

LECCIÓN 2. PROBABILIDADES

SAAVEDRA, P.

1. FENÓMENOS DETERMINÍSTICOS Y FENÓMENOS ALEATORIOS

Uno de los significados del término *fenómeno* es el de *manifestación de actividad que se produce en la naturaleza* (María Moliner). En tal sentido, los fenómenos *deterministas* son aquellos cuyos resultados son *predecibles* mientras que los *aleatorios* son los que no pueden predecirse con certeza. A menudo los fenómenos que se observan en las ciencias físico-químicas son deterministas, mientras que la **complejidad** de los fenómenos tratados en las biociencias hace que éstos sean aleatorios y normalmente sólo sean predecibles en términos de *probabilidad* o *esperanza*. Es por tanto la probabilidad una herramienta esencial en el campo de las ciencias de la vida.

Ejemplo 1. Si se deja caer un objeto desde una cierta altura h , el tiempo t que éste tarda en llegar al suelo es independiente de su masa y está **determinado** por la ley física de la *caída libre de los cuerpos*, más concretamente, $t = \sqrt{2h/g}$, siendo $g \approx 9,8 \text{ m/s}^2$ al nivel del mar. Es éste por tanto, un fenómeno determinista.

Ejemplo 2. Para un gen con alelos $\{A, a\}$, supóngase que se aparean dos sujetos heterocigotos Aa . No es posible desde luego predecir con certeza el genotipo del descendiente, y por tanto, este fenómeno es de naturaleza aleatoria.

En el ejemplo anterior, aunque no sea posible predecir con certeza el genotipo del descendiente, sí que podemos afirmar que la probabilidad de que sea homocigoto AA es de $1/4$. Para obtener esta probabilidad es necesario previamente describir todos los resultados que pueden presentarse. Surge así el concepto de *espacio muestral de probabilidad* que se define en la siguiente sección.

2. ESPACIOS MUESTRALES Y SUCESOS

Definición. El *espacio muestral asociado a un fenómeno aleatorio* es el conjunto de todos los posibles resultados elementales que puede manifestar ese fenómeno.

Ejemplo 3. El espacio muestral asociado al fenómeno aleatorio descrito en el [ejemplo 2](#) es el conjunto de *los pares de alelos* cuyas componentes primera y segunda son los alelos recibidos del padre y la madre respectivamente. Dado que ambos progenitores son heterocigotos Aa, el espacio muestral es el conjunto: $E = \{AA, Aa, aA, aa\}$. Nótese que todos los elementos de E tienen la misma posibilidad de ocurrir.

Ejemplo 4. Considérese una cierta población de sujetos que representamos por \mathcal{P} . Si de esta población se selecciona un sujeto al azar, cada uno de los sujetos de la población tiene la misma posibilidad de ser elegido, de ahí que el espacio muestral sea la propia población; esto es: $E = \mathcal{P}$.

Ejemplo 5. Considérese un paciente afectado por una enfermedad de muy mal pronóstico en relación con la supervivencia. Obviamente, el tiempo que transcurrirá desde el diagnóstico hasta su muerte no es predecible con certeza y por tanto, tal fenómeno (el tiempo de supervivencia) es de naturaleza aleatoria. El espacio muestral es por tanto el conjunto de todos los posibles tiempos de supervivencia; esto es, el conjunto de todos los números positivos ($E = \mathbb{R}^+$).

En el contexto del fenómeno aleatorio descrito en el [ejemplo 2](#) podemos referirnos al *resultado no elemental* consistente en que el descendiente del apareamiento sea heterocigoto. Tal resultado se presentará cuando ocurra alguna de las posibilidades descritas en el conjunto $\{Aa, aA\}$, el cual es un *subconjunto del espacio muestral*. Esto nos lleva a la siguiente definición de suceso:

Definición. En el contexto de un fenómeno aleatorio con espacio muestral asociado E , entenderemos, para los propósitos de este curso, que un **suceso** A es *cualquier subconjunto* de E , lo cual se expresa como: $A \subset E$.

Ejemplo 6. El suceso “sobrevivir al menos cinco años” correspondiente al fenómeno aleatorio descrito en el [ejemplo 5](#) es: $\{t \in \mathbb{R}^+ \mid t \geq 5\}$.

Las siguientes definiciones corresponden a sucesos especiales y a operaciones entre sucesos.

Suceso seguro. Es el propio espacio muestral E . Para el fenómeno de apareamiento descrito en el [ejemplo 2](#), el suceso seguro es: $E = \{AA, Aa, aA, aa\}$.

Suceso imposible. Cualquier resultado que sepamos con seguridad que no se va a producir. Lo representaremos por \emptyset . Así por ejemplo, para un gen con alelos $\{A, a\}$, en un apareamiento de un sujeto AA con otro Aa , podemos escribir: *‘El descendiente es’ aa ’ = \emptyset .

Suceso contrario. Dado un suceso cualquiera A en un espacio muestral E , el suceso contrario, representado por A^C , consiste en la no ocurrencia de A . Estará por tanto formado por todos los resultados de E que no estén en A . Para el fenómeno aleatorio descrito en el [ejemplo 2](#), el suceso contrario de *ser homocigoto* es *ser heterocigoto*, y podemos escribir: $\{AA, aa\}^C = \{Aa, aA\}$.

Unión de sucesos. Dados dos sucesos A y B , el suceso A *unión con* B , que representamos como $A \cup B$, consiste en que ocurra A ó B (al menos uno de los dos). Está formado por tanto por la reunión de todos los resultados de A y B .

Intersección de sucesos. Dados dos sucesos A y B , el suceso A *intersección con* B es otro suceso que representamos por $A \cap B$, y que consiste en que ocurran A y B simultáneamente. Está formado por tanto por los elementos que son comunes a ambos sucesos.

Siguiendo con el fenómeno aleatorio descrito en el [ejemplo 2](#), considérense los sucesos: $A = \text{‘Ser heterocigoto’} = \{Aa, aA\}$ y $B = \text{‘Recibir del padre el alelo } A\text{’} = \{AA, Aa\}$. Se tiene entonces:

- $A \cup B = \text{‘Ser heterocigoto ó recibir del padre el alelo } A\text{’} = \{AA, Aa, aA\}$.
- $A \cap B = \text{‘Ser heterocigoto y recibir del padre el alelo } A\text{’} = \{Aa\}$

Sucesos incompatibles o mutuamente excluyentes. Los sucesos A y B son *incompatibles* si no pueden presentarse simultáneamente; esto es: si $A \cap B = \emptyset$. En el

contexto del fenómeno aleatorio descrito en el [ejemplo 2](#), los sucesos ‘*Ser homocigoto*’ y ‘*Ser heterocigoto*’ son obviamente incompatibles.

Diagramas de Venn. Los sucesos son conjuntos y como tales pueden representarse gráficamente mediante los diagramas de Venn. Tal como veremos, permiten resolver de forma simple diversos problemas de probabilidad. La figura 2.1 corresponde al estudio de Telde ([Boronat et al, 2005](#)) y representa los conjuntos de sujetos con *diabetes mellitus*, *hipertensión arterial* (criterio OMS) y *obesidad central* (criterio ATP-3).

En la [figura 1](#) se muestran los recuentos de sujetos correspondientes a cada intersección. Por ejemplo, 72 es el número de sujetos que pertenecen simultáneamente a los tres subconjuntos; esto es, los que son *diabéticos*, *hipertensos* y tienen *obesidad central*.

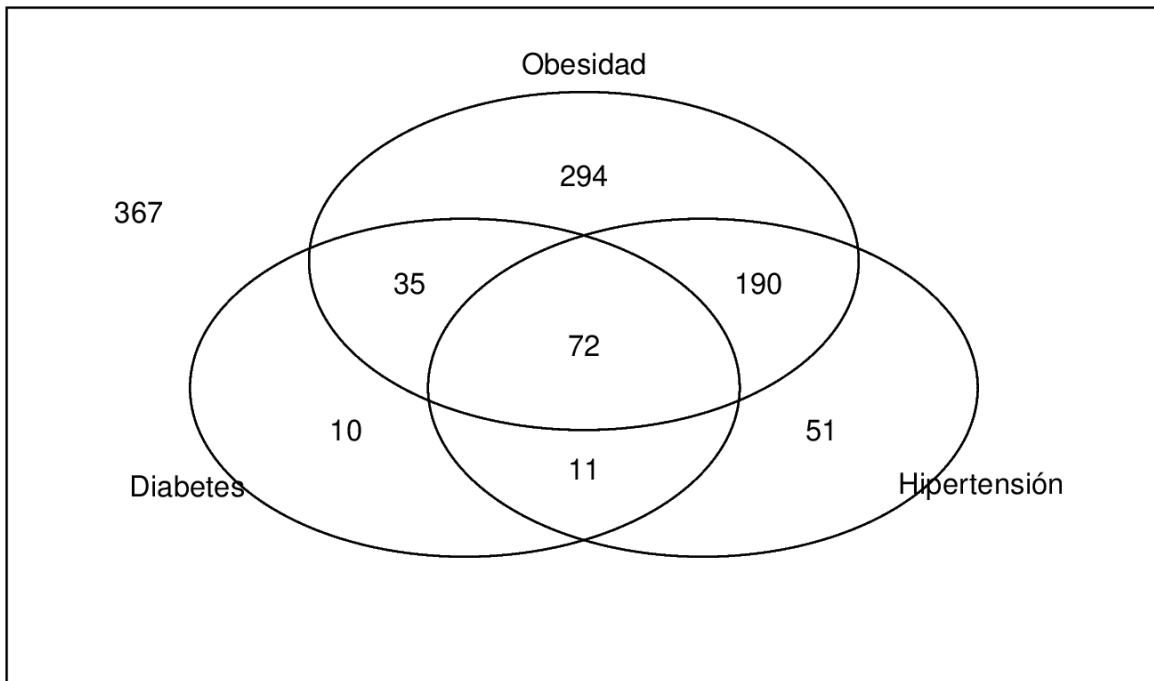


FIGURA 1. Diagrama de Venn correspondiente al estudio de Telde

3. CONCEPTOS DE PROBABILIDAD Y SU AXIOMÁTICA

La **probabilidad** es una medida del grado de creencia acerca de que un suceso pueda ocurrir.

Hay diferentes escenarios en los que ésta puede determinarse:

1. Cuando puede darse de forma totalmente objetiva, la probabilidad recibe el nombre de ***probabilidad exacta***. Así por ejemplo, para el fenómeno aleatorio descrito en el [ejemplo 3](#), puede admitirse con plena objetividad que los cuatro resultados cuentan con las mismas posibilidades de ocurrir, y por tanto la probabilidad de que el descendiente tenga genotipo AA es $1/4$, respondiendo así esta medida al concepto de probabilidad exacta.
2. Cuando no es posible darla con exactitud pero sí realizar una secuencia de observaciones independientes del fenómeno aleatorio subyacente, la probabilidad de un determinado suceso puede basarse en la frecuencia con la que éste se presenta. Este es el concepto de ***probabilidad frecuentista***. Así por ejemplo, si en una amplia muestra de personas de una población se observa que el 50% de los sujetos están infectados por la bacteria *Helicobacter pylori*, podemos entonces admitir que la probabilidad de que un sujeto aleatoriamente seleccionado de tal población esté afectado por la bacteria es $1/2$.
3. La probabilidad puede también usarse para expresar una determinada *creencia personal*. En ocasiones, un clínico puede verse forzado a dar un diagnóstico de forma subjetiva, expresando en términos de probabilidad su creencia de que un paciente tenga una cierta enfermedad. Tal concepto de probabilidad recibe el nombre de ***probabilidad subjetiva***.

Independientemente del modo en el que se defina la probabilidad, ésta debe satisfacer los siguientes axiomas:

A1. $\Pr(E) = 1.$

A2. Si $A \cap B = \emptyset$, entonces $\Pr(A \cup B) = \Pr(A) + \Pr(B)$

De los axiomas anteriores se deducen las siguientes propiedades de la probabilidad:

P1. $\Pr(\emptyset) = 0$

P2. $\Pr(A^C) = 1 - \Pr(A)$

P3. $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$

Las demostraciones se dejan como ejercicio.

Cuando el espacio muestral es finito y todos sus elementos pueden considerarse *equiprobables*, la medida de probabilidad queda determinada por la llamada *ley de Laplace*, que se formula a continuación.

Ley de Laplace. Considérese un espacio muestral finito con N elementos equiprobables. Obviamente, en tal caso la probabilidad de cada elemento (resultado elemental) es $1/N$. De esta forma, si un suceso A tiene k elementos (k resultados favorables al suceso A) su probabilidad será, de acuerdo con (A.2), k/N . Esta es la ley de Laplace que habitualmente se expresa en los siguientes términos:

$$\Pr(A) = \frac{\#(A)}{\#(E)} = \frac{k}{N} = \frac{\text{Casos favorables a } A}{\text{Casos totales}}$$

donde $\#(A)$ representa el número de elementos del suceso A .

4. PROBABILIDAD CONDICIONAL

En diversas ocasiones, la predicción de un fenómeno aleatorio se realiza a través de otros fenómenos observables. Así por ejemplo, la predicción de la enfermedad de Alzheimer (AD) puede hacerse a partir de diversos factores observables, entre ellos, el hábito de realizar ejercicio físico.

Los datos que aparecen resumidos en la tabla 1 corresponden a un estudio prospectivo de cohortes realizado por [Scarmeas et al \(2009\)](#). Los sujetos del estudio se clasificaron según realizaran o no alguna actividad física. Tras un seguimiento a lo largo de un cierto periodo, se observaron las incidencias de AD en cada uno de los grupos considerados. En tal escenario definimos los siguientes sucesos:

$A =$ *Un sujeto aleatoriamente seleccionado realiza actividad física a la inclusión en el estudio*

$B =$ *Un sujeto aleatoriamente seleccionado desarrollará la AD a lo largo del seguimiento*

TABLA 1. Incidencias de la AD según cohorte de actividad física

	Actividad física	
	Alguna (A) N=1360	Ninguna (A^c) N=520
Enfermedad de Alzheimer, n(%)	180 (13.2)	102 (19.6)

La expresión $\Pr(B | A)$ representa la probabilidad de que ocurra el suceso B sabiendo que ocurre (o ha ocurrido) el suceso A . En el ejemplo que nos ocupa es la probabilidad de que, en la población considerada, una persona que practica alguna forma de actividad física llegue a desarrollar la AD. A partir de los datos de la tabla anterior, puede hacerse la aproximación $\Pr(B | A) \approx 0.132$, que se obtiene simplemente como la incidencia de la AD (180) entre aquellos que realizaban alguna forma de actividad física (1360); esto es:

$$\Pr(B | A) \approx \frac{\#(A \cap B)}{\#(A)} = \frac{180}{1360}$$

Si el numerador y denominador del cociente anterior se dividen por el número total de sujetos incluidos en el estudio (1890), se obtiene:

$$\Pr(B | A) \approx \frac{180/1890}{1360/1890} = \frac{\Pr(A \cap B)}{\Pr(A)} \approx 0.132$$

Definición: Para un suceso A tal que $\Pr(A) > 0$, la **probabilidad de un suceso B condicionada por A** se define como:

$$\Pr(B | A) = \frac{\Pr(A \cap B)}{\Pr(A)}$$

De esta expresión se deduce la siguiente *regla multiplicativa de la probabilidad*:

$$\Pr(A \cap B) = \Pr(A) \Pr(B | A)$$

Independencia de sucesos. Un suceso B es **independiente** de A si $\Pr(B | A) = \Pr(B)$; esto es: la ocurrencia o no de A no modifica la probabilidad de que ocurra B .

Es fácil comprobar que si un suceso B es independiente de A , entonces A lo es de B . En efecto, supóngase que B es independiente de A y por tanto que $P(B|A) = P(B)$. Entonces:

$$\Pr(A | B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr(A) \Pr(B | A)}{\Pr(B)} = \frac{\Pr(A) \Pr(B)}{\Pr(B)} = \Pr(A)$$

lo que supone que A es independiente de B . Se tiene por tanto la siguiente regla multiplicativa: *los sucesos A y B son independientes, si y sólo si:*

$$\Pr(A \cap B) = \Pr(A) \Pr(B)$$

En la siguiente definición se extiende el concepto de independencia a cualquier número de sucesos.

Definición: Un conjunto de sucesos A_1, \dots, A_n , son **mutuamente independientes** si la probabilidad de que ocurra simultáneamente cualquier subconjunto de ellos es igual al producto de sus probabilidades.

Una propiedad útil de la independencia de sucesos es la que sigue: *si B es independiente de A , también lo es de su contrario.* En efecto:

$$\Pr(B) = \Pr(B \cap A) + \Pr(B \cap A^C) = \Pr(B) \Pr(A) + \Pr(B \cap A^C)$$

Despejando $\Pr(B \cap A^C)$ se tiene:

$$\Pr(B \cap A^C) = \Pr(B) - \Pr(B) \Pr(A) = \Pr(B) (1 - \Pr(A)) = \Pr(B) \Pr(A^C)$$

Por ser entonces $\Pr(B \cap A^C) = \Pr(B) \Pr(A^C)$, B es independiente de A^C .

Una buena parte de problemas de la probabilidad total pueden resolverse utilizando los teoremas de la *probabilidad total* y de *Bayes*. Estos teoremas se basan en el concepto de sistema completo de sucesos que se define a continuación.

Definición: Los sucesos A_1, \dots, A_n forman un *sistema completo de sucesos* si satisfacen las siguientes propiedades:

1. Son mutuamente excluyentes, esto es: $A_i \cap A_j = \emptyset$, $i \neq j$.
2. Con seguridad ocurrirá alguno de ellos, esto es: $E = A_1 \cup \dots \cup A_n$.

Teorema: (de la probabilidad total) Sea A_1, \dots, A_n un sistema completo de sucesos. Se tiene entonces que:

$$\Pr(B) = \sum_{i=1}^n \Pr(B|A_i) \Pr(A_i)$$

Ejemplo 7: Considérese un gen con alelos $\{A, a\}$ y supóngase que se aparean dos individuos de una población en la que los genotipos AA , Aa y aa se dan en proporciones p , q y r ($p + q + r = 1$) respectivamente. La tabla 2 muestra la distribución de probabilidades sobre el sistema completo de sucesos de todos los posibles apareamientos.

TABLA 2. Sistema completo de apareamientos

M_1 (AA,AA)	M_2 (AA,Aa)	M_3 (AA,aa)	M_4 (Aa,Aa)	M_5 (Aa,aa)	M_6 (aa,aa)
p^2	$2pq$	$2pr$	q^2	$2qr$	r^2

Para el cálculo de las probabilidades se ha tenido en cuenta que los genotipos de los dos elementos que se aparean son independientes. También se ha considerado, por ejemplo,

que el apareamiento $(AA, Aa) = (Aa, AA)$. Por tal motivo, la probabilidad de este apareamiento es $2 \times p \times q$. Denotemos ahora como $\{AA\}_1$, $\{Aa\}_1$ y $\{aa\}_1$ los sucesos consistentes en que el descendiente del apareamiento sea de genotipo AA , Aa y aa respectivamente. Entonces, de acuerdo con el teorema de la probabilidad total se tiene que:

$$\Pr(\{AA\}_1) = \sum_{i=1}^6 \Pr(\{AA\}_1 | M_i) \Pr(M_i) = p^2 + pq + \frac{q^2}{4}$$

$$\Pr(\{Aa\}_1) = \sum_{i=1}^6 \Pr(\{Aa\}_1 | M_i) \Pr(M_i) = pq + 2pr + \frac{q^2}{2} + qr$$

$$\Pr(\{aa\}_1) = \sum_{i=1}^6 \Pr(\{aa\}_1 | M_i) \Pr(M_i) = \frac{q^2}{4} + qr + r^2$$

Teorema: (Bayes) *Para un sistema completo de sucesos A_1, \dots, A_n se verifica:*

$$\Pr(A_i | B) = \frac{\Pr(B | A_i) \Pr(A_i)}{\Pr(B)} = \frac{\Pr(B | A_i) \Pr(A_i)}{\sum_{j=1}^n \Pr(B | A_j) \Pr(A_j)}$$

Ejemplo 8: (Pruebas de paternidad). En el mismo escenario que en el [ejemplo 7](#), supóngase que se observa que el descendiente de un apareamiento es del genotipo Aa y se desea conocer la posibilidad de que descienda del apareamiento $M_2 = (AA, Aa)$. Entonces, de acuerdo con el teorema de Bayes, tenemos:

$$\Pr(M_2 | Aa) = \frac{\Pr(Aa | M_2) \Pr(M_2)}{\Pr(Aa)} = \frac{pq}{pq + 2pr + q^2/2 + qr}$$

5. APLICACIONES EPIDEMIOLÓGICAS DE LA PROBABILIDAD

En esta sección se utiliza la probabilidad para establecer asociaciones entre dos sucesos, principalmente entre un factor de riesgo o exposición y una enfermedad que podría ser causada por aquél. Se definen los conceptos de *riesgo relativo* y *odds-ratio* y se analizan

sus propiedades. Asimismo se discute la validez de la odds-ratio en cualquier tipo de estudio biomédico independiente de su diseño. Se utiliza finalmente el teorema de la probabilidad total para ajustar tasas por grupos de edad.

Riesgo relativo. El *riesgo relativo* (RR) se utiliza normalmente en un escenario en el que un suceso A representa una cierta exposición que puede estar asociada con una enfermedad B . Este se define por:

$$RR = \frac{\Pr(B | A)}{\Pr(B | A^C)}$$

Por tanto esta medida indica cuanto mayor es la probabilidad de que ocurra B en presencia de A que en presencia de A^C .

Los datos de la tabla 3, que corresponden al estudio de Telde, permiten establecer la asociación entre la *obesidad central* (criterio ATP-III) y la *diabetes mellitus*.

TABLA 3. Estudio de Telde: asociación obesidad - diabetes mellitus

	Obesidad central (OC)	
	Sí ($N = 592$)	No ($N = 438$)
<i>Diabetes mellitus</i> (DM), %	18.1	4.8

De acuerdo con estos datos se tiene:

$$RR = \frac{\Pr(DM | OC)}{\Pr(DM | OC^C)} = \frac{0,181}{0,048} \approx 3,77$$

Este resultado indica que la probabilidad de diabetes es 3,77 veces mayor en los sujetos con obesidad central que en los sujetos sin ella.

En la siguiente proposición se da una propiedad esencial del riesgo relativo:

Proposición: B es independiente de A si y solamente si $RR = 1$.

En efecto, si se satisface que B independiente de A , también lo es de A^C y de esta forma se tiene:

$$RR = \frac{\Pr(B | A)}{\Pr(B | A^C)} = \frac{\Pr(B)}{\Pr(B)} = 1$$

Recíprocamente, si $RR = 1$, $\Pr(B | A) = \Pr(B | A^C)$. Entonces, utilizando el teorema de la probabilidad total se tiene:

$$\Pr(B) = \Pr(B | A) \Pr(A) + \Pr(B | A^C) \Pr(A^C) = \Pr(B | A)$$

Por ser $\Pr(B | A) = \Pr(B)$, B es entonces independiente de A , lo que completa la demostración de la proposición.

Cuando $RR > 1$, decimos que existe asociación positiva. En el caso de que A represente una exposición y B una enfermedad, decimos que A es de *riesgo* para B . Si $RR < 1$ se dice entonces que A es de *protección* para B .

Odds-ratio. Una medida más propia de asociación entre dos sucesos A y B es la *odds-ratio* (OR) (razón o cociente de odds), la cual se define por:

$$OR = \frac{\Pr(A \cap B) \Pr(A^C \cap B^C)}{\Pr(A^C \cap B) \Pr(A \cap B^C)}$$

Una propiedad útil de la OR que puede verificarse fácilmente es:

$$OR = \frac{\Pr(B | A) \Pr(B^C | A^C)}{\Pr(B | A^C) \Pr(B^C | A)} = \frac{\Pr(A | B) \Pr(A^C | B^C)}{\Pr(A^C | B) \Pr(A | B^C)}$$

Puede comprobarse también que la independencia de A y B es equivalente a $OR = 1$.

Se analiza a continuación el uso de la odds-ratio para evaluar la asociación entre una exposición F y una enfermedad D según el estudio sea prospectivo o retrospectivo.

1. *Estudios de cohortes.* Considérese una población de *sujetos inicialmente sanos* y clasificados en dos cohortes determinadas por la exposición o no a un cierto factor F . Supóngase ahora que a lo largo de un periodo de seguimiento se observan las incidencias de una enfermedad D en cada una de las dos cohortes. La odds-ratio se puede expresar en la forma:

$$OR = \frac{\Pr(D | F) / \Pr(D^C | F)}{\Pr(D | F^C) / \Pr(D^C | F^C)}$$

El numerador expresa *cuanto más probable es enfermar que no enfermar en los sujetos con la exposición F* . Esta relación recibe el nombre de *superioridad* de la enfermedad en los expuestos. El denominador representa la misma relación pero entre los sujetos no expuestos (F^C). Si la $OR > 1$ significa que la superioridad de la enfermedad es mayor en los expuestos que en los que no lo están, y de ahí, el factor F es de riesgo para D . Recíprocamente, si $OR < 1$, un razonamiento análogo nos lleva a que F es protector para D . Resumimos por tanto la interpretación de la OR en el modo que sigue:

1. $OR = 1$ es equivalente a que D es independiente de F .
 2. $OR > 1$ equivale a que F es de riesgo para D
 3. $OR < 1$ equivale a que F es de protección para D
2. *Estudios de caso-control.* Este es un escenario en el que inicialmente los sujetos de la población están clasificados en enfermos (D) y sanos (D^C) y se quiere investigar retrospectivamente los posibles factores que han podido causar la enfermedad. En este estudio no es posible en general estimar las probabilidades de D en los grupos F y F^C y por tanto, no es posible estimar el riesgo relativo, pero sí se pueden obtener las probabilidades del factor en los grupos de enfermos (D) y sanos (D^C) y de ahí obtener la odds-ratio en la forma:

$$OR = \frac{\Pr(F | D) \Pr(F^C | D^C)}{\Pr(F | D^C) \Pr(F^C | D)}$$

Logroscino *et al* (1996) llevaron a efecto un estudio de caso-control para evaluar la asociación entre las dietas antioxidantes y la enfermedad de Parkinson. En el estudio incluyeron 110 pacientes parkinsonianos y 287 sujetos controles. En la@tbl-parkinson se muestran las tasas de sujetos que superaron el umbral de 1353 calorías según grupo de estudio.

TABLA 4. Ingesta de calorías según grupo de estudio

	Parkinson (D) n=110	Control (D^C) n=287	OR
Factor F : Calorías > 1353, %	70.9	41.8	3.39

De los datos de la tabla 4 se deduce que: $\Pr(F | D) = 0.709$ y $\Pr(F | D^C) = 0.418$. Se tiene entonces:

$$OR = \frac{0.709 \times 0.582}{0.418 \times 0.291} \approx 3.39$$

Nótese que de estos datos no es posible estimar las probabilidades $\Pr(D | F)$ y $\Pr(D | F^C)$, y de ahí, el riesgo relativo. Para que tales probabilidades pudieran determinarse se requeriría conocer la prevalencia de la enfermedad; esto es: $\Pr(D)$. En tal caso, mediante la fórmula de Bayes se obtiene:

$$\Pr(D | F) = \frac{\Pr(F | D) \Pr(D)}{\Pr(F | D) \Pr(D) + \Pr(F | D^C) \Pr(D^C)}$$

Análogamente se obtendría $\Pr(D | F^C)$.

Tasas ajustadas por población. Abordamos ahora el problema de obtener una tasa comparable con otras poblaciones. Para ello consideramos la distribución de edades que aparece en la quinta columna de la tabla. Ésta es una distribución de *consenso*

para ser utilizada como población de referencia (población SEGI). Esto supone que otros estudios similares usarán la misma distribución y de esta forma, las diferentes tasas obtenidas serán comparables. Ajustar la tasa de DM por la población SEGI consiste en reemplazar las probabilidades $\Pr(A_i)$ por las de la población SEGI. De esta forma se obtiene que $\Pr(D) = 0.1381$ (13.81%). Esta tasa es claramente superior a las obtenidas anteriormente. Ello se explica obviamente por el hecho de que la población de Telde es más joven que la población referencial de SEGI. La tasa de DM ajustada por población de SEGI es la tasa real que habría en Telde si la distribución real de edades fuese la de SEGI.

En el estudio de Telde se obtuvo que de las 1030 personas incluidas en el estudio (población con edad superior o igual a 30 años), 128 tenían diabetes mellitus (DM), lo que supone una tasa del 12.43%. Se plantea ahora el problema de la comparabilidad de esta tasa con la de otras poblaciones. Para ello tendremos en cuenta las distribuciones mostradas en la tabla 5.

TABLA 5. Para cada decenio de edad se muestran prevalencia de DM, número total y porcentaje de habitantes de Telde. En la última columna se muestra la distribución de edades para la población referencial de SEGI.

Edad (años)	% de DM Telde	Población Telde, N.º (%)	Muestra N.º (%)	Pobla
30 – 39	2.57	16739 (34.83)	311 (30.19)	
40 – 49	5.90	11878 (24.71)	288 (27.96)	
50 – 59	16.53	8445 (17.57)	236 (22.91)	
60 – 69	33.80	6085 (12.66)	142 (13.79)	
70	30.19	4918 (10.23)	53 (5.15)	

Supongamos en primer lugar que seleccionamos aleatoriamente un sujeto de la población y consideramos los siguientes sucesos:

$D = \text{“El sujeto es diabético”}$

$A_i = \text{“El sujeto pertenece al } i\text{-ésimo grupo de edad”}$ (los que se muestran en la tabla 5).

Obviamente, los sucesos $A_i : i = 1, \dots, 5$ forman un sistema completo de sucesos. Las probabilidades condicionales $\Pr(D | A_i)$ que se muestran en la segunda columna de la tabla (expresadas en porcentajes) se corresponden con las tasas de DM por cada grupo de edad. La probabilidad de DM puede entonces obtenerse también utilizando el teorema de la probabilidad total en la forma:

$$\Pr(D) = \sum_{i=1}^5 \Pr(D | A_i) \Pr(A_i)$$

Antes de decidir los valores $\Pr(A_i)$ (probabilidades de pertenecer a cada uno de los decenios) comparemos las columnas tercera y cuarta de la tabla. Las distribuciones porcentuales de ambas columnas son muy similares. Ello se debe a que los porcentajes que aparecen en la cuarta estiman a los que aparecen en la tercera (estos últimos son los verdaderos de la población). Si los valores $\Pr(A_i)$ se identifican con los porcentajes de la cuarta columna, el lector puede comprobar fácilmente que el valor de $\Pr(D)$ es justamente 0.1243 (12.43%); esto es, la tasa anteriormente obtenida de DM. Sin embargo, se obtendría una mejor estimación considerando los verdaderos valores de $\Pr(A_i)$; esto es, los que se muestran en la tercera columna. De esta forma, $\Pr(D) = 0.1263$ (12.63%). Como puede verse, el resultado es prácticamente el mismo que se obtuvo antes. Esta última tasa o probabilidad recibe el nombre de *tasa ajustada por edad*.

6. DIAGNÓSTICO CLÍNICO

Por lo general, las pruebas diagnósticas concluyentes no se utilizan en una primera fase de un diagnóstico clínico, bien porque son invasivas para el paciente, bien por tener un coste económico elevado, bien porque sus resultados pueden demorarse de forma

inaceptable. En tales casos pueden utilizarse pruebas diagnósticas alternativas (test) que resuelvan las dificultades mencionadas aunque sus resultados puedan ser erróneos. Las situaciones que pueden darse se resumen en la tabla 6.

TABLA 6. Escenarios del diagnóstico clínico

Enfermedad	Test diagnóstico (T)	
	(+)	(-)
Presente (D)	Verdadero positivo	Falso negativo
Ausente (D^c)	Falso positivo	Verdadero negativo

Los parámetros que miden la fiabilidad de cualquier prueba o test diagnóstico son: *sensibilidad*, *especificidad*, *coeficiente falso positivo*, *coeficiente falso negativo*, *valor predictivo positivo*, *valor predictivo negativo*, *razón de verosimilitud positiva* y *razón de verosimilitud negativa*. Estos parámetros se definen en términos de las probabilidades asociadas al experimento aleatorio consistente en seleccionar de la población de estudio un sujeto al azar. Los sucesos de interés son: $D = \text{“Tener la enfermedad”}$ $T = \text{“Ser positivo”}$ (al test diagnóstico).

- La **sensibilidad** de la prueba se define como la *probabilidad de resultar positivo condicionada a estar enfermo*: $\Pr(T | D)$.
- La **especificidad** es la *probabilidad de ser negativo condicionada a estar sano*: $\Pr(T^c | D^c)$.
- El **coeficiente falso-positivo** se define como la *probabilidad de ser positivo condicionada a estar sano*: $\Pr(T | D^c)$.
- El **coeficiente falso-negativo** es la *probabilidad de ser negativo condicionada a estar enfermo*: $\Pr(T^c | D)$.

Es fácil comprobar que se cumplen las siguientes relaciones:

Sensibilidad + coeficiente falso-negativo = 1

Especificidad + coeficiente falso-positivo = 1.

- Si una vez aplicada la prueba diagnóstica el sujeto resulta positivo, interesará saber entonces cuál es la probabilidad de que realmente tenga la enfermedad; esto es: el **valor predictivo positivo** $\Pr(D | T)$.
- Análogamente, cuando la prueba resulte negativa interesará conocer la probabilidad de que el sujeto esté libre de la enfermedad; esto es: el **valor predictivo negativo** $\Pr(D^C | T^C)$.

En aquellos casos en los que se conozca la *sensibilidad*, *especificidad* y la *probabilidad de estar enfermo* $\Pr(D)$, los valores predictivo positivo y predictivo negativo se pueden obtener mediante el teorema de Bayes en las formas:

- Valor predictivo positivo

$$\Pr(D | T) = \frac{\Pr(T | D) \cdot \Pr(D)}{\Pr(T | D) \cdot \Pr(D) + \Pr(T | D^C) \cdot \Pr(D^C)}$$

- Valor predictivo negativo

$$\Pr(D^C | T^C) = \frac{\Pr(T^C | D^C) \cdot \Pr(D^C)}{\Pr(T^C | D^C) \cdot \Pr(D^C) + \Pr(T^C | D) \cdot \Pr(D)}$$

Para la aplicación del teorema de Bayes, téngase en cuenta que los sucesos $\{D, D^C\}$ forman un sistema completo.

La fiabilidad de la prueba requiere que sea a la vez sensible y específica. Una prueba diagnóstica puede tener una excelente sensibilidad pero ser absolutamente inespecífica. En este sentido son útiles los parámetros de evaluación *razón de verosimilitud positiva* y *razón de verosimilitud negativa*:

- La ***razón de verosimilitud positiva*** expresa cuanto mayor es la probabilidad de ser positiva la prueba en un enfermo que en un sano: $LR^+ = \Pr(T|D)/\Pr(T|D^C)$
- La ***razón de verosimilitud negativa***, por su parte, expresa cuanto menor es la probabilidad de ser negativo en un sano que en un enfermo: $(LR^- = \Pr(T^c|D)/\Pr(T^c|D^C))$.

Nótese que una prueba diagnóstica podría tener una sensibilidad del 80% pero si su especificidad es del 20%, la probabilidad que tiene de resultar positiva en los enfermos es la misma que en los sanos ($LR^+ = 1$) lo que equivale a un diagnóstico al azar. La tabla 7 resume los parámetros considerados.

TABLA 7. Parámetros de evaluación de los test diagnósticos

Parámetro	Definición
<i>Sensibilidad</i>	$\Pr(T D)$
<i>Especificidad</i>	$\Pr(T^C D^C)$
<i>Coficiente falso positivo</i>	$\Pr(T D^C)$
<i>Coficiente falso negativo</i>	$\Pr(T^C D)$
<i>Valor predictivo positivo</i>	$\Pr(D T)$
<i>Valor predictivo negativo</i>	$\Pr(D^C T^C)$
<i>Razón de verosimilitud positiva</i>	$\Pr(T D) / \Pr(T D^C)$
<i>Razón de verosimilitud negativa</i>	$\Pr(T^C D) / \Pr(T^C D^C)$

En el año 2008, un comité de expertos redefinió el diagnóstico de diabetes mellitus a través de la *hemoglobina glicada* (A1C). El comité presentó información sobre la historia del diagnóstico de la diabetes y la relación entre los niveles de glucemia y las

complicaciones microvasculares de la diabetes. Señalaron que mientras la aparición de la diabetes de tipo 1 conlleva una brusca elevación de los niveles de glucemia, la diabetes de tipo 2 tenía normalmente un inicio más insidioso. Por tal motivo propusieron como umbral de la A1C para la diabetes mellitus el valor 6.5. En la tabla 8 se muestran los parámetros de evaluación de este criterio ($A1C > 6.5$), utilizando los datos del estudio de Telde. Se consideró como *gold estándar* el diagnóstico de diabetes por glucemia basal y test oral de la glucosa.

TABLA 8. Diagnóstico de la diabetes mellitus por HbA1c %. Los datos son frecuencias y % por columnas

Diabetes mellitus		
HbA1c%	No (N = 902)	Si (N = 128)
6,5%	898 (99.6)	53 (41.4)
> 6,5%	4 (0.4)	75 (58.6)

De la tabla anterior se obtiene:

- *Sensibilidad* = $\Pr(A1C > 6,5 | DM) = 0.586$ (58,6%)
- *Especificidad* = $\Pr(A1C \leq 6,5 | DM^C) = 0,996$ (99,6%)
- *Coficiente falso positivo* = $\Pr(A1C > 6,5 | DM^C) = 0,004$ (0,4%)
- *Coficiente falso negativo* = $\Pr(A1C \leq 6,5 | DM) = 0,414$ (41,4%)

Supóngase ahora que se desean obtener los valores predictivos para una población en la que la probabilidad de la enfermedad es del 20% ($\Pr(D) = 0.20$). Utilizando el teorema de Bayes se tiene:

$$\Pr(DM | A1C > 6,5) =$$

$$\frac{\Pr(A1C > 6,5 | DM) \Pr(DM)}{\Pr(A1C > 6,5 | DM) \Pr(DM) + \Pr(A1C > 6,5 | DM^C) \Pr(DM^C)} \approx$$

$$\frac{0,586 \times 0,20}{0,586 \times 0,20 + 0,004 \times 0,80} \approx 0,9824$$

Se deja como ejercicio obtener el valor predictivo negativo $\Pr(DM^C | A1C \leq 6,5)$.

EJERCICIOS

1. En una cierta población, el 8.1% de las personas son diabéticas e hipertensas, el 12.4% son diabéticas y el 31.6% son hipertensas. Hallar la probabilidad de que una persona seleccionada al azar sea: a. Hipertensa. b. Diabética e hipertensa. c. Diabética y normotensa. d. Que sea hipertensa supuesto que es diabética. e. Que sea hipertensa supuesto que no es diabética. f. Hallar cuanto mayor es el riesgo de que un diabético sea hipertenso en relación con un no diabético (riesgo relativo).
2. Sean los sucesos A , B y C definidos en un mismo espacio muestral. Demostrar que $\Pr(A \cap B \cap C) = \Pr(A) \Pr(B | A) \Pr(C | A \cap B)$.
3. Para dos sucesos A y B , demostrar que $\Pr(B^C | A) = 1 - \Pr(B | A)$.
4. Probar que si $B \cap C = \emptyset$, entonces $\Pr(B \cup C | A) = \Pr(B | A) + \Pr(C | A)$.
5. Demostrar que si dos sucesos son independientes, también lo son sus contrarios.
6. Probar que si dos sucesos son independientes, su odds-ratio es igual a la unidad.
7. En una cierta población de mujeres, el 4% tiene cáncer de mama, el 20% son fumadoras y el 3% son fumadoras y tienen cáncer de mama. Si se selecciona una mujer al azar de esta población, ¿cuál es la probabilidad de que sea fumadora o tenga cáncer de mama?
8. El 13% de las personas de una cierta población son diabéticas. En el grupo de diabéticos, un 80% son hipertensos, mientras que en los no diabéticos lo son el 43%. De esta población se selecciona una persona aleatoriamente. ¿Cuál es la probabilidad de que sea hipertensa?
9. Se sabe que los pacientes con SIDA a menudo presentan tuberculosis. Se llevó a cabo un estudio de los factores de riesgo asociados con el desarrollo de esta enfermedad en los pacientes. Uno de los factores considerados fue la adicción

a drogas intravenosas. De los 1992 pacientes del estudio, 307 habían abusado de las drogas por vía intravenosa. Cuarenta y seis de los pacientes tenían tuberculosis y, de ellos 11 eran adictos a drogas por vía intravenosa.

1. Construir una tabla 2×2 para visualizar los datos.
 2. Hallar e interpretar el riesgo relativo.
10. Si A y B denotan la presencia de dos factores de riesgo A y B (\bar{A} y \bar{B} serían su ausencia), los individuos de una población pueden ser $\bar{A}\bar{B}$, $\bar{A}B$, $A\bar{B}$ y AB con probabilidades del 10%, 20%, 25% y 45% para los enfermos de una patología y del 40%, 30%, 20% y 10% para los sanos. Si hay un 10% de enfermos en la población, ¿qué factor de riesgo es más peligroso?
11. Cuando se examina la asociación entre un factor de riesgo F y una enfermedad D , que puede decirse de la relación entre el riesgo relativo y la odds-ratio cuando la enfermedad es de muy baja probabilidad (*enfermedad rara*).
12. La diferencia entre las prevalencias de una enfermedad entre dos poblaciones puede ser atribuible total o parcialmente a un problema de confusión por edad. Para que tales prevalencias sean comparables se precisa ajustarlas a una misma distribución de edades. Tal distribución correspondería a la de una población ideal consistente en un *promedio* de un amplio número de poblaciones. La siguiente tabla corresponde a la población mundial propuesta por la OMS. Utilizando los datos de la tabla la tabla 5, obtener la tasa de DM ajustadas por la población OMS.

TABLA 9. Distribución de edades correspondiente a la población ideal OMS

Grupo de edad	30 - 39	40 - 49	50 - 59	60 - 69	≥ 70
%	29,96	25,64	20,14	13,56	10,70

13. [Seshadri et al \(1995\)](#) llevaron a efecto un metanálisis en el que se incluyeron ocho estudios de caso-control, siendo el objetivo común analizar la asociación entre el alelo $\epsilon 4$ de la apolipoproteína E y la enfermedad de Alzheimer (AD). El estudio concluyó que en los pacientes con AD, la presencia del alelo $\epsilon 4$ era del 56%, mientras que en los controles era del 24%. Los autores estimaron que en la población general, la probabilidad de desarrollar la AD a los 65 años es del

- 15%. Hallar el riesgo relativo de padecer la AD según presencia o no del alelo $\epsilon 4$.
14. En una cierta población y para un determinado gen con alelos $\{A, a\}$, las frecuencias de los genotipos AA , Aa y aa son p , q y r respectivamente ($p + q + r = 1$). Supongamos que se aparean dos individuos de esta población.
1. Si un sujeto de la siguiente generación es Aa , ¿cuál es la probabilidad de que proceda del apareamiento (AA, Aa)?
 2. Si un sujeto tiene genotipo Aa , ¿cuál es la probabilidad de que un hermano suyo sea también Aa ?
 3. Probabilidad de que dos hermanos tengan el mismo genotipo.
15. Sea a un gen recesivo ligado al sexo, y supongamos que hay un problema de selección que imposibilita el apareamiento de hombres a . Si los genotipos AA , Aa y aa aparecen entre las mujeres con frecuencias p , q y r respectivamente, demuéstrese que, para los descendientes femeninos de la primera generación, $p_1 = p + q/2$, $q_1 = q/2 + r$ y $r_1 = 0$.
16. Una mujer tiene una rara anomalía de los párpados llamada *ptosis*, lo que le hace imposible abrir los ojos por completo. La causa se ha encontrado en un solo gen dominante. El padre de la mujer tenía ptosis, pero su madre tenía los párpados normales. La madre de su padre tenía los párpados normales.
1. ¿Cuáles son los posibles genotipos de la mujer, de su padre y de su madre?
 2. ¿Cuál es la probabilidad de que sus hijos tengan ptosis si ésta se casa con un hombre con los párpados normales?
17. En la sangre humana hay dos alelos S y s y tres fenotipos distintos que pueden ser identificados por medio de los reactivos adecuados. En un estudio realizado en Gran Bretaña sobre 1000 personas se encontró que 99 eran SS , 418 Ss y 483 ss . Calcular la frecuencia de S y s . ¿Hay alguna razón para rechazar la hipótesis de equilibrio de Hardy-Weinberg en esta población?
18. Tomlinson *et al* (2007) analizaron la asociación del polimorfismo de único nucleótido 8q24.21 con el riesgo de cáncer colorrectal. Encontraron que en

una muestra de 1027 sujetos, las frecuencias de los correspondientes genotipos fueron 352, 486 y 189. ¿Cómo es la distribución de los genotipos en la siguiente generación?

19. En las poblaciones humanas, uno de cada 10000 niños es albino. Calcular la frecuencia aproximada del alelo que produce el albinismo sabiendo que es autosómico recesivo.
 1. ¿Qué proporción de individuos portadores se encuentran entre los no albinos?
 2. ¿Cuál es la probabilidad de que matrimonios entre individuos normales produzcan individuos albinos?
20. Se diseñó un test diagnóstico para la *candidiasis invasiva* en pacientes críticos. De un total de 115 pacientes, 31 tenían la candidiasis. Cuando se aplicó la prueba diagnóstica se obtuvieron 66 resultados positivos al test, de los cuales, 28 tenían realmente la candidiasis. Evaluar la prueba diagnóstica.
21. [Guay et al \(2003\)](#) realizaron un estudio para evaluar el valor pronóstico de la tomografía por emisión de positrones (PET) para la recidiva de la enfermedad de Hodgkin. De los 48 pacientes incluidos en el estudio, 14 presentaron recidiva. De éstas, 11 fueron detectadas por la PET mientras que en los 34 pacientes que no recidivaron, 1 resultó ser un falso positivo. Evaluar la prueba diagnóstica indicando la sensibilidad, especificidad y los valores predictivos.
22. El antígeno prostático específico (PSA) se utiliza frecuentemente como prueba de *screening* para la detección precoz del cáncer de próstata. [Smith DS et al](#) llevaron a cabo un estudio en el que se realizó biopsia de próstata a 2411 varones de raza blanca resultando que 580 tenían cáncer de próstata. Se realizó asimismo la determinación del PSA y se consideró de riesgo cuando su valor fue superior a 4 ng/ml. Superaron este umbral 1298 sujetos de los cuales 443 tenían realmente cáncer de próstata. Evaluar la prueba diagnóstica.
23. El diagnóstico de la candidiasis invasiva (CI) desde la perspectiva del paciente y para optimizar la terapia antifúngica en la unidad de cuidados intensivos es

un problema de notable interés. Un marcador utilizado frecuentemente para obtener un diagnóstico precoz es el β – *Glucano*. Diversos trabajos han propuesto que un paciente se considere positivo para la CI cuando se obtengan dos determinaciones consecutivas del β – *Glucano* superiores a 80 pg/ml. Tomando el microcultivo como *gold-estándar* para la CI, León *et al* (2016) obtuvieron para el test diagnóstico una sensibilidad del 76.7% y una especificidad del 57.2%. Obtener los valores predictivo positivo y predictivo negativo para una población en la que la prevalencia de CI sea del 13%.

24. El cáncer de hígado es el sexto cáncer más común y la tercera causa principal de muerte relacionada con el cáncer en todo el mundo. El carcinoma hepatocelular (HCC), que representa entre el 70% y 85% de todos los casos de cáncer de hígado, se diagnostica frecuentemente en estadios avanzados y por tanto se asocia con un pronóstico precario, siendo la tasa de supervivencia general a los cinco años de menos de 15%. La alfafetoproteína (AFP) es el marcador sérico más utilizado para el cribado y el diagnóstico inicial de HCC entre personas con enfermedad hepática. Su sensibilidad para un valor de corte de 20 ng/mls es del 57,7% siendo la correspondiente especificidad del 85,2%. Calcular sus valores predictivos en una población de personas con enfermedad hepática en la que la prevalencia de HCC es del 20%.

7. BIBLIOGRAFÍA

Boronat, M., Varillas, M.V., Saavedra, P., Suárez, V., Bosch, E., Carrillo, A. and Nóvoa, F.J. (2005). [Diabetes mellitus and impaired glucose regulation in the Canary Islands \(Spain\): prevalence and associated factors in the adult population of Telde, Gran Canaria.](#) *Diabetes Care*, **28**, 2388:2393. [pdf](#)

Guay, C., Lépine, M., Verreault, J., and Bénard, F.. (2003). [Prognostic Value of PET Using 18F-FDG in Hodgkin's Disease for Posttreatment Evaluation.](#) *The Journal of Nuclear Medicine*, **44**: 1225-1231. [pdf](#)

International Expert Committee. [International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes.](#) *Diabetes Care*. 2009 Jul;32(7):1327-34. doi: 10.2337/dc09-9033. Epub 2009 Jun 5. PMID: 19502545; PMCID: PMC2699715. [pdf](#)

León C, Ruiz-Santana S, Saavedra P, Castro C, Loza A, Zakariya I, Úbeda A, Parra M, Macías D, Tomás JI, Rezusta A, Rodríguez A, Gómez F, Martín-Mazuelos E; Cava Trem Study Group. [Contribution of Candida biomarkers and DNA detection for the diagnosis of invasive candidiasis in ICU patients with severe abdominal conditions](#). *Crit Care*. 2016 May 16;20(1):149. doi: 10.1186/s13054-016-1324-3. [pdf](#)

Logroscino G, Marder K, Cote L, Tang M-X, Shea S, Mayeux R. (1996). [Dietary lipids and antioxidants in Parkinson's disease: a population-based, case-control study](#). *Ann Neurol* **39**:89-94 [pdf](#)

Moliner, María [Diccionario de uso del español](#). 1a. ed., Ed. Gredos, Madrid, 1967.

Scarmeas, N., Luchsinger, J.A., Schupf, N., Brickman, A.M., Cosentino, S., Tang, M.X., Stern, Y. (2009). [Physical activity, diet, and risk of Alzheimer disease](#). *JAMA*, **302**:627-637. [pdf](#)

Seshadri, S., Drachman, D.A. and Lippa, A.F. (1995). [Apolipoprotein E e4 allele and lifetime risk of Alzheimer's disease. What physicians know, and what they should know](#). *Arch. Neurol*, **52**:1074-1079.

Smith DS, Bullock AD, Catalona WJ. [Racial differences in operating characteristics of prostate cancer screening tests](#). *J Urol*. 1997 Nov;158(5):1861-5; discussion 1865-6. doi: 10.1016/s0022-5347(01)64147-8. PMID: 9334618.

Tomlinson et al. (2007). [A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21](#). *Nature genetics*, **39**:984-988.

