa hasta una probabilidad posterior. Entre 5-10 ó 0,2-0,1 generan cambios moderados. Cocientes entre 2-5 y 0,5-0,2 generan cambios pequeños (pero en ocasiones importantes) de la probabilidad. Y por último, cocientes de 1-2 y 0,5-1 apenas discriminan, alteran la probabilidad en un grado insignificante (y rara vez importante).

El verdadero estado del paciente (enfermo o no) lo determina un procedimiento diagnóstico definitivo, es la prueba de referencia con la que se compara la prueba diagnóstica en estudio, y en la terminología anglosajona se denomina "gold standard". A esta prueba definitiva se le atribuyen una sensibilidad y especificidad del 100%, pero si es tan perfecta, ¿pa-

ra qué cuestionarnos la validez de otras pruebas y no se aplica ésta directamente? La razón es que puede ser cruenta, conllevar sus riesgos (biopsia hepática, punción de médula ósea) o muy cara. Además, en realidad muchas veces no se puede evaluar la eficacia de dichas pruebas de referencia ni asegurar que sean efectivamente definitivas, sencillamente no existe una prueba mejor o ésta no es aplicable.

Errores en la estimación de la sensibilidad y la especificidad

Se ha dicho que la sensibilidad y especificidad de una prueba son características inherentes a la misma y por tanto constantes e independientes de la población a la que se aplica y de la prevalencia de la enfermedad en dicha población. Sin embargo, la fuente de error más importante al medir la validez de una prueba se debe, precisamente, a las diferencias entre la población en la que se realiza el estudio de la validez y aquella a la que se pretenden extrapolar, generalizar los resultados. A continuación se describen los principales sesgos de selección:

- Al comienzo del estudio de la validez de una prueba se espera que resulte positiva en los enfermos y negativa en los no enfermos. Por ello, se suele seleccionar un grupo de individuos realmente enfermos, con la enfermedad avanzada donde casi seguro la prueba resultará positiva, lo que provoca que la sensibilidad sea elevada. En otra muestra de individuos enfermos, pero donde la enfermedad no está tan avanzada, la prueba probablemente no resultaría tan sensible. Un ejemplo real es el caso del antígeno carcinoembrionario (ACE) en el cáncer colorrectal. Cuando se determinó en 36 individuos con el cáncer avanzado, resultó elevado en 35 de ellos. Sin embargo, en estudios posteriores de pacientes con estadios menos avanzados de un cáncer colorrectal (y por lo tanto menos graves) la sensibilidad del ACE cayó en picado, y de hecho se abandonó esta prueba para el diagnóstico del cáncer.
- Para el estudio de la especificidad se suelen seleccionar voluntarios sanos, que generalmente son mucho más jóvenes que la población donde se aplicará la prueba y además están libres de otras enfermedades que podrían dar lugar a falsos positivos, con lo que la especificidad también será más elevada que si la prueba se aplica a la población más representativa de los individuos sin la enfermedad que acuden normalmente a la consulta. Es muy importante analizar la especificidad en individuos que sin padecer la enfermedad en estudio posean los cuadros más usuales en el diagnóstico diferencial de dicha patología, aquellos a los que se les aplicaría la prueba diagnóstica. Si por ejemplo se evalúa una prueba para el diagnóstico del infarto agudo de miocardio

(IAM) en los sujetos que acuden a urgencias con dolor torácico típico y otros síntomas indicativos de IAM, el grupo control lo deben formar individuos con las mismas características clínicas pero que no han sufrido un infarto.

- Otro sesgo importante ocurre cuando la selección de los individuos para la evaluación de la prueba está basada en los resultados de la misma. Esta situación se llama “sesgo de verificación” o “sesgo de orientación diagnóstica”. Por ejemplo: la prueba de referencia para el diagnóstico de embolia pulmonar es la angiografía, pero esta prueba se realiza fundamentalmente en aquellos pacientes en los que la prueba en estudio resulta positiva, por ejemplo la gammagrafía de ventilación-perfusión). Como consecuencia será una población donde habrá muy pocos resultados negativos (ni verdaderos ni falsos) por lo que tanto la tasa de falsos positivos como la de verdaderos positivos serán mayores que en la población clínicamente relevante. Es lo que Diamond denomina “la mayoría silenciosa del Reverendo Bayes”. Debido a los importantes efectos adversos o costes de la prueba de referencia, en muchos estudios la valoración de una prueba diagnóstica se deriva en exclusiva del estudio de aquellos pacientes en los que se pretende la verificación del diagnóstico (la prueba a valorar ya ha resultado positiva), y el resto de pacientes no se tiene en cuenta. Además, la población en estudio se enriquece de individuos enfermos que pueden presentar otras patologías que también pueden provocar un aumento de los falsos positivos. Otro ejemplo es el del diagnóstico de cáncer de mama mediante el estudio anatomopatológico de la pieza quirúrgica, que sólo se realiza en las pacientes en las que las pruebas previas (biopsia, mamografía) han resultado positivas. Una manera de solucionar este problema es utilizando un segundo “gold standard” o patrón de referencia para los pacientes con un resultado negativo de la prueba de estudio. En el caso de la embolia pulmonar se consideraría que no la padece si evolucionan bien sin tratamiento. En el caso del cáncer de mama puede ser el seguimiento al año lo que determinaría el verdadero estado de la paciente.
- Si se conoce el resultado de una de las pruebas (bien sea la prueba en estudio o la prueba de referencia), a la hora de valorar la otra, la concordancia es mucho mayor, aumenta la tasa de verdaderos positivos y verdaderos negativos. Se presta mucha más atención al ECG de un individuo ya diagnosticado de coronariopatía que si se trata de un individuo sano. Una vez que hemos identificado un nódulo pulmonar en el TAC, advertimos la lesión, previamente no detectada, en la radiografía de tórax. Del mismo modo, si en el análisis de sangre ha resultado una anemia ferro-

pénica, se pregunta más acerca de síntomas como el cansancio, la caída del cabello o la intolerancia al frío.

Ajuste de los datos

Hay que estudiar con precaución los resultados publicados de la validez de las pruebas y prestar especial atención a los siguientes aspectos:

1. ¿Existió una comparación ciega, independiente, con un estándar de referencia? En caso contrario, si previa a la interpretación de una prueba ya se conoce el resultado de la otra, la sensibilidad y especificidad serán más elevadas que las reales.
2. ¿Incluyó la muestra de pacientes un espectro adecuado de pacientes a los cuales en la práctica clínica se aplicará el examen diagnóstico? La mayoría de los estudios contienen pacientes en todo el rango de gravedad de la enfermedad y sólo dan un resultado para la tasa de verdaderos positivos y un resultado para la tasa de falsos positivos, cuando lo ideal sería dar un resultado de las tasas para cada uno de los niveles de gravedad, en grupos muy claramente definidos desde el punto de vista clínico. Hay que tener especial cuidado del nivel de gravedad de los enfermos y de los estudios con voluntarios sanos. Si los datos publicados proceden de un hospital de tercer nivel, donde los pacientes son más graves, la sensibilidad es probablemente mayor que si el estudio se hubiese realizado en los pacientes que acuden a una consulta de asistencia primaria.
3. ¿Influyeron los resultados del examen objetivo de la evaluación en la decisión de realizar el estándar de referencia? Si el resultado positivo de la prueba en estudio ha sido el criterio decisor a la hora de aplicar la prueba de referencia, la tasa de falsos positivos real será probablemente más alta que la detectada en este estudio.
4. ¿Se describieron los métodos para realizar el examen con suficiente detalle como para permitir su reproducción? Esta descripción debe abarcar todos los aspectos importantes en la preparación del paciente, la realización de la prueba y el análisis e interpretación de los resultados. No es lo mismo la validez de una prueba diagnóstica en manos del experimentado clínico que publica el artículo, que si la realiza otro médico por primera vez.
5. ¿Es la prueba de referencia una medida definitiva del verdadero estado del paciente? La verdad es que difícilmente la prueba de referencia es definitiva, pero se debe aplicar la mejor prueba disponible.

6. ¿Han sido los datos interpretados por dos o más observadores independientes? Son más fiables los resultados en los que hay acuerdo entre varios investigadores que en aquellos juzgados por una única persona, especialmente cuando el diagnóstico deriva de apreciaciones subjetivas.

Aún sabiendo que los datos se deben ajustar en función de la diferencia entre la muestra en estudio y aquella donde se aplicarán los datos, el verdadero problema consiste en *cuánto* se deben ajustar. Para el caso concreto en el que las características de los pacientes en estudio puede influir en la validez de la prueba, existe una solución que se expone a continuación, y es la de aplicar la regresión logística para modelar la sensibilidad y especificidad en función de parámetros relevantes.

Ajuste mediante la regresión logística

Como hemos visto, la sensibilidad y especificidad no son características inherentes a las pruebas e independientes de las características de la enfermedad, pueden variar según el sexo, la edad, la gravedad, etc., variables que modifican la probabilidad de que la enfermedad esté presente.

Mediante la regresión logística se pueden determinar las probabilidades de determinados sucesos condicionados a un patrón de covariables (combinaciones de factores). En el modelo a tal fin, la variable dependiente debe ser el resultado de la prueba (positivo o negativo), la variable independiente el verdadero estado de los individuos estudiados (enfermo o no) determinado por la prueba de referencia y las covariables aquellos factores que pueden modificar las características de la enfermedad y consecuentemente la sensibilidad y especificidad de la prueba.

El modelo de regresión logística es como sigue:

$$p(T) = \frac{e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}}{1 + e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}}$$

donde:

$p(T)$ es la probabilidad de un resultado positivo.

x_1 = variable dicotómica que indica la presencia de enfermedad (enfermo/no enfermo)

$x_2 \dots x_n$ = covariables que pueden influir en la sensibilidad y especificidad (gravedad de la patología, presencia de síntomas, estadió, edad, sexo...etc).

α = ordenada en el origen.

β_1 = coeficiente de regresión de la variable "presencia de enfermedad".

β_2, \dots, β_n = coeficientes de regresión de las covariables por las que ajustar.

Sustituyendo en la fórmula los valores de las variables, se puede calcular la probabilidad de dar un resultado positivo en los individuos enfermos (sensibilidad) y en los no enfermos (tasa de falsos positivos) según diferentes características de la población.

Por ejemplo, se realizó un modelo para la estimación de la sensibilidad y especificidad del cuestionario de Rose para el diagnóstico de la enfermedad arterial periférica (Tabla 14).

Tabla 14. Estimación de la sensibilidad y especificidad mediante la regresión logística.
Tomada de: Coughlin SS, Trock B, Criqui MH, Pickle LW, Browner D and Tefft MC. *The logistic modeling of sensitivity, specificity, and predictive value of a diagnostic test. J Clin Epidemiol 1992; 45: 1-7.*

Variable	Coeficiente	Error estándar
α	-3,384	0,191
Enfermo	1,805	0,377
Sexo	0,393	0,365

La variable que podía modificar la validez del cuestionario, por lo que se recogió con el fin de ajustar los datos, es el sexo. Si el cuestionario resulta diagnóstico toma el valor 1, y cero en caso contrario. El sexo es 1 en los hombres y cero en las mujeres. Si se trata de un grupo de hombres enfermos, la probabilidad de que la prueba resulte positiva (sensibilidad) es:

$$p(T+) = \frac{e^{-3,384 + 1,805*1 + 0,393*1}}{1 + e^{-3,384 + 1,805*1 + 0,393*1}} = 0,23$$

La especificidad de la prueba en un grupo de mujeres es 1 menos la probabilidad de que la prueba resulte positiva dado que no tienen la enfermedad:

$$p(T+) = \frac{e^{-3,384 + 1,805*0 + 0,393*0}}{1 + e^{-3,384 + 1,805*0 + 0,393*0}} = 0,0327$$

$$\text{Especificidad} = 1 - 0,0327 = 0,967$$

De este modo se pueden calcular la sensibilidad y especificidad según diferentes patrones de covariables.

RESULTADOS DE VARIABLES CONTINUAS. CURVAS ROC

Para el establecimiento de la sensibilidad y especificidad de una prueba diagnóstica las dos variables de interés deben ser dicotómicas (en-

fermo/no enfermo y resultado positivo/negativo). La primera dificultad radica, precisamente en el establecimiento de los límites que demarcan las dos categorías de ambas variables.

El concepto de “normal” para separar individuos enfermos de sanos puede ser tema de una importante discusión. Puede referirse a lo típico, característico de un grupo de la población, pero además, el término normal tiene también un sentido estadístico (Fig. 8): se aceptan como “normales” los valores de la variable incluidos en el intervalo de la media más y menos 1,96 desviaciones estándar, que supone el 95% de los valores centrales. De hecho así es como se calculan los límites de normalidad de la mayoría de los parámetros del laboratorio (hemoglobina, colesterol, glucemia, etc.)

Pero este concepto estadístico de normalidad es a menudo insatisfactorio para propósitos de clasificación, ya que la mayoría de los individuos evaluados son presumiblemente sanos y sin embargo serán etiquetados como patológicos (o se les someterá a más pruebas) un 5% de los mismos (o el 2,5% más alto o más bajo).

Para la resolución de toda esta polémica con las variables cuantitativas, se ha desarrollado una estrategia matemática, que es el análisis de las curvas ROC (cada vez más conocidas como “curvas de rendimiento diagnóstico”). En dichas curvas se demuestra como, conforme varía el punto de corte de la prueba diagnóstica (que define la prueba como positiva o negativa), varían la sensibilidad y especificidad en un sentido inverso.

Por ejemplo, la glucemia como parámetro diagnóstico de diabetes. Si el punto de corte fuese 100 mg/dl de glucosa en sangre, la sensibilidad sería muy alta (prácticamente todos los diabéticos tendrán una glucemia superior a este límite) y una especificidad muy baja (muchas personas no diabéticas pueden tener glucemias superiores a 100 mg/dl). Si se eleva el punto de corte a 140 mg/dl la sensibilidad disminuye mucho y la especificidad aumenta. No todos los diabéticos tendrán glucemias superiores a este valor (habrá falsos negativos), y sin embargo, pocos individuos no diabéticos tendrán la glucemia tan elevada (pocos falsos positivos).

Las curvas ROC se dibujan en un eje de coordenadas (Fig. 9), donde en el eje de las “y” se pone la sensibilidad y en el eje de las “x” la tasa de falsos positivos (o el complementario de la especificidad). Ambos ejes oscilan entre 0 y 100.

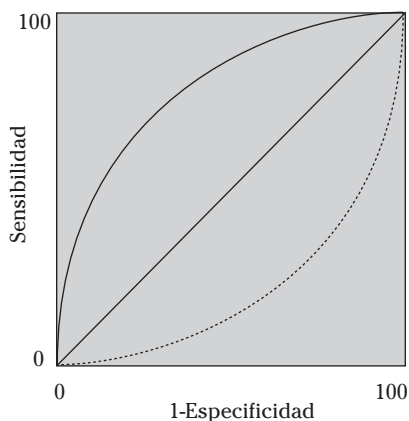


Figura 9. Curva ROC.

Una prueba diagnóstica no válida es aquella cuyo trazado coincide con la diagonal: la razón de verosimilitud (tanto de un resultado positivo como negativo) en dichos puntos es 1. Es tan probable un resultado positivo en los individuos enfermos como en los no enfermos y tan probable un resultado negativo en los no enfermos como en los enfermos (existe el mismo número de verdaderos positivos o negativos que de falsos positivos o negativos respectivamente).

En el área por encima de la diagonal la razón de verosimilitud para un resultado positivo es mayor que 1, es mayor el número de verdaderos que de falsos positivos, y de igual modo, la razón de verosimilitud para un resultado negativo es menor que 1, es mayor el número de verdaderos que de falsos negativos. Es más probable un resultado positivo en los enfermos que en los no enfermos, y más probable un resultado negativo en los no enfermos que en los enfermos. Cuanto mayor es el área bajo la curva, cuanto más se aproxima a la L invertida (ángulo superior izquierdo), mejor discrimina la prueba, mayor es su validez.

En el área por debajo de la diagonal la razón de máxima verosimilitud para un resultado positivo es menor que 1, es más probable un resultado positivo en los no afectados de la enfermedad que en los enfermos. Del mismo modo, la razón de verosimilitud para un resultado negativo es mayor que 1, lo que significa que es más probable un resultado negativo en los enfermos que en los no enfermos. Una prueba cuyos valores de validez se encuentran en esta área no discrimina bien, es preferible hacerlo por azar o mejor aún se corrige cambiando el criterio de positividad y negatividad (lo positivo pasa a ser negativo y viceversa).

Elección de un punto de corte

El punto de corte elegido para decidir si el resultado de una prueba es positivo o negativo depende de las consecuencias que tenga la exis-

tencia de falsos positivos y falsos negativos (Fig. 10). A las consecuencias negativas sobre la salud se las denomina de un modo genérico costes, y así, los costes de un falso positivo se derivan del hecho de que recibirán un tratamiento sin presentar la enfermedad (con los efectos secundarios que ello pueda conllevar), y los costes de un falso negativo se deben a que estos individuos enfermos no se beneficiarán del tratamiento adecuado (en este caso costes equivalen a beneficios no obtenidos). Dependiendo de qué costes sean más elevados se precisa una mayor sensibilidad o especificidad. Cuando el coste de un falso positivo es mayor que el coste de un falso negativo, porque el tratamiento es muy agresivo y no muy beneficioso para los enfermos por ejemplo, interesa una mayor especificidad que sensibilidad y se elige como punto de corte uno que esté en la zona de mayor pendiente, como el punto A de la figura. Cuando, por el contrario, es preferible detectar el mayor número de individuos enfermos, porque éstos se beneficiarán de un tratamiento con escasos efectos adversos, se busca como punto de corte uno con mayor sensibilidad que especificidad, donde la pendiente de la curva es menor, como en el punto C de la figura.

Por otro lado, el número absoluto de falsos positivos y falsos negativos no sólo depende de la sensibilidad y especificidad, sino también de la probabilidad de que la enfermedad esté presente. Si esta probabilidad es baja, dominan los individuos no enfermos y por tanto el número absoluto de falsos positivos puede ser muy elevado. A la hora de disminuir el número de falsos positivos se precisa una elevada especificidad (punto A de la figura). Si por el contrario la probabilidad de la enfermedad es elevada, lo que dominan son los enfermos, y lo que puede aumentar mucho en valores absolutos es el número de falsos negativos, por lo que para disminuirlos interesará una elevada sensibilidad (punto C de la figura).

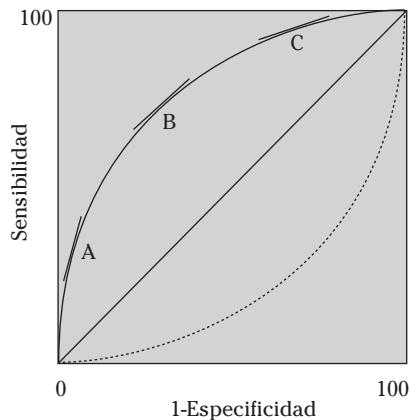


Figura 10. Elección del punto de corte.

Del manejo conjunto de los costes de los falsos positivos y falsos negativos y de la probabilidad de la enfermedad se puede derivar la siguiente fórmula, según la cual, el punto de la curva ROC que mejor discrimina entre enfermos y no enfermos es aquel en que la pendiente coincide con el resultado de la ecuación:

$$\frac{1 - P}{P} \times \frac{C_{FP}}{C_{FN}}$$

Donde:

- P es la probabilidad de la enfermedad
- C_{FP} es el coste neto de tratar a los pacientes no enfermos (coste de los falsos positivos).
- C_{FN} es el coste de no tratar a los pacientes enfermos (coste de los falsos negativos o beneficios asociados a los verdaderos positivos).

Dada una relación de los costes de los falsos positivos y falsos negativos, conforme mayor es la probabilidad de la enfermedad, menor es la pendiente, mayor es la sensibilidad que se precisa. Dada una probabilidad de padecer la enfermedad, cuanto mayores son los riesgos de un tratamiento inadecuado (costes de los falsos positivos), el cociente es mayor, el punto de corte estará en la zona de pendiente elevada, donde es mayor la especificidad; y conforme mayores son los beneficios de tratar a los enfermos (mayores costes asociados a los falsos negativos) más sensibilidad es necesaria (el cociente disminuye, la pendiente es menor). Cuando la probabilidad de presentar la enfermedad es del 50% y tan importante es un falso positivo como un falso negativo, el punto de corte que mejor discrimina es aquel correspondiente a la zona de la curva más próximo al vértice superior izquierdo (punto B de la figura). De hecho, en la literatura, se le ha concedido demasiado valor a este punto. Son muchas las veces que he leído que el punto de corte que mejor discrimina es el correspondiente al punto de la curva más próximo al vértice superior izquierdo. Este error considero que proviene del desconocimiento de los costes de los falsos positivos y falsos negativos (a veces es difícil decidir qué es más peligroso, dejar un enfermo sin tratar o tratar a una persona sin la enfermedad, y ante la duda se da una respuesta salomónica).

Supongamos, por ejemplo, el caso de una enfermedad en la que se precisa confirmar el diagnóstico porque el tratamiento es muy agresivo, de modo que la relación costes/beneficios se ha establecido en 2 a 1 (preocupa el doble aplicar un tratamiento tan agresivo a una persona sin la enfermedad que dejar de tratar a una persona enferma). La probabilidad de la enfermedad es 0,5. El punto de corte es aquel que coincide con una pendiente:

$$0,5/0,5 * 2/1 = 2$$

Una pendiente así ocurre en la primera parte de la curva, donde la especificidad es mayor que la sensibilidad. Si el tratamiento es tan agresivo es preciso disminuir al máximo los falsos positivos, no dar un tratamiento así a un individuo que en realidad no lo requiere.

Ahora imaginemos la situación en la que se aplica una prueba de cribaje para la detección de una infección asintomática que la padecen 1 de cada 1000 personas. El beneficio de detectarlos es muy grande porque se evitaría la transmisión por sangre al evitar que fuesen donantes. El coste de tratar a los falsos positivos es realmente bajo aunque la carga psicológica es importante (creen estar enfermos de una enfermedad fatal). En media, la razón costes/beneficios es 1:50. El punto de corte será:

$$1/50 * 0,999/0,001 = 19,98$$

La pendiente es muy alta, lo que quiere decir que interesa una prueba con una especificidad muy elevada. Aparentemente y dados unos beneficios tan elevados al tratar, no queremos que ningún enfermo se quede sin diagnosticar, por lo que precisaríamos una elevada sensibilidad. Sin embargo, dada la baja prevalencia de la enfermedad, los falsos positivos (que también conllevan sus riesgos) en números absolutos podrían ser muchos, por lo que se precisa mayor especificidad.

Si dicha prueba de cribaje se aplica a una población donde la prevalencia de la enfermedad es bastante más elevada, pongamos por caso del 50%, el punto de corte debería ser:

$$0,5/0,5 * 1/50 = 0,02$$

La pendiente disminuye mucho, por lo que interesa una prueba muy sensible, que ningún individuo enfermo (falso negativo) se prive de los beneficios del tratamiento.

COMPARACIÓN DE LA VALIDEZ DE LAS PRUEBAS DIAGNÓSTICAS

La comparación de la validez de dos pruebas diagnósticas, si son variables dicotómicas, es sencilla, se comparan las proporciones.

Cuando lo que se pretende conocer es si una prueba tiene una sensibilidad o especificidad mayor que por puro azar (50%), el intervalo de confianza de la sensibilidad o especificidad estimada no debe incluir dicho valor esperado por azar. Y el límite del azar es, tanto para la sensibilidad como especificidad, del 50%, porque se espera que una prueba sensible dé más veces resultados positivos que negativos en los pacientes enfermos y, una prueba específica se espera que dé más resultados negativos que positivos en los individuos sin la enfermedad.

El intervalo de confianza más frecuentemente utilizado es del 95%.

$$IC\ 95\% = p \pm 1,96 \sqrt{p \cdot q/n}$$

p = es la sensibilidad o la especificidad

q = es el complementario de la sensibilidad o especificidad según el caso.

n = es el tamaño de la muestra, el número de enfermos para el caso del estudio de la sensibilidad y el número de individuos sin la enfermedad cuando lo que se estudia es la especificidad.

En el ejemplo de la diabetes, el intervalo de confianza del 95% de la sensibilidad es:

$$IC\ 95\% = 0,887 \pm 1,96\sqrt{0,887 \cdot 0,113/150} = 0,836 - 0,938$$

Como el intervalo no incluye el valor nulo, el 50%, se puede decir que la prueba de la glucemia tras una hora de comer superior a 120 mg/100 ml es sensible, detecta a los enfermos significativamente mejor que por puro azar.

El intervalo de confianza del 95% de la especificidad es:

$$IC\ 95\% = 0,835 \pm 1,96\sqrt{0,835 \cdot 0,165/200} = 0,784 - 0,886$$

La conclusión es que esta prueba también discrimina a los no diabéticos significativamente mejor (desde el punto de vista estadístico) que el azar.

Cuando se comparan dos sensibilidades o dos especificidades (en cualquier caso dos proporciones), existe una distribución de diferencias de proporciones cuya media es 0 bajo la hipótesis nula y el intervalo de confianza del 95% es:

$$IC\ 95\% = (p_1 - p_2) \pm 1,96\ EE \quad \text{donde EE} = \text{error estándar}$$

$$EE = \sqrt{\hat{p} \cdot \hat{q} \cdot (1/n_1 + 1/n_2)} \quad \text{donde}$$

$$\hat{p} = (n_1 p_1 + n_2 p_2) / (n_1 + n_2) \quad \hat{q} = 1 - \hat{p}$$

Si el intervalo de confianza del 95% de la diferencia de las proporciones incluye el valor 0 (media bajo la hipótesis nula) no existen diferencias significativas (con un error alfa menor del 5%).

Por ejemplo, si el punto de corte de la glucemia para decidir si una prueba es positiva o negativa, en lugar de 120 mg/100 ml es 140 mg/100 ml (se trata por tanto de otra prueba diagnóstica diferente), la sensibilidad y especificidad son del 76% y 95% respectivamente. El número de individuos

estudiado es el mismo, 150 diabéticos y 200 no diabéticos. ¿Difieren ambas pruebas de un modo estadísticamente significativo?

Para comparar las sensibilidades, el intervalo de confianza del 95% de la diferencia es:

$$IC\ 95\% = 0,887 - 0,760 \pm 1,96 * 0,044 = 0,127 (0,041 - 0,213)$$

Como el intervalo de confianza de las diferencias no incluye el valor nulo (diferencia = 0) se puede concluir que ambas pruebas son significativamente diferentes respecto a la sensibilidad (concretamente, si el punto de corte es 120 mg/100 ml la sensibilidad es mayor).

En cuanto a la especificidad, el intervalo de confianza del 95% de la diferencia es:

$$IC\ 95\% = 0,950 - 0,835 \pm 1,96 * 0,031 = 0,115 (0,054 - 0,176)$$

Como tampoco incluye el valor nulo, la especificidad de la glucemia superior a 140 mg/100ml es significativamente superior a la especificidad de la glucemia cuando el punto de corte para considerar la prueba positiva es 120 mg/100 ml.

Cuando se pretenden comparar la sensibilidad y la especificidad a la vez, una de las maneras de hacerlo es comparando el valor global (VG), que es el porcentaje de individuos correctamente diagnosticados y se sigue la misma dinámica que previamente para la comparación de proporciones, sólo que ahora n es el tamaño de la muestra global, del conjunto de enfermos y no enfermos.

$$VG = \frac{VP + VN}{VP + FN + VN + FP}$$

Este parámetro depende en gran medida de la probabilidad de la enfermedad en estudio, dando más peso a la especificidad si esta probabilidad es baja y a la sensibilidad si la probabilidad es alta, por lo que si se comparan dos pruebas mediante el valor global, los resultados de ambas deben estar basados en una misma probabilidad de presentar la enfermedad (cuanto más parecida a la real mejor), o bien calcular la media entre la sensibilidad y especificidad, que equivale al valor global cuando la probabilidad de la enfermedad es del 50%.

Otra medida unificada de la validez de una prueba es el Índice de Youden, que consiste en la suma de la sensibilidad y especificidad menos 100. Si la prueba es invariablemente correcta, toma el valor de 100. Pero este índice supone implícitamente que la sensibilidad y especificidad tienen igual importancia.

De hecho, para ser más correctos, qué prueba es la mejor depende de los beneficios de tratar a los verdaderos positivos (igual a los costes de

dejar de tratar a los falsos negativos) y de los costes de tratar a los falsos positivos. A su vez, el número absoluto de falsos negativos depende de la probabilidad de presentar la enfermedad y de la sensibilidad de la prueba, y el número de falsos positivos depende también de la probabilidad de la enfermedad y de la especificidad de la prueba:

Número de falsos negativos = Probabilidad enfermedad * (1- Sensibilidad)

Número de falsos positivos = (1-Probabilidad enfermedad) * (1-Especificidad)

Por tanto, será mejor prueba aquella que en valores absolutos menos costes reporte:

$$\text{Costes netos} = C_{FN} * (1-S) * P + C_{FP} * (1-E) * (1-P)$$

donde:

C_{FN} = coste de dejar de tratar a los falsos negativos

1-S = complementario de la sensibilidad o tasa de falsos negativos (TFN)

P = probabilidad de presentar la enfermedad

C_{FP} = coste de tratar a los falsos positivos

1-E = complementario de la especificidad o tasa de falsos positivos (TFP)

Supongamos que dejar de tratar a un paciente diabético tiene las mismas consecuencias negativas que tratar con antidiabéticos orales a un paciente no diabético, y que la probabilidad de presentar diabetes en los pacientes que se les determina la glucemia tras la comida es del 30%. ¿Qué punto de corte es preferible, 120 o 140 mg/100ml? (suponemos una población de referencia de 1000 individuos):

- Punto de corte 120 mg/100ml
 - Sensibilidad = 88,7%
 - Especificidad = 83,5%
 - TFN = 11,3%
 - TFP = 16,5%
 - N° de falsos negativos = $1000 * 0,30 * 0,113 = 33,9$
 - N° de falsos positivos = $1000 * 0,70 * 0,165 = 115,5$
 - Costes netos = $1 * 33,9 + 1 * 115,5 = 149,4$

- Punto de corte 140 mg/100ml
 - Sensibilidad = 76%
 - Especificidad = 95%
 - TFN = 24%
 - TFP = 5%
 - N° de falsos negativos = $1000 * 0,30 * 0,24 = 72$

$$- \text{N}^{\circ} \text{ de falsos positivos} = 1000 * 0,70 * 0,05 = 35$$

$$- \text{Costes netos} = 1 * 72 + 1 * 35 = 107$$

Los costes netos de la primera prueba (punto de corte 120 mg/100 ml) son superiores y por tanto es preferible la segunda prueba (punto de corte 140 mg/100 ml). Esto es así porque al ser la probabilidad de presentar la enfermedad sólo del 30% (o al menos inferior al 50%), en el valor global pesa más la especificidad, y ésta es superior precisamente en la segunda prueba. Estos valores de costes no tienen un sentido propio, sólo son útiles a efectos comparativos.

Si ahora suponemos que los costes de dejar de tratar a un diabético son tres veces superiores a los costes de tratar a los no diabéticos ¿Qué prueba es la preferida?

Con el punto de corte 120 mg/100ml, los costes netos son:

$$3 * 33,9 + 1 * 115,5 = 217,2$$

Con el punto de corte 140 mg/100ml, los costes netos son:

$$3 * 72 + 1 * 35 = 251$$

Ahora es preferible la primera prueba, ya que si los costes de dejar de tratar a los falsos negativos son superiores a los costes de tratar a los falsos positivos, interesa una prueba con una tasa de falsos negativos baja, con una sensibilidad elevada, requisito que cumple la prueba de considerar positivo una glucemia superior a 120 mg/100 ml.

Comparación de áreas bajo la curva ROC

Cuando se comparan las sensibilidades y especificidades como variables continuas, lo que se comparan son las áreas bajo las curvas ROC, ya que éstas son una medida global de la exactitud de una prueba. Aquella curva con mayor área es la mejor, la que tiene un mayor poder discriminatorio.

Por convenio, el área es siempre igual o superior a 0,5 y llega hasta un máximo de 1. Valores entre 0,5 y 0,7 indican baja exactitud, entre 0,7 y 0,9 pueden ser útiles y superior a 0,9 la exactitud es alta.

Para derivar la fórmula del cálculo de esta área, se divide toda la superficie bajo la curva ROC en pequeñas áreas, que a su vez se puede considerar que son la suma de las áreas de un triángulo y de un rectángulo (Fig. 11).

$$\begin{aligned} \text{Área} &= \sum (A_{i+1} - A_i) * B_i + (A_{i+1} - A_i) * (B_{i+1} - B_i) / 2 \\ &\quad \begin{array}{ccc} \leftarrow & & \leftarrow \\ \text{rectángulo} & & \text{triángulo} \end{array} \\ &\quad \downarrow \\ &\quad \text{para todos los valores de } p \\ &= (A_{i+1} - A_i) * (B_i + B_{i+1} / 2 - B_i / 2) = 1/2 \sum (A_{i+1} - A_i) * (B_i + B_{i+1}) \end{aligned}$$

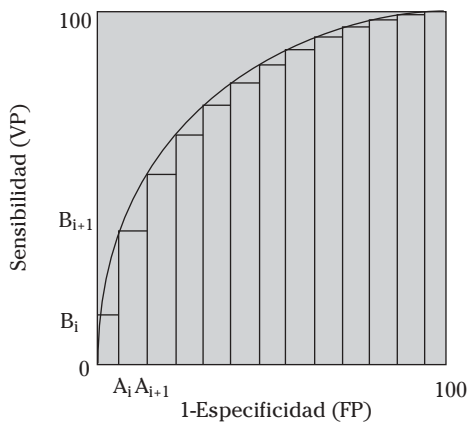


Figura 11. Cálculo del área bajo la curva ROC.

Aunque hoy en día, la mayoría de los paquetes de programas estadísticos permiten calcular el área bajo la curva ROC y comparar así la eficacia de diferentes pruebas.

RESUMEN

Las medidas de la validez de las pruebas diagnósticas son la sensibilidad y la especificidad. La sensibilidad es la probabilidad de que los individuos enfermos resulten positivos en la prueba:

$$\text{Sensibilidad} = \frac{VP}{VP + FN}$$

La especificidad es la probabilidad de que en los individuos no enfermos la prueba resulte negativa:

$$\text{Especificidad} = \frac{VN}{VN + FP}$$

El poder discriminatorio de las pruebas también puede expresarse mediante el cociente de verosimilitud (likelihood ratio = LR). La verosimilitud de un resultado positivo es el cociente entre la probabilidad de un resultado positivo en los enfermos (sensibilidad) y la probabilidad de un resultado positivo en los no enfermos (complementario de la especificidad):

$$LR+ = \frac{\text{Sensibilidad}}{1 - \text{Especificidad}}$$

La verosimilitud de un resultado negativo es el cociente entre la probabilidad de un resultado negativo en los enfermos (complementario de la sensibilidad) y la probabilidad de un resultado negativo en los no enfermos (especificidad):

$$LR- = \frac{1 - \text{Sensibilidad}}{\text{Especificidad}}$$

Cuanto mayor es el cociente de verosimilitud de un resultado positivo o menor el cociente de verosimilitud de un resultado negativo mejor discrimina la prueba evaluada.

La sensibilidad y especificidad dependen no sólo de la prueba sino también de las características de la población en estudio, y consecuentemente de la probabilidad de la enfermedad en dicha población. Lo ideal es evaluar la prueba en un submuestra de la población relevante a la que se extrapolarán los datos, pero si ello no es posible, se pueden ajustar los datos modelando la sensibilidad y la especificidad mediante la regresión logística.

Cuando los resultados se expresan como una variable cuantitativa, conforme varía el punto de corte (que define un resultado positivo o negativo) la sensibilidad y especificidad varían en sentido inverso y esta relación se puede representar gráficamente mediante las curvas ROC. El punto de corte a elegir depende de las consecuencias que tenga el tratar a los falsos positivos y el dejar de tratar a los falsos negativos. A su vez, el nú-

mero absoluto de falsos positivos depende de la probabilidad de la enfermedad y de la especificidad de la prueba y el número absoluto de falsos negativos depende también de la probabilidad de la enfermedad y de la sensibilidad de la prueba. En definitiva, el punto de la curva ROC que mejor discrimina entre enfermos y no enfermos es aquel en que la pendiente coincide con el resultado de la ecuación:

$$\frac{1 - P}{P} \times \frac{C_{FP}}{C_{FN}}$$

donde:

- P es la probabilidad de la enfermedad
- C_{FP} es el coste neto de tratar a los pacientes no enfermos (coste de los falsos positivos)
- C_{FN} es el coste de no tratar a los pacientes enfermos (coste de los falsos negativos)

PROBLEMAS

1. Acude al hospital un paciente en el que se sospecha la presencia de embolismo pulmonar (TEP) con una probabilidad del 20%. Una presión de oxígeno menor de 90 mmHg es un importante criterio diagnóstico ya que este hallazgo se encuentra en el 95% de los pacientes afectados de TEP pero también en el 25% de los pacientes sin TEP. ¿Cuáles son la sensibilidad y especificidad de la prueba? ¿Cuáles son las razones de verosimilitud de un resultado positivo y de un resultado negativo? ¿En qué porcentaje de individuos la prueba resulta positiva? ¿En qué porcentaje de individuos la prueba resulta negativa? Si la presión de oxígeno es inferior a 90 mmHg ¿Qué probabilidades tiene de presentar la enfermedad? Y si la presión de oxígeno es de 100 mmHg ¿Qué probabilidades tiene de presentar la enfermedad?
2. Si un aparato para medir la hemoglobina en los donantes de sangre siempre da como resultado 160 g/l, ¿Cuál es la sensibilidad y especificidad de la prueba para el diagnóstico de anemia (hemoglobina inferior a 120 g/l)?
3. Los signos patognomónicos ¿Qué sensibilidad y especificidad tienen?
4. Usted está estudiando una nueva prueba para detectar el cáncer de próstata y para su sorpresa la prueba es positiva sólo en el 12% de los enfermos con cáncer y en el 92% de los enfermos sin cáncer. ¿Es una prueba válida?
5. En los exámenes de salud laboral se realiza siempre un conjunto de 15 pruebas (el autoanizador da los resultados en cadena). Si la probabilidad de un falso positivo en cualquiera de las 15 pruebas es del 5% ¿Cuál es la probabilidad de que en un individuo sano al menos 1 de las pruebas resulte anormal?
6. El diagnóstico de la infección asintomática por *Chlamydia trachomatis* se puede realizar mediante cultivo celular (sensibilidad del 70% y especificidad del 100%) o por inmunoensayo enzimático (sensibilidad del 80% y especificidad del 95%) (*Genç M, Mardh PA. A cost-effectiveness Analysis of Screening and Treatment for Chlamydia trachomatis Infection in Asymptomatic Women. Ann Intern Med 1996;124:1-7*). La prevalencia de la enfermedad en la población de mujeres sexualmente activas es del 6%. ¿Cuáles deben ser los costes de no diagnosticar (y por tanto no tratar) a las pacientes enfermas respecto a los costes de tratar a las mujeres sin infección para estar indiferentes entre aplicar una prueba u otra? ¿Cuál sería esta relación de costes si la probabilidad de que la infección estuviese presente fuese del 20%?

7. El índice de masa corporal (IMC) se puede utilizar como prueba de cribaje en la infancia, con el fin de detectar aquellos niños con un elevado riesgo de ser obesos en la edad adulta (*Lazarus R, Baur L, Webb K, Blyth F. Body mass index in screening for adiposity in children and adolescents: systematic evaluation using receiver operating characteristic curves. Am J Clin Nutr 1996; 63: 500-506*). La sensibilidad y especificidad de la prueba cambian dependiendo del percentil del IMC que se utilice como punto de corte:

	Percentil 80	Percentil 85
Sensibilidad	85%	71%
Especificidad	91%	95%

No tratar a los niños obesos a temprana edad conlleva sus riesgos, pero etiquetarlos de obesos también tiene efectos psicológicos perjudiciales, de modo que la relación costes/beneficios se ha establecido en 2:1. Por otro lado, independientemente del resultado de la prueba, la probabilidad de que uno de estos niños desarrolle obesidad en la edad adulta es del 18%. ¿Qué percentil del IMC es preferible utilizar como prueba diagnóstica de obesidad?

PROBABILIDAD POSTERIOR

ESTIMACIÓN DE LA PROBABILIDAD POSTERIOR

La probabilidad posterior es la certeza (o incertidumbre, según cómo se mire) acerca de la probabilidad de presentar la enfermedad tras conocer los resultados de la prueba diagnóstica, por tanto es de gran ayuda en la decisión terapéutica. De hecho es con la probabilidad posterior con la que se trabaja cada día; cuando acude un paciente con determinados síntomas, no es cuestión de preguntarse qué sensibilidad y especificidad tienen las pruebas que se le aplican (tema que ya debería estar resuelto), sino qué probabilidad tiene el paciente de presentar la enfermedad una vez que la prueba ha resultado positiva o negativa.

La probabilidad de presentar la enfermedad es previa o posterior según el momento del proceso diagnóstico. Por ejemplo, la prevalencia de coronariopatía en la población general es la probabilidad previa respecto a la probabilidad de presentar esta patología un paciente que acude a la consulta por presentar dolor precordial que irradia hacia el hombro y brazo izquierdos (probabilidad posterior, dados unos síntomas). A su vez, esta probabilidad posterior es la previa al conocimiento del ECG y así sucesivamente.

Intuitivamente parece lógico pensar que si el resultado de la prueba es positivo, con una elevada probabilidad el individuo tiene la enfermedad, y en cambio, si el resultado es negativo lo más probable es que el paciente no tenga la patología en cuestión. En este capítulo tendremos la ocasión de comprobar que la intuición, en estos casos, no siempre funciona.

Dos son los conceptos claves cuando se habla de probabilidad posterior:

- 1) El valor predictivo positivo (VPP): es la probabilidad de tener la enfermedad dado que la prueba ha resultado positiva.
- 2) El valor predictivo negativo (VPN): es la probabilidad de no presentar la enfermedad dado un resultado negativo.

Existen básicamente cuatro métodos para calcular los valores predictivos: tablas 2x2, teorema de Bayes, razón de verosimilitud y regresión logística.

Tablas de 2x2

La forma más sencilla de calcular los valores predictivos es mediante las tablas de 2x2 (Tabla 15) tal y como veíamos en el capítulo anterior, donde se combinan las dos variables, enfermo o no enfermo y resultado positivo o negativo: en las columnas se coloca el verdadero estado del paciente y en las filas el resultado de la prueba. Podemos suponer que trabajamos con una muestra de 1.000 individuos (Tabla 16). Según los datos de la probabilidad previa de la enfermedad se completan los marginales de las columnas. Si por ejemplo, la probabilidad previa de presentar la enfermedad es del 20%, debajo de la columna de enfermos se pone 200 y 800 debajo de la columna de los no enfermos. Dentro de la columna de enfermos se rellenan las casillas según la sensibilidad y dentro de la columna de no enfermos se rellenan las casillas según la especificidad. Si la sensibilidad es del 90%, por ejemplo, quiere decir que en 180 de los 200 individuos enfermos la prueba ha resultado positiva y en 20 negativa. Si la especificidad es del 80% quiere decir que en 640 de los 800 individuos no enfermos la prueba ha resultado negativa y en 160 ha resultado positiva. Ya están todos los datos de las casillas, y así se pueden completar los marginales de las filas. En total la prueba ha resultado positiva en 180 individuos enfermos y 160 no enfermos = 340, y la prueba ha resultado negativa en 20 enfermos y 640 no enfermos = 660.

Tabla 15. Tabla 2x2.

		Enfermedad		
		+	-	
Prueba +	VP	FP	VP + FP	
Prueba -	FN	VN	FN + VN	
	VP + FN	FP + VN		

Tabla 16. Tabla 2x2.

		Enfermedad		
		+	-	
Prueba +	180	160	340	
Prueba -	20	640	660	
	200	800	1.000	

- El valor predictivo positivo es del total de individuos en los que la prueba ha resultado positiva (verdaderos positivos más falsos positivos) cuántos tienen realmente la enfermedad (verdaderos positivos).

$$VPP = \frac{VP}{VP + FP}$$

En el ejemplo concreto, el $VPP = \frac{180}{180 + 160} = 0,53$

Cuando la prueba resulta positiva, la probabilidad de presentar la enfermedad es del 53%.

- El valor predictivo negativo es de todos los individuos en los que la prueba ha resultado negativa (verdaderos negativos más falsos negativos) cuántos realmente no tienen la enfermedad (verdaderos negativos).

$$VPN = \frac{VN}{VN + FN}$$

En el ejemplo concreto, el $VPN = \frac{640}{640 + 20} = 0,97$

Cuando la prueba resulta negativa, la probabilidad de no tener la enfermedad es del 97%.

Teorema de Bayes

En el capítulo segundo se explica cómo calcular probabilidades condicionadas mediante la aplicación del teorema de Bayes, que tiene su aplicación más directa en la estimación de los valores predictivos a partir de la probabilidad previa, la sensibilidad y la especificidad.

Si $p(E+)$ es la probabilidad previa de la enfermedad y $p(E-)$ la probabilidad previa de no tener la enfermedad, $p(T+)$ la probabilidad de un resultado positivo y $p(T-)$ la probabilidad de un resultado negativo, los valores predictivos se pueden expresar en forma de probabilidades condicionadas:

$$VPP = p(E+/T+) \quad \text{y} \quad VPN = p(E-/T-)$$

Por las propiedades de las probabilidades:

$$p(E+/T+) = \frac{p(E+ \cap T+)}{p(T+)} = \frac{p(T+/E+) * p(E+)}{p(T+ \cap E+) + p(T+ \cap E-)}$$

$$p(E+/T+) = \frac{p(T+/E+) * p(E+)}{p(T+/E+) * p(E+) + p(T+/E-) * p(E-)}$$

$p(T+/E+)$ es la sensibilidad de la prueba.
 $p(E+)$ es la probabilidad previa de tener la enfermedad.
 $p(T+/E-)$ es el complementario de la especificidad, o la tasa de falsos positivos.
 $p(E-)$ es la probabilidad previa de no tener la enfermedad.

De aquí se deduce que:

$$VPP = \frac{P*S}{P*S + (1-P)*(1-E)}$$

donde:

VPP = valor predictivo positivo
 P = probabilidad previa de la enfermedad
 S = sensibilidad de la prueba
 E = especificidad de la prueba

Aplicando los datos de la tabla:

$$VPP = \frac{0,20*0,90}{0,20*0,90 + 0,80*0,20} = 0,53$$

El resultado es el mismo que cuando se calcula mediante las tablas de 2x2.

Si se procede del mismo modo con el valor predictivo negativo:

$$p(E-/T-) = \frac{p(E- \cap T-)}{p(T-)} = \frac{p(T-/E-) * p(E-)}{p(T- \cap E-) + p(T- \cap E+)}$$

$$p(E-/T-) = \frac{p(T-/E-) * p(E-)}{p(T-/E-) * p(E-) + p(T-/E+) * p(E+)}$$

$p(T-/E-)$ es la especificidad de la prueba.
 $p(E-)$ es la probabilidad previa de no presentar la enfermedad.
 $p(T-/E+)$ es el complementario de la sensibilidad, o la tasa de falsos negativos.
 $p(E+)$ es la probabilidad previa de tener la enfermedad.

De aquí se deduce que:

$$VPN = \frac{(1-P)*E}{(1-P)*E + P*(1-S)}$$

donde:

VPN = valor predictivo negativo
 P = probabilidad previa de la enfermedad
 S = sensibilidad de la prueba

E = especificidad de la prueba

De nuevo, al aplicar los datos del ejemplo:

$$VPN = \frac{0,80 * 0,80}{0,80 * 0,80 + 0,20 * 0,10} = 0,97$$

Si lo que interesa es conocer la probabilidad de tener la enfermedad, aunque el resultado de la prueba haya resultado negativo, basta con calcular el complementario del valor predictivo negativo, pero por entrenamiento en el manejo del teorema de Bayes vamos a proceder del mismo modo:

$$p(E+/T-) = \frac{p(E+ \cap T-)}{p(T-)} = \frac{p(T-/E+) * p(E+)}{p(T- \cap E-) + p(T- \cap E+)}$$

$$p(E+/T-) = \frac{p(T-/E+) * p(E+)}{p(T-/E-) * p(E-) + p(T-/E+) * p(E+)}$$

$$1 - VPN = \frac{P * (1-S)}{(1-P) * E + P * (1-S)}$$

$$1 - VPN = \frac{0,20 * 0,10}{0,80 * 0,80 + 0,20 * 0,10} = 0,03$$

Cuando la prueba resulta negativa, la probabilidad de tener la enfermedad es del 3%. El 3% es efectivamente el complementario del 97% que había resultado al calcular el valor predictivo negativo.

Razón de verosimilitud

El teorema de Bayes es difícil de calcular sin lápiz y papel, pero existe una forma mucho más sencilla de calcularlo si se sustituyen las probabilidades por odds. Tal y como hemos visto en capítulos anteriores, una odds es el cociente entre la probabilidad de que ocurra un hecho y la probabilidad de que no ocurra. Así, las probabilidades previa y posterior se pueden expresar como odds previa y odds posterior respectivamente, y la capacidad discriminatoria de la prueba (su validez) se expresa en términos de razón de verosimilitud o “likelihood ratio” (LR):

$$LR = \frac{\text{Probabilidad de un resultado en los enfermos}}{\text{Probabilidad del resultado en los no enfermos}}$$

Se puede calcular una razón de verosimilitud para un resultado positivo (LR+) y otra para un resultado negativo (LR-):

$$LR+ = \frac{\text{Probabilidad de resultado (+) en los enfermos}}{\text{Probabilidad de resultado (+) en los no enfermos}} =$$

$$LR+ = \frac{\text{Sensibilidad}}{\text{Tasa de falsos positivos}} = \frac{\text{Sensibilidad}}{1 - \text{Especificidad}}$$

$$LR- = \frac{\text{Probabilidad de resultado (-) en los enfermos}}{\text{Probabilidad de resultado (-) en los no enfermos}} =$$

$$LR- = \frac{\text{Tasa de falsos negativos}}{\text{Especificidad}} = \frac{1 - \text{Sensibilidad}}{\text{Especificidad}}$$

La probabilidad previa se puede transformar en odds previa:

$$\text{Odds previa} = \frac{\text{Probabilidad previa de enfermedad}}{\text{Probabilidad previa de no enfermedad}} = \frac{(p)}{(1-p)}$$

$$\text{Odds posterior} = \frac{\text{Probabilidad posterior de enfermedad}}{\text{Probabilidad posterior de no enfermedad}}$$

La odds posterior para un resultado positivo es:

$$\text{Odds previa} * LR+$$

El resultado de este producto nos indica cuántas veces es más probable presentar la enfermedad que no presentarla una vez que la prueba ha resultado positiva.

La odds posterior para un resultado negativo es:

$$\text{Odds previa} * LR-$$

El resultado de este producto nos indica cuánto menos probable es presentar la enfermedad respecto a no presentarla una vez que la prueba ha resultado negativa. O lo que es lo mismo, el inverso de este valor expresa cuántas veces es más probable no presentar la enfermedad (que presentarla) una vez que la prueba ha resultado negativa.

Aunque en principio no lo parezca (porque no estamos muy acostumbrados a trabajar con odds), es una forma sencilla e intuitiva de calcular la probabilidad posterior. Sólo es preciso recordar un número (la razón de verosimilitud) en lugar de dos (la tasa de verdaderos positivos o sensibilidad y la tasa de falsos positivos o complementario de la especificidad).

Con un ejemplo se verá mejor. Imaginemos una prueba diagnóstica cuya sensibilidad es del 97% y la especificidad del 73%:

$$LR+ = 0,97/0,27 = 3,6 \quad (3,6 \text{ a } 1)$$

Es 3,6 veces más probable un resultado positivo en los enfermos que en los no enfermos. Por cada vez que sale un resultado positivo en los individuos sin la enfermedad, sale 3,6 veces un resultado positivo en los enfermos.

$$LR- = 0,03/0,73 = 0,04 \text{ (1 a 24,3)}$$

Es 24,3 veces más probable un resultado negativo en los no enfermos que en los enfermos. Por cada vez que sale un resultado negativo en los enfermos, sale 24,3 veces un resultado negativo en los sujetos no enfermos.

La probabilidad previa de la enfermedad es del 40%.

$$\text{Odds previa} = 0,4/0,6 = 0,67 \text{ (4 a 6) ó (2 a 3)}$$

Por cada dos enfermos hay tres individuos sin la enfermedad.

$$\text{Odds posterior de resultado positivo} = 0,67 * 3,6 = 2,4 \text{ (2,4 a 1)}$$

Cuando el resultado es positivo es 2,4 veces más probable presentar la enfermedad que no presentarla.

$$\text{Odds posterior de resultado negativo} = 0,67 * 0,04 = 0,027 \text{ (1 a 37)}$$

Si el resultado es negativo son mucho mayores (37 veces más) las probabilidades de no presentar la enfermedad que de presentarla, por cada enfermo habrá 37 pacientes sin la enfermedad.

Posteriormente, las odds pueden ser convertidas en probabilidades:

$$\text{Probabilidad} = \frac{\text{Odds}}{1 + \text{Odds}}$$

Si el resultado es positivo, la probabilidad de estar enfermo es:

$$p = 2,4/3,4 = 0,706$$

o lo que es lo mismo, el valor predictivo positivo es del 70,6%.

Si el resultado es negativo, la probabilidad de estar enfermo es:

$$p = 0,027/1,027 = 0,026$$

El complementario, es decir, la probabilidad de no presentar la enfermedad dado un resultado negativo (valor predictivo negativo), es del 97,4%.

Por fortuna, existe un medio muy simple de estimar la odds postexamen, y es un nomograma propuesto por Fagan (Fig. 12). La primera columna de este nomograma representa la probabilidad previa, preexamen, la segunda representa el cociente de verosimilitud y la tercera la probabilidad posterior, postexamen. Se obtiene la probabilidad postexamen sujetando una regla en la probabilidad previa y girándola hasta que se alinea con la razón de verosimilitud.

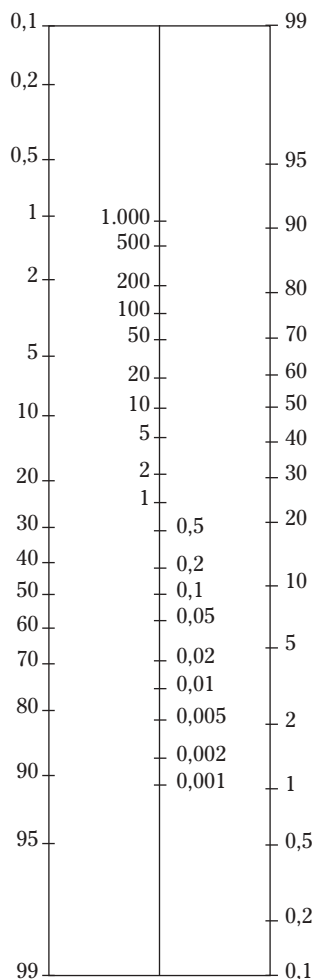


Figura 12. Nomograma de Fagan.

Regresión logística

El valor predictivo de una prueba depende de la probabilidad previa de la enfermedad. Consecuentemente, factores relacionados con la probabilidad de la enfermedad tales como la edad, el sexo o la presencia de determinados signos clínicos influyen en la capacidad de predicción de la prueba. Además, la sensibilidad y especificidad de las pruebas también se influyen por factores como la gravedad de la enfermedad o la presencia de enfermedades concomitantes. Por ello sería deseable determinar los valores predictivos según diferentes combinaciones de los factores (lo que se denominan patrones de covariables) que influyen en la probabilidad de tener la enfermedad y en la validez de las pruebas.

Del mismo modo que se aplicaba la regresión logística para el cálculo de la sensibilidad y especificidad, también se pueden calcular los valores predictivos positivo y negativo. Pero en este caso la variable dependiente es el verdadero estado del paciente (enfermo o no) determinado por la prueba de referencia (gold standard) y las variables independientes son el resultado de la prueba (positivo o negativo) y aquellos factores que pueden influir en la probabilidad previa y en la validez de la prueba, y por tanto en los valores predictivos.

El modelo de regresión logística es como sigue:

$$p(E) = \frac{e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}}{1 + e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}}$$

donde:

$p(E)$ es la probabilidad de estar enfermo.

x_1 = variable dicotómica del resultado de la prueba (positivo/negativo)

$x_2 \dots x_n$ = covariables que pueden influir en los valores predictivos (gravedad de la patología, presencia de síntomas, estadio, edad, sexo...etc).

α = ordenada en el origen.

β_1 = coeficiente de regresión de la variable "resultado de la prueba".

$\beta_2 \dots \beta_n$ = coeficientes de regresión de las covariables.

Supongamos que la variable independiente "resultado de la prueba" toma el valor 1 cuando es positivo y el valor 0 cuando es negativo. Mediante la regresión logística se puede calcular la probabilidad de presentar la enfermedad según distintos valores de las variables independientes (son por tanto probabilidades condicionadas), y así, si el valor predictivo positivo es la probabilidad de presentar la enfermedad cuando el resultado de la prueba es positivo, basta con sustituir en la fórmula x_1 por 1 y dar valores a las otras variables dependiendo de las características del individuo en estudio. El valor predictivo negativo será el complementario de la probabilidad de presentar la enfermedad cuando se sustituye x_1 por 0, ya que es la probabilidad de no presentar la enfermedad cuando el resultado de la prueba es negativo.

Por ejemplo, se realizó un modelo para la estimación de los valores predictivos de la enfermedad arterial periférica, diagnosticada mediante métodos no invasivos (Tabla 17). La prueba utilizada como diagnóstico en los pacientes sintomáticos fue el cuestionario de Rose. El sexo se utilizó como covariable por la que ajustar. Si el cuestionario resulta diagnóstico toma el valor 1, y cero en caso contrario. El sexo es 1 en los hombres y cero en las mujeres.

Tabla 17. Estimación de los valores predictivos mediante la regresión logística. Tomado de: *Coughlin SS, Trock B, Criqui MH, Pickle LW, Browner D and Tefft MC. The logistic modeling of sensitivity, specificity, and predictive value of a diagnostic test. J Clin Epidemiol 1992; 45: 1-7.*

Variable	Coeficiente	Error estándar
α	-2,263	0,191
Enfermo	1,805	0,377
Sexo	0,180	0,264

Si se trata de un hombre en el que la prueba ha resultado positiva, la probabilidad de tener enfermedad arterial periférica es:

$$p(E) = \frac{e^{-2,263 + 1,805*1 + 0,180*1}}{1 + e^{-2,263 + 1,805*1 + 0,180*1}} = 0,43$$

Se puede realizar el ejercicio de comprobar que si se trata de una mujer en la que la prueba resulta positiva, la probabilidad de tener la enfermedad es del 39%. Si en un hombre la prueba resulta negativa, la probabilidad de tener la enfermedad es del 11% y si resulta negativa en una mujer, la probabilidad es del 9%.

PROBABILIDAD PREVIA, SENSIBILIDAD, ESPECIFICIDAD Y PROBABILIDAD POSTERIOR

La probabilidad posterior de una enfermedad depende de la probabilidad previa y de la sensibilidad y especificidad de la prueba diagnóstica. No se pueden extrapolar probabilidades posteriores a otros individuos con diferente probabilidad previa de presentar la enfermedad, aunque las pruebas que se apliquen sean las mismas. Por ejemplo, ante un ECG alterado, la probabilidad de presentar un infarto de miocardio no es la misma si el individuo es asintomático o si presenta el cuadro clínico típico.

En la figura 13 se puede apreciar cómo la probabilidad de la enfermedad (probabilidad posterior) aumenta tras un resultado positivo, sobre todo cuando se parte de una probabilidad previa baja, y a expensas fundamentalmente de la especificidad. Del mismo modo, la probabilidad de la enfermedad disminuye tras un resultado negativo, sobre todo cuando la probabilidad previa es elevada y fundamentalmente a expensas de la sensibilidad. Vamos por partes estudiando a fondo este gráfico.

Cuando la probabilidad previa es igual que la probabilidad posterior (el punto cae en la diagonal del gráfico), representa el hipotético caso en el que la prueba no cambia en absoluto la probabilidad de la enferme-

dad (se trata de una prueba no válida, ni sensible ni específica). Cuanto más se aleja la curva de la diagonal, mayor es el efecto discriminatorio de la prueba (mayor su validez).

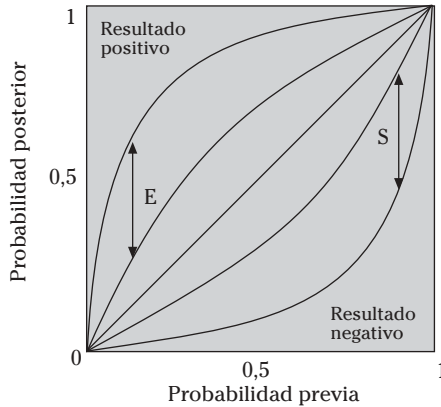


Figura 13. Probabilidad previa, posterior, sensibilidad y especificidad.

Suponiendo una prueba válida, un resultado positivo aumenta la probabilidad de la enfermedad, pero si la probabilidad previa ya es muy elevada, difícilmente podrá aumentar mucho más (hay un máximo que es el 100% y de ahí no pasa). De igual modo, un resultado negativo disminuye la probabilidad de presentar la enfermedad, pero si la probabilidad previa ya es muy baja, difícilmente podrá disminuir mucho más (porque también hay un mínimo que es el 0%). De aquí se deduce una conclusión importante:

“Cuando antes de realizar una prueba, el diagnóstico es casi cierto (en un sentido u otro, para dar un diagnóstico o para descartarlo), un resultado confirmatorio (positivo en el primer caso, negativo en el segundo) tiene poco efecto sobre la probabilidad de la enfermedad”.

Sin embargo, un inesperado resultado negativo en un paciente con una probabilidad previa alta puede modificar nuestra actitud (no dar un tratamiento por ejemplo) al disminuir en gran medida la probabilidad de la enfermedad. Y del mismo modo, un inesperado resultado positivo en un paciente con una probabilidad previa muy baja también puede modificar nuestra actitud en el sentido de ampliar el estudio del paciente porque ahora la probabilidad de que presente la enfermedad es bastante más alta. Pero ahora nos encontramos con un pequeño problema: ese inesperado resultado negativo esperemos que no sea un falso negativo (por lo que se precisa una elevada sensibilidad), al igual que esperamos que el inesperado resultado positivo no sea un falso positivo (elevada especificidad).

De hecho, el valor predictivo positivo (probabilidad de presentar la enfermedad dado un resultado positivo) depende sobre todo de la espe-

cificidad y especialmente cuando la probabilidad previa es baja. Sin nos fijamos en la fórmula, la especificidad sólo afecta al denominador del valor predictivo positivo ($VP/(VP+FP)$) y en cambio al numerador y denominador del valor predictivo negativo ($VN/(VN+FN)$), provocando, por tanto, cambios más importantes en el primero (VPP). Por otro lado, cuando la probabilidad previa de la enfermedad es baja, el número absoluto de falsos positivos puede ser muy alto, y pequeños cambios en la especificidad pueden hacer variar este número en gran medida.

El valor predictivo negativo (probabilidad de no presentar la enfermedad dado un resultado negativo) depende sobre todo de la sensibilidad y especialmente cuando la probabilidad previa es elevada. La sensibilidad sólo afecta al denominador del valor predictivo negativo ($VN/(VN+FN)$) y en cambio al numerador y denominador del valor predictivo positivo ($VP/(VP+FP)$), provocando por tanto cambios más importantes en el primero (VPN). Por otro lado, cuando la probabilidad previa de la enfermedad es elevada, en números absolutos los falsos negativos pueden ser muchos, con lo que pequeños cambios en la sensibilidad pueden modificar este número absoluto de forma importante.

La influencia de la probabilidad previa, la sensibilidad y especificidad sobre la probabilidad posterior pueden comprobarse en los datos de la tabla 18.

De todo este razonamiento se deriva que se requiere una prueba muy específica:

- Cuando se precisa verificar la presencia de una enfermedad con poca evidencia clínica. Por ejemplo, un programa de cribaje que detecta una enfermedad rara en pacientes asintomáticos porque los beneficios de tratarla son muy importantes.
- Cuando no siendo tan rara la enfermedad, es preciso estar muy seguros del diagnóstico antes de tratarla, porque las consecuencias de un tratamiento inadecuado podrían ser nefastas.

Cuando la probabilidad de una enfermedad es baja, la mejor manera de estar seguros de un diagnóstico correcto es diagnosticando bien a la gran mayoría de la población, que en este caso son los individuos sin la enfermedad, y es la especificidad la capacidad de detectar a éstos. Si para detectar a los enfermos buscásemos un signo sensible, lo presentarían todos los enfermos, pero también algunos de los que no presentan la enfermedad (poco específico) que en números absolutos podrían ser muchos, y resultaría difícil saber si al presentar el signo este individuo es enfermo o no. Imaginemos que tenemos una cita a ciegas con un individuo en un bar muy frecuentado. Para reconocerlo ¿qué dato es preferible? ¿que estará tomando una cerveza o que llevará un clavel rojo en la solapa? Si nos fiamos del dato de la cerveza nuestro individuo en cuestión estará tomando

una (es un dato sensible), pero también muchas otras personas del bar (no es un signo específico), sin embargo es difícil que a alguien más se le ocurra llevar el dichoso clavel, que identifica (especifica) a nuestro individuo mucho más fácilmente.

A veces, aún siendo la probabilidad de la enfermedad alta, precisamos una elevada especificidad porque queremos estar muy seguros del diagnóstico, porque el tratamiento tiene importantes efectos adversos y no queremos correr el riesgo de aplicar un tratamiento así (una mastectomía por ejemplo) a un falso positivo.

Tabla 18. Cambios en los valores predictivos dependiendo de la sensibilidad, especificidad y probabilidad previa.

Probabilidad previa baja: 10%

		Enfermos			Aumenta la especificidad			Aumenta la sensibilidad		
		Sí	No		Sí	No		Sí	No	
+		80	180	260	80	90	170	90	180	270
-		20	720	740	20	810	830	10	720	730
		100	900		100	900		100	900	
		S = 0,80			S = 0,80			S = 0,90		
		E = 0,80			E = 0,90			E = 0,80		
		VPP = 0,308			VPP = 0,471			VPP = 0,333		
		VPN = 0,973			VPN = 0,976			VPN = 0,986		

Probabilidad previa alta: 90%

		Enfermos			Aumenta la especificidad			Aumenta la sensibilidad		
		Sí	No		Sí	No		Sí	No	
+		720	20	740	720	10	730	810	20	830
-		180	80	260	180	90	270	90	80	170
		900	100		900	100		900	100	
		S = 0,80			S = 0,80			S = 0,90		
		E = 0,80			E = 0,90			E = 0,80		
		VPP = 0,973			VPP = 0,986			VPP = 0,976		
		VPN = 0,308			VPN = 0,333			VPN = 0,471		

Se requiere una prueba con una elevada sensibilidad:

- Cuando es preciso descartar una enfermedad con una importante sospecha diagnóstica.
- Cuando no siendo tan probable, es necesario estar muy seguro de la ausencia de una enfermedad, porque los beneficios de tratar a los individuos enfermos son elevados.

Cuando la probabilidad de la enfermedad es alta, la mejor manera de no equivocarse es diagnosticando bien a la gran mayoría de la población, que en este caso son los enfermos, y los enfermos se detectan mediante la sensibilidad. La forma de detectar a los individuos sin la enfermedad (los raros en este caso), es mediante la ausencia de un signo sensible. Si para detectar a estos últimos buscásemos la ausencia de un signo específico de los enfermos, efectivamente no lo presentarían, pero tampoco muchos de los enfermos, y ante un resultado negativo no se puede saber si presenta la enfermedad o no. Si, en cambio, buscamos la ausencia de un signo sensible (prácticamente todos los enfermos lo tienen), su ausencia denota ausencia de enfermedad. Imaginemos que queremos identificar a los extranjeros en Escocia, ¿cómo estaremos más seguros de que un individuo no es escocés, porque no lleva la famosa faldita y no toca la gaita o porque no sabe hablar escocés? Lógicamente porque no sabe hablar escocés. Si nos fijamos del detalle de la falda (un signo muy específico pero no sensible), no todos los individuos la llevan, y en cambio, si no sabe hablar escocés difícilmente será de la zona.

Si se realiza el PPD, por ejemplo, un resultado negativo prácticamente excluye el diagnóstico de tuberculosis activa. El positivo, sin embargo, no necesariamente confirma dicho diagnóstico.

Si la probabilidad de la enfermedad es baja pero el tratamiento es muy efectivo, procuraremos que ningún paciente se nos quede sin tratar, que no haya falsos negativos, y la forma de garantizarlo es con una elevada sensibilidad.

En definitiva, la decisión entre sensibilidad y especificidad es un manejo conjunto de los riesgos y beneficios que conlleva el tratamiento de los falsos positivos y verdaderos positivos respectivamente, además de la probabilidad de la enfermedad, que a su vez influirá en el número absoluto de verdaderos y falsos positivos y negativos.

RESUMEN

La probabilidad posterior es la probabilidad de presentar una enfermedad tras conocer el resultado de las pruebas diagnósticas. Concretamente el valor predictivo positivo (VPP) es la probabilidad de presentar la enfermedad tras un resultado positivo y el valor predictivo negativo (VPN) es la probabilidad de no presentar la enfermedad tras un resultado negativo.

Para su cálculo existen cuatro métodos básicos:

1. Tablas de 2x2:

$$VPP = \frac{VP}{VP + FP} \quad VPN = \frac{VN}{VN + FN}$$

donde:

VP = verdaderos positivos
 FP = falsos positivos
 VN = verdaderos negativos
 FN = falsos negativos

2. Teorema de Bayes:

$$VPP = \frac{P \cdot S}{P \cdot S + (1-P) \cdot (1-E)}$$

$$VPN = \frac{(1-P) \cdot E}{(1-P) \cdot E + P \cdot (1-S)}$$

donde:

P = probabilidad previa de la enfermedad
 E = especificidad de la prueba
 S = sensibilidad de la prueba

3. Razón de verosimilitud:

La odds posterior para un resultado positivo es:

Odds previa * LR+

La odds posterior para un resultado negativo es:

Odds previa * LR-

LR = "likelihood ratio". Razón de verosimilitud

4. Regresión logística: donde la variable dependiente debe ser la presencia o ausencia de la enfermedad y la variable independiente el resulta-

do de las pruebas diagnósticas. Además, permite ajustar por otros factores relacionados.

El valor predictivo positivo depende sobre todo de la especificidad y especialmente cuando la probabilidad de la enfermedad es baja. Por el contrario, el valor predictivo negativo depende sobre todo de la sensibilidad, especialmente cuando la probabilidad de la enfermedad es alta.

Se requiere una prueba muy específica cuando se precisa verificar la presencia de una enfermedad con poca evidencia clínica o cuando, no siendo tan rara la enfermedad, es preciso estar muy seguros del diagnóstico antes de tratarla, porque las consecuencias de un tratamiento inadecuado podrían ser nefastas.

Se requiere una prueba con una elevada sensibilidad cuando es preciso descartar una enfermedad con una importante sospecha diagnóstica o cuando no siendo tan probable, es necesario estar muy seguro de la ausencia de una enfermedad, porque los beneficios de tratar a los individuos enfermos son elevados.

PROBLEMAS

1. La probabilidad de la enfermedad en estudio es del 12%. La sensibilidad de la prueba aplicada es del 90% y la especificidad del 95%. Calcular los valores predictivos positivo y negativo mediante las tablas de 2x2, el teorema de Bayes y la razón de verosimilitud.
2. Suponga que tras una evaluación inicial usted cree que su paciente tiene una probabilidad de presentar la enfermedad de 0,25 y decidirá aplicar un tratamiento específico si la probabilidad de la enfermedad es superior al 50%. ¿Cuál es la mínima razón de verosimilitud de la prueba para, en vista de los resultados, decidir administrar el tratamiento?
3. Basado en la historia clínica y en la exploración física, el paciente tiene una probabilidad de presentar tromboembolismo pulmonar del 30%, pero si la probabilidad de la enfermedad es menor del 5% no se le dará un tratamiento anticoagulante. Dispone de dos pruebas diagnósticas, ¿Cuál de ellas resultará más efectiva?

	Sensibilidad	Especificidad
Prueba A	0,85	0,95
Prueba B	0,90	0,90

4. La probabilidad de que un paciente concreto con transplante hepático presente rechazo es del 70%. Se le aumentará la dosis del tratamiento inmunosupresor (con el consiguiente riesgo de infección) si dicha probabilidad es superior al 5%. Tras realizar un análisis (que incluye enzimas hepáticas), todos los resultados están dentro de la normalidad característica de este tipo de pacientes, y entonces usted decide no aumentar el tratamiento inmunosupresor, pero un colega piensa que el hecho de que la analítica esté dentro de la normalidad no es suficiente para descartar el rechazo. El realizaría una biopsia (que también presenta sus riesgos) y si resulta negativa entonces decidiría no aumentar el tratamiento inmunosupresor. ¿Quién tiene razón?

	Tasa de verdaderos positivos	Tasa de falsos positivos
Analítica	0,90	0,10
Biopsia	0,99	0,05

CUÁNDO APLICAR UNA PRUEBA DIAGNÓSTICA

INTRODUCCIÓN

Después de evaluar al paciente mediante la historia clínica, la exploración física y quizás la aplicación de algunas pruebas, y en función de la probabilidad de tener la enfermedad, el médico puede elegir entre las siguientes alternativas:

- 1) no hacer nada más, descartar la enfermedad.
- 2) aplicar otra prueba diagnóstica para obtener más información.
- 3) pasar directamente al tratamiento.

Las pruebas diagnósticas, lo que permiten es un cambio en la probabilidad de tener la enfermedad. De hecho, una prueba es tanto más eficaz cuanto mayor cambio provoca. Por ejemplo, acude a la consulta una paciente que por los síntomas que refiere se estima que la probabilidad de presentar hipotiroidismo es del 30%. Tras la determinación de la TSH y la T4 y habiendo resultado éstas por encima y por debajo, respectivamente, de los límites de normalidad, la probabilidad de que la paciente sea hipotiroidea es del 90%. Ha aumentado mucho la probabilidad de la enfermedad, la TSH y la T4 en conjunto han resultado ser unas pruebas efectivas (se puede hacer el ejercicio y comprobar que este resultado puede equivaler a unas pruebas con una sensibilidad y especificidad del 95%). El ejemplo contrario sería el de una paciente con una imagen en la mamografía muy sugestiva de carcinoma, indicando una probabilidad de tener la enfermedad del 90% y que tras realizar la biopsia y resultando ésta negativa, baja la probabilidad de neoplasia al 8% (corresponde a una sensibilidad y especificidad de la biopsia del 99%).

No es necesario llegar a la certeza absoluta sobre el verdadero estado del paciente, y difícilmente se concluye que la probabilidad de la enfermedad es 1 ó 0. Sin embargo, si llegamos a probabilidades de la enfermedad del 97% o del 2%, las posibilidades de error son pequeñas.

La necesidad de llegar a un diagnóstico más o menos certero deriva del hecho de que si existe la posibilidad de errores en el diagnóstico, existe la posibilidad de errores en el tratamiento, tanto tratando a un falso

positivo como dejando de administrar tratamiento a un falso negativo. Por lo tanto, se precisa una mayor seguridad en el diagnóstico dependiendo de las consecuencias de un tratamiento correcto o equivocado. ¿Enviaríamos a quirófano a la paciente en la que la mamografía era sugestiva de cáncer de mama sin previamente realizarle la biopsia? Las consecuencias de un tratamiento equivocado (mastectomía a una mujer que quizás no tiene una neoplasia) son muy importantes, por lo que es necesario aplicar otra prueba para aumentar o disminuir claramente la probabilidad de la enfermedad. A una paciente de 25 años con anemia ferropénica diagnosticada por hemograma y determinación de ferritina sérica, ¿le practicaríamos una punción de médula ósea para mayor seguridad en el diagnóstico del déficit de hierro? El tratamiento con hierro es eficaz y con pocos efectos secundarios, luego no es preciso buscar tanta seguridad en el diagnóstico antes de decidir tratar a la paciente.

Si de entrada la probabilidad de la enfermedad es tan pequeña, que aunque aplicando una prueba y resultando ésta positiva la decisión es de no tratar al paciente (porque el tratamiento tiene sus riesgos y es preciso mayor seguridad en el diagnóstico antes de aplicarlo) ¿para qué aplicar la prueba? Y si la probabilidad de tener la enfermedad es tan elevada que aunque la prueba resulte negativa se decide tratar al paciente (el tratamiento es beneficioso y con muy pocos efectos secundarios por ejemplo) ¿qué sentido tiene la prueba diagnóstica?

De aquí deriva una de las grandes “normas” de la decisión médica: *sólo se debe aplicar una prueba diagnóstica si la información que aporta puede modificar la actitud ante el paciente.*

De hecho, se ha realizado un diagnóstico cuando el aporte de más información no va a cambiar la decisión de actuar como si el paciente tuviese la enfermedad, y del mismo modo, se descarta una enfermedad cuando el aporte de más información no va a cambiar la decisión de actuar como si el paciente no tuviese la enfermedad.

La probabilidad de tener la enfermedad por debajo de la cual no es preciso aplicar una prueba diagnóstica ni por supuesto tratar al paciente, porque aunque la prueba resultase positiva, no aumentaría tanto la probabilidad como para decidir aplicar el tratamiento, se denomina umbral de diagnóstico (p_1 en la figura 14), de modo que si la probabilidad de la enfermedad está por debajo de este umbral decimos que hemos descartado esta enfermedad.

La probabilidad de tener la enfermedad por encima de la cual ya no se aplica ninguna prueba diagnóstica más, se pasa directamente al tratamiento (porque aunque la prueba resultase negativa, la probabilidad de tener la enfermedad no disminuiría lo suficiente como para decidir no tratar) se denomina umbral de tratamiento (p_2 en la figura 14). Por encima de es-

ta probabilidad decimos que hemos hecho el diagnóstico de la enfermedad.

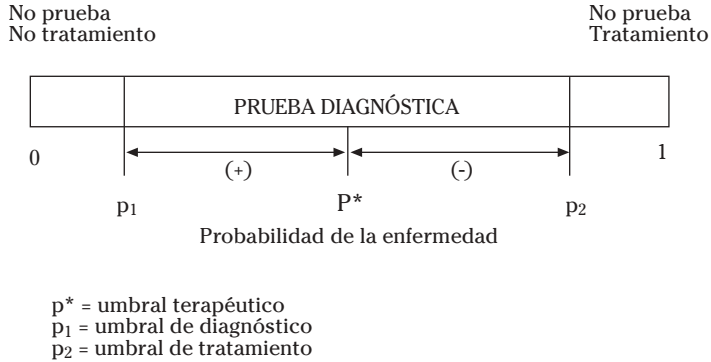


Figura 14. Umbral de diagnóstico y de tratamiento.

Cuando la probabilidad del trastorno en cuestión se encuentra entre estos dos umbrales p_1 y p_2 , son indispensables pruebas adicionales.

Se presenta en la consulta un paciente con el único síntoma de fiebre, ¿sería correcto empezar por realizarle una prueba diagnóstica para descartar una leishmaniasis? No, porque la probabilidad “a priori” es extremadamente baja, y si la prueba resultase positiva, muy fácilmente se trataría de un falso positivo, con lo que la probabilidad de la enfermedad seguiría siendo baja. Del mismo modo, si un paciente acude a la consulta en el mes de noviembre refiriendo sensación de trancazo, dolores musculares, traqueítis, fiebre, tos no productiva, etc. ¿sería preciso determinar los anticuerpos específicos de la gripe?

Continuamente venimos comentando que la probabilidad de la enfermedad es lo suficientemente alta como para decidir tratar o suficientemente baja como para decidir no tratar. Pero ¿cómo de alta o de baja? Esta probabilidad es el umbral terapéutico (p^* en la figura 14), diferente del umbral de tratamiento, aunque el nombre pueda inducir a confusión. El umbral terapéutico es la probabilidad de presentar la enfermedad por encima de la cual es preferible tratar al paciente y por debajo de la cual es preferible no tratarlo, y depende, claro está, de las consecuencias que conlleva el tratamiento correcto o incorrecto, de tratar a un falso positivo o dejar de tratar a un falso negativo.

Si recordamos que sólo se debe aplicar una prueba diagnóstica si la información que aporta puede modificar la actitud ante el paciente, en el rango de aplicación de las pruebas diagnósticas (entre p_1 y p_2):

- Si la probabilidad de la enfermedad es inferior al umbral terapéutico, se debe aplicar la prueba diagnóstica cuando de dar un resultado positivo, la probabilidad de la enfermedad pase a ser superior al umbral terapéutico. Se puede pasar de una actitud inicial de no tratar a la de tratar.
- Si la probabilidad de la enfermedad es superior al umbral terapéutico, la prueba diagnóstica que se aplique debe ser aquella que de dar un resultado negativo, disminuya la probabilidad de la enfermedad hasta niveles, al menos, inferiores al umbral terapéutico. Se puede pasar de una actitud inicial de tratar a la de no tratar.

Acude a la consulta un paciente con sospecha de faringitis estreptocócica. La probabilidad previa de presentar la enfermedad es del 50% y la del umbral terapéutico del 20%. ¿Se le trata directamente con penicilina ya que la probabilidad de presentar la enfermedad es superior al umbral terapéutico o es preferible realizar una prueba diagnóstica como el cultivo faríngeo? En esta enfermedad dejar de tratar a un paciente enfermo tiene sus riesgos (fiebre reumática), pero tratar a un paciente sin la infección también (shock por alergia a la penicilina por ejemplo). Si tras realizar el cultivo, éste resulta negativo y en dicho caso la probabilidad de presentar la enfermedad es del 1% se evitaría tratar a un número importante de pacientes que en realidad no tienen una faringitis estreptocócica.

El caso contrario, un paciente con una ligera sospecha de apendicitis, la probabilidad previa es del 10% y la probabilidad del umbral terapéutico del 20%. ¿Sería correcto enviar al paciente a casa sin operar porque la probabilidad de tener la enfermedad es menor que la del umbral terapéutico? ¿Y si tras aplicar una prueba diagnóstica (como el tacto rectal o la fórmula leucocitaria) da un resultado positivo y en este caso la probabilidad de enfermedad es del 90%?

Aunque a efectos didácticos hemos utilizado el término tratamiento como si fuera una terapia concreta, en realidad tiene un sentido mucho más amplio: puede ser cualquier curso de acción, o bien una terapia propiamente dicha o la decisión de realizar una prueba definitiva para decidir entonces si tratar o no al paciente. De hecho, tratamiento en este segundo sentido es un algoritmo en el que se “prescriben” cada una de las acciones. Por ejemplo el cáncer de mama: la mamografía es la prueba diagnóstica que queremos aplicar y el tratamiento en función de los resultados de la prueba no es la intervención quirúrgica, sino la punción-biopsia de la mama.

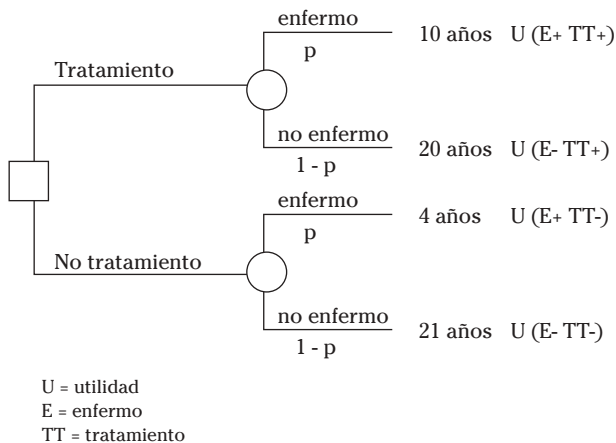
UMBRAL TERAPÉUTICO

Dejar de tratar a un paciente enfermo puede conllevar unos riesgos, pero tratar a uno sin la enfermedad también. Por tanto, se trata de de-

sarrollar una formulación matemática que permita calcular la probabilidad del umbral terapéutico, aquella probabilidad de presentar la enfermedad por encima de la cual es preferible tratar al paciente, y por debajo de ella es preferible no tratar. Para ello se relacionan los costes y beneficios del tratamiento. Costes y beneficios no tienen aquí un sentido monetario: los costes son todos los efectos negativos sobre la salud (riesgos del tratamiento) y beneficios todos los efectos positivos (curación, aumento de la esperanza de vida o de la calidad de la misma...).

La probabilidad umbral será aquella en que la utilidad de tratar (a enfermos y no enfermos) sea igual a la de no tratar (también a enfermos y no enfermos). Debido a que la sensibilidad y especificidad de las pruebas no son del 100%, siempre puede haber falsos positivos y falsos negativos. Tratar a un falso positivo (es un individuo sin la enfermedad en estudio) conlleva unos riesgos (costes del tratamiento), pero dejar de tratar a un enfermo también, son unos beneficios sobre la salud que se dejan de obtener.

Ejemplo: se trata de decidir entre aplicar o no un tratamiento a los pacientes supuestamente afectados de la enfermedad E (Fig. 15). Los resultados sobre la salud se miden en forma de esperanza de vida. De las dos ramas principales del árbol, la primera es tratar a todos los pacientes, los que tengan la enfermedad (con una probabilidad p) y los que no (con una probabilidad $1-p$). La segunda rama del árbol es no tratar a los pacientes, independientes de si tienen o no la enfermedad. La esperanza de vida de los pacientes tratados es de 10 años. De tratar a los que no padecen la enfermedad, éstos viven en media 20 años. En caso de que a los enfermos no se les administre un tratamiento sólo vivirían 4 años, y la esperanza de vida de los no enfermos no tratados es de 21 años.



Como vemos, los pacientes con la enfermedad, de dejarlos a su curso natural vivirían sólo 4 años, y en caso de tratarlos se aumenta su espe-

ranza de vida a 10 años. Estos 6 años de diferencia son los beneficios del tratamiento. En cambio, los pacientes sin la enfermedad pasan de vivir 21 años a vivir 20 en caso de ser tratados. Este año de diferencia son los riesgos, los costes del tratamiento.

Si con la letra E se representa la enfermedad, que puede estar presente (E+) o ausente (E-), con las letras TT el tratamiento (con el signo positivo y negativo según se aplique o no) y con la letra U (de utilidad) los resultados sobre la salud (esperanza de vida, calidad de vida, estado funcional), los costes del tratamiento son la diferencia de la utilidad de los pacientes no enfermos no tratados y tratados, y los beneficios de un tratamiento son la diferencia de la utilidad de los pacientes enfermos tratados y no tratados:

$$\begin{aligned}\text{Coste} &= U(\text{E- TT-}) - U(\text{E- TT+}) = 21 - 20 = 1 \text{ año} \\ \text{Beneficio} &= U(\text{E+ TT+}) - U(\text{E+ TT-}) = 10 - 4 = 6 \text{ años}\end{aligned}$$

La probabilidad del umbral terapéutico será aquella en que la utilidad media de tratar sea igual a la de no tratar, es decir:

$$[p * U(\text{E+ TT+})] + [(1-p) * U(\text{E- TT+})] = [p * U(\text{E+ TT-})] + [(1-p) * U(\text{E- TT-})]$$

Si se despeja de la fórmula:

$$\begin{aligned}\frac{p}{1-p} &= \frac{U(\text{E- TT-}) - U(\text{E- TT+})}{U(\text{E+ TT+}) - U(\text{E+ TT-})} = \frac{\text{Coste}}{\text{Beneficio}} \\ p^* &= \frac{C}{C+B} \quad p = \frac{1}{1+6} = 0,1428\end{aligned}$$

En este ejemplo, si la probabilidad de tener la enfermedad es superior al 14%, es preferible tratar (los beneficios superan a los costes), y si es inferior, los riesgos son superiores a los beneficios, es preferible no tratar.

Si los riesgos del tratamiento fuesen iguales que los beneficios, la probabilidad umbral sería del 50%. Si los beneficios son superiores a los costes, basta una probabilidad inferior al 50% para decidir tratar, y si por el contrario los riesgos son elevados comparados con los beneficios, la probabilidad de tener la enfermedad debe ser superior al 50%, es preciso estar más seguro antes de decidir tratar al paciente. En un extremo, si un tratamiento no tuviese beneficios, sólo efectos secundarios, la probabilidad umbral sería de 1, nunca se tomaría la decisión de tratar, y en cambio, si no tuviese ningún efecto adverso, por mínima que fuese la probabilidad de tener la enfermedad, sería preferible tratar (la probabilidad umbral sería 0).

Umbral terapéutico y ratio beneficio/coste

La fórmula para el cálculo del umbral terapéutico también se puede expresar (tras dividir numerador y denominador por los costes) del siguiente modo:

$$p^* = \frac{1}{1 + B/C}$$

En la figura 16 se observa que el umbral terapéutico es sobre todo sensible a los cambios en la relación beneficio/coste cuando este ratio es muy pequeño. Pequeñas diferencias en el ratio a bajos ratios beneficio/coste (por ejemplo 0,5) tienen mucho más impacto en la decisión que las mismas diferencias cuando los ratios son altos (del orden de 10).

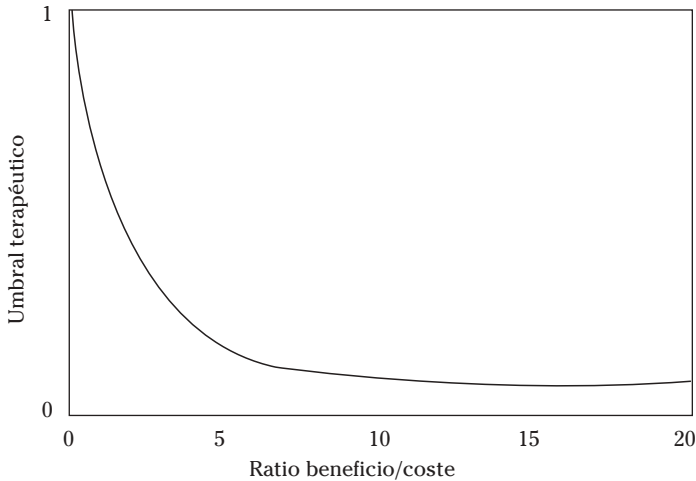


Figura 16. Umbral terapéutico y ratio beneficio/coste.

Cuando los beneficios son muy pequeños (en teoría fácilmente aumentables) o los costes muy altos (en teoría fácilmente disminuibles) el umbral terapéutico es muy sensible a dichos cambios (se puede disminuir). Cuando los beneficios son muy altos (ya no es fácil aumentarlos) o los costes muy pequeños (ya no es fácil disminuirlos) el ratio es elevado y el umbral terapéutico difícilmente modificable (en el sentido de aumentarlo). Y es que los beneficios y los costes tienen un límite, y es la diferencia entre la utilidad de los pacientes sin la enfermedad no tratados y los enfermos no tratados. Si por ejemplo los no enfermos no tratados viven 10 años, éste es el máximo coste del tratamiento (en el caso de que muriesen al día siguiente de iniciado el tratamiento); por otro lado, si los enfermos no tratados vivieran 2 años, el máximo beneficio que se puede esperar es llegar a vivir los 10 años de los no enfermos (8 años de beneficio).

UMBRAL DE DIAGNÓSTICO Y UMBRAL DE TRATAMIENTO

Como venimos comentando, si la probabilidad de la enfermedad es muy baja (menor que el umbral de diagnóstico p_i) no se debe aplicar una

prueba diagnóstica porque los resultados de ésta no van a modificar el curso de acción, se ha descartado la enfermedad. Lo mismo ocurre cuando la probabilidad de la enfermedad es muy elevada (por encima del umbral de tratamiento p_2), se ha diagnosticado la enfermedad y se pasa directamente al tratamiento.

Para el cálculo de p_1 se puede construir un árbol como el de la figura 17. Si no se trata al paciente (ni se le aplica una prueba), éste puede estar enfermo con una probabilidad p_1 y dar unos resultados sobre la salud que serían $U(E+TT-)$, o puede no estar enfermo con una probabilidad $1-p_1$, en cuyo caso los resultados sobre la salud serían $U(E-TT-)$.

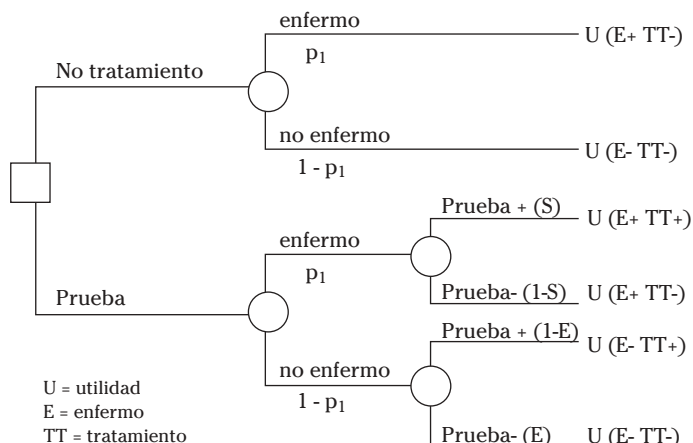


Figura 17. Cálculo de p_1 . Umbral de diagnóstico.

Si el paciente está enfermo con una probabilidad p_1 y se le aplica una prueba diagnóstica, ésta puede resultar positiva (sensibilidad de la prueba, representada por una S) en cuyo caso se trata al paciente y los resultados sobre la salud serían $U(E+TT+)$, o la prueba puede resultar negativa ($1-S$), no se trata al paciente y los resultados sobre la salud serían $U(E+TT-)$. Si el paciente no presenta la enfermedad con una probabilidad $1-p_1$ y se le aplica una prueba diagnóstica, ésta puede resultar positiva (es el complementario de la especificidad, representado por $1-E$), en cuyo caso se trata al paciente y los resultados sobre la salud serían $U(E-TT+)$. Si la prueba resulta negativa (especificidad = E), el paciente no se trata y los resultados sobre la salud son $U(E-TT-)$.

El cálculo de p_1 se realiza igualando los resultados sobre la salud de ambas ramas principales del árbol de decisiones, cuando se es indiferente entre no tratar y no aplicar la prueba a cuando se decide aplicar la prueba y tratar en consecuencia.

$$U(TT-) = p_1 * U(E+ TT-) + (1-p_1) * U(E- TT-)$$

$$U(\text{PRUEBA})= p_1 * S * U(E + TT+) + p_1 * (1-S) * U(E + TT-) + (1-p_1) * (1-E) * U(E- TT+) + (1-p_1) * E * U(E- TT-) + [U(\text{prueba})].$$

$$p_1 * U(E+ TT-)+(1-p_1) * U(E- TT-) = p_1 * S * U(E+ TT+)+p_1 * (1-S) * U(E+ TT-) + (1-p_1) * (1-E) * U(E- TT+) + (1-p_1) * E * U(E- TT-) + [U(\text{prueba})].$$

Tras despejar p_1 resulta:

$$p_1 = \frac{(1-E) * (C) - U(\text{prueba})}{(1-E) * (C) + (S * B)}$$

E = especificidad de la prueba

S = sensibilidad de la prueba

C = coste neto del tratamiento = $U(E-TT-) - U(E-TT+)$

B = beneficio neto del tratamiento = $U(E+TT+) - U(E+TT-)$

$U(\text{prueba})$ = utilidad de la prueba propiamente dicha

Para el cálculo de p_2 se realiza el mismo proceso pero igualando los resultados promedios sobre la salud de tratar directamente a todos los pacientes o primero aplicar una prueba diagnóstica y tratar según los resultados (Fig. 18).

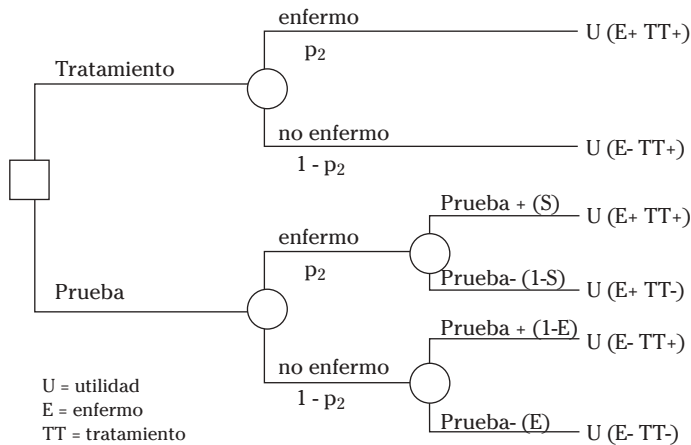


Figura 18. Cálculo de p_2 . Umbral de tratamiento.

$$U(TT+) = p_2 * U(E+ TT+) + (1-p_2) * U(E- TT+)$$

$$U(\text{PRUEBA})= p_2 * S * U(E+ TT+) + p_2 * (1-S) * U(E+ TT-) + (1-p_2) * (1-E) * U(E- TT+) + (1-p_2) * E * U(E- TT-) + [U(\text{prueba})].$$

$$p_2 * U(E+ TT+) + (1-p_2) * U(E- TT+) = p_2 * S * U(E+ TT+) + p_2 * (1-S) * U(E+ TT-) + (1-p_2) * (1-E) * U(E- TT+) + (1-p_2) * E * U(E- TT-) + [U(prueba)].$$

Despejando p_2 de la igualdad resulta:

$$p_2 = \frac{E * (C) + U(prueba)}{E * (C) + (1-S) * B}$$

Analizando la ecuación para la determinación del umbral de diagnóstico (p_1), se observa que cuanto mayores son los costes del tratamiento, mayor es el umbral de diagnóstico, no hacen falta probabilidades muy bajas para descartar la enfermedad, ya que los riesgos de tratarla son muy elevados. Sin embargo, cuanto mayores son los beneficios asociados a los verdaderos positivos menor es el umbral de diagnóstico, más baja tiene que ser la probabilidad de la enfermedad para que sea descartada y no se beneficien los enfermos del tratamiento correcto.

En la fórmula del umbral de tratamiento (p_2), se observa que cuanto mayor es el coste del tratamiento (efectos adversos de tratar a los individuos sin la enfermedad) más seguros debemos estar de que el paciente está realmente enfermo para tratar directamente, mayor es la probabilidad umbral. En cambio, cuanto mayores son los beneficios del tratamiento, menor es la probabilidad umbral, menos seguridad hace falta a la hora de decidirse a tratar directamente.

Respecto a la validez de las pruebas y en términos generales, conforme aumentan la sensibilidad y especificidad de una prueba aumenta el rango de probabilidades de aplicación de dicha prueba, disminuye p_1 y aumenta p_2 (Fig. 19). Antes de descartar o diagnosticar una enfermedad, aplicamos una prueba válida que discrimine bien a los verdaderos de los falsos (positivos y negativos).

Los umbrales de diagnóstico y de tratamiento toman un valor que es propio para cada conjunto de enfermedad, prueba diagnóstica y tratamiento.

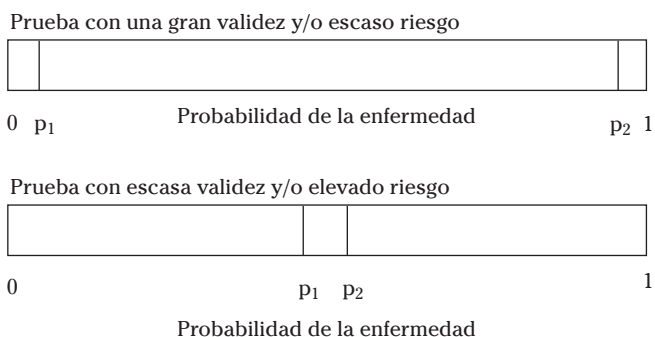


Figura 19. Umbrales de una prueba diagnóstica según la validez de la prueba.

Utilidad de la prueba

En cuanto al término $U(\text{prueba})$, se refiere a la utilidad de la prueba diagnóstica en sí misma. Depende del coste monetario de las pruebas, los efectos indeseables que pueda conllevar (sangrado, sensación desagradable, dolor, estrés psicológico), el sentimiento del paciente respecto a lo beneficioso que pueda resultar realizar dicha prueba (necesidad de que le confirmen un diagnóstico, además de cierto efecto placebo por el hecho de sentirse “más cuidado”) y de la actitud del médico ante lo que se denomina medicina defensiva (puede ordenar una prueba aunque la probabilidad de un hallazgo anormal sea prácticamente nula).

La utilidad de la prueba puede tener un signo negativo o un signo positivo, según dominen o no los efectos indeseables. Así se observa que para la estimación de p_1 , a mayor utilidad de la prueba menor es la probabilidad de la enfermedad a la que se decide realizar la prueba. Al contrario ocurre para la estimación de p_2 .

Si la utilidad de la prueba es nula, ni positiva ni negativa, se compensan los efectos indeseables con los beneficiosos, se puede comprobar que si la probabilidad previa de la enfermedad coincide con el umbral de diagnóstico (p_1), la probabilidad posterior tras un resultado positivo coincide con el umbral terapéutico (p^*). Del mismo modo, si la probabilidad previa es la misma que el umbral de tratamiento (p_2), la probabilidad post-prueba tras un resultado negativo es también el umbral terapéutico. (Fig. 20).

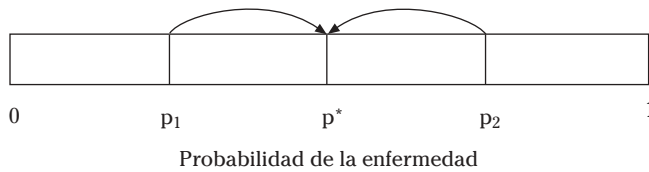


Figura 20. Cambio de la probabilidad previa a la posterior. Utilidad nula de la prueba.

Pero ahora imaginemos que la prueba tiene una utilidad positiva, dominan los efectos beneficiosos. Desplazará p_1 y p_2 hacia los extremos, con lo que si la probabilidad previa coincide con p_1 o p_2 , la probabilidad posterior (tras un resultado positivo y negativo respectivamente) “ya no llega” a coincidir con el umbral terapéutico (Fig. 21).

Si la utilidad de la prueba es negativa, dominan los efectos adversos, desplaza los umbrales p_1 y p_2 hacia el centro, con lo que con probabilidades previas inferiores a p_1 o superiores a p_2 , tras un resultado positivo o negativo respectivamente, se consigue “superar” el umbral terapéutico (Fig. 22).

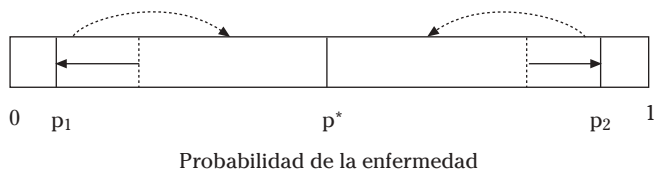


Figura 21. Cambio de la probabilidad previa a la posterior. Utilidad positiva de la prueba.

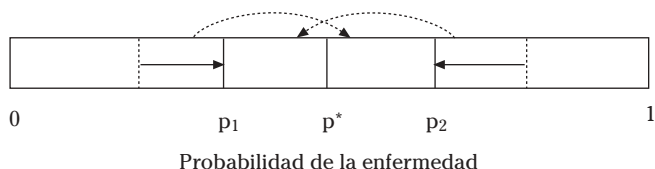


Figura 22. Cambio de la probabilidad previa a la posterior. Utilidad negativa de la prueba.

Una de las formas de estimar la utilidad de la prueba es, tras explicar todas las consecuencias de la misma, someter al paciente al juego de la lotería estándar (Fig. 23). Si la utilidad de la prueba es negativa, se trata de elegir entre dos alternativas, la de certeza, presentar la enfermedad y realizarse la prueba (alternativa de utilidad desconocida) o jugársela a entrar en quirófano por ejemplo, sabiendo que o se muere con una probabilidad p (y utilidad igual a 0) o sobrevive (y se ha ahorrado la prueba diagnóstica) con una probabilidad $1-p$ (y utilidad igual a 1). Se varía la probabilidad p hasta que el paciente está indiferente entre realizarse la prueba o jugársela. De ahí que el método se llame “lotería”.

Si el paciente está, por ejemplo, indiferente entre realizarse la prueba o no cuando la probabilidad de morir en quirófano es 0,005, la utilidad de la alternativa de certeza sería:

$$U? = 0,005 \cdot 0 + 0,995 \cdot 1 = 0,995$$

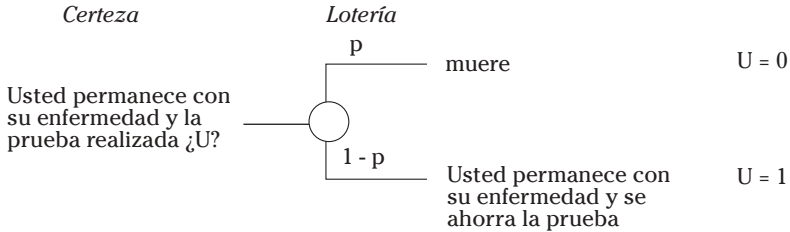
Realizarse la prueba tiene una utilidad de 0,995, pierde 0,005 respecto a la unidad, que es la utilidad negativa de la prueba.

Si la utilidad de la prueba es positiva, ¿cuál es la utilidad de no realizársela? Aquí la alternativa de certeza es permanecer con la enfermedad y sin realizarse la prueba. En la lotería se la juega a morir en quirófano (con una probabilidad p y utilidad 0) o vivir (con una probabilidad $1-p$ y utilidad 1) y realizarse la prueba. Imagine que contesta que acepta el reto cuando la probabilidad de morir es 0,002:

$$U? = 0,002 \cdot 0 + 0,998 \cdot 1 = 0,998$$

La utilidad de la prueba es 0,002 y positiva.

Utilidad negativa de la prueba



Utilidad positiva de la prueba

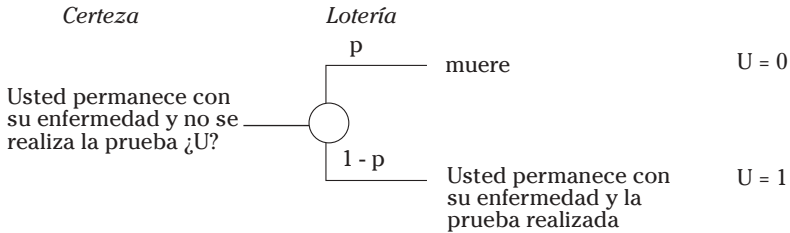


Figura 23. Utilidad de la prueba diagnóstica.

DIAGNÓSTICO PRECOZ

Hasta ahora, el proceso diagnóstico se ha enfocado casi exclusivamente sobre los pacientes sintomáticos. Sin embargo, existe un tema clave en la medicina preventiva, y es el diagnóstico precoz, el diagnóstico de la enfermedad en los pacientes asintomáticos.

Básicamente, los criterios que ayudan a decidir cuándo realizar una prueba de cribaje son los mismos que cuando se decide aplicar una prueba a un paciente con síntomas: la probabilidad de tener la enfermedad debe ser mayor que el umbral de diagnóstico p_1 y menor que el umbral de tratamiento p_2 . Estos umbrales dependen a su vez de la sensibilidad y especificidad de la prueba diagnóstica aplicada, de la aceptabilidad (utilidad) de la prueba y de los beneficios y riesgos que conlleva el tratamiento en esta fase precoz.

Sin embargo, se dedica un apartado especial a este tema por la gran confusión que hay al respecto y la exagerada buena fama que ha ganado esta parcela de la medicina; por puro sentido común parece que el diagnóstico precoz siempre tiene que ser bueno, cuando en realidad, a veces, conlleva más riesgos que beneficios.

Las principales estrategias del diagnóstico precoz son dos: el cribaje (se invita al público a someterse a una serie de pruebas) y la búsqueda de casos (aprovechando la visita al médico por otras causas se le aplican determinadas pruebas para el diagnóstico de otras enfermedades no sospechadas por la clínica). La búsqueda de casos capta una proporción

mayor de la población que el cribaje, ya que éste sólo se realiza esporádicamente y rara vez capta más del 10% de los posibles candidatos.

Independientemente del método, para valorar la factibilidad y efectividad del diagnóstico precoz es imprescindible entender el curso natural de la enfermedad, que se puede dividir en cuatro estados:

- a) Comienzo biológico: es el momento en el que interactúa el factor de riesgo con el individuo.
- b) Diagnóstico precoz posible: la enfermedad ya ha producido cambios estructurales (aunque el individuo no manifieste ningún síntoma) que pueden ser detectados con la prueba apropiada.
- c) Diagnóstico clínico habitual: la enfermedad ha progresado hasta que aparecen los síntomas y el individuo afecto, entonces enfermo, acude al médico que puede diagnosticarle.
- d) Resultado: la enfermedad sigue su curso, acabando en curación, incapacidad o la muerte.

Esta evolución ordenada de la enfermedad, de forma aislada, no hace posible el diagnóstico precoz, es necesario que exista también un “punto crítico” en la historia natural de la enfermedad, antes del cual el tratamiento es o más eficaz o más fácil de aplicar, ya que el objetivo último del diagnóstico precoz es la mejora de la salud de los individuos a los que se aplica.

En la figura 24 se observa que si una enfermedad tuviese un único punto crítico, entre el comienzo biológico y el tiempo en el que el diagnóstico precoz sería posible, el cribaje o la búsqueda de casos siempre llegaría tarde, ya habría pasado el punto crítico. Si el único punto crítico estuviese tras el diagnóstico clínico habitual, no se adelantaría nada, ya no se trataría de un diagnóstico “precoz”. Sólo cuando la enfermedad tiene un punto crítico entre el diagnóstico precoz posible y el diagnóstico clínico habitual, es cuando o el cribaje o la búsqueda de casos pueden dar lugar a una mejora de los resultados sobre la salud.

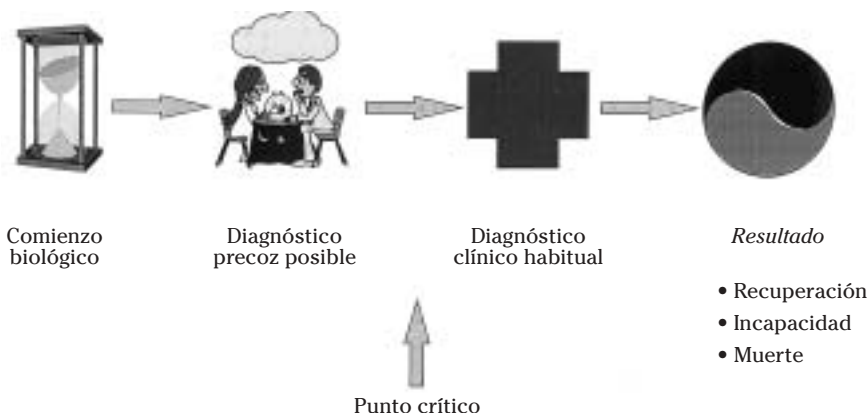


Figura 24. Historia natural de la enfermedad.

Esta mejora respecto al diagnóstico clínico habitual (algo que parece que siempre tiene que ocurrir) es indispensable para aplicar la prueba de diagnóstico precoz, sin embargo su estimación puede verse afectada por tres sesgos fundamentalmente:

- a) *Sesgo del voluntario sano*. Los pacientes que voluntariamente acuden a un programa de cribaje son generalmente más sanos que los que no acuden. Así, aparentemente, la supervivencia, por ejemplo, será mayor, pero no por el cribaje en sí mismo, sino por su condición previa.
- b) *Sesgo de adelanto diagnóstico* (Fig. 25). Imaginemos que los pacientes con un determinado cáncer son diagnosticados habitualmente (por su sintomatología) a los 50 años y tras el tratamiento su esperanza media de vida es de 10 años. Mediante un programa de cribaje son diagnosticados a los 40 años, y la esperanza de vida media desde el diagnóstico es de 20 años. Si sólo se presentan los datos de supervivencia tras el diagnóstico, aparentemente el diagnóstico precoz alarga la vida 10 años más que el diagnóstico habitual, sin embargo, ambos grupos acabarán muriendo en media a los 60 años. Se ha movido el tiempo cero a partir del cual se cuenta la supervivencia, “no se ha alargado la vida hacia adelante, sino que se ha alargado el tiempo de enfermedad hacia atrás”, empeorando incluso la calidad de vida de aquellos pacientes que fueron etiquetados de enfermos, operados y tratados 10 años antes de lo que hubiese sido lo habitual.
- c) *Sesgo de larga duración*. Los pacientes en los que la duración preclínica de la enfermedad es larga, también tienden a tener una duración clínica larga, y los pacientes en los que la duración preclínica es corta, también tienden a tener una duración clínica corta y un curso fatal. Al primer tipo de pacientes, por su larga duración, es más fácil identificarlos en alguna estrategia de diagnóstico precoz que aquellos en los que la duración preclínica es más corta, viviendo más aquellos que éstos.

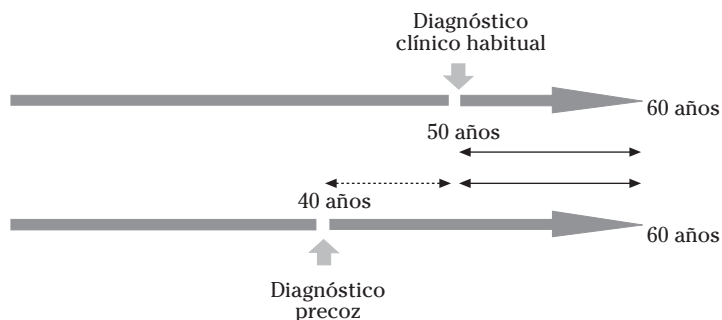


Figura 25. Sesgo de adelanto diagnóstico.

Da la sensación de que el diagnóstico precoz, lo peor que puede provocar es la no mejora de la salud, sin embargo, no hay que olvidar que incluso puede provocar daño, que el diagnóstico precoz también conlleva sus riesgos, de los que los más importantes son los siguientes:

- a) En primer lugar el diagnóstico puede estar equivocado. La probabilidad de un falso positivo puede ser muy elevada, sobre todo cuando se trata de enfermedades de muy baja prevalencia. Ejemplo: la hemocromatosis es una enfermedad hereditaria (autosómica recesiva) cuya prevalencia oscila entre 4 por 1.000 (en las zonas donde la endogamia es importante) hasta 3 por cada 10.000 personas en la población general (*Wiggers P, Dalhoj J, Kier H, Ring-Larsen H, Hyltoft Petersen P, Balabjerg O, Horder M. Screening for haemochromatosis: prevalence among Danish blood donors. J Int Med 1991; 230: 265-270*) El diagnóstico definitivo de la enfermedad se establece por biopsia hepática, una prueba cruenta que conlleva sus riesgos. Se puede realizar el diagnóstico precoz mediante la determinación de la ferritina sérica y la saturación de la transferrina, pero aunque estas pruebas conjuntamente tuviesen una sensibilidad del 100% y una especificidad del 99%, dada la baja prevalencia (3 por 10.000), el valor predictivo positivo sería del 3%, lo que equivale a decir que por cada 3 individuos con hemocromatosis que se detectan, se está realizando la biopsia hepática a 97 individuos sin hemocromatosis.
- b) En segundo lugar, el tratamiento puede conllevar más riesgos que beneficios. Algunos estudios sobre el cribaje del cáncer de próstata mediante el antígeno específico de próstata (PSA) revelan que tras el diagnóstico precoz se aumenta la esperanza de vida, pero cuando los efectos sobre la salud se ajustan por calidad de vida, en realidad disminuyen (*Drahn MD, Mahoney JE, Eckman MH, Trachtenberg J, Pauker SG, Detsky AS. Screening for Prostate*

Cancer: a decision analytic view. JAMA 1994;272:773-780). Ello se debe a que el tratamiento en fase precoz no es mucho más efectivo que cuando el diagnóstico se realiza por la sintomatología, y además la cirugía conlleva sus riesgos (impotencia, incontinencia, mortalidad, etc.).

- c) El tercer tipo de daño es el del etiquetado de los pacientes, el efecto placebo también puede tener un sentido negativo, aunque es difícil cuantificar este “trauma psicológico”. Se ha observado cómo aumenta el absentismo laboral, por ejemplo, al comunicarle a un paciente que es hipertenso, aunque siga igual de asintomático que antes de recibir esta información.

Por todo ello, antes de aplicar una prueba de cribaje (o búsqueda de casos) es preciso asegurarse de que se produce una mejora real en la salud de los individuos, medida en cantidad y calidad de vida, donde también se debe contemplar la aceptabilidad de la prueba y los recursos necesarios para llevar a cabo el programa.

RESUMEN

Las pruebas diagnósticas se aplican con el fin de aumentar la información sobre la probabilidad de que el paciente tenga o no una determinada enfermedad. El nivel de certidumbre que se precisa depende de las consecuencias de un error diagnóstico y por consiguiente error en el tratamiento.

Sólo se debe aplicar una prueba diagnóstica si la información que aporta puede modificar la actitud ante el paciente. De hecho, se ha realizado un diagnóstico cuando el aporte de más información no va a cambiar la decisión de actuar como si el paciente tuviese la enfermedad (tratar), y del mismo modo, se descarta una enfermedad cuando el aporte de más información no va a cambiar la decisión de actuar como si el paciente no tuviese la enfermedad (no tratar).

El umbral terapéutico es la probabilidad de la enfermedad por encima de la cual es preferible tratar y por debajo es preferible no tratar.

$$p^* = \frac{C}{C + B}$$

Se denomina umbral de diagnóstico (p_1) a la probabilidad de la enfermedad por debajo de la cual no es preciso aplicar una prueba diagnóstica (ni por supuesto tratar), porque aún resultando la prueba positiva, la probabilidad de la enfermedad no supera el umbral terapéutico. Se denomina umbral de tratamiento (p_2), la probabilidad de la enfermedad por encima de la cual no se realiza ninguna prueba diagnóstica más y se pasa directamente al tratamiento, porque aunque la prueba resultase negativa, la probabilidad de la enfermedad todavía sería superior al umbral terapéutico.

Estos dos umbrales toman un valor propio para cada conjunto de enfermedad, prueba diagnóstica y tratamiento, y dependen de:

- sensibilidad y especificidad de la prueba diagnóstica.
- riesgos (costes) y beneficios que conlleva el tratamiento
- utilidad de la prueba en sí misma.

$$p_1 = \frac{(1-E) \cdot (C) - U(\text{prueba})}{(1-E) \cdot (C) + (S \cdot B)}$$

$$p_2 = \frac{E \cdot (C) + U(\text{prueba})}{E \cdot (C) + (1-S) \cdot B}$$

E = especificidad de la prueba

S = sensibilidad de la prueba

C = coste neto del tratamiento. Es el empeoramiento en la salud de los individuos no enfermos por el hecho de ser tratados.

B = beneficio neto del tratamiento. Mejora en la salud de los individuos enfermos por el hecho de ser tratados.

U(prueba) = utilidad de la prueba.

El diagnóstico precoz, además de someterse a las mismas leyes que el diagnóstico habitual en presencia de síntomas (superar los umbrales de las pruebas diagnósticas para aplicar la prueba), precisa de una condición básica: la existencia de un punto crítico en la historia natural de la enfermedad antes del cual el tratamiento es o más eficaz o más fácil de aplicar.

PROBLEMAS

1. Se trata de un varón de 65 años que presumiblemente presenta cudicia. La cudicia es una enfermedad metabólica degenerativa que afecta por igual a ambos sexos y su prevalencia aumenta con la edad. De no tratarla, causa la muerte por afectación hepática, digestiva o renal. En este caso se ha presentado de forma violenta con dolor abdominal y diarrea. En el enema opaco se observan inflamación y úlceras en la mucosa, especialmente en el colon izquierdo, con lo que la probabilidad de presentar la enfermedad es del 15%. Dadas las características que se exponen a continuación ¿Se tiene la información suficiente como para decidir el tratamiento o es preferible realizar antes una colonoscopia?

Sensibilidad de la colonoscopia = 0,90

Especificidad de la colonoscopia = 0,95

Beneficios del tratamiento = supervivencia del 70%

Riesgos del tratamiento inapropiado = mortalidad del 2%

Riesgos de la colonoscopia = mortalidad del 0,001

2. Acude a la consulta un paciente con cefalea y hemiplejía progresiva. La probabilidad preprueba de tumor cerebral es 0,05. El umbral terapéutico (el tratamiento es la intervención quirúrgica) es de 0,15. La prueba diagnóstica que no se sabe si aplicar o no es la resonancia nuclear magnética (RNM), cuya sensibilidad y especificidad son del 95% y se le atribuye una utilidad intrínseca igual a cero. ¿Qué es lo correcto, decidir el tratamiento directamente o realizar primero la RNM?

RESULTADOS SOBRE LA SALUD

INTRODUCCIÓN

En prácticamente todas las decisiones médicas que se realizan el objetivo final es mejorar la salud de los individuos, un concepto bastante difícil de medir, aunque en este capítulo se explican algunas aproximaciones.

Supongamos que al comparar dos tratamientos observamos que con el tratamiento A mueren sólo el 10% de los individuos afectados de la enfermedad, mientras que con el tratamiento B mueren el 12%, y sin tratamiento morirían prácticamente todos. ¿Qué tratamiento sería el de elección? Lógicamente el primero, el A, y en este caso la medida de los resultados sobre la salud para decidir la alternativa terapéutica ha sido “vidas salvadas” (con el tratamiento A se salvan un 2% de vidas más que con el tratamiento B).

Complicamos más el ejemplo y suponemos que el 10% que se mueren con el primer tratamiento lo hacen instantáneamente, en quirófano, mientras que con el segundo tratamiento se mueren el 12% pero a los 5 años. En ambos casos, los individuos que no mueren viven 20 años más. ¿Sigue siendo mejor el primer tratamiento? La medida a utilizar más correcta en este caso concreto sería “años de vida ganados, esperanza de vida”. Se puede realizar el ejercicio y comprobar que en media, tras el tratamiento A los pacientes viven 18 años ($0,90 \cdot 20 + 0,10 \cdot 0$), y tras el tratamiento B 18,2 años ($0,88 \cdot 20 + 0,12 \cdot 5$). Dados estos resultados, la alternativa de elección es la B.

Ahora supongamos que con el tratamiento A los pacientes que sobreviven lo hacen en perfecto estado de salud, mientras que con el segundo tratamiento les queda un dolor residual ¿Elegiríamos sin duda alguna el segundo tratamiento porque en media viven más tiempo? No, porque tener buena salud es tener una gran expectativa de vida saludable. En este caso la mejor medida de los resultados sobre la salud serían los años vida ajustados por calidad (AVAC o QALY en la terminología inglesa). Al realizar el ajuste por calidad de vida se enlaza directamente con el concepto de utilidad o preferencia que muestran los individuos por un determinado estado de salud. Los sentimientos sobre las experiencias de la enfermedad son

muy subjetivos y personales, y puede resultar presuntuoso intentar medirlos, pero puesto que años de vida en diferentes estados de salud no son estrictamente comparables, se han desarrollado diferentes métodos en un intento de aproximación a dicha medida.

VIDAS SALVADAS, CASOS EVITADOS

Tanto para cuantificar el número de vidas salvadas como el de casos evitados (enfermedades prevenidas por una vacuna por ejemplo) es preciso el conocimiento de dos conceptos claves de la epidemiología: la incidencia y la prevalencia.

Incidencia y prevalencia

La incidencia de una enfermedad es la proporción de individuos que a lo largo de un período de tiempo desarrolla dicha enfermedad.

$$I = \frac{\text{Casos nuevos en un período de tiempo}}{\text{Población en riesgo}}$$

Es muy importante entender que en el numerador se trata de casos nuevos, que no presentaban la enfermedad al comienzo del período de seguimiento, y el denominador es la población en riesgo. Así, en las enfermedades crónicas, en el estudio de la incidencia se deben excluir del denominador los individuos que al comienzo del estudio ya tenían la enfermedad. Por ejemplo, si queremos estudiar la incidencia de SIDA a lo largo de un año, el numerador estará constituido por los pacientes que a lo largo del año desarrollen la enfermedad (casos nuevos) y del denominador no deben formar parte aquéllos que al inicio del estudio ya tenían SIDA, puesto que no están en riesgo.

La mortalidad no es más que una medida especial de incidencia en la que el numerador es el número de muertos durante un periodo determinado.

La prevalencia es la proporción de casos con la enfermedad en un momento determinado. A la prevalencia no le importa el tiempo que los individuos llevan con la enfermedad, es lo mismo si el paciente lleva enfermos 10 años que 10 días.

$$P = \frac{\text{Casos con la enfermedad en un momento determinado}}{\text{Población en ese momento dado}}$$

Es como si la incidencia fuese la película (durante un período de tiempo) y la prevalencia la foto (instantánea).

La prevalencia de una enfermedad depende de la incidencia, la duración y la letalidad. En general, conforme aumenta la incidencia, al aparecer más casos, mayor será la prevalencia en un momento dado. Conforme

mayor sea la duración de la enfermedad también será mayor la prevalencia, más fácil es captar a estos individuos en el momento de la foto, pero si se curan pronto o se mueren (letalidad) la prevalencia disminuirá.

Una medida preventiva que disminuye la exposición a los factores de riesgo de una enfermedad (prevención primaria) disminuye su incidencia y por consiguiente su prevalencia. Una medida preventiva que detecta precozmente la enfermedad y mejora su supervivencia (prevención secundaria), no modifica la incidencia pero, de no curar a los pacientes, aumentará la prevalencia. Una medida preventiva rehabilitadora (prevención terciaria) también puede aumentar la duración de la enfermedad y por tanto su prevalencia.

Los estudios de incidencia, que por lógica precisan de un seguimiento, son más caros que los de prevalencia, y por ello abundan menos en la literatura. Pero si se precisan datos de incidencia y éstos no están disponibles, se pueden estimar a partir de los datos de prevalencia por grupos de edad, creando pseudocohortes bajo la suposición de que la diferencia de prevalencias de un grupo de edad a otro superior es indicativa de la incidencia en ese período. Esta suposición sólo puede realizarse en el caso de enfermedades crónicas que no curan y no matan (por ejemplo el glaucoma). La estimación de la incidencia se realiza del siguiente modo:

$$I = (P_{j+1} - P_j) / (1 - P_j)$$

I = incidencia

P_{j+1} = es la prevalencia en el grupo de edad superior

P_j = es la prevalencia en el grupo de edad inferior

Supongamos que la prevalencia de glaucoma a los 50 años es del 10% y a los 60 años del 15%. El 5% (numerador) de los individuos ha adquirido el glaucoma en los 10 años, pero en riesgo de adquirirlo sólo estaban los que a los 50 años no lo tenían (denominador). La incidencia anual es por lo tanto:

$$(0,15 - 0,10) / (1 - 0,10) = 5,5\% \text{ en } 10 \text{ años} = 0,55\% \text{ por año}$$

Medida de los resultados sobre la salud

La terminología utilizada para referirse a las medidas de asociación y más concretamente a las medidas de impacto, puede inducir a cierta confusión, ya que hay muchos términos relacionados con un nombre algo diferente: riesgo atribuible, proporción de riesgo atribuible, fracción atribuible, fracción etiológica en expuestos, riesgo atribuible en la población, proporción de riesgo atribuible en la población, fracción etiológica en la población. En cualquier caso, lo importante es entender el concepto en sí y sobre todo saber interpretarlo.

Las medidas que vamos a ver a continuación, reflejan el cambio en la incidencia de la enfermedad por el hecho de estar expuesto o no a un factor.

En relación con una enfermedad, un factor de exposición y una población determinados, el riesgo relativo (RR) o razón de riesgos indica la probabilidad de desarrollar la enfermedad en el grupo expuesto en relación al grupo no expuesto y su fórmula es la siguiente:

$$RR = I_1 / I_0$$

I_1 = incidencia en los expuestos

I_0 = incidencia en los no expuestos

Ejemplo: si la incidencia de cáncer de pulmón entre los fumadores es 160/100.000 y en los no fumadores es 10/100.000, el riesgo relativo es 16. El riesgo de desarrollar cáncer de pulmón es 16 veces mayor en los fumadores que en los no fumadores.

Un riesgo relativo igual a 1 indica que la incidencia de la enfermedad en los expuestos y no expuestos es idéntica, no se asocia la exposición a la enfermedad. Si es superior a 1, se trata de un factor de riesgo de la enfermedad, y cuanto más alejado de 1 mayor es el incremento del riesgo. Y por el contrario, si es inferior a 1 indica un efecto protector de la exposición, la incidencia en los expuestos es menor que en los no expuestos (cuanto más cercano a cero más protector).

El riesgo atribuible (RA) es el exceso de riesgo de enfermar en los expuestos comparados con los no expuestos. Nos da información sobre el efecto absoluto de la exposición. Es la diferencia de incidencias del grupo expuesto y no expuesto.

$$RA = I_1 - I_0$$

Siguiendo con el ejemplo, el riesgo de cáncer de pulmón atribuible al tabaco es del 150/100.000 (160/100.000 – 10/100.000).

La proporción de enfermos en el grupo expuesto que pueden atribuir su enfermedad al hecho de haber estado expuestos al factor de interés, y que por tanto, podría prevenirse si desapareciera la exposición, es la proporción de riesgo atribuible (RA%) o fracción atribuible o fracción etiológica en los expuestos.

La fórmula es la siguiente:

$$RA \% = \frac{I_1 - I_0}{I_1} \times 100$$

Con los datos del ejemplo, en el grupo de fumadores, el 93,75% de los cánceres de pulmón son atribuibles al tabaco.

$$RA \% = \frac{160 - 10}{160} \times 100 = 93,75\%$$

Pero también es importante conocer la proporción de casos de enfermedad en la población que es atribuible a dicha exposición. Esta medida es el riesgo atribuible en la población (RAP) y es útil para determinar qué exposiciones son las de mayor importancia para la salud de una población o comunidad. Se calcula como la incidencia total en la población menos la incidencia en los no expuestos y representa la tasa o cantidad de enfermedad que podría eliminarse en toda la población si la exposición desapareciera.

$$RAP = I_r - I_0$$

I_r = incidencia total en la población

I_0 = incidencia en los no expuestos

También puede calcularse multiplicando el riesgo atribuible por la proporción de individuos expuestos en la población.

$$RAP = RA \times P_e \quad (P_e \text{ es la proporción de expuestos en la población})$$

Imaginemos que en el ejemplo del cáncer de pulmón la incidencia total en la población es 70/100.000 y la proporción de fumadores es del 40%. En la población, el riesgo de cáncer de pulmón atribuible al tabaco es 60/100.000 (70/100.000 – 10/100.000), resultado que coincide si se aplica la segunda fórmula $150/100.000 \times 40\% = 60/100.000$.

También se puede calcular la proporción atribuible en la población (RAP%) o fracción etiológica en la población, para expresar la proporción de la enfermedad bajo estudio que es atribuible a la exposición y que, por tanto, podría evitarse si tal exposición se eliminara.

$$RAP\% = \frac{I_r - I_0}{I_r} \times 100$$

RAP% = proporción de riesgo atribuible en la población

I_r = incidencia total en la población

I_0 = incidencia en los no expuestos

$$RAP\% = \frac{70 - 10}{70} = 85,7\%$$

La incidencia total en la población es la media de la incidencia en los expuestos y la incidencia en los no expuestos, ponderadas por la proporción de individuos expuestos y no expuestos respectivamente.

$$I_r = P_e I_1 + (1 - P_e) I_0$$

P_e = proporción de expuestos en la población

I_1 = incidencia en los expuestos

I_0 = incidencia en los no expuestos

Si se sustituye en la fórmula de la proporción atribuible en la población resulta:

$$RAP\% = \frac{P_e I_1 + (1-P_e)I_0 - I_0}{P_e I_1 + (1-P_e)I_0} \times 100$$

Aplicamos los datos del ejemplo:

$$I_1 = 160/100.000 \quad I_0 = 10/100.000 \quad P_e = 40\%$$

$$I_r = (160/100.000) \cdot 0,4 + (10/100.000) \cdot 0,6 = 70/100.000$$

$$RAP\% = (70-10) / 70 = 0,857$$

El 85,7% de los cánceres de pulmón de la población son debidos al tabaco, o lo que es lo mismo, si se eliminase completamente el tabaco, se reduciría la incidencia de cáncer de pulmón en un 85,7%.

Si en lugar de eliminar completamente el hábito tabáquico, sólo se lograra reducir en un 50%, ¿en qué porcentaje se reduciría el cáncer de pulmón?

Reducir el tabaquismo en un 50% equivale a decir que el hábito tabáquico sería del 20%. Pasaría de la incidencia total cuando el 40% de la población fuma ($I_r=70/100.000$) a la incidencia en caso de fumar el 20%.

$$I_{20\%} = (160/100.000) \cdot 0,2 + (10/100.000) \cdot 0,8 = 40/100.000$$
$$(70-40)/70 = 0,429$$

Se reduciría la proporción de cánceres de pulmón en un 42,9%.

Si una enfermedad es prevenible mediante la exposición a un factor (que en este caso es protector, ya no es de riesgo), por ejemplo la gripe mediante la vacunación, la incidencia de enfermedad en los expuestos es menor que en los no expuestos, el efecto absoluto es negativo. Aunque conceptualmente es lo mismo, los nombres que reciben las medidas de impacto en estos casos son diferentes.

El riesgo relativo es menor de 1. El equivalente al riesgo atribuible es la reducción absoluta del riesgo (RAR), la diferencia del riesgo de la enfermedad (incidencia) en los no expuestos y los expuestos, y representa el riesgo de la enfermedad que es atribuible a la no exposición.

$$RAR = I_0 - I_1$$

Ejemplo: La incidencia de la enfermedad E en los individuos no vacunados es del 18% y en los vacunados (expuestos) es del 1,8%.

El riesgo de la enfermedad E atribuible al hecho de no vacunarse es del 16,2% (18%-1,8%).

La fracción prevenida (FP) o reducción relativa del riesgo (RRR) es el porcentaje en el que se reduce el riesgo de la enfermedad debido a la exposición.

$$FP = \frac{I_0 - I_1}{I_0} \times 100$$

$$FP = \frac{18 - 1,8}{18} \times 100 = 90\%$$

La incidencia de la enfermedad E ha disminuido en un 90% gracias a la vacunación. Este resultado es la eficacia de la vacuna.

La fracción prevenible en la población (FP_T) es la proporción de la incidencia total atribuible a la no exposición.

$$FP_T = \frac{I_r - I_1}{I_r} \times 100$$

La incidencia total es la media de la incidencia en los expuestos y la incidencia en los no expuestos, ponderadas por la proporción de expuestos y no expuestos respectivamente.

$$I_r = I_0(1-P_e) + I_1P_e$$

Si se sustituye en la fórmula resulta:

$$FP_T = \frac{I_0(1-P_e) + I_1P_e - I_1}{I_0(1-P_e) + I_1P_e} \times 100$$

Cuando nadie en la población ha sido vacunado previamente, la incidencia total coincide con la incidencia en los no vacunados, y la fracción prevenible coincide con la fracción prevenible en la población.

Si sólo se vacuna al 70% de la población ¿en qué porcentaje se reduciría la incidencia de la enfermedad?

La incidencia si se vacunan el 70% de la población es la media de la incidencia en los vacunados y la incidencia en los no vacunados, ponderadas por el porcentaje de vacunados y no vacunados respectivamente.

$$I_{70\%} = 0,018 \times 0,70 + 0,18 \times 0,30 = 0,0666$$

luego

$$FP_T = \frac{0,18 - 0,0666}{0,18} \times 100 = 63\%$$

La incidencia de la enfermedad disminuiría en un 63% (pasa del 18% al 6,66%) al vacunar al 70% de la población. Este resultado también se obtiene al multiplicar la eficacia de la vacuna por el porcentaje de población vacunada ($0,9 * 0,7 = 0,63$).

Cuando se trata de factores protectores, la utilización de unos términos u otros está relacionado con el ámbito de estudio. Así, en salud pública se utiliza la fracción prevenible y la eficacia vacunal, por ejemplo, y en cambio, en los ensayos clínicos los resultados se expresan generalmente mediante la reducción absoluta del riesgo (RAR), reducción relativa del riesgo (RRR) o número que es necesario tratar (NNT), que como la misma palabra indica, es el número de pacientes que es necesario tratar para evitar un evento (de los que se pretenden evitar con el tratamiento experimental). Por una simple regla de 3, se deduce que el NNT es el inverso de la reducción absoluta del riesgo.

$$\text{NNT} = 1 / \text{RAR}$$

Ejemplo: en un ensayo clínico se prueba un tratamiento para la disminución del riesgo de neuropatía en los pacientes diabéticos. La incidencia de neuropatía en el grupo control (sin el tratamiento) es del 9,6%, mientras que en el grupo experimental (el que recibe el tratamiento) es del 2,8%.

$\text{RAR} = 9,6\% - 2,8\% = 6,8\%$ (El riesgo de neuropatía atribuible al hecho de no tratarse es del 6,8%).

$\text{NNT} = 1 / 0,068 = 14,7$ (Es preciso tratar 15 pacientes diabéticos para evitar una neuropatía).

De todas estas medidas ¿cuál es preferible, la reducción relativa del riesgo, la reducción absoluta del riesgo o el número que es necesario tratar?

Los datos de la tabla 19 representan la incidencia de neuropatía en el grupo control y en el experimental bajo tres supuestos.

Tabla 19. Incidencia de neuropatía bajo tres supuestos.

	Grupo control	Grupo experimental
A	9,6%	2,8%
B	96%	28%
C	0,096%	0,028%

En la tabla 20 se observa que si los resultados se expresan mediante la reducción relativa del riesgo (RRR), aparentemente los tres tratamientos son igualmente eficaces. Y, en cambio, cuando se utilizan la reducción absoluta del riesgo (RAR) o el número que es necesario tratar (NNT), se pone de manifiesto que la alternativa terapéutica más eficaz es la B, disminuye mucho el riesgo de neuropatía por el hecho de tratarse. Es más, por cada dos pacientes tratados (1,47 para ser más exactos) se evita una neuropatía.

Tabla 20. Reducción relativa del riesgo, reducción absoluta del riesgo y número que es necesario tratar. RRR, RAR y NNT.

	RRR	RAR	NNT
A	(9,6-2,8)/9,6=71%	9,6-2,8=6,8%	1/0,068=14,7
B	(96-28)/96=71%	96-28=68%	1/0,68=1,47
C	(0,096-0,028)/0,096=71%	0,096-0,028=0,068%	1/0,00068=1470

No es lo mismo si el riesgo inicial del que se parte es elevado (y entonces relativamente fácil de disminuir) que si ya el riesgo basal es muy bajo, donde cualquier disminución en términos absolutos va a ser muy pequeña.

ESPERANZA DE VIDA

Muchas veces lo importante no es si son más pacientes los que viven o no, sino en media cuánto tiempo viven. La esperanza de vida es el promedio de años de vida de una persona a una determinada edad, y se habla de promedio porque no es un número concreto, sino que presenta una distribución.

DEALE

Generalmente existen tres formas de presentar los datos de supervivencia de los individuos:

- Fracción de supervivencia. Es la proporción de la cohorte original que permanece viva en un momento determinado. Por ejemplo, el 50% sobreviven a los 5 años.
- Tasa media anual de mortalidad: por ejemplo, el 14% mueren cada año.
- Esperanza de vida: por ejemplo 7 años.

En algunas enfermedades la curva de la supervivencia no sigue una distribución concreta, varía según el tiempo de seguimiento y por tanto se debe estimar por métodos no paramétricos (Kaplan Meier).

Pero generalmente, en una población sana, la tasa de mortalidad anual aumenta exponencialmente con la edad y la supervivencia puede así representarse mediante una relación matemática compleja, que se conoce como la función de Gompertz (Fig. 26), y donde la edad es una variable.

$$\text{Tasa de mortalidad} = 0,000185 * e^{(0,071 * \text{edad})}$$

Sin embargo, si un grupo de personas está afectado por una enfermedad grave, la esperanza de vida es lo suficientemente corta como para que la tasa de mortalidad pueda considerarse constante, y la relación en-

tre supervivencia y tiempo se describe por una función exponencial decreciente (Fig. 26). El método se denomina DEALE (*Declining exponential approximation of life expectancy*) y el cálculo de la esperanza de vida es relativamente fácil:

$$S = S_0 * e^{-mt}$$

S = pacientes vivos en el tiempo t.

S₀ = pacientes vivos en el tiempo 0 (al inicio del estudio).

m = tasa de mortalidad.

t = tiempo.

$$S/S_0 = e^{-mt} \quad (S/S_0 = \text{es la fracción de supervivencia})$$

Despejando de la fórmula anterior resulta

$$m = (-1/t) * \ln(S/S_0)$$

y la esperanza de vida es el inverso de la tasa de mortalidad.

$$\text{Esperanza de vida} = 1/m$$

Supongamos que a los 5 años del tratamiento sobreviven el 25% de los pacientes iniciales.

$$m = (-1/5) * \ln(0,25) = 0,277$$

$$\text{Esperanza de vida} = 1/0,277 = 3,6 \text{ años.}$$

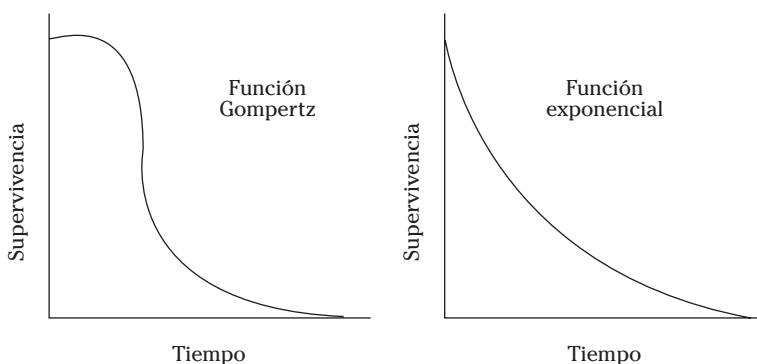


Figura 26. Función Gompertz y DEALE.

Dos son los factores que provocan mortalidad, la edad (el implacable tiempo) y las enfermedades. Entonces, a la mortalidad específica por causas es preciso sumarle la tasa de mortalidad específica por edad, y ésta es una de las características principales del DEALE, que permite la suma de diferentes tasas (lo mismo se aplica cuando el paciente presenta diferentes enfermedades).

$$m_{\text{total}} = m_{\text{específica-edad}} + m_1 + m_2 + m_3 \dots$$

Pero esta suma también conlleva sus problemas, y es que muchas veces una segunda enfermedad modifica la tasa de mortalidad de la primera, la tasa total no es simplemente la suma (no es lo mismo la gripe en un broncópata que en un paciente sano previamente, aunque ambos tengan la misma edad).

Además, las tasas de mortalidad específicas por enfermedad pueden cambiar con la edad (por ejemplo, el cáncer de mama suele ser más grave en las mujeres más jóvenes). Sería preferible, si se dispone de estos datos, utilizar tasas de mortalidad específicas por enfermedad y según estratos de edad. En realidad se debería estratificar por todas aquellas variables que influyen significativamente en el pronóstico y pueden interactuar entre ellas.

Si la enfermedad es muy común, las estadísticas vitales de las tasas de mortalidad por edades están reflejando también la mortalidad por esta enfermedad. Por ello, si se suma la tasa de mortalidad específica por la enfermedad a la específica por edad, en realidad se está sobreestimando la tasa de mortalidad total.

El método DEALE, comparado con el Gompertz (considerado el patrón de referencia, el gold standard) es más exacto cuando la tasa de mortalidad específica por edad es mayor del 10% y sobre todo en los pacientes mayores de 50 años (a mayores tasas de mortalidad mejores estimadores produce).

Modelos de Markov

Los modelos de Markov son otro sistema de cálculo de la esperanza de vida y que obvia algunos de los problemas que presenta el método DEALE.

En los modelos de Markov, los individuos pueden encontrarse en diferentes estados de salud; cada determinados períodos de tiempo (días, meses o años según lo relevante en cada cuestión en estudio) que se denominan ciclos, los individuos pasan de un estado a otro, siendo dichos pasos determinados por unas probabilidades de transición. Sirve también para estados recurrentes.

Un ejemplo sencillo es el de un individuo que se puede encontrar sano, enfermo o muerto (Fig. 27). Estos estados pueden ser no absorbentes (sano o enfermo) o absorbentes (la muerte), cuando al llegar a ellos acaba el proceso de seguimiento.

Cuando las probabilidades de transición de un estado a otro son constantes el modelo se denomina “Cadenas de Markov” y cuando varían de acuerdo a alguna regla se denomina “Proceso de Markov”, en el que las probabilidades son dependientes del tiempo (depende del número de ci-

culos hasta el momento), lo que puede ser equivalente a decir que dependen de la edad. Pero las probabilidades de transición constantes sólo son realistas en aquellas enfermedades con un horizonte temporal corto; en los procesos crónicos no se puede evitar el efecto de la edad.

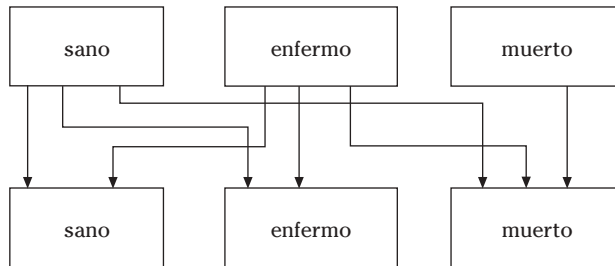


Figura 27. Modelos de Markov.

El proceso no tiene memoria de los estados previos, no sabe cuánto tiempo lleva el individuo enfermo, por ejemplo, basta con conocer su estado actual para predecir cuál será el siguiente estado. Todos los pacientes de un determinado estado en el presente ciclo tienen el mismo pronóstico, independientemente de cómo llegaron al presente estado, o lo que es lo mismo, el futuro está entera y solamente determinado por el estado actual, el pasado del individuo no le importa al proceso Markoviano.

Para la construcción de los modelos de Markov son precisos los siguientes pasos:

- 1) Lo primero es enumerar todos los distintos estados de salud que se pueden dar. Estos estados se deben definir claramente, se distinguen por su diferente pronóstico, y debe ser posible colocarlos en una escala continua de valores relativos sobre la salud (en general estar muerto es peor que enfermo y enfermo es peor que sano).
- 2) Definir las transiciones de un estado a otro. Como veíamos en el ejemplo, el paciente sano puede continuar sano, puede enfermar y puede morir; el paciente enfermo puede continuar enfermo, pero también puede curar o morir; y el paciente muerto sólo puede continuar muerto, es un estado absorbente, con él acaba el seguimiento.

Se pueden distinguir entre estados largos (aquellos en los que se puede permanecer en más de un ciclo) y temporales (la transición a otro ciclo es obligatoria, la probabilidad de permanecer en el mismo estado es cero). El tiempo de permanencia en un estado es igual al recíproco de la probabilidad de pasar a otro estado (del mismo modo que la esperanza de vida es el recíproco

de la tasa de mortalidad). Por ejemplo, si la probabilidad de permanecer en un estado es 0,2, la probabilidad de pasar a otro es el complementario 0,8 y en media permanecerá en dicho estado 1,25 ciclos (1/0,8). Si dicha probabilidad es 0,5 permanecerá en dicho estado 2 ciclos.

- 3) Asignar probabilidades a cada una de las transiciones. Es uno de los pasos más difíciles, ya que no se dispone fácilmente de estos datos, pero generalmente proceden de la literatura o del juicio de expertos. Una de las formas de definir las probabilidades de transición es mediante tasas: para cualquier tasa m , la probabilidad de que ocurra un hecho en un período t es:

$$P_{(t)} = 1 - e^{-mt}$$

Si en el modelo las probabilidades fuesen tiempo-dependientes (proceso de Markov en lugar de cadenas de Markov) se deben utilizar funciones específicas. La tasa de mortalidad general espontánea puede ser modelada mediante la función de Gompertz (donde la edad es una variable):

$$\text{Tasa de mortalidad} = 0,000185 * e^{(0,071 * \text{edad})}$$

Este modelo produce una tasa de mortalidad exponencialmente creciente. La probabilidad de transición asociada viene dada por:

$$P_{(t)} = 1 - e^{-[0,000185 * e^{(0,071 * \text{edad})}]t}$$

El cálculo de la esperanza de vida se puede realizar mediante la simulación de Montecarlo y la simulación probabilística de una cohorte de Markov.

1) Simulación de Montecarlo

En este método, un gran número de pacientes (suele ser del orden de 10.000) atraviesan el proceso de Markov de uno en uno, y mediante un generador de números aleatorios se determina qué ocurrirá con el individuo en cada uno de los ciclos. Se suman los ciclos por los que ha pasado cada individuo antes de llegar a un proceso absorbente (como la muerte).

La esperanza de vida es la media del número de ciclos de todos los individuos, los 10.000 que se han simulado. Al conocer en qué estados ha estado, cuándo y durante cuánto tiempo, los años de vida se pueden actualizar y ajustar por calidad de vida. Por ejemplo, el individuo de la figura 28 ha estado tres ciclos sano, dos enfermo y al final ha muerto. Ha vivido 5 años.

Los pacientes no tienen por qué empezar todos en el mismo estado, se puede establecer una distribución de frecuencias, la de la población de estudio.

La exactitud de este método depende del número de individuos estudiados, la calidad del generador de números aleatorios y sobre todo del conocimiento sobre los estados de salud y las probabilidades de transición. Debido a que los individuos se simulan de uno en uno, requiere mucho tiempo, pero es el método que más flexibilidad y detalle proporciona, dando también medidas de variabilidad (como desviación estándar).

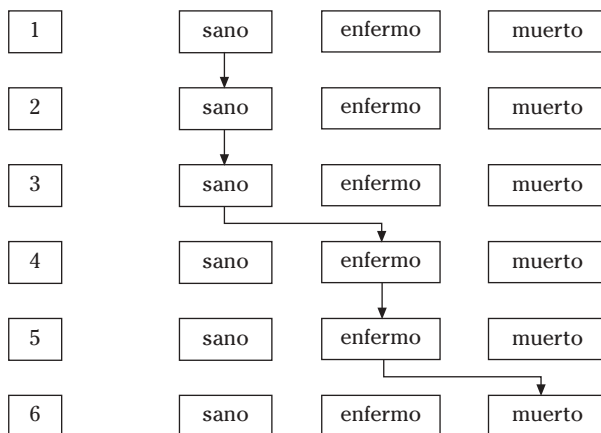


Figura 28. Simulación de Montecarlo.

2) Cohorte de Markov

En este método, un gran número de pacientes se sigue como una cohorte (Tabla 21). La cohorte comienza con una distribución de estados, y en cada ciclo, toda la cohorte se recoloca en cada uno de los estados de acuerdo a unas probabilidades de transición. Se van registrando para cada estado y cada ciclo el número de pacientes.

Tabla 21. Cálculo de la esperanza de vida en la cohorte de Markov.

Tiempo	Sano	Enfermo	Muertos
0	100.000	0	0
1	30.000	50.000	51.000
2	9.000	40.000	51.000
3	2.700	24.500	72.800
-	-	-	-
-	-	-	-
-	-	-	-
Suma	142.860	142.860	-
Medida de ciclos (Suma 100.000)	1,43	1,43	-
Esperanza de vida $1,43 + 1,43 = 2,86$ ciclos.			

Después de un número importante de ciclos prácticamente todos los individuos estarán muertos, y los pocos que quedan (la cola de la curva de supervivencia) puede arreglarse añadiendo una pequeña cantidad de esperanza de vida a cada estado o bien truncando la esperanza de vida. El proceso de la simulación de la cohorte de Markov acaba cuando la cohorte disminuye hasta un número en el que el error añadido es muy pequeño.

El cálculo de la esperanza de vida se realiza sumando el número de individuos en cada ciclo (en estados no absorbentes) y dividiendo por el número de individuos inicial de la cohorte.

La esperanza de vida es muy alta o bien porque la mortalidad es baja y en cada ciclo el número de individuos en los estados no absorbentes es alto, o porque el número de ciclos es muy elevado antes de que se extinga la cohorte. Que 30.000 individuos de los 100.000 iniciales estén otro ciclo más es como que en media todos los individuos estén 0,3 ciclos más (si yo me como dos naranjas y tú no te comes ninguna, es como si los dos nos comiéramos una).

Mediante este método se pueden incorporar probabilidades dependientes del tiempo (funciones), se pueden actualizar los estados y ajustar por calidad de vida. Hasta aquí presenta las mismas características que la simulación de Montecarlo, el problema es que no incorpora medidas de variabilidad de los resultados (sólo el valor más frecuentemente esperado), aunque como ventaja requiere menos tiempo que la simulación de Montecarlo.

Es muy conveniente comprender el cálculo de la esperanza de vida mediante la simulación de Montecarlo y la cohorte de Markov, pero hoy en día existen programas informáticos que realizan todos los cálculos. Estos programas, en cualquier caso, no solucionan todos los problemas, es preciso definir muy bien todos los estados, las transiciones y estimar las probabilidades de transición. Aquí es donde entra la sabiduría y el saber hacer del analista.

Kaplan Meier

Este método de estimación de la supervivencia se utiliza cuando la tasa de mortalidad varía en el tiempo de seguimiento y la supervivencia no sigue una distribución concreta, no se puede aplicar una función. Se trata de un método no paramétrico, también llamado producto-límite o estimador de Kaplan Meier.

Con un ejemplo la comprensión es más fácil. Los datos de la tabla 22 son el resultado del análisis de supervivencia de 96 pacientes con cáncer de pulmón. Los pacientes han sido ordenados de menor a mayor período de seguimiento y la supervivencia se calcula para cada período de tiem-

po. El seguimiento acaba por varios motivos: el paciente ha muerto por la causa que se estudia (en este caso cáncer de pulmón), por otras causas, se ha perdido la pista en el seguimiento, o sencillamente ha finalizado el estudio.

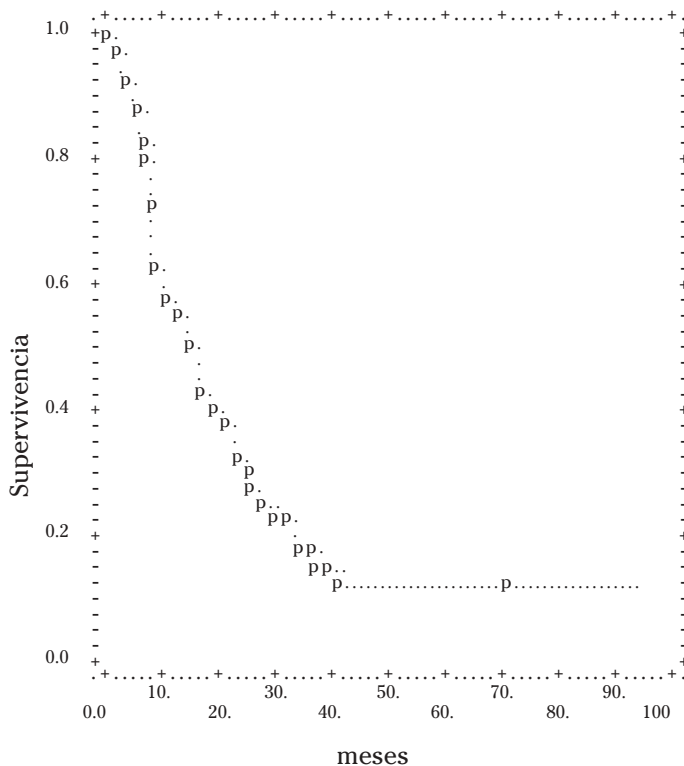
Tabla 22. Estimación de la supervivencia por el método de Kaplan Meier.

PRODUCT-LIMIT SURVIVAL ANALYSIS TIME VARIABLE IS SUPERV							
CASE NUMBER	TIME	STATUS	CUMULATIVE SURVIVAL	STANDARD ERROR	CUM muerto	CUM LOST	REMAIN AT RISK
16	2.00	muerto			1	0	95
88	2.00	muerto			2	0	94
90	2.00	muerto	0.9688	0.0178	3	0	93
18	3.00	muerto			4	0	92
31	3.00	muerto			5	0	91
75	3.00	muerto	0.9375	0.0247	6	0	90
64	4.00	muerto	0.9271	0.0265	7	0	89
11	5.00	muerto			8	0	88
12	5.00	muerto			9	0	87
33	5.00	muerto			10	0	86
61	5.00	muerto			11	0	85
69	5.00	muerto	0.8750	0.0338	12	0	84
4	6.00	muerto			13	0	83
7	6.00	muerto			14	0	82
39	6.00	muerto			15	0	81
82	6.00	muerto			16	0	80
95	6.00	muerto	0.8229	0.0390	17	0	79
51	7.00	muerto	0.8125	0.0398	18	0	78
26	8.00	muerto			19	0	77
45	8.00	muerto			20	0	76
54	8.00	muerto			21	0	75
58	8.00	muerto			22	0	74
71	8.00	muerto			23	0	73
72	8.00	muerto			24	0	72
85	8.00	muerto			25	0	71
91	8.00	muerto	0.7292	0.0454	26	0	70
2	9.00	muerto			27	0	69
-	-	-			-	-	-
-	-	-			-	-	-
-	-	-			-	-	-
-	-	-			-	-	-
-	-	-			-	-	-
76	39.00	muerto	0.1354	0.0349	83	0	13
35	40.00	vivo			83	0	12
41	56.00	vivo			83	0	11
24	58.00	vivo			83	0	10
9	61.00	vivo			83	0	9
21	61.00	vivo			83	0	8
60	68.00	muerto	0.1185	0.0344	84	0	7
6	72.00	vivo			84	0	6
52	72.00	vivo			84	0	5
92	72.00	vivo			84	0	4
1	74.00	vivo			84	0	3
49	77.00	vivo			84	0	2
40	91.00	vivo			84	0	1
59	94.00	vivo			84	0	0
QUANTILE	ESTIMATE						
75TH	8.00						
MEDIAN (50TH)	14.00						
25TH	24.00						

Las columnas de la tabla 22 indican lo siguiente:

- *CASE NUMBER*: número de registro del paciente
- *TIME*: período de seguimiento (en este caso en meses). Los pacientes han sido ordenados de menor a mayor período de seguimiento.
- *STATUS*: aquí se puede indicar si es muerto por cáncer de pulmón, muerto por otras causas, perdido o ha finalizado el estudio vivo.
- *CUMULATIVE SURVIVAL*: supervivencia acumulada
- *STANDARD ERROR*: error estándar de la supervivencia
- *CUM muerto*: número de muertos por cáncer de pulmón hasta ese momento
- *CUM LOST*: número de pacientes perdidos en el seguimiento.
- *REMAIN AT RISK*: número de pacientes que permanecen en riesgo en cada período. Es la base del cálculo de la tasa de mortalidad instantánea.

Figura 29. Curva de supervivencia.



A los dos meses de seguimiento, de los 96 pacientes iniciales habían muerto 3, y la supervivencia es el complementario de la mortalidad:

$$\text{Supervivencia a los 2 meses} = 1 - (3/96) = 1 - 0,3125 = 0,6875$$

Durante el tercer mes de seguimiento, han muerto 3 de los 93 pacientes (96 iniciales – 3 muertos a los dos meses), luego la supervivencia durante este período es:

$$\text{Supervivencia durante el tercer mes} = 1 - (3/93) = 1 - 0,0323 = 0,9677$$

La probabilidad de sobrevivir a los 3 meses, es la probabilidad de sobrevivir durante el tercer mes por la probabilidad de que haya sobrevivido hasta entonces:

$$\text{Probabilidad de sobrevivir a los tres meses} = 0,6875 \times 0,9677 = 0,6641$$

Podemos seguir y comprobar que la probabilidad de sobrevivir a los cuatro meses es la probabilidad de sobrevivir durante el cuarto mes, condicionado a que haya sobrevivido hasta entonces. Durante el cuarto mes de seguimiento, ha muerto 1 de los 90 pacientes que permanecen vivos (96 iniciales – 3 muertos a los dos meses – 3 muertos a los 3 meses), luego la supervivencia durante este período es:

$$\text{Supervivencia durante el cuarto mes} = 1 - (1/90) = 1 - 0,0111 = 0,9889$$

La probabilidad de sobrevivir a los 4 meses, es la probabilidad de sobrevivir durante el cuarto mes por la probabilidad de que haya sobrevivido hasta entonces:

$$\text{Probabilidad de sobrevivir a los cuatro meses} = 0,6641 \times 0,9889 = 0,6566$$

En resumen, la formulación general es la siguiente (Tabla 23): si los individuos de un estudio son seguidos por períodos t_1, t_2, \dots, t_k , donde los tiempos no están igualmente espaciados, la probabilidad de sobrevivir hasta un tiempo t es:

Tabla 23. Estimación de la supervivencia por el método de Kaplan Meier.

$S(t_i) = \text{Prob (sobrevivir hasta } t_i) = \text{Prob. (sobrevivir hasta } t_1) \\ \times \text{Prob. (sobrevivir hasta } t_2 \mid \text{ sobrevive a } t_1) \\ \cdot \\ \cdot \\ \cdot \\ \times \text{Prob. (sobrevivir hasta } t_j \mid \text{ sobrevive a } t_{j-1}) \\ \cdot \\ \cdot \\ \cdot \\ \times \text{Prob. (sobrevivir hasta } t_i \mid \text{ sobrevive a } t_{i-1})$
--

En realidad, el estudio de supervivencia ideal sería aquel en el que siguiésemos a todos los pacientes hasta que muriesen de la causa que estamos estudiando. En este caso, para estimar la probabilidad de supervivencia en cada período, bastaría con dividir el número de muertos acumulados hasta ese momento por el número de pacientes al inicio del estudio.

Supervivencia a los 2 meses: $1 - 3/96 = 1 - 0,0312 = 0,9688$

Supervivencia a los 3 meses: $1 - 6/96 = 1 - 0,0625 = 0,9375$

Supervivencia a los 4 meses: $1 - 7/96 = 1 - 0,0729 = 0,9271$

Supervivencia a los 5 meses: $1 - 12/96 = 1 - 0,125 = 0,8750$

Sin embargo, no es lo habitual poder seguir a los pacientes hasta su muerte, algunos siguen vivos en el momento de acabar el estudio, o se les ha perdido la pista en el seguimiento. Este es el motivo por el que se calcula la probabilidad de sobrevivir a un período como la probabilidad de sobrevivir durante ese período condicionado a que haya llegado vivo hasta entonces, y la probabilidad de sobrevivir durante un período determinado sólo se puede calcular sobre aquellos pacientes que permanecen en riesgo.

Si observamos los datos de la tabla, a los 68 meses la supervivencia no se puede calcular directamente como $1 - 84/96 = 1 - 0,875 = 0,125$.

La probabilidad de sobrevivir durante el mes 68 es : $1 - 1/8 = 1 - 0,125 = 0,875$. Esta tasa de mortalidad (0,125) se ha calculado en base a los 8 pacientes que permanecían en riesgo en ese momento (cinco pacientes finalizaron el seguimiento vivos en un período inferior a 68 meses).

La probabilidad de sobrevivir a los 68 meses es la probabilidad de sobrevivir durante el mes 68 por la probabilidad de que haya sobrevivido hasta entonces: $0,875 \times 0,1354 = 0,1185$.

Además de expresar los resultados mediante la supervivencia en cada período, también puede hacerse con los percentiles. El percentil 75 es el momento en el que viven el 75% de los pacientes de la muestra (han muerto el 25%). El percentil 50 o mediana es el momento en el que el 50% de los pacientes de la muestra han muerto y el 50% restante viven. El percentil 25 es el momento en el que viven el 25% (y el 75% ha muerto). En este ejemplo concretamente los percentiles 25, 50 y 75 son: 8, 14 y 24 meses. A los 8 meses han muerto el 25% de los pacientes, a los 14 meses el 50% y a los 24 meses el 75%.

El análisis de supervivencia no sólo se aplica para el cálculo de la proporción de individuos que viven tras un período de tiempo, sino también para estimar el tiempo libre de recidiva de una enfermedad (en oncología por ejemplo) o la proporción de individuos libres de complicaciones en un período de tiempo.

AÑOS DE VIDA AJUSTADOS POR CALIDAD (AVAC)

En un principio, estar sano parecía equivalente a presentar un análisis dentro de los límites de normalidad, y sin alteraciones fisiológicas ni anatómicas. Y en este sentido, la medida de los resultados sobre la salud más extendida era la esperanza de vida. Sin embargo, no sólo importa la cantidad, sino la calidad de estos años de vida, y esta calidad no viene dada por unos parámetros bioquímicos o anatómicos, es mucho más, es el perfecto estado de salud, que según la O.M.S. incluye el bienestar físico, mental y social. De lo que se habla en este capítulo es de la calidad de vida relacionada con la salud (CVRS), aunque en la calidad de vida propiamente dicha también intervienen otros aspectos como la economía, la educación, el medio ambiente, etc.

En un principio, los instrumentos de medida de la calidad de vida relacionada con la salud no tenían en cuenta la percepción del paciente. En un intento de ser objetivos, era únicamente el médico quien daba un valor a cada uno de los estados de salud. Sin embargo, parecía lógico que se tuvieran en cuenta los juicios de los pacientes sobre estados de la salud que ellos mismos podían experimentar. Y así, hoy en día, la mayoría de los instrumentos de medida de la calidad de vida introducen expresiones propias de los pacientes.

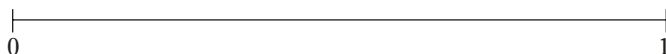
El desarrollo de estos instrumentos ha sido y está siendo muy importante, generando instrumentos cada vez más específicos para determinadas enfermedades, muy precisos, y que permiten detectar pequeños cambios en la salud.

A continuación se exponen algunos métodos de obtención de AVAC (años de vida ajustados por calidad) o QALY en la terminología inglesa (quality adjusted life years). A lo largo del texto se utiliza indistintamente calidad de vida y utilidad, ya que la utilidad es un concepto que refleja la preferencia de los individuos por determinados estados de salud.

Técnicas de obtención de los AVAC

Escala de categorías (EC)

Es uno de los métodos más sencillos, y consiste en una línea con dos puntos finales bien definidos a los que se les da el valor 0 y 1 ó 0 y 100. El 0 representa el peor estado de salud imaginable y el 1 el mejor. Normalmente la línea se divide en 10 intervalos iguales, aunque puede no tener más marcas que la inicial y final o por el contrario estar milimetrada, en cuyo caso se denomina termómetro.



Se trata de pedir al individuo que ponga una marca allí donde considere que se localizan (respecto a la calidad de vida) estados intermedios de salud. Para la obtención de los AVAC se multiplica el valor que ha dado al estado de salud por los años de vida. Si por ejemplo, al hecho de presentar dolor en la espalda le ha dado un valor de 0,9; vivir 10 años con dolor es como vivir 9 años ($10 \times 0,9$) en perfecto estado de salud (son 9 AVAC, 9 años de vida ajustados por calidad).

Es una técnica sencilla, visual, pero presenta algunos problemas metodológicos como es el de los “efectos contextuales”, y es que la puntuación depende del contexto en el que se presentan los diferentes estados de salud. Por ejemplo, cuando un estado poco preferido se presenta con otros estados menos preferidos todavía la puntuación es más alta que si se presentase con estados que para el paciente presentan mayor calidad de vida. Todo depende de con qué se compare.

Estimación de la magnitud (EM)

En este método se describe un estado de salud de referencia, que suele ser el perfecto estado de salud, y se le da el valor de 100. Entonces se le pide al encuestado que asigne una proporción subjetiva a los otros estados de salud, de manera que puede ser “la mitad de bueno”, “la cuarta parte de bueno” que el estado de referencia.

Esta técnica, aunque también parece sencilla, puede producir inconsistencias importantes al comparar los resultados con la escala de categorías, en el sentido de que un método puede decir que el estado de salud A es mejor que el B y el otro método dice lo contrario.

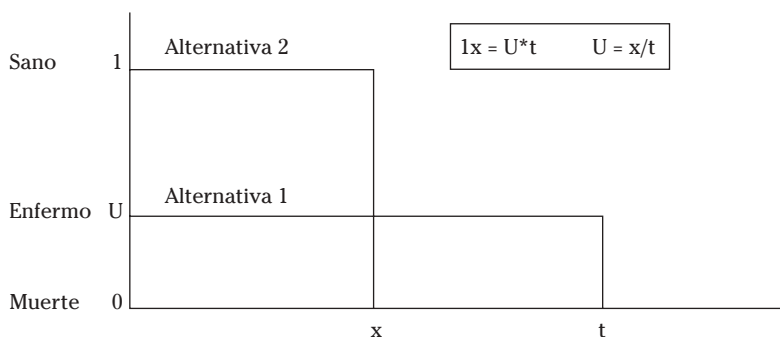
Fue el método utilizado para la construcción de la matriz de Rosser y Kind (Tabla 24), desarrollada en el Reino Unido en 1976. Esta matriz resulta de la combinación de 8 niveles de incapacidad y 4 de sufrimiento o dolor, que da lugar a 29 estados diferentes (en tres no es aplicable). El resultado es la media de las valoraciones obtenidas de 70 personas encuestadas. Dejo al lector que saque sus propias conclusiones de la observación de los resultados.

Tabla 24. Matriz de Rosser y Kind.

Niveles de incapacidad	Niveles de sufrimiento			
	Ausente A	Leve B	Moderado C	Grave D
1. Ausencia de incapacidad	1	0,995	0,990	0,967
2. Ligeras incapacidad social	0,990	0,986	0,973	0,932
3. Incapacidad social grave. Capaz de realizar tareas domésticas (no pesadas)	0,980	0,972	0,956	0,912
4. Limitación grave en el trabajo. Sólo tareas domésticas ligeras	0,964	0,956	0,942	0,870
5. Incapacidad total para el trabajo. Confinados hogar	0,946	0,935	0,900	0,700
6. Confinado en una silla (incapaz de moverse solo)	0,875	0,845	0,680	0,000
7. Confinado en cama	0,677	0,564	0,000	-1,486
8. Inconsciente	-1,028	No aplicable		

Equivalencia temporal (ET)

En este método al paciente se le ofrecen dos alternativas: la primera la de estar enfermo durante t años y la segunda la de estar en perfecto estado de salud x años (siempre menor x que t) seguido de muerte. Se varía el valor de x hasta que el paciente está indiferente entre una alternativa y la otra, de modo que los AVAC = años de vida * x/t (Fig. 30).

**Figura 30.** Equivalencia temporal.

Por ejemplo, supongamos que el paciente tiene una esperanza de vida de 20 años pero con dolor, y está dispuesto a cambiarlos por 15 años en perfecto estado de salud: la utilidad será $15/20 = 0,75$ y por lo tanto AVAC = $20 \cdot 15/20 = 15$.

La ventaja principal de esta técnica es que enfrenta directamente a los entrevistados con la duración de los diferentes estados de salud, y es más sencilla que el juego estándar, considerado el patrón de oro.

Como inconveniente está el hecho de que supone que la calidad de vida es proporcional al tiempo, que vale lo mismo un año dentro de 10 años que ahora, lo que no es así, porque en general se valora más la supervivencia actual que la futura (más vale pájaro en mano).

Equivalencia de personas (EP)

Esta es una de las técnicas propuestas como la más válida para la estimación de los valores sociales de los diferentes estados de salud en la asignación de los recursos sanitarios.

Consiste en preguntar a los encuestados cuántos “Y” individuos en un determinado estado de salud los considera equivalentes en valor social a “X” individuos en otro estado de salud diferente (Fig. 31). Por ejemplo, mejorar la salud de 10 personas en el estado 7 es como mejorar la salud ¿de cuántas personas en el estado 3? ¿Cuántas caries dentales son equivalentes a una apendicitis?

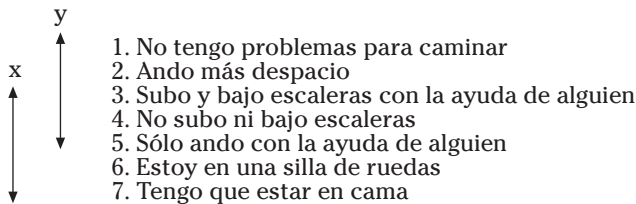


Figura 31. Equivalencia de personas.

El procedimiento es muy complejo y sólo ha sido utilizado en pequeños grupos de expertos. Una dificultad importante, común a la mayoría de los métodos, es la descripción del problema, de la situación y la elección de los límites inicial y final de la escala.

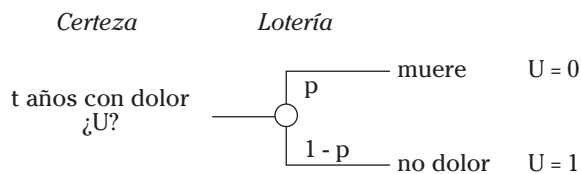
Juego estándar (JE)

Este método, desarrollado por Von Neuman y Morgenstern, está basado en la teoría de decisiones en condiciones de incertidumbre y es la técnica de obtención de la calidad de vida relacionada con la salud considerada patrón de referencia.

Al paciente se le ofrecen dos alternativas (Fig. 32):

- una de certeza, la de continuar en la condición de enfermo durante t años.

- una segunda alternativa de lotería, donde el paciente se la juega a morir inmediatamente (con una probabilidad p de producirse) o vivir en perfecto estado de salud (con una probabilidad $1-p$). Se trata de variar este valor de p hasta que el paciente está indiferente entre las dos alternativas.



U = utilidad

$$U? = 0 \cdot p + 1 \cdot (1-p)$$

Figura 32. Juego estándar.

Supongamos que un individuo puede vivir 10 años con dolor de espalda y se le ofrece la probabilidad de jugársela a la lotería y esta lotería es la entrada en quirófano: si se opera puede morir en quirófano (con una probabilidad p) o sobrevivir esos mismos 10 años sin el dolor (con una probabilidad $1-p$). ¿Con qué probabilidad de morir en quirófano el individuo estaría dispuesto a ser intervenido? Imaginemos que contesta que la probabilidad debería ser del 1% o menor. Esto quiere decir que la utilidad que le reporta 10 años con dolor es la misma que la de jugar, y la de jugar es la media de la utilidad de morir (0) y la de vivir (1) ponderadas por la probabilidad de que ocurran estas dos posibilidades:

$$\text{Utilidad ?} = p \cdot 0 + (1-p) \cdot 1 = 0,01 \cdot 0 + 0,99 \cdot 1 = 0,99$$

Los AVAC se calculan multiplicando los años de vida por la utilidad estimada.

$$\text{AVAC} = 10 \cdot 0,99 = 9,9$$

Vivir 10 años con dolor es como vivir 9,9 años sin dolor.

En resumen, los pasos a seguir para la estimación de la calidad de vida (utilidad) por este método son los siguientes:

- 1) se listan todos los posibles resultados (en el ejemplo eran 10 años vivo sin dolor de espalda, 10 años con dolor de espalda o muerto).
- 2) se ordenan por orden de preferencia (siguiendo con el ejemplo, el peor estado es muerto y el mejor vivo sin dolor).
- 3) se da el valor 1 al mejor estado de salud y el 0 al peor.

- 4) se realiza el juego estándar para cada resultado intermedio, cada uno con su escenario (lo más realista posible). El individuo se enfrenta a una elección entre un estado intermedio seguro y la lotería, donde pueden darse el peor y mejor estado de salud.
- 5) se determina cuál es la probabilidad p de que ocurra el peor estado de salud en la lotería, con la que el individuo está indiferente entre jugársela o no. Conviene comenzar con probabilidades extremas por arriba y por abajo, y se va estrechando el intervalo hasta que el paciente sea indiferente (es lo que se denomina variar la probabilidad en ping-pong).
- 6) se resuelven las utilidades intermedias en función de las utilidades conocidas.

$$U(\text{estado intermedio}) = p \times U(\text{peor estado}) + (1-p) \times U(\text{mejor estado})$$

Si al peor estado se le ha asignado la utilidad cero y al mejor una utilidad de 1, despejando de la fórmula se obtiene que la utilidad del estado intermedio es $1-p$.
- 7) Los AVAC se calculan multiplicando la esperanza de vida por la utilidad estimada.

Se trata de un método difícil, complejo, donde las respuestas de los individuos pueden ser poco fiables. Los individuos están poco acostumbrados a tratar con probabilidades, sobre todo cuando son extremas (o muy altas o muy bajas, por encima de 0,9 o por debajo de 0,1). Además el juego estándar es internamente inconsistente, porque puede dar diferentes resultados dependiendo del modo en que se prepara el escenario (los efectos contextuales de los que hablábamos previamente).

Comparación de los métodos de medida de la calidad de vida

Supongamos que mediante la técnica de escala de categorías se pide a un individuo que puntúe la calidad de vida de dos estados de salud A y B. El estado de salud A ha recibido la puntuación de 0,4 y el estado B 0,8. Entonces, vivir 10 años en el estado A es como vivir 4 años en perfecto estado de salud y vivir 10 años en el estado B es como vivir 8 años en perfecto estado de salud (equivalencia temporal). Si el resultado se midiese mediante la técnica de estimación de la magnitud, el resultado se expresaría diciendo que el estado A es la mitad de bueno que el B. A su vez, el estado A es el 40% de bueno que el perfecto estado de salud y el estado B representa el 80% del perfecto estado de salud. Curar a una persona en el estado A (equivaldría a mejorar la calidad de vida en $1 - 0,4 = 0,6$ puntos) es como curar a tres personas en el estado B (cada una produce una mejora de $1 - 0,8 = 0,2$ puntos), siendo éste el resultado de medir la calidad de vida mediante la técnica de equivalencia de personas. Si un individuo decide

entrar en quirófano para curarse de la enfermedad A, la probabilidad de morir que aceptaría sería de 0,6 (juego estándar) y si se trata de la enfermedad B aceptaría una probabilidad de 0,2.

Es difícil la comprobación de la validez de las técnicas, debido a la ausencia de un estándar de comparación, un patrón de referencia. Pero en general, el método de la equivalencia temporal es de los más simples y fiables y el que produce los resultados más comparables al juego estándar.

En cuanto a la fiabilidad de las técnicas, es relativamente buena en períodos de tiempo cortos.

Por último y pese al generalizado uso de los AVAC, éstos presentan importantes problemas, que no sólo dependen del instrumento de medida de la utilidad, porque:

- La utilidad que los enfermos asignan a los diferentes estados depende en gran medida del modo de preparar los escenarios (es diferente si se habla de mortalidad que de supervivencia, de esperanza de vida o años hasta la muerte, etc.). Este problema se puede obviar presentando diferentes escenarios para un mismo juego estándar y observando la consistencia de los resultados. Por ejemplo ¿Cuál sería la máxima probabilidad de muerte que usted estaría dispuesto a aceptar para someterse a una intervención? y contesta 0,1. Por otro lado ¿Cuál sería la mínima probabilidad de sobrevivir para aceptar la intervención quirúrgica? y si es consistente debería contestar 0,9.
- Los individuos tienden a ser aversores o amantes del riesgo dependiendo de si los resultados del escenario que se presenta son positivos o negativos. Por ejemplo, no se tiende a probar un medicamento nuevo (aversión al riesgo), aunque prometa ser más eficaz, si el actual ya es efectivo (más vale pájaro en mano). Ahora, si se presenta el nuevo medicamento con menores efectos adversos que el ya conocido, se tiende a ser más arriesgado (de perdidos al río).
- Los enfermos responden de manera inconsistente a las preguntas realizadas según sea su estado de ánimo, la experiencia previa sobre los resultados de un tratamiento, etc.
- La opinión de los enfermos varía con el tiempo.
- Se valora idénticamente un AVAC con independencia de quien lo reciba, su edad, enfermedades concomitantes, nivel socioeconómico, etc.
- Existen diferencias entre los individuos en la valoración de la calidad. Este problema se suele obviar encuestando una muestra amplia de sujetos y obteniendo la media de sus valoraciones. Ello

convierte una valoración en principio subjetiva, en una medida casi objetiva de valor general.

- La utilidad de los AVAC no es proporcional al tiempo, porque, por lo general, los enfermos valoran más los años que esperan vivir en un futuro próximo que los que esperan vivir en un futuro lejano. Soluciones a este problema son la actualización de los años (explicado en el capítulo 10), realizar las elecciones teniendo en cuenta calidad y cantidad de vida a un mismo tiempo o la utilización del método en los tres pasos (primero equivalencia temporal de los años de vida enfermo, segundo utilidad de los años de vida sano y por último utilidades de los años de vida enfermo) descrito más adelante en este capítulo.
- Trabajar con probabilidades es difícil, especialmente si son extremas.

Instrumentos de medida de la calidad de vida

Entre los instrumentos de medida de la calidad de vida están los genéricos (perfiles de salud y medidas de utilidad o preferencia) y los específicos.

Los instrumentos genéricos, dado que cubren un amplio espectro de dimensiones (como mínimo las dimensiones física, mental y social) pueden ser aplicados tanto a población general como a grupos específicos de pacientes. El principal inconveniente que presentan es que el contenido puede no adecuarse en aquellos aspectos de la CVRS más importantes de la enfermedad concreta que se estudia y, por tanto, puede ser un instrumento poco sensible a los cambios antes y después del tratamiento.

Entre los perfiles de salud validados en España están el Perfil de consecuencias de la enfermedad (*Sickness impact profile*), el Perfil de salud de Nottingham (*Nottingham health profile*) o el SF-36.

Respecto a los instrumentos de medida de la utilidad o preferencias, los más utilizados son la matriz de Rosser y Kind, la Escala de la calidad del bienestar (*Quality of Well Being Scale*) y el *EuroQol*.

Los instrumentos específicos incluyen sólo aquellos aspectos importantes de una determinada enfermedad. Tienen la ventaja de presentar una alta sensibilidad a los cambios ante el problema específico de salud que se evalúa y su mayor desventaja es que no permiten comparaciones entre diferentes afecciones y, por lo tanto, no son útiles para conocer la eficiencia relativa de diferentes intervenciones sanitarias. Se han desarrollado sobre todo en pacientes con artritis reumatoidea, síntomas gastrointestinales, insuficiencia respiratoria crónica, diálisis, pacientes oncológicos, epilépticos, etc.

RESULTADOS MULTIDIMENSIONALES SOBRE LA SALUD

A menudo, el efecto de una alternativa sobre la salud es tanto sobre la longitud de la vida como sobre la calidad de la misma. Pero la utilidad de un estado de salud puede depender del tiempo que se va a estar en dicho estado y además, los pacientes no dan el mismo valor a los años vividos en un futuro próximo que en futuro lejano.

Imaginemos que un cirujano se enfrenta ante la amputación de un pie. Tiene dos alternativas: amputar relativamente poco (el paciente podrá caminar mejor pero las probabilidades de supervivencia son menores) o amputar más (el paciente caminará peor pero tendrá más probabilidades de sobrevivir).

El método más sencillo consistiría en multiplicar la cantidad de vida por la utilidad en cada estado, pero este método está basado en suposiciones poco realistas, la de que la utilidad de la vida es proporcional al tiempo.

Existe un método para obviar este problema, y consiste en establecer la utilidad de cada estado teniendo en cuenta a la vez calidad y cantidad de vida. Se procede igual que con el método de la lotería estándar (se listan los estados, se ordenan por orden de preferencia, se crean escenarios de juego para calcular la utilidad de los estados intermedios...). Por ejemplo: si se opera puede morir en la intervención, vivir 36 meses en un determinado estado (generalmente el de peor calidad), 24 en otro de mejor calidad ó 12 meses en otro de mejor calidad todavía. ¿Qué utilidad tiene para el paciente cada una de estas cuatro combinaciones de calidad y cantidad de vida? Este procedimiento tiene la ventaja de que los escenarios de la lotería estándar son muy similares a las experiencias con las que se puede enfrentar el paciente en la vida real, pero presenta el inconveniente de que pensar en calidad de vida y esperanza de vida a la vez resulta difícil.

Por último, otro método muy poderoso pero también costoso (en el sentido de la dificultad que entraña), es medir la utilidad de un período de tiempo enfermo, primero valorando la calidad y segundo la cantidad de vida. Se miden las dos dimensiones de forma independiente y luego se combinan en un sólo número.

Ejemplo: Supongamos un paciente de 50 años con un dolor de espalda intenso. La esperanza de vida normal es de 25 años (según las estadísticas vitales). El tratamiento es la intervención quirúrgica, que conlleva cierto riesgo de muerte.

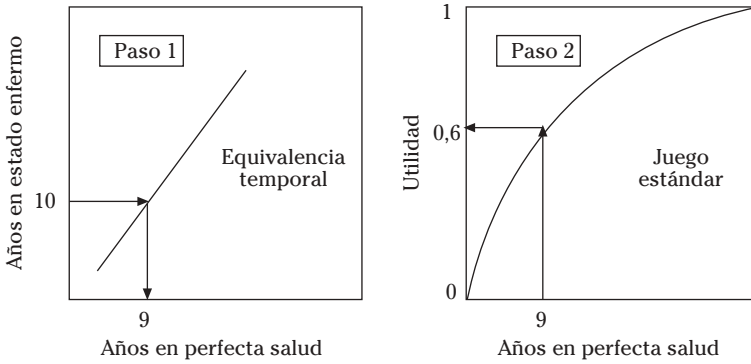


Figura 33. Resultados multidimensionales sobre la salud.

Este procedimiento consta de tres pasos (Fig. 33):

- 1) Determinar la calidad de vida. Cuántos años en perfecto estado de salud el paciente considera equivalentes a un período en estado enfermo (método de equivalencia temporal). Han resultado los datos de la tabla 25.
- 2) Determinar la utilidad que asignan los pacientes a las diferentes longitudes de vida en estado sano (método de lotería estándar), siendo el mejor estado la esperanza de vida normal (25 años en el ejemplo) y el peor la muerte inmediata. Se calcula así la utilidad para longitudes de vida intermedias.

En estos modelos (Fig. 34), las probabilidades de indiferencia se especifican y el paciente debe decir a qué período de vida es equivalente el juego. Supongamos que el paciente responde que si la probabilidad de muerte inmediata es 0,05, la mínima esperanza de vida para evitar la lotería es de 21 años (Tabla 26). Si recordamos que la utilidad de vivir es 1 y la de morir es 0, para este paciente la utilidad de vivir 21 años es:

$$U[21\text{años}] = 0,5 \cdot 0 + 0,95 \cdot 1 = 0,95$$

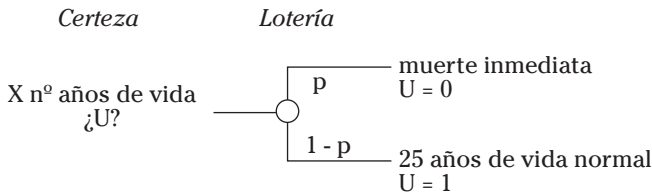


Figura 34. Utilidad de la longitud de vida (juego estándar).

Se realiza el mismo ejercicio cambiando las probabilidades del juego, para determinar las utilidades de diferentes cantidades de vida sana. Siguiendo con el ejemplo, si la probabilidad de muerte inmediata fuese del 60%, y el paciente contesta que es equivalente a 5 años seguros, la utilidad de 5 años de vida es 0,4.

$$U[5\text{años}] = 0,6*0 + 0,4*1 = 0,4$$

- 3) El último paso consiste en determinar la utilidad de los pacientes para varias longitudes de vida en estado enfermo combinando las dos medidas anteriores (Tabla 27). A la vista de los resultados se comprueba que los incrementos en la utilidad no son proporcionales al tiempo, son cada vez menores.

Tabla 25. Paso 1. Equivalencia temporal.

Años de vida enfermo	Años equivalentes de vida sano
5	5
10	9
15	13
20	17
25	21

Tabla 26. Paso 2. Juego estándar.

Años equivalentes de vida sano	Utilidad
5	0,40
9	0,60
13	0,80
17	0,90
21	0,95

Tabla 27. Paso 3. Combinación de calidad y cantidad de vida.

Años de vida enfermo	Años equivalentes de vida sano	Utilidad
5	5	0,40
10	9	0,60
15	13	0,80
20	17	0,90
25	21	0,95

Curva de utilidad de la esperanza de vida

La curva de la utilidad de la longitud de la vida nos enseña sobre la actitud de los pacientes hacia el riesgo de muerte. Para comprender mejor la siguiente explicación es preciso entender el concepto de “utilidad marginal de un año de vida” que es el cambio producido en la utilidad al aumentar la esperanza de vida en un año.

La mayoría de las curvas de utilidad de los años de vida son curvas cóncavas hacia abajo (Fig. 35), como la del ejemplo anterior, de modo que la utilidad marginal es decreciente. Al principio de la curva la pendiente es elevada y ésta va disminuyendo conforme pasa el tiempo, reflejando la menor utilidad de los años en un futuro lejano. Podría ser el caso de una persona que espera morir en un futuro lejano y no está dispuesta a jugársela en el momento actual. Esta actitud es la de un “aversor al riesgo”.

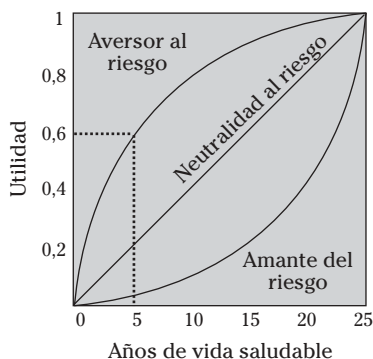


Figura 35. Utilidad de la esperanza de vida.

Ocasionalmente la curva de utilidad es cóncava hacia arriba, con una pendiente inicial muy pequeña y ésta va aumentando con el tiempo (Fig. 35). La utilidad marginal aumenta. Podría ser el caso de una persona que espera morir pronto y está dispuesto a arriesgarse a morir inmediatamente (con la lotería estándar) sin con ello puede ganar años de vida. Sin embargo, conforme supera años puede pensar que está curado y la utilidad de esos años de vida aumenta. Esta actitud es la de un “amante del riesgo”.

Los individuos cuya curva de utilidad es una recta adoptan una actitud de “neutralidad ante el riesgo”. La utilidad aumenta de un modo proporcional a la longitud de la vida (la utilidad marginal es constante).

RESUMEN

La forma más sencilla de medir los resultados sobre la salud es mediante el cálculo de las muertes evitadas o los casos de una enfermedad prevenidos. Para ello existen medidas que comparan el riesgo de la enfermedad entre los expuestos o no expuestos a un factor. Aunque conceptualmente son iguales, el nombre de dichas medidas cambia según se trate de factores de riesgo (riesgo atribuible, proporción de riesgo atribuible o fracción etiológica, riesgo atribuible en la población, proporción de riesgo atribuible en la población) o factores protectores (disminución absoluta del riesgo, fracción prevenible o reducción relativa del riesgo).

Sin embargo lo importante no es vivir o morir en sí mismo, sino cuánto tiempo se vivirá o lo que es lo mismo, cuánto se tardará en morir. Para ello la medida de los resultados sobre la salud idónea es la esperanza de vida. Esta esperanza de vida se puede obtener por métodos no paramétricos (kaplan Meier) o por métodos paramétricos como el DEALE (declining exponential approximation life expectancy):

$$S = S_0 * e^{-mt} \quad m = (-1/t) * \ln(S/S_0) \quad \text{Esperanza de vida} = 1/m.$$

S = pacientes vivos en el tiempo t.

S₀ = pacientes vivos en el tiempo = 0 (al inicio del estudio).

m = tasa de mortalidad.

t = tiempo.

Otro método de cálculo de la esperanza de vida es mediante los modelos de Markov: se parte de que los individuos pueden encontrarse en diferentes estados de salud y pueden pasar en cada ciclo a otros estados según unas probabilidades de transición. Hay dos modos de calcular la esperanza de vida: mediante la simulación de Montecarlo, en el que se hace pasar individuo a individuo por el proceso de Markov (y así no sólo se estima la media de la esperanza de vida sino que también proporciona medidas de variabilidad), y mediante las cohortes de Markov, en donde toda la cohorte se va recolocando en los diferentes estados en cada ciclo, hasta que todos han muerto o son tan pocos los supervivientes que apenas modifican la esperanza de vida. La esperanza de vida es la media de ciclos por los que pasan los individuos.

Por último, no sólo se espera vivir muchos años, sino en perfecto estado de salud. Para ello es preciso añadir la matización de la utilidad o calidad de vida. Utilidad es la preferencia que los individuos manifiestan por los diferentes estados de salud. Hay diversos métodos para su cálculo: escala de categorías, estimación de la magnitud, equivalencia temporal, equivalencia de personas y juego estándar (considerado el patrón de oro).

La mayoría de las veces los resultados sobre la salud son tanto en forma de cantidad como calidad de vida. La forma más sencilla de tener en

cuenta ambos factores es multiplicando la cantidad de vida por la calidad estimada por cualquiera de los métodos anteriores. Sin embargo, la utilidad de los años de vida no es proporcional en el tiempo, en general se valoran más los años de vida en un futuro próximo que los que se espera vivir en un futuro más lejano (se tiende a ser aversores al riesgo).

PROBLEMAS

1. Si la incidencia de la gripe es de un 25% anual en la población no vacunada y de un 2,5% en la población vacunada. ¿Cuál es la eficacia de la vacuna? ¿En qué proporción disminuye la incidencia de la gripe si la cobertura vacunal es del 10%? Si la población a la que se vacuna (cubriendo sólo al 10%) es de 5 millones de habitantes ¿cuántos casos de gripe, en números absolutos, se habrán evitado?
2. Imaginemos que en un grupo determinado de pacientes, el riesgo de muerte en el primer año tras un infarto es del 10% (sin tratamiento). Este riesgo, si toma beta bloqueantes es del 7,5%. ¿Cuál es la efectividad de los betabloqueantes en la reducción de la mortalidad. ¿Cuántos pacientes es necesario tratar para evitar una muerte? ¿Cómo varían estos resultados si la mortalidad con y sin tratamiento fuesen 0,75% y 1% respectivamente?
3. La tasa de mortalidad por colecistectomía varía según el sexo, la edad y la presencia de complicaciones, de modo que en los hombres la tasa de mortalidad es el doble que en las mujeres, la presencia de complicaciones multiplica por 4 el riesgo de morir en quirófano y respecto a la edad, de los 30 a los 70 años aumenta la tasa un 8,45% anual. Si en los hombres de 40 años la probabilidad de morir tras una colecistitis aguda intervenida es de 0,96%, ¿Cuál será la tasa de mortalidad operatoria de una mujer de 45 años, en la que se practica la colecistectomía de forma profiláctica?
4. Asumiendo que la supervivencia se describe por un modelo exponencial decreciente (DEALE) con una tasa de mortalidad constante, ¿cuál de los siguientes tratamientos maximiza la supervivencia del paciente?
Tratamiento A: tasa de mortalidad anual del 8%
Tratamiento B: mortalidad a los 5 años del 40%
Tratamiento C: esperanza de vida de 10 años.
5. Se pretende estudiar la supervivencia de un individuo de 55 años que padece una enfermedad específica. En la literatura sólo se dispone de información de la supervivencia de un grupo de individuos con la misma enfermedad pero cuya edad media es de 40 años. La supervivencia en este grupo es 20% a los 5 años. Las tasas de mortalidad específicas por edad (según las estadísticas vitales) son 0,6% para el grupo de 40 años y 0,9% para el grupo de 55. ¿Cuál es la esperanza media de vida de un individuo de 55 años afecto de la enfermedad en estudio?
6. Se ha realizado un modelo de Markov con el fin de comparar la eficacia de dos tratamientos contra la depresión: imipramina y terapia interpersonal frente al placebo. Los tres estados en los que se puede en-

contrar el paciente son: (A) paciente no deprimido y con un tratamiento de mantenimiento, (B) paciente deprimido y (C) paciente muerto por suicidio. La calidad de vida del estado deprimido, mediante la técnica del juego estándar, se ha estimado en 0,45. Los resultados de la esperanza de vida son los que se muestran en la tabla 28. Suponiendo que la calidad de vida es proporcional al tiempo ¿Es eficaz el tratamiento médico?

Tabla 28. Cálculo de la esperanza de vida en la cohorte de Markov.

Grupo de tratamiento con imipramina y terapia interpersonal

Tiempo (años)	No deprimido	Deprimido	Muerto
0	100.000	0	0
1	80.000	19.900	100
2	50.000	49.500	500
3	10.000	88.500	1.500
.	.	.	.
.	.	.	.
Suma	249.800	641.200

Grupo placebo

Tiempo (años)	No deprimido	Deprimido	Muerto
0	100.000	0	0
1	60.000	39.800	200
2	30.000	69.000	1.000
3	5.000	92.000	3.000
.	.	.	.
.	.	.	.
Suma	198.100	698.200

7. En la tabla 29 se muestran los datos de 30 pacientes con cáncer de pulmón estadio IIIA. ¿Cuáles son los percentiles 25, 50 y 75 de supervivencia?

Tabla 29. Análisis de supervivencia. Cáncer de pulmón estadio IIIA.

PRODUCT-LIMIT SURVIVAL ANALYSIS						
CASE NUMBER	TIME	STATUS	CUM muerto	CUM LOST	REMAIN AT RISK	
31	3.00	muerto	1	0	29	
61	5.00	muerto	2	0	28	
69	5.00	muerto	3	0	27	
95	6.00	muerto	4	0	26	
72	8.00	muerto	5	0	25	
91	8.00	muerto	6	0	24	
67	10.00	muerto	7	0	23	
53	11.00	muerto	8	0	22	
38	13.00	muerto	9	0	21	
50	13.00	muerto	10	0	20	
74	15.00	muerto	11	0	19	
86	15.00	muerto	12	0	18	
44	16.00	muerto	13	0	17	
62	16.00	muerto	14	0	16	
96	16.00	muerto	15	0	15	
15	18.00	muerto	16	0	14	
93	20.00	muerto	17	0	13	
81	21.00	muerto	18	0	12	
89	26.00	muerto	19	0	11	
37	30.00	muerto	20	0	10	
94	30.00	muerto	21	0	9	
83	31.00	muerto	22	0	8	
35	40.00	vivo	22	0	7	
41	56.00	vivo	22	0	6	
21	61.00	vivo	22	0	5	
6	72.00	vivo	22	0	4	
92	72.00	vivo	22	0	3	
49	77.00	vivo	22	0	2	
40	91.00	vivo	22	0	1	
59	94.00	vivo	22	0	0	

8. Mediante la técnica de escala de categorías se pide a un paciente que puntúe la calidad de vida de dos estados A y B. Ha resultado 0,25 y 0,75 respectivamente. ¿Cómo debería responder el paciente si las técnicas aplicadas fuesen la de equivalencia temporal, equivalencia de personas, estimación de la magnitud y juego estándar, con el fin de que las respuestas resulten consistentes?
9. Suponga que un paciente con cáncer, cuya supervivencia es de 6 años, se enfrenta a la decisión de elegir entre el tratamiento médico y quirúrgico con el fin de eliminar o reducir los síntomas de su enfermedad.
- El tratamiento médico no tiene efectos adversos pero sólo es parcialmente efectivo, de modo que el paciente continuará teniendo síntomas durante los 6 años que le restan de vida.
 - El tratamiento quirúrgico tiene un 60% de probabilidades de eliminar completamente los síntomas, si sobrevive a la intervención, porque también hay un 1% de probabilidades de que el paciente muera en quirófano. Además, si sobrevive a la intervención, se puede acortar la vida del paciente a 2 años (libre de síntomas) con una probabilidad

del 20%, o aún no disminuyendo la esperanza de vida puede fracasar la intervención en eliminar los síntomas en otro 20% de los casos.

Como lo que está en juego es tanto la calidad como la cantidad de vida, primero se estudia la actitud del paciente hacia los síntomas y se observa (mediante la técnica de equivalencia temporal) que para el paciente, vivir con síntomas es como acortar en un tercio su vida. En cuanto a la utilidad de diferentes longitudes de vida sana, mediante la técnica de la lotería estándar han resultado los datos de la tabla 30.

¿Qué tratamiento elige, el médico o el quirúrgico?

Tabla 30. Juego estándar.

Años libres de síntomas	Utilidad
Muerte inmediata	0,00
1	0,50
2	0,70
3	0,85
4	0,95
5	0,99
6	1,00

EVALUACIÓN ECONÓMICA

INTRODUCCIÓN

Hasta ahora, prácticamente toda la preocupación de los clínicos se ha basado en buscar las mejores pruebas y tratamientos (independientemente del coste), porque la salud de su paciente es lo primero. Se tiene la falsa idea de que el considerar los costes disminuye la libertad del médico y además no es ético, por lo tanto se debe ignorar el tema.

Esta actitud cambiará cuando el médico se convenza de que la evaluación económica sólo es una herramienta más que le ayuda en la toma de decisiones, y cuyo objetivo es mejorar la salud de los pacientes, no disminuir el gasto sanitario.

“La salud no tiene precio” es otra de las frases conocidas, sin embargo, cada día, cualquiera de nosotros realiza elecciones que pueden perjudicar nuestra salud (fumar, beber, conducir a gran velocidad), con lo que una vida larga y saludable sólo es una de las muchas metas deseables en la vida, y para la mayoría no es precisamente la más importante.

El argumento clave en favor de la evaluación económica es que los recursos son escasos o cuando menos limitados. Si se dispone de una cantidad limitada de recursos su utilización en una determinada actividad no permite su utilización en otras actividades también beneficiosas. Imaginemos que acude a urgencias un paciente con infarto agudo de miocardio. La unidad de coronarias está completa, pero entre los pacientes ingresados en dicha unidad hay uno con un pequeño infarto y muy pocas probabilidades de complicación. El médico decide, entonces, enviar este paciente a planta para poder ingresar en la unidad de coronarias al que acaba de acudir a urgencias. Esta es una de las formas de maximizar el beneficio global de todos los pacientes, exponiendo a uno a un pequeño riesgo, a costa de beneficiar en gran medida a otro.

La evaluación económica es el análisis comparativo de las acciones alternativas tanto en términos de costes como de beneficios o, dicho de otro modo, mide la eficiencia de las diferentes opciones. Por costes no sólo se entienden pesetas sino todos los efectos negativos (entre los que se incluye el dinero que es necesario desembolsar para llevar a cabo un determinado programa), y como beneficios todos los resultados positivos

(entre los que se incluyen la mejora en la salud y el dinero ahorrado por ejemplo).

Aquí es conveniente aclarar que el coste real de cualquier programa no es la cantidad que aparece en el presupuesto del mismo, sino más bien los resultados que se podrían haber alcanzado con otro programa no realizado porque los recursos se han dedicado al primero. De aquí surge un concepto clave, el de “coste de oportunidad”: son los resultados que se podrían haber obtenido de haber aplicado los recursos a la mejor alternativa en comparación. El coste real de ampliar una unidad de cuidados coronarios es una reducción del programa de cuidados paliativos para los enfermos con cáncer, por ejemplo. Este coste de oportunidad (también llamado coste alternativo) es el que pretende valorar la evaluación económica. ¿A qué es preciso renunciar para llevar a cabo un programa?

TÉCNICAS DE EVALUACIÓN ECONÓMICA

Todas las técnicas de evaluación económica miden los costes de la misma manera, en unidades monetarias, la diferencia estriba en la forma de medir los resultados sobre la salud (Tabla 31). Si dichos resultados se miden en unidades monetarias se trata del análisis coste-beneficio (ACB); si se miden en las unidades consustanciales al programa en estudio (vidas salvadas, enfermedades prevenidas, casos diagnosticados, años de vida ganados) se trata de un análisis coste-efectividad (ACE); si los resultados se expresan en calidad de vida, el análisis se denomina coste-utilidad (ACU); por último, cuando los resultados de las diferentes alternativas en comparación son similares, sólo se comparan los costes y entonces se trata de un análisis de minimización de costes (AMC).

Tabla 31. Tipos de evaluación económica.

TIPO DE ANÁLISIS	MEDIDA DE LOS COSTES	MEDIDA DE LOS EFECTOS
Coste-beneficio (ACB)	Unidades monetarias	Unidades monetarias
Coste-efectividad (ACE)	Unidades monetarias	Unidades clínicas habituales (años de vida ganados p.e.)
Coste-utilidad (ACU)	Unidades monetarias	Calidad de vida (AVAC p.e.)
Minimización de costes	Unidades monetarias	Efectos equivalentes

AVAC = años de vida ajustados por calidad.

Medida de los costes

Los costes que se tienen en cuenta dependen del punto de vista del análisis (del Ministerio de Sanidad, otros ministerios, gobierno regional, el del paciente, empresario, el del centro que proporciona el programa, etc.). Cuando la evaluación la realiza un organismo determinado, éste puede y debe proporcionar los puntos de vista clave desde los que abordar el pro-

blema, sin embargo, cuando existan dudas, lo ideal es adoptar el punto de vista de la sociedad, que es el más amplio de todos y siempre es relevante.

Si la comparación se realiza sólo con los programas en estudio, no es preciso tener en cuenta los costes comunes, ya que no afectarán a la elección entre las alternativas. Sólo se deben cuantificar aquellos que afectan de un modo diferente a uno y otro programa, por ello se denominan costes diferenciales. Por ejemplo: si se trata de comparar un programa de rehabilitación hospitalaria con otro de rehabilitación domiciliaria en los pacientes con prótesis de cadera, los costes referentes a los primeros 15 días de estancia hospitalaria (incluyendo la intervención quirúrgica, los análisis, los cuidados de enfermería, etc.) así como las visitas médicas de control, son comunes a las dos alternativas, por lo que no es necesario tenerlos en cuenta: tanto si son muy elevados como si no, afectan por igual a los dos programas, y en este caso no se trata de saber cuánto cuesta la rehabilitación de estos pacientes, sino cuál de las dos formas de hacerlo es más eficiente.

Otra clasificación de los costes es en variables y fijos según varíen o no con el nivel de producción. Por ejemplo: se evalúa un programa de vacunación contra la gripe en un servicio médico de empresa. La efectividad de programa serán los casos de gripe evitados. Los costes de mantenimiento de las instalaciones, del personal sanitario que atienden el servicio, de la luz, agua...etc no varían según el número de personas vacunadas, por lo que son costes fijos. En cambio, el coste de la vacuna propiamente dicha sí varía según la población vacunada, por lo que se denominan costes variables. De todos modos, los costes son fijos hasta un cierto límite, porque el número de personas a vacunar puede ser tan elevado y la sobrecarga de trabajo tan importante que sea preciso contratar a más personal, por ejemplo, y el hasta entonces coste fijo del personal sanitario se ha convertido en variable.

Una advertencia útil antes de comenzar con la cuantificación de los costes es que no resulta práctico invertir grandes cantidades de tiempo y esfuerzo en considerar todos y cada uno de los costes, incluso los de muy poca cuantía, ya que es poco probable que modifiquen de forma importante los resultados del estudio.

La siguiente clasificación ayuda a tener en cuenta los costes más relevantes:

- *Costes directos*: son aquellos que se pueden atribuir directamente a una determinada actividad. Son costes de organización y funcionamiento, que a su vez pueden ser variables o fijos. Son ejemplos de éstos el coste de estancia hospitalaria, del tiempo médico, de una determinada prueba diagnóstica, de los medicamentos y otros costes del cuidado de la salud. También hay cos-

tes directos “no sanitarios”, entre los que se cuentan el transporte al hospital, servicios sociales, cuidados en casa, etc.

- *Costes indirectos*: son aquellos indirectamente relacionados con la alternativa que se estudia, como costes de pérdida de producción porque los pacientes participan en el programa. Por ejemplo, el tiempo extra fuera del trabajo para recibir sesiones de hemodiálisis.
- *Costes intangibles*: son aquellos de muy difícil traducción monetaria, como el dolor y el sufrimiento del paciente y su familia debido a la participación en el programa. Aunque no se puedan cuantificar, sí suelen citarse expresamente, pues pueden ser de tal importancia, que a pesar del resultado de la evaluación económica, orienten la decisión en sentido contrario.

Los costes (y los beneficios) no son directos o indirectos en sí mismos, sino dependiendo del marco en el que se analizan. Además existe controversia sobre la terminología en el sentido de que tan directamente están relacionados con un proyecto los costes de un programa de vacunación, por ejemplo, como los de las pérdidas de producción por participar en el programa. Si los individuos no acudieran a vacunarse no se podría llevar a cabo tal programa. Por todo ello, se tienden a sustituir los conceptos de costes directos e indirectos por “variaciones en la disponibilidad de recursos”. En cualquier caso, lo importante no es el nombre que se dé a los costes, sino el acuerdo respecto a su inclusión en el análisis.

La valoración de los costes es una de las tareas más complejas. Si existe mercado para los bienes y servicios que se están midiendo, se acepta como coste el precio de mercado (el precio de un medicamento, por ejemplo), pero los servicios hospitalarios públicos no tienen mercado, y ante esta situación se suelen utilizar costes medios, contabilizando de forma detallada todos los recursos utilizados por el paciente durante el proceso que se pretende medir: consumo de medicamentos, pruebas diagnósticas, costes de personal, etc.

Otro tema controvertido es la contabilización del tiempo de los pacientes, de su capacidad productiva. Si se basa en el cálculo del salario medio de cada uno de los individuos de la población en estudio, no se valorarían determinados colectivos como los jubilados o las amas de casa, que no perciben un salario. Una solución parcial al problema es adoptar un criterio general, valorar la hora de todo el mundo según el salario mínimo interprofesional.

Análisis coste-beneficio (ACB)

¿Cómo se podría comparar la eficiencia de un programa de vacunación contra la hepatitis B con un programa de detección precoz del cán-

cer de mama? Es necesario tener una unidad común de los resultados, y es más, si se pretenden comparar los resultados con los costes (para saber si un programa es rentable en sí mismo), todos ellos deben medirse en las mismas unidades (en este caso pesetas).

El análisis coste-beneficio mide tanto los costes como los beneficios sobre la salud en unidades monetarias. La expresión que resulta de un análisis coste-beneficio es la siguiente: “el programa X tiene una rentabilidad de n pesetas por cada peseta invertida” o bien “el beneficio neto del programa X es de Y pesetas”. La alternativa de elección es aquella que mayores beneficios reporta dado un coste. El resultado se puede expresar tanto en forma de la diferencia entre costes y beneficios (valor actual neto VAN) como mediante un cociente (índice coste-beneficio):

$$\begin{aligned} \text{VAN (valor neto actualizado)} &= \sum(B-C) \\ \text{Índice Beneficio/Coste bruto} &= \sum B / \sum C \\ \text{Índice Beneficio/Coste neto} &= \sum(B-C) / \sum C \end{aligned}$$

Un programa es eficiente cuando el VAN es mayor que 0, el índice bruto mayor que 1 o el índice neto mayor que 0, y tanto más eficiente cuanto más se aleje del 0 ó del 1 según el caso.

De modo similar a la clasificación de los costes se puede actuar con los beneficios y así se distinguen:

- *Beneficios directos*: son los ahorros en el coste de los cuidados de la salud porque el programa hace a los individuos más saludables y ellos, en consecuencia, usan menos recursos para el cuidado de la salud (acuden menos al hospital, toman menos medicinas, etc.). También suelen llamarse costes negativos.
- *Beneficios indirectos*: son las ganancias de producción para la sociedad, porque más personas están bien, viven o son capaces de volver a trabajar. Es el valor monetario de la actividad neta ganada por la sociedad en su conjunto (no exclusivamente la del sector sanitario) gracias al aumento de las productividades laborales que deberían seguirse (en una situación de pleno empleo) a las incapacidades reducidas o a las muertes evitadas gracias a las prestaciones sanitarias. Hay que tener en cuenta el hecho de que al alargar la vida a una persona, no sólo puede producir más, sino también consumir más (más alimentos, más ropa, más cuidados sanitarios).
- *Beneficios intangibles*: son aquellos que no pueden ser valorados fácilmente en unidades monetarias. Representan beneficios de este tipo el tiempo de ocio ganado gracias a la prevención de una enfermedad o al evitar una incapacidad, la reducción del dolor y sufrimiento del paciente y su familia como consecuencia de su mejora en la salud.

Existe la misma polémica que con respecto a los costes, y es que la terminología de beneficios directos o indirectos puede inducir a confusión, por lo que ahora se adopta más la denominación de “variaciones en la salud”.

La principal ventaja de este tipo de análisis radica en que además de agregar en una única dimensión monetaria múltiples consecuencias de diferente naturaleza, permite comparaciones con cualesquiera otros proyectos públicos, siempre y cuando hayan utilizado la misma metodología: se pueden comparar diversos programas sanitarios con efectos diferentes o incluso comparar con otros programas fuera del campo sanitario que también han utilizado el análisis coste-beneficio y se puede comparar un programa con la alternativa de no realizarlo (lo que no quiere decir que no tenga consecuencias) y conocer de este modo su eficiencia neta.

La limitación más importante de este tipo de análisis es la dificultad de expresar en dinero cuestiones relacionadas con la salud (muchas de las cuales son intangibles), y por ello puede dar lugar a inequidades, ya que favorece aquellos programas que tienen gran impacto sobre el desarrollo económico.

“Francamente, no creo que tal procedimiento sea posible, ciertamente yo no lo veré. El problema es que diferentes productos no guardan relación entre sí. Estar muerto no es, ni nunca será, diez veces, o cien veces, peor que sufrir algún tipo de incomodidad en la marcha de la actividad normal” (Doll, 1974).

Sin embargo, lo que parece incomodar a algunas personas no es tanto que la valoración monetaria de la vida humana tenga un límite, sino la explicitación de dicha limitación. Si un programa que podría reducir la mortalidad a un coste de un millón por un año de vida ganado no se lleva a cabo por falta de presupuesto, se puede inferir que la valoración de un año de vida ganado implícita en dicha decisión es de menos de un millón de pesetas.

Un ejemplo de análisis coste-beneficio es la evaluación económica del programa de detección precoz de enfermedades metabólicas (fenilcetonuria e hipotiroidismo congénito) en la Comunidad Autónoma Vasca (Mugarra I, Cabasés J. *Análisis coste-beneficio del programa de detección precoz de enfermedades metabólicas en la Comunidad Autónoma Vasca. Gaceta Sanitaria* 1990;19:140-144). Respecto a los costes se han tenido en cuenta los relacionados con la detección, diagnóstico y tratamiento precoz de las dos enfermedades estudiadas. No se han tenido en cuenta los costes de pérdida de trabajo productivo o en el hogar, y el tiempo de ocio perdido por los padres en las actividades relacionadas con el programa. Respecto a los beneficios, se ha estimado el ahorro de recursos educativos y asistenciales (beneficios directos), así como el tiempo de trabajo productivo ganado por los individuos que han logrado evitar la deficiencia men-

tal como consecuencia del programa (beneficios indirectos). Se enumeran (aunque no se cuantifican) los beneficios intangibles, como el valor que la sociedad otorga a prevenir la subnormalidad. El resultado en forma de índice beneficio-coste bruto fue de 1,71, lo que significa que por cada peseta invertida se ganan 1,71 pesetas, resultaba ser un programa eficiente. Si el resultado se expresa en forma de VAN, éste es de 45.634 pesetas, lo que quiere decir que los beneficios exceden a los costes en 45.634 pesetas. Si la tasa de descuento aplicada fuese del 7% (tema que se explica más adelante en este capítulo), el índice beneficio-coste bruto es 0,80 y el VAN -12.430, dando a entender que se trataría de un programa no rentable, ineficiente, porque los costes exceden a los beneficios en 12.430 pesetas, o lo que es lo mismo, por cada peseta invertida sólo se recuperan 0,80 pesetas.

Análisis coste-efectividad (ACE)

Es la forma de evaluación económica más frecuentemente utilizada en el sector sanitario.

En este tipo de análisis los beneficios son medidos en unidades clínicas habituales: reducción en la incidencia de una enfermedad (casos prevenidos por una vacuna), reducción de los factores de riesgo (milímetros de mercurio que disminuye la tensión arterial), casos correctamente diagnosticados, la reducción de la mortalidad general o específica por causas, años de vida ganados, años libres de enfermedad o de incapacidad ganados (días de dolor evitados).

Mediante este tipo de análisis sólo se pueden comparar programas con un resultado común, que puede variar en magnitud entre las diferentes alternativas. Aquí radica el principal inconveniente de este tipo de análisis: no se puede conocer si un programa sanitario es eficiente en sí mismo, no se puede comparar un programa con la alternativa de no llevarlo a cabo, sólo se puede decir que una alternativa es más eficiente que otra siempre y cuando los efectos se midan del mismo modo. Mediante el ACE no se puede decidir si iniciar, continuar o detener programas, ni sobre el nivel de provisión óptimo. Los resultados de los estudios coste-efectividad simplemente ofrecen el método o camino menos costoso para lograr el objetivo deseado.

El resultado de un análisis coste-efectividad se expresa como un índice, el cociente entre los costes extras de una alternativa respecto a otra y el cambio en los resultados (análisis incremental). Si por ejemplo, la efectividad se mide en esperanza de vida (EV), el índice resultante de comparar la alternativa A con la B sería:

$$\text{Índice C/E} = \frac{\text{Costes de A} - \text{Costes de B}}{\text{EV de A} - \text{EV de B}}$$

Cuando se trata de efectos logrados en condiciones ideales de actuación médica (pruebas de laboratorio o ensayos clínicos) se utiliza la denominación análisis coste-eficacia, y el término efectividad se limita a la consecución de unos objetivos en condiciones reales.

Un ejemplo de este tipo de análisis es la evaluación de la eficacia del tratamiento con anticuerpos monoclonales antiendotoxina en la sepsis por gramnegativos (*Badía X, Segú L, García F, Rovira J. Análisis coste-eficacia del tratamiento con anticuerpos monoclonales antiendotoxina en la sepsis por gramnegativos. Med Clin (Barc) 1993;100: 84-89*). Las alternativas a comparar son: 1) no tratar a los pacientes con sepsis (se aplica el tratamiento convencional), 2) tratar a todos los pacientes con sepsis y 3) tratar sólo a los pacientes que presentan shock séptico. Las probabilidades asignadas al árbol de decisiones se recogen de un ensayo aleatorio controlado (por ello es un análisis coste-eficacia más que un análisis coste-efectividad). Los resultados sobre la salud son los años de vida ganados. Los costes que se han tenido en cuenta son los del fármaco y costes adicionales por estancias hospitalarias. El resultado fue que el coste por año de vida ganado adicional de tratar a todos los pacientes con sepsis fue de 859.288 pesetas, y el de tratar a los pacientes con shock séptico de 293.810 pesetas, ambos tomando como referencia la alternativa del tratamiento convencional (sobre qué alternativa utilizar como referencia se discute más adelante, en los apartados “análisis incremental” y “asignación de recursos”).

Análisis coste-utilidad (ACU)

Un inconveniente importante del análisis coste-efectividad es que muchas veces los programas tienen efectos clínicos diversos, como mortalidad y morbilidad evitada por ejemplo, y además no tiene en cuenta los aspectos subjetivos de los pacientes. Ambos problemas pueden solucionarse en gran medida mediante el análisis coste-utilidad.

En el análisis coste-utilidad, la medida de los resultados sobre la salud, como su nombre indica, es la utilidad. La utilidad expresa la preferencia de los individuos por los diferentes estados de salud. Es una técnica relativamente nueva, y que permite ajustar por calidad de vida a la vez que proporciona un denominador común para comparar costes y resultados de programas diferentes. La medida más comúnmente utilizada han sido los AVAC (años de vida ajustados por calidad) o QALY en la terminología anglosajona (*Quality adjusted life years*).

El resultado de un análisis coste-utilidad, al igual que en el ACE, se expresa como un índice, el cociente entre los costes extras incurridos y el cambio en los AVAC. Si se comparan los programas A y B el resultado puede ser:

$$\text{Índice C/U} = \frac{\text{Costes de A} - \text{Costes de B}}{\text{AVAC de A} - \text{AVAC de B}}$$

El análisis coste-utilidad, al ser derivado del análisis coste-efectividad cuenta con algunos de sus inconvenientes: no se puede conocer si un programa es eficiente en sí mismo, sólo se puede decir que una alternativa es más eficiente que la otra; la utilidad es con frecuencia difícil de determinar y la sociedad hipervalora la utilidad de los servicios sanitarios directamente relacionados con la supervivencia.

En una evaluación económica del trasplante de pulmón, se comparan costes y resultados sobre la salud de un grupo transplantado con un grupo control en lista de espera (*Ramsey SD, Patrick DL, Albert RK, Larson EB, Wood DE, Raghu G. The cost-effectiveness of Lung Transplantation. Chest 1995; 108:1594-1601*). Cuando la efectividad se mide en esperanza de vida (análisis coste-efectividad), ambas alternativas no difieren significativamente (5,89 versus 5,32 años), con lo que no sería eficiente realizar el trasplante, en cambio, cuando lo que se comparan son años de vida ajustados por calidad (análisis coste-utilidad), estas diferencias se tornan significativas (4,71 vs 3,46 AVAC), dando un resultado de 177 mil dólares por AVAC ganado tras el trasplante.

Análisis de minimización de costes (AMC)

En este tipo de análisis se consiguen los mismos resultados sobre la salud con las diferentes alternativas, por lo que lo único que determinará la elección es el coste de cada una. El AMC puede considerarse un tipo especial de ACE en el que los efectos de las diversas opciones se consideran similares.

La limitación más importante de este tipo de análisis es que, en la práctica, las efectividades de las diferentes opciones difícilmente son equivalentes. Por otro lado, al ser un derivado del ACE o del ACU, cuenta con algunos de sus problemas: no informa sobre si los costes exceden o no el valor monetario de las consecuencias sobre la salud, no se puede responder a la pregunta de si es más eficiente llevar o no llevar a cabo un programa, y no es posible hacer comparaciones con otros proyectos cuya efectividad sea medida de un modo diferente.

Por ejemplo, se plantea si la administración domiciliaria de inmunoglobulina endovenosa puede ahorrar costes frente a su administración hospitalaria (*Rodríguez M, Procupet A, Heras J. Análisis coste-efectividad de la administración domiciliaria de inmunoglobulina intravenosa frente a su administración hospitalaria. Med Clin (Barc) 1991; 96: 47-51*). Puesto que la administración de la inmunoglobulina es igualmente efectiva independientemente de dónde se administre, sólo entran en comparación los costes (costes directos de la administración del medicamento, estancia en el hos-

pital, materiales y equipo sanitario y costes indirectos de las horas de trabajo perdidas por los acompañantes). Se concluye que la administración domiciliaria puede ahorrar 250.000 pesetas por paciente el primer año y 265.000 pesetas en años sucesivos. Además, esta opción tiene importantes beneficios intangibles: mayor independencia, situación más confortable, menor sensación de enfermedad y evita el estrés de acudir al hospital.

Es necesario diferenciar este tipo de análisis del “análisis de costes” en el que no se tienen en consideración las consecuencias (los resultados sobre la salud) de las diferentes alternativas, sino sólo los costes.

VALOR MONETARIO DE LA VIDA HUMANA

El objetivo ideal de toda evaluación económica es saber si un programa es eficiente en sí mismo, si reporta unos beneficios superiores a los costes. Esto sólo se consigue mediante el análisis coste-beneficio, para el cual es necesario dar un valor monetario a los resultados sobre la salud. En este sentido se han desarrollado diferentes métodos:

- *Valor que la sociedad da a la vida humana.* Se puede observar qué se ha gastado la sociedad en los programas encaminados a salvar vidas humanas, y cuántas vidas ha logrado salvar. Si mediante una campaña publicitaria sobre el cinturón de seguridad que ha costado ‘x’ pesetas se han logrado salvar ‘y’ vidas, equivaldría a decir que la sociedad valora en x/y pesetas una vida humana.
- *Capital humano.* Se cuantifica lo que el paciente podría llegar a ganar en el resto de su vida. Este método es muy simple y de fácil cálculo, pero está siendo rechazado por varias razones:
 - La salud y la vida tienen valor en sí mismos, independientemente de que los años ganados o vividos con mejor salud se dediquen a actividades productivas.
 - Si se acepta el valor productivo de la vida, lo correcto sería determinar su valor neto, restando de la productividad bruta esperada la inversión y el consumo necesarios para asegurar dicha capacidad (alimentación por ejemplo).
 - Discrimina a los grupos sociales inactivos (¿no vale nada la vida de un jubilado?). Una solución relativa a este problema es valorar la vida, la hora productiva de todas las personas por igual, según el salario mínimo interprofesional, pero este método también asume que la gente no está dispuesta a pagar por salvar su vida más de lo que gana (podrían estar dispuestos a pagar por su vida, por ejemplo, su casa o sus ahorros).
- *Disponibilidad a pagar.* Se trata de preguntar a todos los individuos susceptibles cuánto estarían dispuestos a pagar hoy para evitar una enfermedad, dadas unas características de la misma, y

una probabilidad de que ocurra a lo largo de su vida. Por ejemplo, se le dice a un individuo: suponga que la probabilidad de que usted muera de infarto de miocardio a lo largo de su vida es de 0,01. ¿Cuánto es el máximo que usted estaría dispuesto a pagar para reducir este riesgo a la mitad? Supongamos que la persona responde 200.000 pesetas. Un riesgo del 0,01 equivale a decir que por cada 1000 personas, 10 morirán de infarto de miocardio. Reducir el riesgo a la mitad es evitar 5 muertes por cada 1000 personas. Si la persona entrevistada responde que pagaría 200.000 pesetas por reducir la probabilidad de muerte por infarto a la mitad, equivale a decir que está dispuesto a pagar 40 millones de pesetas ($200.000/0,005$) por salvar su vida. Si la misma pregunta se realiza a un gran grupo de personas se puede estimar el valor medio que la población da a salvar una vida.

Este método cuenta también con varios problemas:

- Se precisa un importante manejo de probabilidades que resulta difícil sobre todo si éstas son pequeñas.
- La disponibilidad a pagar aumenta conforme aumenta el riesgo de la gente a morir.
- También varía con el nivel de renta (la gente rica está dispuesta a pagar más por salvar su vida que la gente pobre).
- Se está dispuesto a pagar más cuando se trata de un programa que los beneficia a ellos mismos que cuando beneficia a otros.

ACTUALIZACIÓN O DESCUENTO

Los efectos de los programas sanitarios ocurren en distintos momentos del tiempo; por ejemplo, tras un programa de vacunación los beneficios (casos de la enfermedad evitados) se producen a lo largo de varios años, y de forma similar, el tratamiento de los pacientes con insuficiencia renal crónica mediante hemodiálisis requiere el uso de recursos a lo largo de toda la vida del paciente.

Si tuviéramos que cobrar un millón de pesetas ¿cuándo preferimos hacerlo, ahora o dentro de 4 años? La respuesta clara es ahora. Un beneficio presente es preferible a otro de la misma cuantía en el futuro. Si tenemos que pagar un millón de pesetas ¿cuándo preferimos hacerlo, ahora o dentro de 4 años? Dentro de 4 años, ya que un coste que tiene lugar en el futuro es preferible al mismo coste en el momento actual. Dos aspectos hacen que se ponderen los costes y beneficios futuros en menor medida que los presentes. Uno es la inflación: con la misma cantidad de dinero se pueden comprar más cosas en 1996 que en el año 2001, por ejemplo, y en segundo lugar y más importante, es que un dinero no gastado ahora puede ser invertido para producir mayor cantidad de dinero en un futuro.

Por lo tanto, costes y beneficios que ocurren en distintos períodos de tiempo no son estrictamente conmensurables. El valor actual de un beneficio (o un coste) futuro será inferior en una determinada proporción al mismo beneficio (o coste) en el momento actual. El factor de corrección que se aplica es la “tasa marginal de preferencia temporal” (o tasa de descuento) y el proceso por el cual se hacen homogéneos valores referidos a diferentes intervalos de tiempo “actualización o descuento”. La formulación general es del siguiente modo:

$$A = \sum^n \frac{X}{(1+r)^t}$$

donde A es el beneficio o coste actualizado,
 X es el beneficio o coste dentro de un período t,
 t es el período de tiempo en obtener el beneficio,
 r es la tasa de descuento y
 n son los diferentes períodos de tiempo de obtención de los beneficios.

En la tabla 32 se ilustra un ejemplo: el programa A tiene unos costes de 20.000 pts. anuales durante 10 años y el programa B tiene unos costes anuales de 15.000 pts. durante 15 años. Aparentemente es más costoso el programa B porque la suma total es superior (225.000 vs 200.000). En cambio, si la tasa de actualización es del 6%, 20.000 pts. anuales durante 10 años son equivalentes a 147.202 pts. actuales, y 15.000 pts. anuales durante 15 años sería lo mismo que 145.683 pts. actuales, con lo que la alternativa B pasa a ser más barata.

Tabla 32. Actualización de los costes de dos proyectos alternativos.

	Alternativa A	Alternativa B
Coste anual	20.000 pts./año	15.000 pts./año
Duración del programa	10 años	15 años
Suma de costes anuales	200.000 pts.	225.000 pts.
Tasa de descuento	6%	6%
Valor actual	147.202 pts.	145.683 pts.
Valor actual alternativa A =	$\frac{20.000}{(1+0,06)^1} + \frac{20.000}{(1+0,06)^2} + \dots + \frac{20.000}{(1+0,06)^{10}} = 147.202 \text{ pts.}$	
Valor actual alternativa B =	$\frac{15.000}{(1+0,06)^1} + \frac{15.000}{(1+0,06)^2} + \dots + \frac{15.000}{(1+0,06)^{15}} = 145.683 \text{ pts.}$	

El problema se suele presentar a la hora de elegir la tasa de descuento. Para hacer los resultados comparables se ha acordado una tasa del 6% para los estudios a nivel nacional. Otro importante dilema es que

conforme el futuro es más lejano, la incertidumbre respecto a la tasa de descuento a utilizar es mayor. En cualquier caso siempre es conveniente realizar un análisis de sensibilidad ofreciendo diferentes resultados según un rango de tasas.

Un tema controvertido es el de la actualización de los beneficios cuando éstos no son medidos en unidades monetarias. La razón de que los años de vida ganados, por ejemplo, se descuenten, no es el hecho de que tales años puedan ser invertidos para producir más años. Tampoco se trata de que la vida en un futuro sea menos valiosa que en el presente. Las razones más importantes para descontar los beneficios futuros son, en primer lugar, que se han puesto en relación a unos costes que sí han sido actualizados, y en segundo lugar, por coherencia con otras formas de análisis que valoran los beneficios en unidades monetarias: si en el ACB los efectos se miden en pesetas y se descuentan, el hecho de que en el ACE y en el ACU no se midan en unidades monetarias (aunque se podría hacer a través de la disponibilidad a pagar, por ejemplo) no justifica que no se descuenten. De hecho, aplicando un criterio de consistencia, se debe aplicar la misma tasa de descuento para costes y beneficios.

ANÁLISIS INCREMENTAL

Un error muy frecuentemente encontrado en la literatura, cuando se realiza un análisis coste-efectividad (o coste-utilidad), es la utilización como índice de eficiencia del ratio medio en lugar del ratio incremental.

El ratio coste-efectividad medio es el cociente entre los costes totales de un programa y la efectividad, no precisa una alternativa de comparación, que es la base de toda evaluación económica. El ratio incremental divide el incremento en costes por el incremento en la efectividad que resulta de pasar de una alternativa de referencia (generalmente ya en ejercicio) a otra nueva.

En la tabla 33 se presentan los resultados del análisis coste-efectividad de dos alternativas de prevención de la neumonía por *Pneumocystis Carinii* (PCP) en pacientes con SIDA: el tratamiento con trimetropín-sulfametoxazol (TMP-SMX) y el tratamiento con pentamidina. Se comparan los resultados según de dónde procedan los datos, de la bibliografía o de un ensayo clínico. La alternativa de referencia es la de no hacer prevención, sino tratar a los pacientes cuando desarrollen la neumonía. Si se comparan los ratios medios, aparentemente todas las alternativas son igualmente eficientes, pero si se comparan los ratios incrementales, la profilaxis con TMP-SMX es claramente más eficiente que la profilaxis con pentamidina. Los costes del tratamiento de la neumonía cuando aparezca (alternativa de referencia) ya ocurren y no se trata de eliminarlos, del mismo modo que estos pacientes tienen una determinada esperanza de vida. Lo que se trata de averiguar es en qué medida mejoran (y la palabra mejorar ya implica

un incremento) los resultados sobre la salud (en este caso aumento de los años de vida) y a expensas de qué coste extra. Aún se puede ir más lejos, y es que si se dispusiera de los recursos necesarios para llevar a cabo la profilaxis con pentamidina, ya que es la alternativa que más aumenta la esperanza de vida (según los datos de la literatura), el coste por año de vida ganado sería de 110.880 dólares, porque las comparaciones se deben hacer con la mejor alternativa a la que es preciso renunciar al hacer una determinada elección (coste de oportunidad) y en este caso es la profilaxis con TMP-SMX. Según los datos del ensayo clínico, la alternativa de profilaxis con pentamidina es más costosa y menos efectiva que la profilaxis con TMP-SMX y, por tanto, no crea problemas de elección.

Tabla 33. Análisis coste-efectividad de la profilaxis de la PCP secundaria mediante trimetropin-sulfametoxazol y pentamidina.

	Coste total ^a	Esperanza de vida ^b	Ratio C/E medio	Ratio C/E ^c incremental	Ratio C/E ^d incremental
Bibliografía					
No profilaxis	42.080	1,430	29.427	-	-
TMP-SMX	42.300	2,051	20.624	350	350
Pentamidina	43.960	2,066	21.278	2.950	110.880
Ensayo clínico					
No profilaxis	42.080	1,430	29.427	-	-
TMP-SMX	42.550	2,076	20.496	720	720
Pentamidina	44.000	1,935	22.739	3.800	No procede

(a) coste total en dólares, (b) los años de vida han sido actualizados,

(c) la base de referencia es la no profilaxis, (d) la base de referencia es la alternativa inmediatamente inferior.

Tomado de: Freedberg KA, Hardy WD, Holzman RS, Tosteson A, Craven DE. Validating Literature-based Models with Direct Clinical Trial Results: The Cost-Effectiveness of Secondary Prophylaxis for PCP in AIDS patients. *Med Decis Making* 1996;16:29-35.

En términos generales, si la alternativa de referencia es barata y muy efectiva, aunque el incremento en el coste sea importante, éste se diluye entre toda la efectividad, y el coste medio puede no ser muy elevado, dando la falsa sensación de que la alternativa en estudio es eficiente. Así, si sólo se comparan ratios medios puede estar disparándose el gasto público aunque apenas se añada bienestar a la población. Si los ratios incrementales son estimados a partir de un programa basal caro y con escasa efectividad, se está, artificialmente, aumentando el incremento en la efectividad y disminuyendo los costes incrementales, pero en la evaluación económica, la comparación se realiza con la mejor alternativa en la que se podrían haber empleado los recursos.

En la bibliografía se suelen utilizar indistintamente los términos coste incremental y coste marginal, aunque existen diferencias de matices: el coste incremental es el coste adicional que un programa o servicio impone respecto a otro y el coste marginal es el cambio en el coste total producido por el incremento o decremento de una unidad del servicio o programa. En la tabla 34 se presentan los resultados del análisis coste-efec-

tividad de la determinación de ferritina sérica de cribaje en los exámenes periódicos de salud en las mujeres en edad fértil. Suponiendo que el programa de cribaje comienza a los 20 años, éste se puede aplicar anualmente durante los primeros 5 años, 10 años, 15 años, 20 años, 25 años, o durante toda la vida fértil, hasta la menopausia (a la edad media estimada de 50 años). La medida de efectividad son los años con ferropenia evitados. El ratio medio utiliza como alternativa de referencia la realización del hemograma (alternativa en uso), en cambio, el ratio marginal utiliza como referencia la alternativa de duración inmediatamente menor. Así se observa que al pasar de un programa de 15 años de duración a uno de 20, los años con ferropenia evitados sólo son 5 (6.470,3 – 6.465,3) y en cambio los costes extras son de 1.933.178 pesetas (13.719.825 – 11.786.647), por lo que el ratio marginal es muy elevado ($1.933.178/5 = 388.578$ pesetas por año con ferropenia evitado).

Tabla 34. Análisis coste-efectividad según la duración del programa de cribaje de ferropenia mediante la determinación de ferritina sérica.

Duración del programa	Años con ferropenia evitados	Coste incremental	Ratio medio	Ratio marginal
5 años	5.075,9	5.740.374	1.131	1.131
10 años	6.175,6	9.192.822	1.489	3.140
15 años	6.465,3	11.786.647	1.823	8.954
20 años	6.470,3	13.719.825	2.120	388.578
25 años	6.616,3	15.181.598	2.295	10.014
30 años	6.618,2	16.070.009	2.428	461.992

La efectividad (años con ferropenia evitados) y el coste incremental (en pesetas) son por cada 1.000 pacientes.

Tomado de: Bermejo B, Rovira J, Olona M, Serra M, Soriano B, Vaqué J. Análisis coste-efectividad de la determinación de ferritina sérica de cribado en los exámenes periódicos de salud en las mujeres en edad fértil. *Med Clin* 1996;106:445-450.

ANÁLISIS DE SENSIBILIDAD

Rara vez se conocen con exactitud los valores necesarios para realizar la evaluación económica, por lo que los resultados de la misma tampoco son exactos. A menudo se hacen suposiciones o estimaciones sobre la incidencia de una enfermedad, sobre la probabilidad de curación, sobre el coste de la estancia hospitalaria, sobre la tasa de descuento, etc.

El análisis de sensibilidad consiste, en síntesis, en calcular los resultados de la evaluación económica según diferentes valores de los parámetros inciertos, pero debido a que los parámetros que no se conocen con certeza pueden ser muchos, se realiza este análisis con los más relevantes.

Se suelen distinguir cuatro diferentes formas de realizar el análisis de sensibilidad:

- a) *análisis de sensibilidad simple*, en el que dentro de un rango razonable se varía uno o más de los parámetros de la evaluación, para ver cómo afecta a los resultados. Si las variables inciertas

son relativamente pocas, y los resultados son sensibles a sus variaciones, es conveniente realizar el análisis de sensibilidad para dichas variables simultáneamente, por ejemplo de dos en dos.

- b) *análisis de umbral*, que trata de identificar el valor crítico de los parámetros por encima o por debajo de los cuales (valores de cruce) puede una alternativa dejar de ser más eficiente en favor de otra, y por tanto variar las conclusiones de un estudio. Por ejemplo, ¿a partir de qué prevalencia de infección por *Clamidia Trachomatis* es rentable realizar el cribaje de la enfermedad en la población susceptible?
- c) *análisis de valores extremos*, en donde las alternativas se comparan tratando de identificar los casos extremos, el pesimista (alto coste, baja efectividad) y el optimista (bajo coste, elevada efectividad).
- d) *análisis de sensibilidad probabilístico*, que asigna rangos y distribuciones a las variables sometidas a incertidumbre. Por ejemplo, si se trata de una variable que sigue una distribución normal, se realiza el análisis de sensibilidad con los valores correspondientes a la media más y menos dos desviaciones estándar. En caso de que sea posible, es el método preferible por los analistas.

En el análisis coste-efectividad de la determinación de ferritina sérica de cribaje, comentado previamente, las variables utilizadas en la evaluación son el coste de la determinación de ferritina sérica, el coste del hemograma, el coste del tratamiento de las mujeres ferropénicas, la prevalencia de ferropenia en este grupo de la población, la sensibilidad y especificidad del hemograma para el diagnóstico de ferropenia, el grado de cumplimiento del tratamiento, el porcentaje de curación tras el mismo y la tasa de actualización. Se ha realizado el análisis de sensibilidad simple, sumando y restando a cada uno de los parámetros su 25%. El resultado es que el ratio coste/efectividad es sobre todo sensible (presenta la mayor variación) a los cambios en el coste de las pruebas diagnósticas y la prevalencia de ferropenia.

En resumen, el análisis de sensibilidad se pregunta si las conclusiones cambian (si los resultados son sensibles) al cambiar las estimaciones dentro de determinado rango. Si los resultados no cambian mucho, va en favor de una robustez del análisis de decisión.

PRESENTACIÓN DE RESULTADOS

A la hora de presentar los resultados siempre es conveniente hacerlo de forma desagregada, presentando además de los índices finales

de eficiencia, los componentes de costes y de los resultados sobre la salud.

A su vez, va bien identificar todos los componentes del coste que se han tenido en cuenta, y del mismo modo, si un programa presenta diversos efectos sobre la salud, conviene calcular diferentes índices de eficiencia, coste por sintomatología evitada, coste por año de vida ganado y coste por AVAC por ejemplo.

Es preferible presentar los flujos anuales de costes y beneficios a lo largo del período considerado en el estudio, además de los valores ya actualizados.

Los resultados del análisis de sensibilidad, si las variables inciertas son muchas, sólo es necesario presentarlos para las más relevantes y aquéllas que modifican especialmente los resultados.

Además de presentar los resultados sobre una hipotética cohorte (o un individuo), es conveniente estimar también el efecto sobre la población real (la relevante) a la que van dirigidos.

Esta presentación desagregada de los resultados facilita el análisis crítico de las suposiciones (sobre la incidencia de una enfermedad, la mortalidad, la tasa de descuento, etc.), así como la reproducibilidad del análisis o su adaptación a contextos diferentes, sólo modificando los parámetros distintos y recalculando los resultados.

ASIGNACIÓN DE RECURSOS

La decisión de a qué programa se destinan los recursos disponibles depende de dos factores fundamentalmente:

1. Si las alternativas en comparación son o no excluyentes: un programa de rehabilitación domiciliaria en los pacientes con prótesis de cadera excluye el programa de rehabilitación hospitalaria (se trata de elegir uno de los dos), en cambio, un programa de vacunación de la hepatitis B no excluye un programa de detección precoz del cáncer de mama.
2. De la técnica de evaluación económica:
 - En el análisis de minimización de costes (AMC) es evidente a qué programa se destinan los recursos, al menos costoso. Además, las opciones comparadas serán, normalmente, excluyentes.
 - En el análisis coste-beneficio (ACB), si las opciones son excluyentes (van encaminadas a los mismos resultados sobre la salud), se debe elegir aquella que reporte mayores beneficios (mayor valor actual neto). Si las opciones no son excluyentes, se debe, en teoría, asignar recursos a todos los programas que tienen un beneficio neto positivo, pero si el presupuesto es fijo,

se seleccionan las opciones por orden de cuantía de los beneficios netos hasta agotar el presupuesto.

- En el caso del análisis coste-efectividad (ACE) y del análisis coste-utilidad (ACU), si las opciones no son excluyentes, se deben ordenar según el valor creciente del índice coste/efectividad, y asignar los recursos disponibles por este orden hasta agotarlos. Si las opciones son excluyentes, primero se eliminan las opciones dominadas (de menor efectividad y mayor coste que otra opción), dejando sólo las dominantes (de mayor efectividad y menor coste que otra opción). Dentro de éstas, la elección depende de lo que se esté dispuesto a pagar por la ganancia en la efectividad (la alternativa de referencia debe ser la más efectiva de las comparables), siempre y cuando entre dentro del presupuesto disponible. Por ejemplo, en la tabla 33 y basados en los datos de la literatura, si se elige la opción de la pentamidina por ser la más efectiva, quiere decir que se está dispuesto a pagar 110.880 dólares por un año de vida ganado y no los 2.950 dólares que resultan de la comparación con la alternativa de no profilaxis. En el caso del análisis coste-efectividad de la determinación de ferritina sérica de cribaje (Tabla 34), si se está dispuesto a pagar 10.000 pesetas por año con ferropenia evitado, quiere decir que para ser eficiente el programa no puede durar más de 15 años.

Por último, es importante recordar que, aunque la asignación de los recursos se base en criterios de eficiencia, siempre hay quien gana y quien pierde con cada decisión. Existen muchos factores que influyen en el valor relativo de un beneficio en la salud entre los diferentes segmentos de la población. Por ejemplo, los programas orientados a mejorar la salud de los niños son más valorados que cuando los beneficios se distribuyen entre la gente mayor; los programas con claros beneficiarios o víctimas también se valoran más que los programas cuyos beneficiarios no son nunca identificados (por ejemplo prevención de patologías o programas de cribaje). La posición de la institución que dispone de los recursos también es importante en la valoración de la eficiencia de un programa: una institución de cuidados de nivel terciario es más proclive a la implantación de un programa de transplante o una unidad de cuidados intensivos que un programa de cribaje o inmunización. Una institución que tiene a su cuidado un segmento de la población con un nivel de salud precario, puede preferir mejorar el estado de salud de estos individuos que mejorar la de otros con un nivel de salud aceptable, aunque el programa sea menos eficiente.

A pesar de todas las dificultades conceptuales y prácticas que presenta una evaluación económica no debe cundir el desánimo, puesto que las decisiones se toman en todo caso y las decisiones informadas cuentan con una mayor probabilidad de ser eficientes que las no informadas.

RESUMEN

La misión principal del médico es indicar a sus pacientes las mejores pruebas o los mejores tratamientos, sin embargo, cuando los recursos son escasos, lo ético es maximizar el beneficio de todos los posibles pacientes, de la sociedad en su globalidad.

La evaluación económica es la técnica que permite determinar la eficiencia de diferentes alternativas de actuación, comparándolas tanto en términos de costes como de resultados.

Los costes (directos e indirectos) se miden en unidades monetarias, en cambio, los beneficios pueden medirse también en unidades monetarias (análisis coste-beneficio), en unidades clínicas habituales (análisis coste-efectividad) o en calidad de vida (análisis coste-utilidad).

La única forma de determinar la eficiencia de una opción en sí misma es mediante el análisis coste-beneficio, por ello, se han ideado formas de dar un valor monetario a los resultados sobre la salud que se miden en el análisis coste-efectividad o coste-utilidad y son principalmente la disponibilidad a pagar y la capacidad productiva de los individuos.

Costes y beneficios que se producen en diferentes momentos no son comparables, ya que los costes futuros son preferibles a costes de la misma cuantía en el presente, y ante beneficios de igual cuantía son preferibles los presentes que los futuros. El problema se soluciona mediante la actualización o descuento, que consiste en aplicar un factor de corrección a los costes y beneficios futuros con el fin de buscar su equivalencia en el momento actual. Los beneficios sobre la salud, aunque no se midan en unidades monetarias también deben actualizarse, y de hecho se aplica la misma tasa de descuento que para los costes.

Se debe realizar un análisis incremental, comparando la efectividad añadida por una alternativa respecto a otra con el incremento en los costes que se ha producido. La alternativa de referencia debe ser la más efectiva, aquella a la que es preciso renunciar al llevar a cabo la elección (coste de oportunidad).

Por último, ante todo ejercicio de evaluación económica se debe realizar un análisis de sensibilidad para ver cuan robustos son los resultados dependiendo de diferentes valores de los datos inciertos utilizados en el análisis.

Los resultados del análisis se deben presentar de forma desagregada, con el fin de que se pueda realizar un análisis crítico de los mismos y que el estudio sea reproducible en otros contextos.

PROBLEMAS

- Las dos alternativas de tratamiento de la insuficiencia renal crónica son la hemodiálisis y el trasplante de riñón. Suponga que el coste total de la primera alternativa es de 1.900.000 pesetas anuales por paciente dializado, mientras que el coste del trasplante (incluidas las revisiones) es de 8.900.000 pts. por paciente.

Asumiendo las siguientes hipótesis:

- todos los pacientes dializados son susceptibles de trasplante y hay órganos suficientes para todos.
 - la esperanza de vida para los tratados con diálisis es de 10 años, en tanto que para los transplantados se estima en 15 años.
 - la calidad de vida para los transplantados se estima en 0,70 para el primer año, reduciéndose progresivamente hasta 0,60 al final de su vida.
 - la calidad de vida para los dializados se ha estimado en 0,65 durante los primeros 9 años, reduciéndose en un 50% durante el último año de su vida.
 - los individuos transplantados tienen unos ingresos esperados de 1.200.000 pts. anuales, y los individuos dializados de 1 millón.
 - la tasa de descuento es del 6%
- ¿Cuál es el resultado del análisis coste-beneficio?
 - ¿Cuál es el resultado del análisis coste-efectividad?
 - ¿Cuál es el resultado del análisis coste-utilidad?

- En la tabla 35 se presentan el número de casos de cáncer de colon detectados y los costes totales (en dólares de 1975) en 10.000 individuos según el número de pruebas de sangre oculta en heces realizadas. ¿Cuál es el coste de detectar un caso de cáncer de colon al realizar 6 pruebas de sangre oculta en heces?

Tabla 35. Análisis coste-efectividad de la detección del cáncer de colon.

Número de pruebas	Casos detectados	Coste total
1	65,9469	77.511
2	71,4424	107.690
3	71,9003	130.199
4	71,9385	148.116
5	71,9417	163.141
6	71,9420	176.331

Tomado de: *Newhauser D, Lewicki AM. What do we gain from the sixth stool guaiac? N Eng J Med 1975; 293: 226-228.*

- Suponga que el riesgo de morir por cáncer es del 3% y usted pregunta a un grupo de individuos cuánto es el máximo que estarían dispuestos a pagar por reducir este riesgo al 1%. Si en media, el grupo contesta que estarían dispuestos a pagar 2 millones de pesetas ¿En cuánto estiman el valor monetario de su vida?

4. En la tabla 36 se presentan los costes (en millones de pesetas) y la efectividad (años de vida ganados) de 6 alternativas. Si se trata de programas no excluyentes y se dispone de un presupuesto de 100 millones de pesetas. ¿Qué programas se llevarían a cabo? Si, por el contrario, son alternativas excluyentes ¿Qué alternativa es la elegida?

Tabla 36. Asignación de recursos.

Alternativa	Coste neto	Años de vida ganados	Relación coste-efectividad
A	30	30	1
B	30	20	1,50
C	55	70	0,79
D	20	40	0,50
E	30	50	0,60
F	50	75	0,67

SOLUCIONES A LOS PROBLEMAS

PROBABILIDAD

1. El sexo de un hijo es independiente del de los otros hijos nacidos de la pareja. Si la probabilidad de que nazca una niña es 0,5 la probabilidad de que sean 4 niñas es $(0,5)^4 = 0,0625$.

Como son sucesos independientes la probabilidad de que el cuarto hijo sea un niño dado que las tres primeras son niñas es igual a la probabilidad de que sea un niño = 0,5.

2. Como una alarma es independiente de la otra, la probabilidad de que fallen las dos (intersección) es igual al producto de probabilidades = $0,001*0,001 = 0,000001$.
3. Concordancia diagnóstica entre el Dr. A y el Dr. B.

a) Cuando dos sucesos son independientes la probabilidad de la intersección es el producto de probabilidades:

$$p(A \cap B) = p(A) * p(B)$$

y en este caso $0,08 \neq 0,10 * 0,17 = 0,017$ luego no son independientes.

b) La probabilidad de que el doctor B confirme el resultado positivo del doctor A es:

$$p(B/A) = p(B \cap A) / p(A) = 0,08 / 0,10 = 0,8 \text{ (80\%)}$$

c) El porcentaje de casos en que el Dr. B contradice el diagnóstico negativo del Dr. A es:

$$p(B) = p(B \cap A) + p(B \cap \bar{A}) \text{ de donde } p(B \cap \bar{A}) = 0,17 - 0,08 = 0,09$$

$$p(B/\bar{A}) = p(B \cap \bar{A}) / p(\bar{A}) = 0,09 / 0,9 = 0,1 \text{ (10\%)}$$

d) El porcentaje de pacientes cuya mamografía es positiva (unión), son remitidas para realizar más pruebas y éstas son:

$$p(A \cup B) = 0,10 + 0,17 - 0,08 = 0,19 \text{ (19\%)}$$

4. La transformación de odds en probabilidades es la siguiente:

$$p = \text{odds}/1+\text{odds} = \frac{1/9}{1/9 + 1} = 0,1$$

Una odds de 1: 9 es lo mismo que una probabilidad del 10%

5. Del mismo modo que en el ejercicio anterior, la odds se puede calcular en función de las probabilidades:

$$\text{Odds} = p/1-p = \frac{0,8}{1-0,8} = 4$$

Se expresa como 4:1 queriendo decir que por cada cuatro veces que ocurre un suceso, una vez no ocurre.

6. Se dispone de los siguientes datos:

$$p(E+) = 0,10 \quad p(E-) = 0,90$$

$$p(T+/E+) = 0,80$$

$$p(T+/E-) = 0,05$$

Lo que pide el problema es $p(E+/T+)$

Por el teorema de Bayes se pueden invertir las probabilidades condicionadas de tal modo que:

$$p(E+/T+) = \frac{p(T+/E+) * p(E+)}{p(T+/E+) * p(E+) + p(T+/E-) * p(E-)} =$$

$$= \frac{0,80 * 0,10}{0,80 * 0,10 + 0,05 * 0,90} = 0,64$$

En un individuo en el que la prueba diagnóstica ha resultado positiva, la probabilidad de tener la enfermedad es del 64%.

También se puede construir una tabla 2x2 (Tabla 37) con una hipotética muestra de 1000 individuos de los que el 10% (100) presentan la enfermedad y el resto (900) no la presentan; de los enfermos, en el 80% (80) la prueba resulta positiva y de los no enfermos en el 5% resulta positiva ($900 * 0,05 = 45$). El resto de las casillas se calculan restando de los totales.

Por la tabla se observa que la prueba ha resultado positiva en 125 personas, de las que 80 presentan la enfermedad.

$$p(E+/T+) = 80/125 = 0,64$$

Tabla 37. Tabla 2x2.

		Enfermedad		
		+	-	
Prueba +	80	45	125	
Prueba -	20	855	875	
	100	900	1000	

ÁRBOLES DE DECISIÓN

1. El tratamiento preferible será aquel que, en término medio, presente una esperanza de vida mayor. Se puede elaborar un árbol de decisiones como el de la figura 36. Como los pacientes mueren a lo largo del año, se supone que lo hacen justo en medio, el que muere a lo largo del primer año vive medio año, el que muere a lo largo del segundo año vive un año y medio, y así sucesivamente. La esperanza de vida promedio de cada alternativa es la suma de los productos de los resultados finales por la probabilidad de que ocurran:

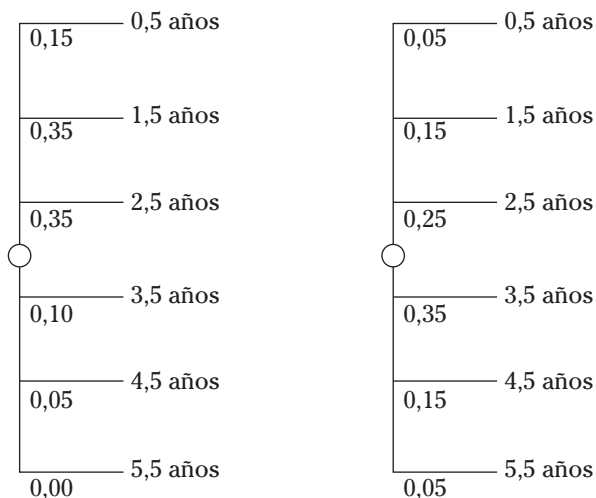


Figura 36. Árbol de decisión.

Tratamiento A:

$$0,5 \cdot 0,15 + 1,5 \cdot 0,35 + 2,5 \cdot 0,35 + 3,5 \cdot 0,10 + 4,5 \cdot 0,05 + 5,5 \cdot 0 = 2,05 \text{ años en media.}$$

Tratamiento B:

$$0,5 \cdot 0,05 + 1,5 \cdot 0,15 + 2,5 \cdot 0,25 + 3,5 \cdot 0,35 + 4,5 \cdot 0,15 + 5,5 \cdot 0,05 = 3,05 \text{ años de esperanza de vida media.}$$

2. Para solucionar este problema se puede elaborar un árbol como el representado en la figura 37. Se está indiferente entre operar y no operar cuando la probabilidad de sobrevivir es igual en una alternativa que en la otra.

$$(p \cdot 0,95) + ((1-p) \cdot 0,95) = (p \cdot 0,85) + (1-p) \cdot 1$$

Si de aquí se despeja la p

$$0,95p + 0,95 - 0,95p = 0,85p + 1 - 1p$$

$$0,95p - 0,95p - 0,85p + 1p = 1 - 0,95$$

$$0,15p = 0,05$$

$$p = 0,05/0,15 = 0,33$$

Si la probabilidad de presentar la enfermedad de Sosika dado que el paciente presenta las manchas de Pinbian, es superior al 33% será preferible operar, si es menor será preferible no operar.

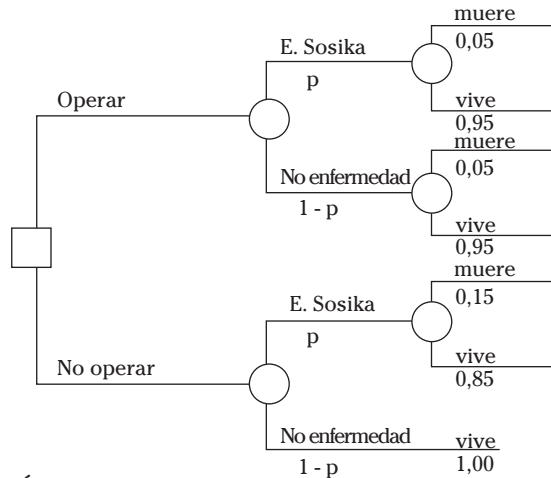


Figura 37. Árbol de decisión.

3. Si se supone una cohorte de 1000 individuos, el coste de prevenir es el coste del tratamiento preventivo a toda la cohorte, y el coste del tratamiento curativo es el del 10% de la población que desarrollará la enfermedad.

$$\text{Coste de prevenir} = 1000 * 100 = 100.000$$

$$\text{Coste de curar} = 1000 * 0,1 * 1.100 = 110.000, \text{ luego es preferible prevenir.}$$

Como se observa en la figura 38, si la eficacia del tratamiento preventivo es sólo del 80%, el coste de la prevención es el del coste preventivo para todos los individuos más el coste del tratamiento curativo para los individuos no protegidos (el 20% en los que la vacuna no resultó efectiva) que desarrollan la enfermedad (el 10%):

$$(1000 * 100) + (1000 * 0,2 * 0,1 * 1.100) = 122.000$$

luego en este caso es preferible curar.

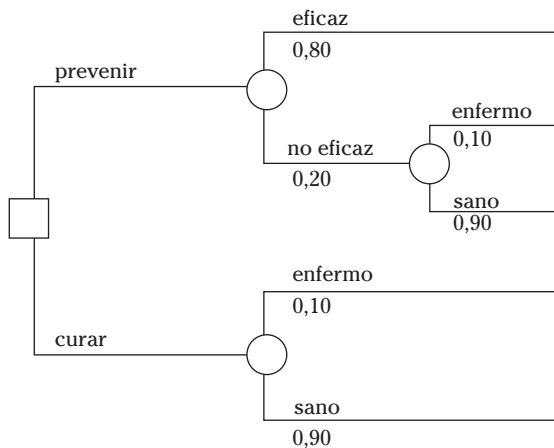


Figura 38. Árbol de decisión.

PROBABILIDAD PREVIA

1. Le inducirá menos a equivocarse si el dato erróneo se lo dice al final que si se lo dice al principio. Se tiende a ser muy conservador con la probabilidad estimada al comienzo del proceso del diagnóstico.
2. La enfermedad no es tan frecuente y estos síntomas bien podrían ser explicados por una viriasis o una infección de orina.
3. Todos los signos son redundantes y no hacen más que reflejar un aumento del gasto cardiaco y un aporte insuficiente de oxígeno a los tejidos.
4. Está utilizando el proceso mental de la disponibilidad. El hecho de haber visto recientemente casos similares que resultan fáciles de recordar influyen en su estimación de la probabilidad. Podría cometerse el error de anclaje, partiendo de una probabilidad más alta de la real. En este caso concreto podría tratarse de un proceso alérgico.
5. Prevalencia. Cuando el paciente es estudiado ya tiene o no la psoriasis. Para estudiar la incidencia de psoriasis en individuos con la enfermedad de Crohn debemos estudiar pacientes con la enfermedad de Crohn sin psoriasis, seguirlos durante un período de tiempo y medir en cuántos pacientes aparece psoriasis.
6. Falso. En primer lugar, el denominador que se ha utilizado no es la población cubierta por el hospital, sino el total de pacientes ingresados en el hospital. En segundo lugar, el autor no ha descartado del numerador a aquellos pacientes que al ser ingresados ya tenían tuberculosis pulmonar (presumiblemente la mayoría, ya que muy pocos pacientes contraen la enfermedad durante la estancia hospitalaria). El autor debería haber concluido que la prevalencia de tuberculosis entre los pacientes ingresados en 1986 fue del 12,7%.

FIABILIDAD

1. En $14+81 = 95$ radiografías, los dos expertos radiólogos se ponen de acuerdo. Ello supone un índice de concordancia global del 86,4% ($95/110$). Pero si los dos radiólogos decidieran por azar en qué radiografías están presentes los signos de neumoconiosis, los resultados serían los de la tabla 38.

Como norma general, los valores esperados se calculan multiplicando los marginales (totales de filas y columnas) y dividiendo por el total. El esperado de la celda 'a' es $22 \cdot 21/110 = 4,2$. El esperado de la celda 'b' es $88 \cdot 21/110 = 16,8$. El esperado de la celda 'c' es $22 \cdot 89/110 = 17,8$ y el esperado de la celda 'd' es $88 \cdot 89/110 = 71,2$. En realidad, en una tabla de 2×2 sólo es preciso calcular uno de los valores esperados, el resto salen por la diferencia.

La concordancia esperada por azar es $= (4,2 + 71,2)/110 = 68,5\%$

$$\text{El índice kappa} = \frac{86,4 - 68,5}{100 - 68,5} = 56,8\%$$

Tabla 38. Concordancia entre radiólogos.

Datos observados

		Radiólogo 1		
		Sí	No	
Radiólogo 2	Sí	14	7	21
	No	8	81	89
		22	88	110

Datos esperados por azar

		Radiólogo 1		
		Sí	No	
Radiólogo 2	Sí	4,2 (a)	16,8 (b)	21
	No	17,8 (c)	71,2 (d)	89
		22	88	110

2. Por azar se esperan los datos de la tabla 39. Sólo es preciso calcular 9 valores, el resto se deducen por la diferencia con los totales de filas o columnas (marginales).

El desacuerdo observado, ponderado por el cuadrado de la diferencia de categorías es:

$$4 \cdot 1 + 3 \cdot 4 + 0 \cdot 9 + 7 \cdot 1 + 6 \cdot 1 + 1 \cdot 4 + 2 \cdot 4 + 5 \cdot 1 + 2 \cdot 1 + 1 \cdot 9 + 0 \cdot 4 + 3 \cdot 1 = 60$$

El desacuerdo esperado por azar, ponderado por el cuadrado de la diferencia de categorías es:

$$7,3 \cdot 1 + 9,3 \cdot 4 + 4,2 \cdot 9 + 9,8 \cdot 1 + 8,1 \cdot 1 + 3,7 \cdot 4 + 9,1 \cdot 4 + 6 \cdot 1 + 3,3 \cdot 1 + 4,9 \cdot 9 + 3,2 \cdot 4 + 4,1 \cdot 1 = 252$$

El índice kappa ponderado es:

$$K_p = 1 - \frac{q_o}{q_e}$$

donde q_o = proporción de desacuerdos real
 q_e = proporción de desacuerdos debidos al azar.

$$\text{luego } K_p = 1 - \frac{60}{252} = 76\%$$

Cuando el sistema de ponderación es el de la diferencia de categorías al cuadrado, el resultado del kappa ponderado coincide con el coeficiente de correlación intraclases.

Valores observados

Tabla 39. Concordancia en el estadiaje del cáncer de mama.

Datos observados

		1ª vez				
		I	II	III	IV	
2ª vez	I	25	7	2	1	35
	II	4	14	5	0	23
	III	3	6	17	3	29
	IV	0	1	2	10	13
		32	28	26	14	100

Datos esperados por azar

		1ª vez				
		I	II	III	IV	
2ª vez	I	11,2	9,8	9,1	4,9	35
	II	7,3	6,4	6,0	3,2	23
	III	9,3	8,1	7,5	4,1	29
	IV	4,2	3,7	3,3	1,8	13
		32	28	26	14	100

3. La diferencia absoluta entre las dos determinaciones, la media del resultado de los dos métodos, la desviación estándar y el coeficiente de variación se presentan en la tabla 40. Un coeficiente de variación medio (de los 5 individuos) del 4,69% indica una fiabilidad muy buena, ya que los datos no difieren mucho según se determinen con uno u otro método.

Tabla 40. Concordancia en la determinación de ferritina sérica.

	Diferencia	Media	DE	CV%
1	2,0	21,0	1,414	6,73
2	-1,0	12,5	0,707	5,66
3	0,0	4,0	0,000	0,00
4	4,0	39,0	2,828	7,25
5	-1,0	18,5	0,707	3,82
				4,69

CV= coeficiente de variación; DE= desviación estándar.

4. El coeficiente de correlación de Pearson es +1. La correlación es perfecta y positiva, conforme aumenta la puntuación con un cuestionario aumenta con el otro, concretamente el segundo cuestionario siempre es 10 puntos superior. En cambio, el coeficiente de correlación intraclases es menor que 1 (0,72) porque en cierto modo tiene en cuenta la diferencia entre las dos puntuaciones. Con este ejemplo se explica por qué no se puede estudiar la fiabilidad mediante el coeficiente de correlación de Pearson.

VALIDEZ DE LAS PRUEBAS DIAGNÓSTICAS

1. Si construimos la tabla de 2x2 (Tabla 41) sobre una población de referencia de 1000 individuos, 200 presentan TEP y 800 no lo presentan (probabilidad del 20%). De los 200 individuos con TEP, la prueba resulta positiva en el 95% (190 casos) y negativa en el 5% (10 casos). De los 800 individuos sin TEP la prueba resulta positiva en el 25% (200 casos) y negativa en el resto (600).

La sensibilidad es la probabilidad de un resultado positivo en los pacientes con la enfermedad:

$$\text{Sensibilidad} = \frac{190}{190 + 10} = 95\%$$

La especificidad es la probabilidad de un resultado negativo en los pacientes sin la enfermedad:

$$\text{Especificidad} = \frac{600}{600 + 200} = 75\%$$

La razón de verosimilitud de un resultado positivo es la probabilidad de un resultado positivo en los enfermos (sensibilidad) dividido por la probabilidad de un resultado positivo en los no enfermos (tasa de falsos positivos o complementario de la especificidad):

$$\text{LR+} = \frac{190/200}{200/800} = 3,8$$

Es 3,8 veces más probable un resultado positivo en un paciente realmente afecto de TEP que en uno sin TEP.

Tabla 41. Tabla 2x2.

	TEP sí	TEP no	
$\text{PO}_2 \leq 90 \text{ mmHg}$	190	200	390
$\text{PO}_2 > 90 \text{ mmHg}$	10	600	610
	200	800	1000

La razón de verosimilitud de un resultado negativo es la probabilidad de un resultado negativo en los enfermos (tasa de falsos negativos o complementario de la sensibilidad) dividido por la probabilidad de un resultado negativo en los no enfermos (especificidad):

$$\text{LR-} = \frac{10/200}{600/800} = 0,067 \quad (1 \text{ a } 15)$$

Es 15 veces más probable un resultado negativo en un paciente sin TEP que en uno con TEP.

La probabilidad de un resultado positivo es la suma de la probabilidad de un resultado positivo en los enfermos (sensibilidad) ponderada por la probabilidad de ser enfermo más la probabilidad de un resultado positivo en los no enfermos (1-especificidad) ponderada por la probabilidad de no presentar la enfermedad:

$$\text{Resultado positivo} = (0,95 \cdot 0,20) + (0,25 \cdot 0,80) = 0,39$$

De todos modos, la solución también se puede hallar mirando la tabla, pues de los 1000 individuos la prueba resulta positiva en 390 (39%).

La probabilidad de un resultado negativo se puede calcular del mismo modo, es la suma de la probabilidad de un resultado negativo en los enfermos (1-sensibilidad) ponderada por la probabilidad de presentar la enfermedad más la probabilidad de un resultado negativo en los no enfermos (especificidad) ponderada por la probabilidad de no presentar la enfermedad:

$$\text{Resultado negativo} = (0,05 \cdot 0,20) + (0,75 \cdot 0,80) = 0,61$$

Y de nuevo este resultado es el mismo que observando la tabla, $610/1000 = 61\%$.

Si la prueba resulta positiva, la probabilidad de presentar la enfermedad es del 48,7% ($190/390$) y si la prueba resulta negativa, la probabilidad de presentar la enfermedad disminuye al 1,64% ($10/610$).

2. La sensibilidad es del 0%, porque nunca da unos valores de hemoglobina suficientemente bajos como para sospechar la presencia de anemia. La especificidad, por el contrario, es del 100%, no puede haber falsos positivos (individuos no anémicos con niveles de hemoglobina bajos).
3. Los signos patognomónicos son aquellos que cuando están presentes indican enfermedad, no hay falsos positivos luego la especificidad es del 100%. Sin embargo no en todos los enfermos están presentes estos signos, la sensibilidad puede tener cualquier valor.
4. Parece que la sensibilidad es sólo del 12% y la especificidad del 8%, pero si desde ahora se decide llamar resultado positivo a lo que antes se denominaba resultado negativo y viceversa, la sensibilidad pasa a ser del 88% y la especificidad del 92%, luego puede ser una prueba muy válida.
5. La probabilidad de que al menos una de las pruebas resulte anormal (puede ser una, dos, tres...hasta 15 pruebas anormales) es uno menos la probabilidad de que todas resulten normales. Debido a que son pruebas independientes, la probabilidad de que todas resulten normales es

el producto de las probabilidades de que cada una de las 15 resulte normal.

$$1 - (0,95 * 0,95 * 0,95 \dots) = 1 - 0,95^{15} = 0,54$$

Llamo la atención del lector sobre la elevada probabilidad de que aparezca al menos un asterisco en una determinación analítica rutinaria.

6. Los costes de aplicar una prueba son:

$$C_{FN} * P * (1-S) + C_{FP} * (1-P) * (1-E)$$

donde:

C_{FN} = costes de no diagnosticar ni tratar a las pacientes enfermas

P = probabilidad de presentar la enfermedad

S = sensibilidad

C_{FP} = coste de tratar a las pacientes sin la infección

E = especificidad

La relación de costes que hace indiferente aplicar una prueba u otra es aquella en la que se igualan los costes netos de las dos pruebas.

Si en la fórmula anterior dividimos todo por C_{FP} resulta que los costes netos de aplicar una prueba son:

$$\frac{C_{FN}}{C_{FP}} * P * (1-S) + (1-P) * (1-E)$$

$$(C_{FN}/C_{FP}) * 0,06 * 0,30 + 0,94 * 0 = (C_{FN}/C_{FP}) * 0,06 * 0,20 + 0,94 * 0,05$$

Si de aquí se despeja $C_{FN}/C_{FP} = 7,8$

Si es mucho más importante (más de 7,8 veces) no dejar de tratar a los pacientes enfermos (C_{FN}) que administrar antibióticos a las mujeres sin infección (C_{FP}) es preferible aplicar la prueba de la determinación enzimática que es la más sensible (menor tasa de falsos negativos). En caso contrario es preferible el cultivo celular.

Si la prevalencia de la enfermedad es baja (6%), sólo por este hecho se precisa una prueba específica (porque el número absoluto de falsos positivos puede ser muy elevado), y para decidir aplicar una prueba más sensible los costes de los falsos negativos deben ser muy importantes (7,8 veces superiores a los costes de los falsos positivos). Si, en cambio, la prevalencia es más elevada (20%), no se precisa tanta especificidad y para decidir aplicar una prueba más sensible los costes de los falsos negativos no deben ser tan importantes. En este caso la relación de costes es:

$$(C_{FN}/C_{FP}) * 0,20 * 0,30 + 0,80 * 0 = (C_{FN}/C_{FP}) * 0,20 * 0,20 + 0,80 * 0,05$$

Si de aquí se despeja $C_{FN}/C_{FP} = 2$

Es suficiente con que el dejar de tratar a las pacientes con la infección sea dos veces más importante que tratar a las pacientes sin la infección

para decidir aplicar la prueba más sensible que es la determinación enzimática.

7. Dado que la probabilidad de desarrollar obesidad en la edad adulta es del 18%, con el fin de diagnosticar bien a la gran mayoría (los no obesos) se precisaría mayor especificidad. Por otro lado, es dos veces más arriesgado dejar de tratar a un niño obeso que etiquetar y tratar a uno que no desarrollaría obesidad, y por este hecho se precisaría más sensibilidad. Manejando conjuntamente todos estos datos, será preferible aquella prueba que en valores absolutos menos costes reporte:

$$\text{Costes netos} = C_{FN} * T_{FN} * P + C_{FP} * T_{FP} * (1-P)$$

donde:

C_{FN} = coste por dejar de tratar a los falsos negativos

T_{FN} = tasa de falsos negativos

P = probabilidad de presentar la enfermedad

C_{FP} = coste de tratar a los falsos positivos

T_{FP} = tasa de falsos positivos

Supongamos una muestra de 100 niños:

Percentil 80 del IMC

- Sensibilidad = 85%

- Especificidad = 91%

- T_{FN} = 15%

- T_{FP} = 9%

- N° de falsos negativos = $100 * 0,18 * 0,15 = 2,7$

- N° de falsos positivos = $100 * 0,82 * 0,09 = 7,38$

- Costes netos = $2 * 2,7 + 1 * 7,38 = 12,78$

Percentil 85 del IMC

- Sensibilidad = 71%

- Especificidad = 95%

- T_{FN} = 29%

- T_{FP} = 5%

- N° de falsos negativos = $100 * 0,18 * 0,29 = 5,22$

- N° de falsos positivos = $100 * 0,82 * 0,05 = 4,1$

- Costes netos = $2 * 5,22 + 1 * 4,1 = 14,54$

Es más costosa la segunda prueba, luego es preferible utilizar el percentil 80 del IMC como punto de corte para el diagnóstico de obesidad.

PROBABILIDAD POSTERIOR

1. Los datos de la tabla 2x2 serían los siguientes (Tabla 42):

suponiendo una población de 1000 individuos, si la probabilidad de la enfermedad es del 12%, 120 de los 1000 individuos presentan la enfermedad y 880 no la presentan. Si la sensibilidad de la prueba es del 90%, en 108 de los 120 individuos enfermos la prueba resulta positiva y en 12 negativa. Si la especificidad es del 95%, en 836 de los 880 individuos no enfermos la prueba resulta negativa y en 44 positiva. En total, la prueba ha resultado positiva en 108 enfermos y 44 no enfermos (152) y ha resultado negativa en 12 enfermos y 836 no enfermos (848).

El valor predictivo positivo es la probabilidad de presentar la enfermedad dado que la prueba ha resultado positiva:

$$VPP = 108/152 = 71,1\%$$

El valor predictivo negativo es la probabilidad de no presentar la enfermedad dado que la prueba ha resultado negativa:

$$VPN = 836/848 = 98,6\%$$

Tabla 42. Tabla 2x2.

		Enfermedad		
		+	-	
Prueba +	108	44	152	
Prueba -	12	836	848	
	120	880	1000	

Si se aplica el teorema de Bayes:

$$VPP = \frac{P*S}{P*S + (1-P)*(1-E)} = \frac{0,12*0,90}{0,12*0,90 + 0,88*0,05} = 71,1\%$$

$$VPN = \frac{(1-P)*E}{(1-P)*E + P*(1-S)} = \frac{0,88*0,95}{0,88*0,95 + 0,12*0,10} = 98,6\%$$

Si se aplica la razón de verosimilitud:

$$\text{Odds previa} = 0,12/0,88 = 0,136 \quad (1 \text{ a } 7,3 \text{ ó } 3 \text{ a } 22)$$

Por cada individuo enfermo hay 7,3 individuos no enfermos, o utilizando números enteros, por cada tres enfermos hay 22 no enfermos.

$$LR_+ = \frac{\text{Sensibilidad}}{1 - \text{Especificidad}} = \frac{0,90}{0,05} = 18$$

Por cada vez que resulta la prueba positiva en un paciente no afecto de la enfermedad resulta 18 veces en uno enfermo.

$$LR- = \frac{1 - \text{Sensibilidad}}{\text{Especificidad}} = \frac{0,10}{0,95} = 0,11 \quad (1 \text{ a } 9,5)$$

Por cada vez que sale un resultado negativo en un enfermo resulta negativo 9,5 veces en uno no enfermo.

La odds posterior de un resultado positivo es:

$$\text{odds previa} * LR+ = 0,136 * 18 = 2,45$$

Cuando el resultado de la prueba es positivo, es 2,45 veces más probable que se trate de un individuo enfermo que uno no enfermo. Si se expresa en forma de probabilidades:

$$\text{probabilidad} = \frac{\text{odds}}{1 + \text{odds}} = \frac{2,45}{3,45} = 0,71 \quad (71\%)$$

La odds posterior de un resultado negativo es:

$$\text{odds previa} * LR- = 0,136 * 0,11 = 0,015 \quad (1 \text{ a } 67)$$

Cuando el resultado de la prueba es negativo, es 67 veces más probable que se trate de un individuo no enfermo que de uno enfermo. Si se expresa en forma de probabilidades:

$$\text{probabilidad} = \frac{0,015}{1,015} = 0,0147$$

Esta es la probabilidad de presentar la enfermedad dado un resultado negativo, luego el complementario es el valor predictivo negativo:

$$VPN = 1 - 0,0147 = 0,985 \quad (98,5\%)$$

2. Son los resultados positivos los que aumentan la probabilidad de la enfermedad. Por otro lado, la odds posterior que corresponde a una probabilidad del 0,5 es 1.

$$\text{odds} = p/(1-p) = 0,5/0,5 = 1$$

La odds posterior de un resultado positivo es el producto de la odds previa y la razón de verosimilitud de un resultado positivo:

$$\text{Odds previa} = 0,25/0,75 = 0,33 \quad (1 \text{ a } 3)$$

$$\text{Odds posterior} = \text{odds previa} * LR+$$

$$1 = 0,33 * LR+$$

$$\text{de donde se despeja que } LR+ = 1/0,33 = 3,03.$$

Los resultados positivos tienen que ser por lo menos tres veces más frecuentes en los enfermos que en los no enfermos para que dada una probabilidad previa del 25%, la probabilidad posterior sea al menos del 50%. Es lo mismo que afirmar que la sensibilidad debe ser, al menos, tres veces superior a la tasa de falsos positivos.

3. Se trata de ver cuánto disminuye la probabilidad de la enfermedad (en caso de que fuera inferior al 5% se evitaría dar tratamiento anticoagulante a un buen número de falsos positivos), y son los resultados negativos de las pruebas los que disminuyen la probabilidad de la enfermedad. A su vez, el valor predictivo negativo depende sobre todo de la sensibilidad, luego a simple vista, parece que es más útil la prueba B.

Aplicando el teorema de Bayes para el cálculo del complementario del valor predictivo negativo:

$$\text{Probabilidad de enfermedad si la prueba A es negativa} = \frac{0,30 \cdot 0,15}{0,30 \cdot 0,15 + 0,70 \cdot 0,95} = 0,063$$

$$\text{Probabilidad de enfermedad si la prueba B es negativa} = \frac{0,30 \cdot 0,10}{0,30 \cdot 0,10 + 0,70 \cdot 0,90} = 0,045$$

Con la prueba A no disminuye la probabilidad de la enfermedad lo suficiente como para decidir no dar el tratamiento anticoagulante. Es más efectiva la prueba B (como ya habíamos deducido previamente).

4. Puesto que la probabilidad previa es mayor que la posterior exigida para no dar el tratamiento inmunosupresor, interesa saber en qué medida disminuye la probabilidad de la enfermedad tras un resultado negativo y si es por debajo del 5% se evitaría aumentar el tratamiento inmunosupresor. Si se aplica el teorema de Bayes:

$$\text{Probabilidad de enfermedad si la analítica está dentro de la normalidad} = \frac{0,70 \cdot 0,10}{0,70 \cdot 0,10 + 0,30 \cdot 0,90} = 0,21$$

Como la probabilidad sigue siendo superior al umbral para decidir tratar o no a los pacientes, en vista de estos resultados la decisión debería ser tratar al paciente. Los resultados de la analítica no cambian la decisión en este caso de tratar al paciente con inmunosupresores.

Sin embargo, si la probabilidad previa ahora es del 21% y tras la realización de la biopsia, ésta resultase normal, la probabilidad de la enfermedad podría disminuir por debajo del 5% que representa el umbral. Aplicando el teorema de Bayes:

$$\text{Probabilidad de enfermedad si la biopsia es normal} = \frac{0,21 \cdot 0,01}{0,21 \cdot 0,01 + 0,79 \cdot 0,95} = 0,003$$

Efectivamente, la probabilidad de 0,003 es menor que el umbral terapéutico (probabilidad de presentar la enfermedad por encima de la cual se decide administrar el tratamiento), por lo tanto se debería hacer esta prueba ya que de sus resultados depende el evitar aumentar el tratamiento inmunosupresor a los falsos positivos y el riesgo de infección que ello conlleva.

El problema de estas afirmaciones es que no se sabe si la sensibilidad y la especificidad de la biopsia se ha estudiado sobre individuos que previamente tenían la analítica normal o no, y este hecho puede cambiar los parámetros de la validez de la prueba.

CUÁNDO APLICAR UNA PRUEBA DIAGNÓSTICA

1. Si la probabilidad preprueba de cudicia es inferior al umbral de diagnóstico (p_1), aunque la colonoscopia resultase positiva, la probabilidad de la enfermedad seguiría siendo inferior al umbral terapéutico, por lo que la colonoscopia no resulta efectiva. Si la probabilidad previa es superior al umbral de tratamiento (p_2), aunque la colonoscopia resultase negativa, la probabilidad de la enfermedad seguiría siendo superior al umbral terapéutico, y la decisión sería tratar, por lo que tampoco en este caso es efectiva la prueba. Es entre estos dos umbrales, cuando la colonoscopia nos puede aportar información que ayuda a disminuir la incertidumbre.

$$p_1 = \frac{(1-E)*(C) - U(\text{prueba})}{(1-E)*(C) + (S*B)} \quad p_2 = \frac{E*(C) + U(\text{prueba})}{E*(C) + (1-S)*B}$$

$$P_1 = \frac{(0,05)*(0,02) + (0,001)}{(0,05)*(0,02) + (0,9)*(0,7)} = 0,0032$$

$$P_2 = \frac{(0,95)*(0,02) - (0,001)}{(0,95)*(0,02) + (0,1)*(0,7)} = 0,2022$$

El umbral de diagnóstico es extremadamente bajo, por lo que ante la mínima sospecha de cudicia es preferible realizar una colonoscopia (los beneficios del tratamiento son muy altos y los riesgos tanto del tratamiento como de la prueba comparativamente son muy bajos). No se descarta la enfermedad hasta no estar bastante seguro de que el paciente no la tiene.

El umbral terapéutico es muy bajo ($0,02/0,72 = 0,028$), precisamente por lo que hemos dicho, porque los beneficios de tratar a los enfermos son bastante elevados comparados con los riesgos de tratar a los no enfermos. Esto motiva que el umbral de tratamiento p_2 también sea bajo. No hace falta tanta seguridad en el diagnóstico para decidirse a tratar.

Al estar la probabilidad preprueba entre las dos probabilidades umbral p_1 y p_2 , es preferible realizar la colonoscopia, que nos pueda aportar más información. En este caso concreto, al estar la probabilidad preprueba por encima del umbral terapéutico, de resultar la colonoscopia negativa, la probabilidad de la enfermedad pasa a ser inferior al umbral terapéutico, y nos evitaríamos enviar a quirófano (con el riesgo que ello conlleva) a un determinado número de pacientes sin cudicia.

Otra manera de enfocar el problema y aplicando el teorema de Bayes es:

1. Determinar el umbral terapéutico.
2. Estimar la probabilidad previa de cudicia.
3. Decidir si el resultado de la colonoscopia puede afectar a la decisión de tratar o no.

Hemos visto que el umbral terapéutico es 0,028.

La probabilidad preprueba es 0,15.

Al estar la probabilidad preprueba por encima del umbral terapéutico, el resultado de la colonoscopia afectaría a la decisión de tratar o no si tras resultar negativa, la probabilidad de la enfermedad es inferior al umbral terapéutico.

El complementario del valor predictivo negativo (probabilidad de tener la enfermedad dado un resultado negativo) es el siguiente:

$$1 - p(E-/T-) = \frac{P^*(1-S)}{P^*(1-S) + (1-P)*E}$$

donde:

$p(E-/T-)$ = valor predictivo negativo o probabilidad de no tener la enfermedad dado un resultado negativo.

P = probabilidad de la enfermedad

S = sensibilidad de la prueba

E = especificidad de la prueba

$$1 - VPn = \frac{0,15*0,1}{0,15*0,1 + 0,85*0,95} = 0,018$$

Efectivamente esta probabilidad es inferior al umbral terapéutico, con lo que se llega a la misma conclusión, se debe realizar primero la colonoscopia y en función de los resultados se trata o no al paciente.

- Como la probabilidad preprueba es inferior al umbral terapéutico, se evita el tratamiento quirúrgico a menos que la probabilidad de un tumor cerebral tras una RNM positiva sea superior al umbral terapéutico. Si la probabilidad preprueba es inferior al umbral de diagnóstico p_1 , la probabilidad de tener un tumor cerebral tras una RNM positiva seguirá siendo inferior al umbral terapéutico, por lo que no será preciso realizar la prueba (no cambiará la actitud de tratar al paciente como si no tuviera el tumor cerebral).

La fórmula del umbral p_1 , y dado que la utilidad de la prueba en sí misma es cero se puede expresar tanto en función de los costes y beneficios del tratamiento, como en función del umbral terapéutico:

$$p_1 = \frac{(1-E)*(C)}{(1-E)*(C) + (S*B)}$$

si se divide cada factor por $C+B$ el resultado es:

$$p_1 = \frac{(1-E)*(p^*)}{(1-E)*(p^*) + S*(1-p^*)} \quad \text{donde } p^* \text{ es el umbral terapéutico.}$$

Aplicando los datos del problema:

$$p_1 = \frac{0,05*0,15}{0,05*0,15 + 0,95*0,85} = 0,0092$$

Siguiendo el mismo razonamiento para el umbral de tratamiento p_2 :

$$p_2 = \frac{E^*(C)}{E^*(C) + (1-S)*B}$$

si se divide cada factor por $C+B$ el resultado es:

$$p_2 = \frac{E^*(p^*)}{E^*(p^*) + (1-S)*(1-p^*)}$$

Aplicando los datos del problema:

$$p_2 = \frac{0,95*0,15}{0,95*0,15 + 0,05*0,85} = 0,7703$$

La probabilidad preprueba es 0,05, superior a p_1 e inferior a p_2 , por lo que es conveniente realizar primero la RNM y en función de los resultados operar o no al paciente. En realidad es suficiente con saber que es superior a p_1 , puesto que si la probabilidad previa es inferior al umbral terapéutico, lógicamente también será inferior a p_2 .

También se puede solucionar el problema mediante el teorema de Bayes. Los pasos a seguir son los siguientes:

1. Determinar el umbral terapéutico.
2. Estimar la probabilidad previa de tumor cerebral.
3. Decidir si el resultado de la RNM puede afectar a la decisión de intervenir o no.

El umbral terapéutico es 0,15.

La probabilidad preprueba es 0,05.

Al estar la probabilidad preprueba por debajo del umbral terapéutico, el resultado de la RNM afectaría a la decisión de tratar o no si tras resultar positiva, la probabilidad de la enfermedad fuese superior al umbral terapéutico.

Aplicamos el teorema de Bayes para el cálculo de la probabilidad de la enfermedad tras un resultado positivo (valor predictivo positivo):

$$p(E+/T+) = \frac{S*P}{S*P + (1-E)*(1-P)}$$

donde:

$p(E+/T+)$ = probabilidad de tener la enfermedad dado un resultado positivo o valor predictivo positivo.

S = sensibilidad de la prueba

P = probabilidad previa

E = especificidad de la prueba

Aplicando los datos del problema:

$$p(E+/T+) = \frac{0,95*0,05}{0,95*0,05 + 0,05*0,95} = 0,5$$

Como es lógico, se llega a la misma conclusión, la probabilidad de la enfermedad tras un resultado positivo es superior al umbral terapéutico, con lo que es preferible realizar primero la RNM y en función de los resultados tratar o no quirúrgicamente al paciente.

RESULTADOS SOBRE LA SALUD

1. La eficacia vacunal (o fracción prevenida) es la proporción en que disminuye la incidencia de la gripe al vacunar a la población.

$$FP = (0,25 - 0,025)/0,25 = 0,9 = 90\%$$

Si la cobertura vacunal fuese del 100% se podrían evitar el 90% de las gripes, lo que equivale a decir que la eficacia de la vacuna es del 90%.

La incidencia si se vacunan el 10% de la población es la media de la incidencia en los vacunados y la incidencia en los no vacunados, ponderadas por el porcentaje de vacunados y no vacunados respectivamente:

$$I_{10\%} = 0,25 \cdot 0,9 + 0,025 \cdot 0,1 = 0,2275$$

$$FP = (0,25 - 0,2275)/0,25 = 0,09 = 9\%$$

Si la cobertura vacunal es del 10% se podrían evitar el 9% de las gripes. El mismo resultado da al multiplicar la eficacia vacunal por el porcentaje de individuos vacunados (90% x 10% = 9%).

El número total de casos de gripe antes de la vacunación era de 5 millones x 0,25 = 1.250.000. Si de éstas, gracias a la vacunación se evitan el 9%, en números absolutos son:

$$1.250.000 \cdot 0,09 = 112.500 \text{ casos de gripe evitados.}$$

2. Si el riesgo de muerte sin tratamiento es del 10% y con betabloqueantes del 7,5%, la diferencia, 2,5%, es la reducción absoluta del riesgo (RAR) que ha provocado el tratamiento (efectividad en términos absolutos). Por tanto, los betabloqueantes han disminuido el riesgo inicial en un 25% (efectividad en términos relativos o reducción relativa del riesgo RRR).

$$RRR = \frac{0,10 - 0,075}{0,10} \times 100 = 25\%$$

El número absoluto que es necesario tratar (NNT) es el inverso de la reducción absoluta del riesgo, por lo tanto, por cada 40 pacientes (1/0,025) tratados con betabloqueantes se evita una muerte.

Si la mortalidad de los grupos sin y con tratamiento fuese del 1% y 0,75% respectivamente, se reduce el riesgo en términos absolutos un 0,25% (RAR = 1% - 0,75% = 0,25%). La efectividad en términos relativos (RRR) sigue siendo del 25%.

$$RRR = \frac{0,01 - 0,0075}{0,01} \times 100 = 25\%$$

En cambio el número que es necesario tratar para evitar una muerte en este caso es de 400 pacientes (1/0,0025).

3. Si las mujeres presentan la mitad de riesgo que los hombres, y la presencia de complicaciones multiplica por 4 el riesgo de morir en quiró-

fano, la tasa de mortalidad de las mujeres de 40 años por colecistectomía profiláctica es del 0,12%

$$0,96 : 2 : 4 = 0,12$$

Si aumenta el riesgo un 8,45% anual, la tasa de mortalidad a los 45 años es del 0,18%.

40 años	0,1200
41 años	$0,1200 + 0,1200 * 0,0845 = 0,1301$
42 años	$0,1301 + 0,1301 * 0,0845 = 0,1411$
43 años	$0,1411 + 0,1411 * 0,0845 = 0,1531$
44 años	$0,1531 + 0,1531 * 0,0845 = 0,1660$
45 años	$0,1660 + 0,1660 * 0,0845 = 0,1800$

4. La esperanza de vida es el inverso de la tasa de mortalidad, luego el tratamiento A tiene una esperanza de vida de $1/0,08 = 12,5$ años.

Si tras el tratamiento B, la mortalidad a los 5 años es del 40%, la supervivencia es del 60%. La tasa de mortalidad entonces es:

$$\text{tasa de mortalidad anual} = (-1/5) * \ln(0,6) = 0,10$$

$$\text{esperanza de vida} = 1/0,10 = 10 \text{ años.}$$

El tratamiento C tiene una esperanza de vida de 10 años.

Los tratamientos B y C son igualmente efectivos, pero menos que el A, cuya esperanza de vida es la mayor.

Si los resultados se expresan mediante la tasa de mortalidad, el mejor tratamiento será aquel cuya tasa sea menor:

Tratamiento A: tasa de mortalidad de 8%.

Tratamiento B: tasa de mortalidad del 10%

Tratamiento C: $1/\text{esperanza de vida} = 1/10 = 0,1 = 10\%$.

Si los resultados se expresan en forma de supervivencia a los 5 años, aquel con mayor supervivencia será el tratamiento más efectivo:

Tratamiento A: $e^{-0,08*5} = 0,67$ (67% sobreviven a los 5 años)

Tratamiento B: 60% sobreviven a los 5 años

Tratamiento C: $e^{-0,1*5} = 0,60$ (60% sobreviven a los 5 años)

Independientemente de la forma de expresar los resultados, el tratamiento A es el más efectivo.

5. Para el grupo de 40 años, una supervivencia del 20% a los 5 años corresponde a una tasa de mortalidad de 0,32.

$$m = (-1/t) * \ln(S/S_0)$$

$$m = -1/5 * \ln 0,20 = 0,322$$

$$\text{esperanza de vida} = 1/0,322 = 3,11 \text{ años}$$

Para conocer la tasa de mortalidad específica por enfermedad, basta restar a esta tasa global la tasa de mortalidad específica por edad.

$$m_{\text{especifica-enfermedad}} = 0,322 - 0,006 = 0,316$$

Y esta es la tasa que se suma a la específica por edad en el grupo de 55 años para conocer la tasa de mortalidad global.

$$m_{\text{total 55 años}} = 0,316 + 0,009 = 0,325$$

$$\text{esperanza de vida} = 1/0,325 = 3,077 \text{ años}$$

6. La esperanza de vida en el grupo de tratamiento es:

$$249.800/100.000 + 641.200/100.000 = 8,91 \text{ años}$$

La esperanza de vida en el grupo placebo es:

$$198.100/100.000 + 678.200/100.000 = 8,76 \text{ años}$$

Sólo comparando la esperanza de vida es mejor el grupo de tratamiento que el placebo, pero si además se ajusta por calidad de vida, los años de vida ajustados por calidad en el grupo de tratamiento son:

$$2,498 * 1 + 6,412 * 0,45 = 5,3834 \text{ AVAC}$$

Los años de vida ajustados por calidad en el grupo placebo son:

$$1,981 * 1 + 6,782 * 0,45 = 5,0329 \text{ AVAC}$$

Los resultados sobre la salud siguen siendo mejores en el grupo sometido al tratamiento con imipramina y terapia interpersonal.

7. Los resultados del análisis de supervivencia son los de la tabla 43.

A los tres meses ha muerto un paciente, luego la supervivencia es el complementario:

$$\text{Supervivencia a los 3 meses} = 1 - 1/30 = 1 - 0,0333 = 0,9667$$

Durante el quinto mes han muerto 2 pacientes de los 29 que estaban vivos, luego

$$\text{Supervivencia durante el quinto mes} = 1 - 2/29 = 1 - 0,0689 = 0,9310$$

$$\text{Supervivencia a los 5 meses} = 0,9310 \times 0,9667 = 0,9000$$

Durante el sexto mes ha muerto un paciente de los 27 que estaban vivos, luego

$$\text{Supervivencia durante el sexto mes: } 1 - 1/27 = 1 - 0,0370 = 0,9630$$

$$\text{Supervivencia a los seis meses: } 0,9630 \times 0,9000 = 0,8667$$

Y así sucesivamente, de modo que el percentil 25 es 11 meses (aunque la supervivencia no es exactamente del 75%), la mediana es 16 meses (han muerto el 50%) y no hay percentil 75 porque no llegan a morir el 75% de la muestra.

En realidad, como todos los pacientes mueren, excepto los 8 últimos, que ya no cambian la supervivencia, para calcular la supervivencia en cada período es suficiente con dividir el número de muertos acumulados por el total de pacientes iniciales.

$$\text{Supervivencia a los 3 meses: } 1 - 1/30 = 1 - 0,0333 = 0,9667$$

Supervivencia a los 5 meses: $1 - 3/30 = 1 - 0,1000 = 0,9000$

Supervivencia a los 6 meses: $1 - 4/30 = 1 - 0,1333 = 0,8667$

Supervivencia a los 8 meses: $1 - 6/30 = 1 - 0,2000 = 0,8000$

.....

Supervivencia a los 31 meses: $1 - 22/30 = 1 - 0,7333 = 0,2667$

Tabla 43. Análisis de supervivencia. Cáncer de pulmón estadio IIIA.

PRODUCT-LIMIT SURVIVAL ANALYSIS							
CASE NUMBER	TIME	STATUS	CUMULATIVE SURVIVAL	STANDARD ERROR	CUM muerto	CUM LOST	REMAIN AT RISK
31	3.00	muerto	0.9667	0.0328	1	0	29
61	5.00	muerto			2	0	28
69	5.00	muerto	0.9000	0.0548	3	0	27
95	6.00	muerto	0.8667	0.0621	4	0	26
72	8.00	muerto			5	0	25
91	8.00	muerto	0.8000	0.0730	6	0	24
67	10.00	muerto	0.7667	0.0772	7	0	23
53	11.00	muerto	0.7333	0.0807	8	0	22
38	13.00	muerto			9	0	21
50	13.00	muerto	0.6667	0.0861	10	0	20
74	15.00	muerto			11	0	19
86	15.00	muerto	0.6000	0.0894	12	0	18
44	16.00	muerto			13	0	17
62	16.00	muerto			14	0	16
96	16.00	muerto	0.5000	0.0913	15	0	15
15	18.00	muerto	0.4667	0.0911	16	0	14
93	20.00	muerto	0.4333	0.0905	17	0	13
81	21.00	muerto	0.4000	0.0894	18	0	12
89	26.00	muerto	0.3667	0.0880	19	0	11
37	30.00	muerto			20	0	10
94	30.00	muerto	0.3000	0.0837	21	0	9
83	31.00	muerto	0.2667	0.0807	22	0	8
35	40.00	vivo			22	0	7
41	56.00	vivo			22	0	6
21	61.00	vivo			22	0	5
6	72.00	vivo			22	0	4
92	72.00	vivo			22	0	3
49	77.00	vivo			22	0	2
40	91.00	vivo			22	0	1
59	94.00	vivo			22	0	0

8. Si en la escala analógica ha respondido que el estado A presenta una calidad de vida de 0,25 y el estado B de 0,75 es como decir que vivir 10 años en el estado A es como vivir 2,5 años en perfecto estado de salud y vivir 10 años en el estado B es como vivir 7,5 años en perfecto estado de salud (equivalencia temporal).

Mediante la técnica de estimación de la magnitud la respuesta debe ser que el estado B es el triple de bueno que el A. El estado A es el 25% de bueno respecto al perfecto estado de salud y el estado B es el 75% de bueno respecto al perfecto estado de salud.

Curar a una persona en el estado A (equivaldría a mejorar la calidad de vida en $1 - 0,25 = 0,75$ puntos) es como curar a tres personas en el estado B (donde la mejoría es de $1 - 0,75 = 0,25$ puntos), siendo éste el resul-

tado de medir la calidad de vida mediante la técnica de equivalencia de personas.

Si un individuo decide entrar en quirófano para curarse de la enfermedad A, la probabilidad de morir que aceptaría sería de 0,75 (juego estándar) y si se trata de la enfermedad B aceptaría una probabilidad de 0,25 (Fig. 39).

Estado A Utilidad = $0,25 = p \cdot 0 + (1-p) \cdot 1$ $p = 0,75$

Estado B Utilidad = $0,75 = p \cdot 0 + (1-p) \cdot 1$ $p = 0,25$

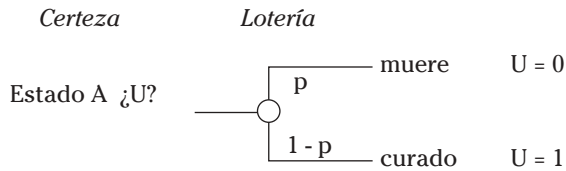
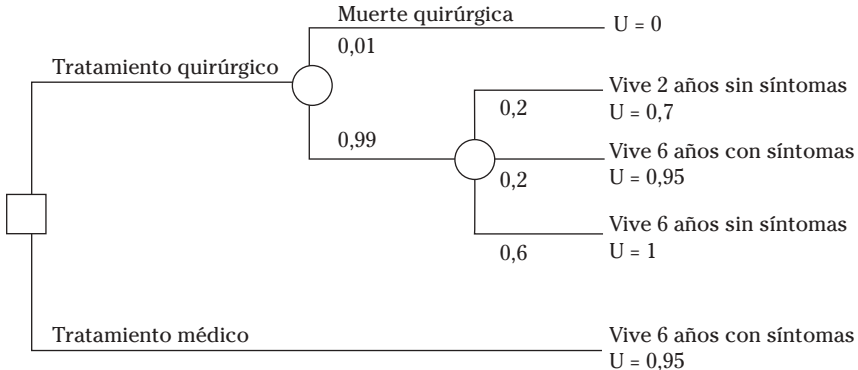


Figura 39. Juego estándar.

9. Para facilitar el ejercicio se puede realizar un árbol de decisión como el de la figura 40.



La utilidad de los años libres de síntomas se determina directamente tras la aplicación de la técnica del juego estándar. A la muerte le corresponde la utilidad igual a cero y uno al mejor estado de salud imaginable, que en este caso es vivir los seis años sin síntomas.

Para el cálculo de la utilidad de los años con síntomas es preciso combinar datos de calidad y cantidad de vida. Primero, y mediante la técnica de equivalencia temporal, los síntomas acortan la longitud de vida en un tercio, es decir, vivir 6 años con síntomas es como vivir cuatro años sin síntomas. Posteriormente, y mediante la técnica del juego estándar vemos que la utilidad que corresponde a 4 años de vida sano es 0,95.

Para saber qué tratamiento es el idóneo se debe calcular la utilidad media de cada alternativa, ponderando los resultados finales por la probabilidad de que ocurran:

Utilidad del tratamiento quirúrgico:

$$0,01*0 + 0,99*0,2*0,7 + 0,99*0,2*0,95 + 0,99*0,6*1 = 0,921$$

Utilidad del tratamiento médico = 0,95

Es preferible el tratamiento médico.

EVALUACIÓN ECONÓMICA

1. En el análisis coste-beneficio los resultados se miden en unidades monetarias.

Los costes del transplante se producen en el presente y son de 8.900.000 pts. por paciente.

Los costes de la diálisis son de 1.900.000 durante 10 años y para hacerlos comparables, primero es necesario actualizarlos:

$$\frac{1,9 \cdot 10^6}{(1,06)^1} + \frac{1,9 \cdot 10^6}{(1,06)^2} + \dots + \frac{1,9 \cdot 10^6}{(1,06)^{10}} = 13.984.190 \text{ pts.}$$

Los beneficios del transplante (medidos en productividad) son 1.200.000 pts. anuales durante 15 años, que actualizados son igual a:

$$\frac{1,2 \cdot 10^6}{(1,06)^1} + \frac{1,2 \cdot 10^6}{(1,06)^2} + \dots + \frac{1,2 \cdot 10^6}{(1,06)^{15}} = 11.654.640 \text{ pts.}$$

Los beneficios correspondientes a los pacientes dializados son de 1 millón anual durante 10 años, que actualizados corresponden a:

$$\frac{1 \cdot 10^6}{(1,06)^1} + \frac{1 \cdot 10^6}{(1,06)^2} + \dots + \frac{1 \cdot 10^6}{(1,06)^{10}} = 7.360.100 \text{ pts.}$$

El VAN (valor actualizado neto) de cada alternativa es la diferencia entre los beneficios y los costes:

VAN transplante = 11.654.640 - 8.900.000 = 2.754.640 pts.

VAN diálisis = 7.360.100 - 13.984.190 = - 6.624.090 pts.

El transplante es una opción rentable, pues los beneficios son superiores a los costes, al contrario que ocurre con la diálisis.

En el análisis coste-efectividad la medida de los resultados sobre la salud son los años de vida ganados. Los costes se miden de la misma manera. La esperanza de vida de 10 años en los pacientes dializados corresponde a 7,3601 años actualizados:

$$\frac{1}{(1,06)^1} + \frac{1}{(1,06)^2} + \dots + \frac{1}{(1,06)^{10}} = 7,3601 \text{ años de vida}$$

En los pacientes transplantados, la esperanza de vida de 15 años corresponde a 9,7122 años de vida actualizados:

$$\frac{1}{(1,06)^1} + \frac{1}{(1,06)^2} + \dots + \frac{1}{(1,06)^{15}} = 9,7122 \text{ años de vida}$$

El índice coste/efectividad resulta de dividir los costes por la efectividad conseguida, y puesto que se trata de opciones excluyentes, la opción elegida será la de menor índice, en este caso el transplante.

Además la diálisis es una opción de mayor coste y menor efectividad que el trasplante, luego la elección no da lugar a dudas.

Índice C/E (trasplante) = $8.900.000/9,7122 = 916.373$ pts. por año de vida.

Índice C/E (diálisis) = $13.984.190/7,3601 = 1.900.000$ pts. año de vida.

En el análisis coste-utilidad, los resultados sobre la salud se miden en AVAC.

En la alternativa de trasplante, los AVAC durante el primer año son 0,70 y durante los restantes 14 años disminuye progresivamente hasta 0,60, luego se puede asumir que en media son de 0,65 cada uno de estos 13 años centrales. Si se actualizan los AVAC el resultado es el siguiente:

$$\frac{0,70}{(1,06)^1} = 0,6604 \text{ AVAC} \quad \frac{0,60}{(1,06)^{15}} = 0,2504 \text{ AVAC}$$

$$\frac{0,65}{(1,06)^2} + \frac{0,65}{(1,06)^3} + \dots + \frac{0,65}{(1,06)^{14}} = 5,4285 \text{ AVAC}$$

En total, con el trasplante son $0,6604 + 5,4285 + 0,2504 = 6,3393$ AVAC

En la alternativa de diálisis, durante los primeros 9 años la calidad de vida es de 0,65 y 0,325 (el 50%) durante el último año de vida. Estos AVAC actualizados son:

$$\frac{0,65}{(1,06)^1} + \frac{0,65}{(1,06)^3} + \dots + \frac{0,65}{(1,06)^9} = 4,4211 \text{ AVAC}$$

$$\frac{0,325}{(1,06)^{10}} = 0,1815 \text{ AVAC}$$

En total $4,4211 + 0,1815 = 4,6026$ AVAC

El índice coste/utilidad resulta de dividir los costes por la efectividad conseguida:

Índice C/U (trasplante) = $8.900.000/6,3393 = 1.403.940$ pts./AVAC

Índice C/U (diálisis) = $13.984.190/4,6026 = 3.038.324$ pts./AVAC

Como es lógico, corrobora los resultados del análisis coste-efectividad, y es que la opción de diálisis es más costosa y menos efectiva que el trasplante.

Los resultados se resumen en la tabla 44.

Tabla 44. Resultados según la técnica de evaluación económica.

ANÁLISIS COSTE-BENEFICIO			
Alternativa	Costes	Beneficios	VAN
Transplante	8.900.000	1.654.640	2.754.640
Diálisis	13.984.190	7.360.100	-6.624.090
ANÁLISIS COSTE-EFECTIVIDAD			
Alternativa	Costes	Años de vida	Índice C/E
Transplante	8.900.000	9,7122	916.373
Diálisis	13.984.190	7,3601	1.900.000
ANÁLISIS COSTE-UTILIDAD			
Alternativa	Costes	AVAC	Índice C/U
Transplante	8.900.000	6,3393	1.403.940
Diálisis	13.984.190	4,6026	3.038.324

2. Si se analiza el coste medio, el coste por caso detectado es $176.331/71,9420 = 2.451$ dólares. De hecho, el coste medio aumenta muy gradualmente conforme aumenta el número de pruebas.

Pero en la evaluación económica lo que se valora es el coste de oportunidad, el coste de aquello a lo que se ha tenido que renunciar al hacer una elección. ¿Por qué hacer 6 pruebas de sangre oculta en heces y no 5, qué aporta realizar una prueba más? Realizar 6 pruebas (respecto a 5) aumenta el número de casos detectados, concretamente en 0,0003, pero también aumenta el coste en 13.190 dólares, lo que equivale a decir que con 6 pruebas el coste de detectar un caso de cáncer de colon es de 44 millones de dólares.

$$\text{Índice C/E (incremental)} = \frac{176.331 - 163.141}{71,9420 - 71,9417} = \frac{13.190}{0,0003} = 43.966.666$$

3. Si están dispuestos a pagar 2 millones de pesetas por reducir el riesgo en un 2% (3%-1%) quiere decir que estiman el valor monetario de su vida en 100 millones de pesetas. Por una simple regla de tres, si por salvar 0,02 vidas se pagan 2 millones de pts., por salvar 1 vida se pagan 100 millones (2/0,02).
4. En la tabla 45 se presentan los datos de las 6 alternativas, ordenados por el índice coste/efectividad creciente.
- Si las alternativas no son excluyentes, una vez ordenadas por el índice coste-efectividad, se van eligiendo hasta que se agota el presupuesto. En este caso, con un presupuesto de 100 millones de pesetas, se eligen las opciones D, E y F.
 - Si son opciones excluyentes (por ejemplo, 6 medicamentos diferentes para tratar la hipercolesterolemia), sólo se puede elegir una de ellas. Las opciones A, B y E son igualmente costosas, pero la más efectiva es la E (opción dominante), luego ésta es la de elección y la A y la B

se descartan (opciones dominadas). La opción C también se descarta, ya que es más costosa y menos efectiva que la F (dominante). Entre las opciones que quedan (D, E y F) la elección puede guiarse por un límite presupuestario o estableciendo un límite al valor del coste por año de vida ganado. Si se elige la opción F, supone un coste por año de vida ganado de 800.000 pesetas. Recordemos que la comparación se debe hacer con la mejor alternativa a la que se renuncia al hacer la elección, la de mayor efectividad, que en este ejemplo es la E.

$$\frac{\text{Costes de F} - \text{Costes de E}}{\text{Años ganados en F} - \text{Años ganados en E}} = \frac{50 - 30}{75 - 50} = 800.000 \text{ pts. /año}$$

Tabla 45. Asignación de recursos.

Alternativa	Coste neto	Coste neto acumulado	Años de vida ganados	Relación coste-efectividad
D	20	20	40	0,50
E	30	50	50	0,60
F	50	100	75	0,67
C	55	155	70	0,79
A	30	185	30	1
B	30	215	20	1,50

BIBLIOGRAFÍA

ANÁLISIS DE DECISIONES MÉDICAS

- JAESCHKE R, GUYATT GH, SACKETT DL. Evidence-Based Medicine Working Group. Guías para usuarios de la literatura médica. III. Cómo utilizar un artículo sobre un examen diagnóstico. JAMA 1994; 271: 389-391. Forma parte de una serie de artículos de la revista JAMA que ayudan, de una forma muy esquemática y clara, a realizar una lectura crítica sobre la evidencia científica.
- PAUKER SG, KASSIRER JP. Therapeutic decision making: a cost-benefit analysis. N Engl J Med 1975; 293: 229-234. Estos autores, clave en cuestiones de análisis de decisión, explican un método para el cálculo del umbral terapéutico, basado en la relación coste-beneficio.
- PAUKER SG, KASSIRER JP. The threshold approach to clinical decision making. N Engl J Med 1980; 302: 1109-1117. Mediante el desarrollo de árboles de decisión, ilustran la obtención de los umbrales de diagnóstico y de tratamiento, así como la interpretación de los mismos.
- RANSOHOFF DF, FEINSTEIN AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978; 299: 926-930. Artículo clave que muestra un amplio rango de sesgos que se pueden cometer al evaluar la validez de una prueba diagnóstica.
- SACKETT DL, HAYNES RB, TUGWELL P. Epidemiología Clínica: una ciencia básica para la medicina clínica. Madrid. Ed. Díaz de Santos, 1989. Es un libro entretenido para los "usuarios" de las investigaciones hechas por los demás, por lo que está explicado más en términos clínicos que epidemiológicos.
- SACKETT DL, RICHARDSON WS, ROSENBERG W, HAYNES RB. Medicina Basada en la Evidencia; Cómo ejercer y enseñar la MBE. Madrid: Churchill Livingstone, 1997. Libro diseñado para ayudar al clínico (en cualquier fase de su formación) en la búsqueda de la mejor evidencia científica.
- SOX HC, BLATT MA, HIGGINS MC, MARTON KI. Medical Decision Making. Boston. Ed. Butterworth Publishers, 1988. Referencia indiscutible de este monográfico. Se trata de un libro muy didáctico, incluso divertido, con muchas ilustraciones, y cuyo objetivo, que considero cumple sobradamente, es "revelar la elegancia del análisis de decisiones sin llegar a ser abstracto". Para ello utiliza múltiples ejemplos clínicos, muchos de los cuales se han presentado también en este monográfico.

Referencias complementarias

- BURGUEÑO MJ, GARCÍA-BASTOS JL, GONZÁLEZ-BUITRAGO JM. Las curvas ROC en la evaluación de las pruebas diagnósticas. *Med Clin* 1995; 104: 661-670.
- COUGHLIN SS, TROCK B, CRIQUI MH, PICKLE LW, BROWNER D, TEFFT MC. The logistic modeling of sensitivity, specificity, and predictive value of a diagnostic test. *J Clin Epidemiol* 1992; 45: 1-7.
- DIAMOND GA. Clinical epistemology of sensitivity and specificity. *J Clin Epidemiol* 1992; 45: 9-16.
- DUBILET P. A mathematical approach to interpretation and selection of diagnostic tests. *Med Dec Making* 1983; 3: 177-196.
- GORRY GA, PAUKER SG, SCHWARTZ WB. The diagnostic importance of the normal finding. *N Engl J Med* 1978; 298: 486-489.
- KASSIRER JP. Diagnostic reasoning. *Ann Intern Med* 1989; 110: 893-900.
- KASSIRER JP, MOSKOWITZ AJ, LAU J, PAUKER SG. Decision analysis: a progress report. *Ann Intern Med* 1987; 106: 275-291.
- KNILL-JONES RP. Diagnostic systems as an aid to clinical decision making. *Br Med J* 1987; 295: 1392-1396.
- LEAPE LL. Error in medicine. *JAMA* 1994; 272: 1851-1857.
- MCDONALD CJ. Medical heuristics: the silent adjudicators of clinical practice. *Ann Intern Med* 1996; 124: 56-62.
- MCNEIL BJ, KEELER E, ADELSTEIN SJ. Primer on certain elements of medical decision making. *N Engl J Med* 1975; 293: 211-215.
- MCNEIL BJ, HANLEY JA. Statistical approaches to clinical predictions. *N Engl J Med* 1981; 304: 1292-1294.
- POZO F. La eficacia de las pruebas diagnósticas (I). *Med Clin (Barc)* 1988; 90: 779-785.
- POZO F. La eficacia de las pruebas diagnósticas (II). *Med Clin (Barc)* 1988; 91: 177-183.
- REDELMEIER DA, SHAFIR E. Medical decision making in situations that offer multiple alternatives. *JAMA* 1995; 273: 302-305.
- ROSE G. *The Strategy of Preventive Medicine*. Oxford. Ed Oxford University Press, 1992.
- ROSS JM, SOX HG. Clinical problem solving. *N Eng J Med* 1995; 333: 1557-1560.
- SACKETT DL. A primer on the precision and accuracy of the clinical examination. *JAMA* 1992; 267: 2638-2644.
- SCHWARTZ WB, GORRY GA, KASSIRER JP, ESSIG A. Decision analysis and clinical judgment. *Am J Med* 1973; 55: 459-472.
- SOX HC. Probability theory in the use of diagnostic tests. *Ann Intern Med* 1986; 104: 60-66.
- SOX HC, MARGULIES I, SOX CH. Psychologically mediated effects of diagnostic tests. *Ann Intern Med* 1981; 95: 680-685.
- WASSON JH, SOX HC, NEFF RK, GOLDMAN L. Clinical prediction rules. Applications and methodological standars. *N Engl J Med* 1985; 313: 793-799.

WEINSTEIN MC, FINEBERG HV. Clinical decision analysis. WB Saunders, Philadelphia, 1980.

FIABILIDAD

ROSNER B. The Kappa Statistic. En: Fundamentals of biostatistics. Belmont. Ed. Duxbury Press, 1995:423. Referencia para todos los temas relacionados con la estadística (índice kappa, coeficiente de correlación intraclases, coeficiente de variación, intervalos de confianza). Es claro, sencillo, muy didáctico, donde todas las explicaciones van acompañadas de ejemplos muy ilustrativos.

STREINER DL, NORMAN GR. Health Measurement scales. A practical guide to their development and use. Oxford. Oxford University Press, 1989. Como su mismo título indica, se trata de una guía práctica para el desarrollo de medidas de la salud donde hay un componente subjetivo, y por tanto, tiene cabida la variabilidad.

Referencias complementarias

BLAND JM, ALTMAN DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1988; 1: 307-310.

BRAVO G, POTVIN L. Estimating the reliability of continuous measures with cronbach's alpha or the intraclass correlation coefficient: toward the integration of two traditions. J Clin Epidemiol 1991; 44: 381-390.

FEINSTEIN AR, CICHETTI DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990; 43: 543-549.

CICHETTI DV, FEINSTEIN AR. High agreement but low kappa: II. Resolving the paradoxes. J Clin Epidemiol 1990; 43: 551-558.

HERNÁNDEZ AGUADO I, PORTA SERRA M, MIRALLES M, GARCÍA BENAVIDES F, BOLÚMAR F. La cuantificación de la variabilidad en las observaciones clínicas. Med Clin (Barc) 1990; 95: 424-429.

MEDIDA DE LOS RESULTADOS SOBRE LA SALUD

BADÍA X. La medida de la calidad de vida relacionada con la salud en la evaluación económica: técnicas de obtención de preferencias en los estados de salud. En: Sacristán JA, Badía X, Rovira J. Farmacoeconomía: Evaluación Económica de Medicamentos. Madrid. Editores Médicos, 1995: 77-99. El autor realiza una clara explicación de diferentes sistemas de ajuste por calidad de vida de los resultados sobre la salud.

GUYATT GH, SACKETT DL, COOK DJ. Evidence-Based Medicine Working Group. Guías para usuarios de la literatura médica. II. Cómo utilizar un artículo sobre tratamiento o prevención. A. ¿Son válidos los resultados del estudio? JAMA 1993; 270: 2598-2601.

GUYATT GH, SACKETT DL, COOK DJ, Evidence-Based Medicine Working Group. Guías para usuarios de la literatura médica. II. Cómo utilizar un artículo sobre tratamiento o prevención. B. ¿Cuáles son los resultados? ¿Me ayudarán en la asistencia a mis pacientes? JAMA 1993; 271: 59-63. Junto al anterior, forma parte de

una serie de artículos de la revista JAMA cuyo objetivo es ayudar a realizar una lectura crítica sobre la evidencia científica de la eficacia clínica. En este caso concreto, sobre la medida de los resultados sobre la salud, fundamentalmente de los ensayos clínicos.

Referencias complementarias

- BECK JR, KASSIRER JP, PAUKER SG. A Convenient Approximation of Life Expectancy (The DEALE). I. Validation of the method. II. Use in medical decision making. *Am J Med* 1982; 73: 883-888; 889-897.
- BECK JR, PAUKER SG. The Markov proces in medical prognosis. *Med Decis Making* 1983; 3: 341-358.
- BOWE TR. Measuring patient preferences: rating scale versus standard gamble. *Med Decis Making* 1995; 15: 283-285.
- KASSIRER JP. Incorporating patients' preferences into medical decisions. *N Engl J Med* 1994; 330: 1895-1896.
- NORD E. The Person-trade-off approach to valuing health care programs. *Med Decis Making* 1995; 15: 201-208.
- READ JL, QUINN RJ, BERWICK DM, FINEBERT HV, WEINSTEIN MC. Preferences for health outcomes: comparisons of assesment methods. *Med Decis Making* 1984; 4: 315-329.
- REDELMEIER DA, SHAFIR E. Medical decision making in situations that offer multiple alternatives. *JAMA* 1995; 273: 302-305.
- ROTHMAN KJ. Medidas del efecto. En: *Epidemiología moderna*. Madrid. Ed: Díaz de Santos, 1987: 43-48.
- ROSNER B. Hypothesis testing: Person-Time Data. En: *Fundamentals of biostatistics*. Belmont. Ed Duxbury Press, 1995: 585-631.
- SOX HC, BLATT MA, HIGGINS MC, MARTON KI. Measuring the outcome of health and decision making when the outcomes have several dimensions. En: *Medical Decision Making*. Boston. Ed: Butterworth Publishers, 1988: 167-200; 201-237.
- TORRANCE GW. Utility approach to measuring health-related quality of life. *J Chronic Dis* 1987; 40: 593-600.
- Decision about life-threatening risks. *N Engl J Med* 1994; 331: 193-196.
- Discounting health care: only a matter of timing? Editorial. *Lancet* 1992; 340: 148-149.

EVALUACION ECONÓMICA

- BADÍA X, ROVIRA J. Evaluación económica de medicamentos. Un instrumento para la toma de decisiones en la práctica clínica y la política sanitaria. *DuPont Pharma*, 1994. Es un libro que consigue convertir, tal y como titula el prólogo, la evaluación económica en un instrumento cotidiano. Realiza una presentación de casos reales que ilustran de forma muy clara las diferentes técnicas de evaluación económica.

- DRUMMOND MF, STODDART GL, TORRANCE GW. Métodos para la evaluación económica de los programas de atención de la salud. Madrid. Ed: Díaz de Santos 1991. Es un texto clásico, lleno de ejemplos prácticos y muy didáctico, muy fácil de leer. De hecho está orientado para su utilización en talleres y diversas experiencias académicas.
- SACRISTÁN JA, BADÍA X, ROVIRA J. Farmacoeconomía: Evaluación Económica de Medicamentos. Madrid. Editores Médicos, 1995. Aunque en principio va dirigido a la evaluación económica de los medicamentos, la metodología es aplicable a cualquier tecnología sanitaria. De hecho, destacaría el capítulo 6, "Temas controvertidos en la evaluación económica de tecnologías sanitarias" de Joan Rovira Forns y la propuesta de estandarización de la metodología de la evaluación económica, realizada por un grupo multidisciplinario y patrocinado por el Ministerio de Sanidad y Consumo.

Referencias complementarias

- BIRCH S, GAFNI A. Cost-effectiveness/utility analyses. Do current decision rules lead us to where we want to be? *J Health Econ* 1992; 11: 279-296.
- DETSKY AS, NAGLIE IG. A clinicians's guide to cost-effectiveness analysis. *Ann Inter Med* 1990; 113: 147-154.
- DOUBILET P, WEINSTEIN MC, McNEIL BJ. Use and misuse of the term cost effective in Medicine. *N Engl Med J* 1986; 314: 253-256.
- GAFNI A. Willingness-to-pay as a measure of benefits. *Medical Care* 1991; 29: 1246-1252.
- HURLEY S. A review of cost-effectiveness analyses. *Med J Aust* 1990; 153: 20-23.
- RUBIO S. Glosario de Economía de la Salud. Madrid. Ed. Díaz de Santos, 1994.
- WEINSTEIN MC, STASON, WB. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med* 1977; 296: 716-721.
- Discounting health care: only a matter of timing? Editorial. *Lancet* 1992; 340: 148-149.

PÁGINAS WEB DE INTERÉS

Agencias de evaluación de tecnologías

ÁMBITO INTERNACIONAL

INAHTA: International Network of Agencies for Health Technology Assesment
<http://www.inahta.org>

ESPAÑA

Agencia de Evaluación de Tecnologías Sanitarias (AETS). Instituto de Salud Carlos III.

<http://www.isciii.es/unidad/aet>

Agencia de Evaluación de Tecnologías Sanitarias de Galicia (AVALIA-T):

<http://www.sergas.es>

Agencia d'Avaluació de Tecnologia i Recerca Mèdiques (AATM)

<http://www.aatm.es>

Agencia de Evaluación de Tecnologías Sanitarias de Andalucía (AETSA)

<http://www.csalud.junta-andalucia.es/org-dep/AETSA/default.htm>

Osasunerako Teknologien Ebaluaketa (Osteba)

http://www.euskadi.net/sanidad/osteba/indice_c.htm

SOCIEDADES DE EVALUACIÓN DE TECNOLOGÍAS SANITARIAS

ISTAHC: International Society for Technology Assesment Community

<http://www.istahc.org>

BASES DE DATOS SOBRE REVISIONES SISTEMÁTICAS

DARE: Database of Abstracts of Reviews of Effectiveness

<http://nhscrd.ac.uk/darehp.htros>

Cochrane Library

<http://www.cochranelibrary.net>

REVISTAS DE EVALUACIÓN DE TECNOLOGÍAS

ACP Journal Club

<http://www.acponline.org>

Bandolier Journal

<http://www.jr2.ox.ac.uk/Bandolier>

Evidence Based Medicine Journal

<http://www.bmjpg.com/data/ebm.htm>

Evidence Based Purchasing

<http://www.epi.bris.ac.uk/rd/publicat/ebpurch/index.htm>

Journal of Evidence-Based Health Care

<http://www.ihs.ox.ac.uk/ebhc>

