

El uso del método epidemiológico para la evaluación de la calidad de la asistencia (I)

Javier Moliner

Dirección Gerencia. Hospital Clínico Universitario Lozano Blesa. Zaragoza.

Correspondencia

Javier Moliner.

Coordinador del Programa de Calidad. Dirección Gerencia.

Hospital Clínico Universitario Lozano Blesa.

San Juan Bosco, 15. 50009 Zaragoza.

Correo electrónico: esta-moliner@hcu-lblesa.es

Introducción

La calidad asistencial se concreta en fenómenos relacionados con la salud, que se estudian clásicamente^{1,2} con dos enfoques: evaluación y monitorización. Evaluar la calidad significa efectuar un juicio, comparando con los objetivos esperados; generalmente se evalúa el proceso de atención, pero también se evalúan los resultados o las estructuras, siempre que exista una relación bien demostrada entre éstos (estructuras, procesos y resultados) con el grado de calidad. La otra estrategia, monitorizar, consiste en vigilar de forma constante la aparición de señales que permitan detectar tempranamente los problemas. Los dos métodos, evaluar y monitorizar, llevan implícita la medición.

Para conocer cómo son las características de calidad de un proceso de atención sanitaria es necesario disponer de herramientas de medida y de métodos para usar esas herramientas. El paso inicial para evaluar la calidad es determinar cuáles son las dimensiones y los criterios que se eligen como representativos, para seguidamente seleccionar los métodos y herramientas de medida adecuados.

La epidemiología se ocupa de estudiar los patrones de distribución de las enfermedades y de los fenómenos relacionados con la salud en las poblaciones humanas, así como de los factores que influyen en esos patrones.

Las utilidades de las técnicas epidemiológicas en el ámbito sanitario se han clasificado³ en:

- Recogida y análisis de datos demográficos y de morbi-mortalidad.
- Vigilancia epidemiológica y registro de los problemas de salud.
- Identificar a personas o grupos sometidos a riesgo elevado de enfermar.
- Identificar a problemas de salud pública y valorar su importancia.
- Diseño y ejecución de estudios de investigación.
- Cribado de enfermedades.
- Evaluación de la eficacia de tratamientos, programas y servicios.

- Evaluación de la utilidad de las pruebas diagnósticas.
- Descripción de la historia natural de la enfermedad.
- Lectura crítica de la bibliografía científica médica.

El objetivo de este trabajo es repasar algunos métodos y herramientas usados en epidemiología que se emplean también en evaluación de la calidad de la atención sanitaria. Trataremos de encontrar ejemplos del uso de las técnicas epidemiológicas en los informes o trabajos que manejan habitualmente los profesionales de la calidad.

Otro propósito es analizar dichas técnicas renunciando de antemano a lo más complejo (p. ej., demostraciones matemáticas o teorías). Las dificultades de la materia nos hacen mantener la recomendación al usuario medio de: "En los casos de duda, consulte con su epidemiólogo."

En términos esquemáticos, la epidemiología proporciona herramientas para tres objetivos: cuantificar la aparición de las enfermedades, comparar esos datos entre grupos de personas y diseñar estrategias para averiguar las causas de las enfermedades.

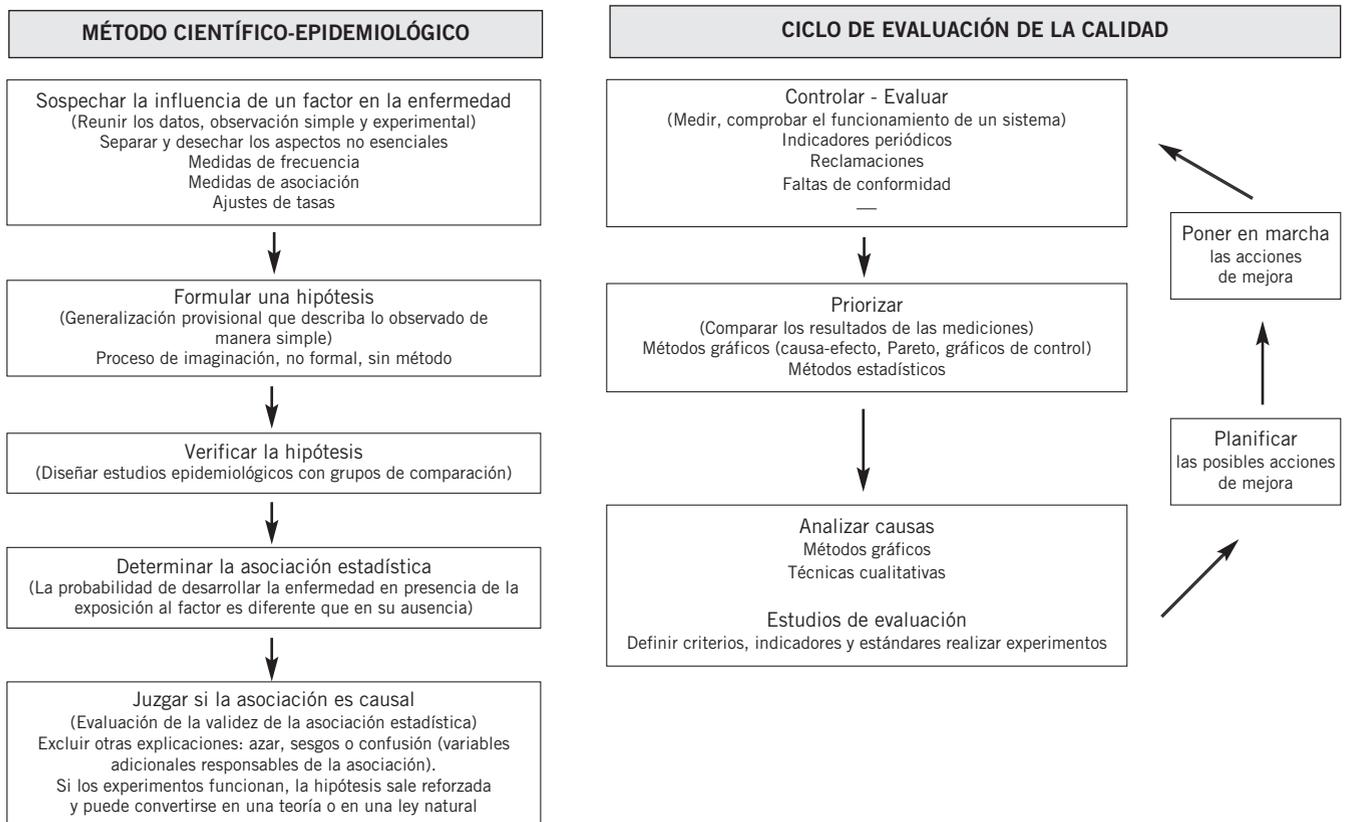
Haciendo un repaso breve de los índices de los libros de epidemiología⁴, para el primer objetivo, cuantificar, será necesario saber cuáles son las medidas de frecuencia de las enfermedades (incidencia, prevalencia, tasas) y cuáles son más convenientes para lo que se pretende medir. Todo lo que tiene que ver con medición se incluye en este capítulo: tipos de medidas de frecuencia, métodos de recogida de datos, errores en la medición, estudios transversales, etc.

En el segundo objetivo, comparar los datos de distintas poblaciones, se analizan las medidas de asociación e impacto, la estandarización de tasas, los estudios ecológicos, los estudios de utilidad de pruebas diagnósticas, etc.

Para el tercer objetivo, buscar las causas, se necesita conocer cómo se diseñan y analizan los estudios de cohortes, los de casos y controles, y los estudios experimentales. En este punto viene bien recordar las fases del método epidemiológico, en esencia las mismas que las del método científico^{5,6}, y analizar el paralelismo con el ciclo de evaluación de la calidad (fig. 1).

Antes de repasar los métodos y los instrumentos de medida, podemos enumerar algunas características o peculiaridades

Figura 1. Correspondencias entre el método científico-epidemiológico y el de evaluación de la calidad.



que se plantean al evaluar la calidad de la atención sanitaria y algunas similitudes con el análisis tradicional de las enfermedades.

En primer lugar, es necesario tener claro cuál es el fenómeno a medir, la definición de caso. La calidad es un concepto que se puede descomponer en muchas dimensiones y criterios, que deben ser definidos cuidadosamente.

Para este trabajo, podemos aceptar que la calidad en atención sanitaria equivale al resultado de las interacciones personales que se dan entre profesionales y pacientes y que se miden tanto con evaluaciones subjetivas como con datos objetivos. Todos estos resultados deben ser medidos de forma sistemática para poder tomar decisiones que acerquen la organización de los servicios hacia la implantación de modelos de calidad total⁶.

Puede ser necesario un gran esfuerzo para establecer los criterios y las normas que indiquen cuál es el nivel de calidad aceptable. Lo aceptable variará con el conocimiento científico y con otras variables, entre las que intervienen de forma decisiva las preferencias y las expectativas de cada paciente. Establecer estándares es también una tarea subjetiva y, por tanto, expuesta a la variabilidad.

Entre los factores que afectan a la calidad se encuentran todas las actividades de una organización y todas las personas. La organización, el sistema de trabajo, resulta muy importante

como factor causal en evaluación de calidad. Las oportunidades de mejora más importantes suelen pasar por cambiar sistemas organizativos. Si sobre las enfermedades se demuestra la influencia de factores genéticos, ambientales o de estilos de vida, también en la calidad de la atención se puede hablar de factores "genéticos" o sistémicos (los atribuidos a la variabilidad propia de cada sistema) y distinguirlos de otros factores externos o extrasistémicos. En los problemas de calidad, como en muchas enfermedades, aunque no llegue a aclararse completamente la etiopatogenia, se interviene para modificar los signos o los síntomas.

Las personas son el otro gran factor explicativo. En muchos casos, la calidad es el resultado de evaluaciones subjetivas individuales, son los clientes los que determinan cuáles son las características de calidad. Los proveedores son también seres humanos, y en su comportamiento, requisito para conseguir la calidad aceptable, influyen variables tan distintas como la motivación, las condiciones de trabajo, la formación, la comunicación, etc. Todos estos factores deben ser considerados como explicativos en los análisis de causas.

El instrumento de medida es un componente básico de la medición. Cada aspecto a medir debe ser susceptible de ser medido de forma fiable. Los datos generados deben ser lo precisos y oportunos que sea necesario para poder tomar decisiones sobre la calidad del producto. Del trabajo habitual, de los

procesos y de sus resultados, se obtiene información (entradas, productos intermedios, salidas) que manejamos en forma de indicadores y que pueden equivaler a los signos y síntomas de una enfermedad, los fenómenos a comparar y a investigar.

Como ventaja de las técnicas epidemiológicas se puede señalar que los resultados de la epidemiología son fáciles de comprender, por lo que se utilizan de forma generalizada como lenguaje de comunicación entre profesionales.

Una reflexión más sobre la necesidad de medir la calidad. Todo el mundo tiene opiniones personales sobre la calidad. Sin embargo, las impresiones personales por sí solas suelen ser un instrumento inadecuado para juzgar la calidad y, paso siguiente, para justificar acciones en atención de salud. El estudio de la calidad requiere herramientas más objetivas, que permitan conocer la efectividad de los servicios, valorar los cambios, comunicarse, comparar los sistemas de servicios (entre ellos y consigo mismos) y que sirvan para la variedad de usuarios que van a utilizar esas herramientas.

A pesar de las dificultades *a priori*, sobre todo de la dificultad para estandarizar instrumentos de medida, hay que avanzar hacia la medición objetiva de las características de la calidad. En la evaluación de cualquiera de nuestros inextricables servicios de salud (con escasas "reglas de juego" poco cumplidas, histórica opacidad e infinitos hechos diferenciales) hay que poner un poco de rigor metodológico, y saberse la epidemiología básica puede ser de ayuda.

Medidas de frecuencia y medidas de asociación. Cuantificar los fenómenos relacionados con la calidad

En epidemiología se calculan distintos tipos de medidas cuantitativas. Se diferencian según la estrategia que se emplea para el cálculo, y su utilidad depende del tipo de fenómeno que se quiere medir.

Frecuencias absolutas

La medida de frecuencia de enfermedad más básica es la frecuencia absoluta, el número de casos sin referencia temporal ni espacial. Una medida de este tipo usada como indicador de calidad son los "sucesos centinela" (*sentinel events*): hechos o problemas que se caracterizan por una baja probabilidad de presentación y una alta probabilidad de ser atribuibles a un tipo de atención dada (o no dada). Son datos que tratan de identificar situaciones que probablemente hayan sido el resultado de una atención deficiente y que deben ser investigadas más a fondo para comprobarlo. Constituyen una medida típica de monitorización de la calidad. La mortalidad potencialmente evitable, las reacciones a medicamentos, algunas complicaciones tras un procedimiento quirúrgico pueden considerarse sucesos centinela.

La mortalidad innecesariamente prematura y sanitariamente evitable (MIPSE) es un ejemplo de "suceso centinela" del que la REVISTA DE CALIDAD ASISTENCIAL ha publicado un estudio recientemente⁷. Este estudio, realizado en cuatro hospitales de Cádiz, utilizó una definición de MIPSE restringida a sólo algunos diagnósticos principales en pacientes que fallecieron en

el hospital (colecistitis, colelitiasis, apendicitis, hernia abdominal y complicaciones de embarazo, parto y puerperio, todo en pacientes entre 5 y 64 años de edad). Con esos criterios, entre las altas de un año tan sólo encontraron 5 casos (el 0,02% de los fallecimientos).

Otros ejemplos de este tipo de medidas se pueden encontrar en el boletín Sentinel Events Alert, elaborado por la Joint Commission on Accreditation of Healthcare Organizations⁸, que proporciona información sobre sucesos centinela y recomendaciones para prevenirlos. En la publicación (disponible en www.jcaho.org) se analizan las bases de datos de que dispone la Joint Commission sobre sucesos centinela, con estudios recientes sobre problemas asociados al uso de equipos de ventilación mecánica, sobre errores en los procedimientos quirúrgicos (operar al paciente equivocado, en un órgano equivocado), los errores de medicación relacionados con el uso de abreviaturas en las órdenes de tratamiento, las caídas que resultan mortales en pacientes hospitalizados, etc. Recientemente, también la revista *International Journal for Quality in Health Care* (accesible en <http://intqhc.oupjournals.org>) ha incluido una sección periódica con información sobre incidencia y prevención de sucesos centinela.

Razón o ratio

De menos a más complejidad, la siguiente medida es la razón o *ratio*. En sentido estricto, una razón es un cociente entre dos números comparables entre sí. Por tanto, la razón es un término general en el que están incluidos la proporción, el porcentaje y la tasa, que son tipos de razones. En epidemiología el término "razón" se suele aplicar a un cociente en el que el numerador no está incluido en el denominador. Son razones el índice "consultas sucesivas/consultas primeras" o la relación "número de varones/número de mujeres". También son razones otras medidas frecuentes en epidemiología como la *odds ratio* (razón de *odds*, OR), o la *likelihood ratio* (cociente de probabilidades, CP) que se repasan más adelante.

Incidencia acumulada

Existen otros tres tipos de medidas de frecuencia en epidemiología:

- La tasa de incidencia (densidad de incidencia, *incidence rate*) es una medida de la "fuerza" con que se manifiesta una enfermedad en una población.
- La incidencia acumulada (proporción de incidencia, *cumulative incidence*) es la proporción de personas que, en un tiempo especificado, pasan de no estar enfermas a estarlo.
- La prevalencia es la proporción de gente que tiene la enfermedad en un tiempo concreto.

Los datos necesarios para el cálculo de estas medidas son el número de personas que han desarrollado la enfermedad, la cantidad de tiempo que todos los posibles casos han estado expuestos al riesgo, y todo ello especificando siempre el período temporal al que se refiere la medida (un instante, un día, un año, etc.).

De estas tres medidas, la que se va a utilizar más en evaluación y monitorización de la calidad es la incidencia acumulada, que informa del riesgo de que se produzca un hecho. Es la probabilidad de que una persona desarrolle una enfermedad en un período de tiempo, es una medida del riesgo medio que tienen las personas de desarrollar la enfermedad. Es una proporción, por lo que se puede expresar también como porcentaje. Para interpretarla resulta imprescindible especificar el tiempo al que se refiere el cálculo. Muchos indicadores de calidad tendrán este formato: el 10% de ingresos inadecuados en una semana, el 20% de partos por cesárea en el mes anterior, o el 0,5% de complicaciones infecciosas tras cirugía de catarata en el año pasado, etc.

En el trabajo de Andrews et al⁹ sobre la aparición de casos de efectos adversos en un hospital en Chicago, se encontró que de los 1.047 pacientes que fueron ingresados durante un período de 9 meses en tres salas (de cirugía y de cuidados intensivos) del hospital, 185 casos (el 17,7%) sufrieron un efecto adverso grave. Se puede decir que ése es el riesgo medio que tuvo cada paciente durante los 12 días de estancia promedio que duró el seguimiento.

Es muy importante ser específico en la definición de numerador y denominador. Por ejemplo, la incidencia de efectos adversos de un fármaco podría cuantificar el número de personas que lo reciben o el número de veces que se ha pautado (y darse la posibilidad de que en un mismo paciente se prescriba más de una vez). En el denominador de esta medida se debe excluir del cálculo a las personas que dejan de estar a riesgo porque han desarrollado la enfermedad.

Densidad de incidencia

En el cálculo de la incidencia acumulada se asume, con una probabilidad de error, que toda la población a la que se refiere el denominador ha estado expuesta al riesgo durante el mismo tiempo. Sin embargo, esto no suele ser así. En el citado estudio de Andrews et al⁹, por ejemplo, se calculó que la probabilidad de sufrir un efecto adverso se incrementaba en un 6% por cada día que el paciente permaneció hospitalizado.

Para tener en cuenta los distintos tiempos de exposición de cada persona se calcula la tasa de incidencia. Se obtiene de dividir el número de casos de enfermedad entre la suma de los períodos de tiempo que han estado expuestos a un riesgo los individuos de la población. Es una medida de la velocidad de aparición de nuevos casos o de la fuerza de los agentes causales de la enfermedad. Para un número de casos, cuanto menor sea el denominador, es decir, el tiempo de exposición, mayor es la fuerza de la enfermedad. El denominador es una suma de tiempos, y se expresa en unidades de personas-tiempo (personas-año, personas-mes). La tasa de incidencia se nombra también en unidades de tiempo elevado a la menos uno; la unidad de medida es el recíproco del tiempo.

Un ejemplo que ilustra la diferencia entre incidencia acumulada y densidad de incidencia es la forma de calcular el indicador: "tasa de neumonía asociada a ventilación mecánica". De todos los pacientes sometidos a ventilación se cuentan los días con ventilación y la aparición de casos de neumonía. La fórmula es: número de casos partido por la suma de días de exposi-

ción, y el valor medio de este indicador en el año 2001 en los hospitales grandes del INSALUD (datos no publicados) fue de 27 casos por cada 1.000 días de ventilación (0,027 días⁻¹).

Para calcular esta tasa, lo lógico es que cada paciente haya tenido una experiencia distinta, con variaciones en el número de días con ventilación. Además, cuando uno de los pacientes desarrolla una neumonía, deja de estar en riesgo y, por tanto, a partir de entonces hay que excluirlo para el cálculo del denominador de esta tasa. Para evaluar el problema en el hospital, no es lo mismo que las neumonías aparezcan casi siempre en los primeros días tras la intubación que se den sobre todo en los casos de intubaciones prolongadas.

Los datos del National Nosocomial Infection Surveillance System¹⁰ para este indicador en el período 1997-1999 calculados en casi 680.000 días de ventilación mecánica indican que el valor medio de las tasas de incidencia era de 14,4 por 1.000 días en unidades de cuidados intensivos quirúrgicas y 7,8 por 1.000 en unidades médicas.

La densidad de incidencia es más útil para estudiar los efectos de un factor de riesgo que se mantiene durante bastante tiempo (riesgos ambientales, sistemas de organización de los servicios, profesionales poco competentes) y cuando hay variabilidad importante en la cantidad de riesgo a que están sometidas unas y otras personas. Los factores que influyen en algunas medidas usadas como indicadores de calidad (accidentes, errores, uso inadecuado) comparten estas características, sobre todo la diferente cantidad de riesgo a que se someten unos pacientes y otros.

Calcular tasas de incidencia puede resultar caro, en tiempo y en colaboración de otras personas, sobre todo cuando los tiempos de exposición son largos (años), cuando es difícil determinar el comienzo de la enfermedad, cuando hay pérdidas de pacientes por otras causas (mortalidad), o cuando hay que hacer seguimientos diarios. En la práctica, algunos servicios de epidemiología, sobre todo los dedicados a epidemiología comunitaria, se convierten en servicios "calculadores" de personas-tiempo, para determinar con exactitud las tasas de incidencia (de cáncer poblacional, por ejemplo) y poder diferenciar lo normal de lo patológico.

Cuando el período causa-efecto es corto o mínimo, la cuantificación del tiempo en riesgo pierde interés. Los estudios de estancias inadecuadas o algunas complicaciones posquirúrgicas, por ejemplo, se pueden analizar bien con datos de incidencia acumulada.

Relación matemática entre incidencia acumulada y tasa de incidencia

Si se conoce la tasa de incidencia se puede calcular la incidencia acumulada, el riesgo para un paciente individual. Si conocemos la tasa de incidencia, ésta es constante en el tiempo y los riesgos son pequeños (menos del 10%), por medio de una demostración matemática (que incluye integrales y antilogaritmos) se demuestra que la incidencia acumulada es igual a la tasa de incidencia multiplicada por el período de tiempo.

En el caso de la neumonía asociada a ventilación, si se cumple la condición de que la tasa sea constante, una tasa de incidencia de 27 por cada 1.000 días equivale a un riesgo dia-

rio del 2,7% o a un riesgo semanal del 19%. Otra condición para que este cálculo sea válido es que no existan riesgos competidores, lo que casi nunca pasa, por lo que ésta va a ser una medida hipotética. La incidencia acumulada calculada de esta forma se denomina también tasa acumulada o riesgo acumulado.

Prevalencia

En las medidas de prevalencia no se tiene en cuenta el tiempo que la población ha estado en riesgo. Se cuentan únicamente los casos que existen en un determinado momento o en un determinado período. La tasa de prevalencia es una proporción entre los casos de enfermedad y los individuos que en el momento estudiado podrían ser casos. Se habla también de proporción de prevalencia, prevalencia puntual o prevalencia de período.

La prevalencia es una medida útil para situaciones relativamente estables, como presentar una discapacidad, ser portador de un marcapasos o formar parte de una lista de espera para cirugía, pero no suele ser útil para estudiar enfermedades agudas.

Un conjunto de datos de prevalencia frecuentemente usados en monitorización de calidad son los que produce el Proyecto EPINE¹¹, realizado desde 1990, en la actualidad en más de 200 hospitales, con el fin de analizar la frecuencia de infecciones hospitalarias en nuestro país. Al analizar los datos del proyecto EPINE conviene tener en cuenta que presentan las limitaciones derivadas de que las infecciones hospitalarias suelen originar problemas de evolución aguda. Como ejemplo, la prevalencia de infección de herida quirúrgica superficial (9% del total) y de herida quirúrgica profunda (10%) es similar a pesar de que las primeras son mucho más frecuentes.

La prevalencia en un período, el porcentaje de población que son casos en cualquier momento de un período, combina una medida de prevalencia puntual (estado al inicio) con la incidencia a lo largo del período de observación. Es, por tanto, una medida difícil de interpretar, poco usada.

La prevalencia depende tanto de la aparición de nuevos casos (incidencia) como de la duración de la enfermedad. Si en el fenómeno que se estudia, la velocidad de salida –la tasa de abandono– es pequeña y la cronicidad es alta (pertenecer a una lista de espera, ser afectado por un error médico), incluso con baja incidencia se producirá una prevalencia alta.

La tasa de enfermedad detectada en autopsias es también una medida de prevalencia. Es la proporción de casos que tienen una enfermedad entre aquellos en que se realizó la autopsia. Aunque estos estudios no suelen ser representativos de la población de la que proceden, se encuentran excepciones como en el estudio de Ebbesen et al¹², que analizaron las historias clínicas y las autopsias de 732 pacientes fallecidos en un hospital noruego y encontraron que en el 18,2% de los casos existió asociación directa o indirecta (la mitad de cada) entre la muerte del paciente y los efectos de un fármaco.

Relación entre incidencia y prevalencia

Cuando tanto la incidencia como la duración de la enfermedad son estables (no hay epidemias ni cambios bruscos en

la historia de la enfermedad), los datos de prevalencia se pueden usar como alternativa a los de incidencia. Se demuestra que para prevalencias pequeñas (menores de 10%), se cumple la fórmula de:

$$[\text{Prevalencia} = \text{Densidad de incidencia} \times \text{Duración media}]$$

En los estudios que buscan causas, es importante recordar que la selección de casos prevalentes, más a mano, tiene el inconveniente de que en la prevalencia influyen tanto la incidencia como la duración. Los casos que duran más (más supervivencia) tienen más probabilidades de ser elegidos para el estudio, con lo que se corre el riesgo de confundir causas con factores pronósticos.

Los cambios en la prevalencia se pueden deber a cambios en la incidencia, en la duración o en ambas. Un ejemplo claro es el caso de la epidemia de sida en nuestro medio, que en los últimos años ha experimentado un descenso de la incidencia y un aumento de la supervivencia, lo que convierte los datos de prevalencia en poco o nada informativos.

Comparación de datos. Medidas de asociación

Las tres cuestiones que se pueden responder con ayuda de la epidemiología son: *a)* cuantificar problemas, evaluar tendencias; *b)* comprobar si un cambio ha producido un impacto, comparar entre poblaciones, y *c)* analizar la relación entre un efecto y sus posibles causas.

Las medidas de frecuencia vistas hasta ahora sirven para ser comparadas, entre poblaciones distintas o en una misma población en distintos momentos. El 5% de riesgo anual de infección urinaria no dice mucho; la diferencia entre dos datos, el 6% en un grupo de pacientes frente al 4% en otro grupo, o un exceso de riesgo del 50%, son el tipo de medidas con las que se empieza a manejar información.

Con las medidas de asociación e impacto se trata de medir la relación entre una exposición y el riesgo de desarrollar una enfermedad. Se usan para resumir las comparaciones de tasas entre las poblaciones de los que están expuestos y los no expuestos a un factor de riesgo.

Existen varias medidas para comparar tasas en epidemiología: las más usadas son el riesgo relativo, la *odds ratio* y el riesgo atribuible.

Para entenderlas, el caso más simple consiste en analizar los datos de dos grupos de población, y de su exposición (sí/no) a un factor, en una tabla 2 por 2 (tabla 1). Las letras (a, b, c y d) indican el número de casos en cada situación. Una variación de esta tabla se da en los estudios en que los denominadores se miden como personas-tiempo (tasas de incidencia); en este último caso la columna de “no enfermos” es sustituida por la cantidad de personas-tiempo que han acumulado los expuestos y los no expuestos

Riesgo relativo

El riesgo relativo (RR) es la medida que se calcula en los estudios de seguimiento para comprobar la asociación entre un factor y una enfermedad. Es el cociente entre la tasa de enfer-

Tabla 1. Presentación de resultados de la relación entre exposición y enfermedad

	Enfermedad	No enfermedad	Total
Exposición	a	b	a + b
No exposición	c	d	c + d
Total	a + c	b + d	a + b + c + d

medad en expuestos dividida por la tasa en los no expuestos. Es la probabilidad de desarrollar la enfermedad en el grupo de expuestos, en relación con la probabilidad en el grupo de no expuestos.

Existen dos posibilidades de cálculo del riesgo relativo, según la medida de frecuencia disponible. Si estamos ante un estudio en el que los denominadores de las tasas son números de personas (datos de incidencia acumulada), el riesgo relativo (*risk ratio*) es el cociente entre la incidencia acumulada en los expuestos partida por la incidencia en los no expuestos. Indica el exceso de riesgo, cuántas veces más riesgo tiene un grupo que otro.

En los estudios que tienen datos de personas-tiempo (densidades de incidencia), el riesgo relativo (*rate ratio*) se calcula con el cociente entre las dos densidades de incidencia.

En cualquier caso, un riesgo relativo de 1 indica no asociación entre factor y enfermedad. Los valores superiores a 1, indican exceso de riesgo, número de veces, porcentaje de exceso de riesgo; los valores inferiores a 1 indicarían reducción del riesgo en el grupo de expuestos.

Los datos de incidencias acumuladas en el cálculo del riesgo relativo son más o menos útiles según el tiempo al que se refieran. Para períodos cortos no existe demasiado problema, ya que la incidencia acumulada es equivalente a la densidad de incidencia multiplicada por el tiempo, pero si son períodos más largos pueden influir otros riesgos que afecten a los datos. Para interpretar bien el riesgo relativo hay que especificar el período de tiempo en que se han calculado las tasas.

Kollef et al¹³ realizaron un estudio de seguimiento de 2.000 pacientes ingresados en unidades de cuidados intensivos, con un 26% que, o bien al ingreso o bien durante la estancia, sufrieron un proceso infeccioso, y encontraron que la probabilidad de muerte en los pacientes con infección a los que se había pautado un tratamiento antibiótico inadecuado era del 52,1%, frente al riesgo del 23,5% en quienes recibieron un tratamiento adecuado (tabla 2). El riesgo relativo en este caso era de 2,22, es decir, más del doble de riesgo en un grupo que en otro.

Odds ratio

En los estudios en que no se realiza seguimiento de los sujetos, como los estudios de casos y controles, no suele ser posible calcular las tasas de enfermedad en expuestos y en no expuestos. Sin embargo, se puede aproximar un cálculo del riesgo relativo haciendo un cociente entre la *odds* de exposición entre los casos y la *odds* de exposición entre los controles. Este cociente es la *odds ratio* (OR) o razón de *odds*.

Tabla 2. Relación entre riesgo de muerte y tratamiento antibiótico inadecuado (Kollef et al¹³)

	Muerte	No muerte	Total
Tratamiento antibiótico inadecuado	88	81	169
Tratamiento antibiótico adecuado	114	372	486
Total	202	453	655

Para entender la *odds ratio*, primero hay que entender qué es una *odds*: es el cociente entre una probabilidad (P) y su complementario (1-P). Si entre los pacientes que fallecen en un hospital la probabilidad de haber sufrido un suceso adverso es del 10%, la *odds* de sufrir ese suceso es 0,1/0,9, o sea, 1/9 (hay una *odds* o una ventaja de 9 a uno de no sufrirlo). Una forma fácil de hacer el cálculo automático es recordar que una proporción de 1 de cada x equivale a una *odds* (de no presentar el evento) de x menos 1, a 1.

Con los datos de la tabla 1, la *odds* de exposición entre los casos (probabilidad de que un caso haya estado expuesto partida por la probabilidad de que ese caso no lo haya estado) es a/c y la *odds* de exposición en los controles es b/d.

Por tanto, la *odds ratio* es el cociente de esas dos *odds*, o lo que es lo mismo:

$$[a \times d / b \times c]$$

El riesgo relativo y la *odds ratio* son medidas que indican la magnitud de la asociación entre un factor y un efecto, y si el estudio con el que se han calculado no tiene sesgos, permiten hacer juicios sobre la probabilidad de que esa asociación observada sea realmente causal. Su interpretación es la misma.

El valor de la OR es en muchos casos una buena aproximación del RR. Esta aproximación se demuestra de forma intuitiva. Si volvemos a la tabla 1 y cambiamos los títulos de las columnas por casos y controles, para enfermedades poco frecuentes, es lógico que la casilla b sea mucho más grande que la a, de manera que a/a + b sea parecido a a/b. Lo mismo con c/c + d que será similar a c/d. Con esta suposición, la fórmula de la razón de riesgos será: a/b partido por c/d, lo que es lo mismo que la razón de *odds*. La experiencia demuestra que en la mayoría de los casos la OR es un estimador válido del RR.

García Martín et al¹⁴ compararon en un estudio los casos de 524 pacientes fallecidos en un hospital con otros 524 controles (no fallecidos) similares en diagnóstico y fecha de ingreso. La probabilidad de sufrir un efecto adverso entre los pacientes fallecidos fue del 57%. La OR que relacionaba la presencia de un efecto adverso con el fallecimiento del paciente tenía un valor de 1,75. La interpretación de este dato es que existió un 75% más de riesgo de muerte en los pacientes que habían presentado un efecto adverso.

Cuando los datos se presentan en tablas de más de dos filas, cuando los niveles de exposición son más de dos (en el caso del consumo de tabaco, las categorías de no fumador, de

menos de una cajetilla, de más de una cajetilla), los valores del RR o de la OR se interpretan como si cada nivel formara una sola tabla de 2 por 2 con la categoría de comparación inicial, o basal, que generalmente es la categoría de no expuestos. En los resultados de los análisis de regresión logística, muy empleados para analizar el peso de distintos factores causales en un efecto, a cada nivel de exposición de cada variable le corresponde un valor que se interpreta como una OR ajustada.

El estudio de Kollef et al¹³ citado más arriba, que realizó un seguimiento de pacientes para analizar la influencia de determinados factores en el riesgo de mortalidad, calculó OR ajustadas para distintas variables y encontró valores de OR, por ejemplo, de 4,26 para el tratamiento antimicrobiano inadecuado, de 2,2 para el uso de vasopresina o de 1,05 por cada unidad de aumento en el índice APACHE II. La significación estadística de cada una de estas OR ajustadas permite considerar si la variable actúa o no como factor de riesgo independiente para el efecto estudiado, en este caso la mortalidad.

Riesgo atribuible

El riesgo atribuible (RA) o diferencia de riesgos es una medida de comparación de tasas que indica el impacto que sobre la salud de la población tienen las diferencias de tasas observadas. Para un factor cuya asociación causal con un efecto ya se conoce, el riesgo atribuible cuantifica el exceso de riesgo que ocasiona la exposición a ese factor. Es la resta entre las tasas de incidencia de la población de expuestos menos la tasa de incidencia de los no expuestos. Según el diseño del estudio (con datos de personas-tiempo o de personas) se calcula restando tasas de incidencia (*rate difference*) o incidencias acumuladas (*risk difference*).

Con los datos de la tabla 2 empleados como ejemplo, el riesgo atribuible en el caso de la mortalidad asociada a tratamiento antibiótico inadecuado es la diferencia de incidencias: el 52,1% menos el 23,5%, es decir un 28,6%.

Las medidas relativas tienen el peligro de ser mal interpretadas, ya que dependen de cuan grandes sean las tasas iniciales. Sin embargo, el riesgo relativo es más utilizado que el riesgo atribuible, entre otras razones porque se calcula de forma aproximada con varios tipos de estudio, entre ellos, los de casos y controles, mientras que el riesgo atribuible no.

Fracción etiológica

Existen otras medidas para manejar la información sobre el impacto de un factor causal. Para asignar una parte de un problema a una determinada causa, el riesgo atribuible se expresa como un porcentaje y se calcula la proporción atribuible o fracción etiológica (FE). Se ha llamado también fracción atribuible y porcentaje del riesgo atribuible. Se calcula dividiendo el riesgo atribuible entre la tasa de enfermedad en los expuestos y se usa para comparar el impacto potencial de distintas medidas de control.

En los datos del ejemplo anterior, la fracción etiológica resulta de dividir el riesgo atribuible, 28,6%, entre la tasa en expuestos (la mortalidad en los casos de tratamiento inadecuado), 52,1%. El resultado, 55%, es la parte del problema que se eli-

minaría en la población del estudio si la incidencia se redujera hasta el nivel del grupo de no expuestos.

Con una pequeña demostración, y utilizando los mismos datos, se obtiene que la fracción etiológica (FE) se puede calcular a partir del riesgo relativo (RR):

$$FE = (RR - 1) / RR \times 100$$

$$FE = (2,22 - 1) / 2,22 \times 100 = 55\%$$

Para la mayor parte de los estudios de casos y controles, el RA no puede ser calculado puesto que las tasas de incidencia en expuestos y en no expuestos no suelen ser conocidas (la cantidad de controles por cada caso la decide el investigador, luego no puede decir que "a/a + b" sea una tasa de incidencia). Sin embargo, sí se puede calcular la fracción etiológica con la siguiente fórmula:

$$FE = (OR - 1) / OR \times 100$$

En el trabajo de García Martín et al¹⁴ que analizaba la relación entre fallecimiento y efectos adversos, la OR de haber sufrido un efecto adverso era de 1,75. La fracción etiológica calculada de esta forma era:

$$FE = (1,75 - 1) / 1,75 \times 100 = 42,8\%$$

Riesgo atribuible poblacional

Para extrapolar los datos a toda la población y determinar cuál es el exceso de enfermedad que en una población produce una determinada causa, se calcula el riesgo (o tasa) atribuible poblacional (RAP) o diferencia de riesgos poblacional. Es la proporción de todos los casos que ocurren en una población que es achacable a la exposición. Es una medida del impacto potencial de las medidas de control, indica qué relevancia tienen distintos factores sobre la salud en una comunidad. Se calcula como la tasa de incidencia en la población (el conjunto de personas del estudio, la tasa de incidencia global) menos la tasa de incidencia en los no expuestos. En el ejemplo de estudio de seguimiento¹³:

$$RAP = 30,8\% - 23,5\% = 7,3\%$$

Que indica que en toda la población estudiada, el riesgo de fallecimiento atribuible al tratamiento antibiótico inadecuado es el 7,3%.

Esta diferencia de riesgos poblacional se puede usar para calcular la fracción atribuible poblacional (RAP%) dividiendo la diferencia de riesgos entre la incidencia global:

$$RAP\% = 7,3\% / 30,8\% = 23,7\%$$

Es decir, casi la cuarta parte del total del problema (los casos de mortalidad en la población estudiada) se podría evitar si se eliminara el factor de riesgo, en este caso el tratamiento antibiótico inadecuado.

En un número reciente de *Gaceta Sanitaria*¹⁵ se repasa el significado e interpretación de la fracción atribuible poblacional

utilizando ejemplos de cálculo tanto en estudios de seguimiento como en estudios de casos y controles.

Medidas de asociación en los ensayos clínicos

Cuando en lugar de factores de riesgo y enfermedades, lo que se valora es la eficacia de varios tratamientos, como es el caso de los ensayos clínicos, las medidas de asociación empleadas son otras. Si la magnitud del efecto de los tratamientos se valora con datos cualitativos (muerte/supervivencia, síntomas/no síntomas) las medidas de asociación son la reducción del riesgo relativo (RRR), la reducción del riesgo absoluto (RRA) y el número de pacientes que se necesita tratar (NNT). En la tabla 3 se ordenan los datos y se calculan las medidas.

En el Helsinki Heart Study¹⁶ se siguió a una población durante 5 años, a lo largo de los cuales se realizó un ensayo de tratamiento con gemfibrozil frente a placebo y se analizó la aparición de infartos de miocardio. Los resultados con los que se calcularon las medidas de asociación se recogen en la tabla 4.

La forma como se presentan los resultados, sobre todo cuál de las anteriores es la medida elegida, puede ser determinante para conseguir uno u otro efecto en el receptor de la información. Es habitual que, para "arrimar el ascua a su sardina", los fabricantes de fármacos prefieran las medidas relativas (RRR), que han demostrado¹⁷ tener el efecto más favorable para quien presenta los resultados.

El NNT es el número de pacientes que se calcula que sería necesario someter a un nuevo tratamiento, en lugar de al tratamiento anterior, para que uno de esos pacientes obtenga el beneficio descrito.

Un efecto de magnitud importante tendrá un pequeño NNT. Como ocurre con otras medidas de estimación, es conveniente acompañar el NNT de su intervalo de confianza, que resulta más válido e interpretable cuando las diferencias entre los tratamientos comparados son significativas¹⁸. Si el NNT fuera negativo, indicaría que el tratamiento experimentado tendría un

Tabla 3. Tabla de resultados de un ensayo clínico

	Resultado clínico		
	Sí	No	Total
Grupo control	a	b	a + b
Grupo experimental	c	d	c + d
Total	a + c	b + d	a + b + c + d
Tasa de episodios en el grupo control (TEC):	TEC = a / (a + b)		
Tasa de episodios en el grupo experimental (TEE):	TEE = c / (c + d)		
Reducción relativa del riesgo (RRR):	RRR = (TEC - TEE) / TEC		
Reducción absoluta de riesgo (RAR):	RAR = TEC - TEE		
Número de pacientes que se necesita tratar (NNT):	NNT = 1 / RAR, en tantos por uno		

Tabla 4. Ensayo de gemfibrozil contra placebo en el Helsinki Heart Study¹⁶

	Resultado clínico		
	Infarto	Sin infarto	Total
Grupo placebo	84	1.946	2.030
Grupo gemfibrozil	56	1.995	2.051
Total	140	3.941	4.081
Tasa de episodios en el grupo control (TEC):	TEC = 84 / 2.030 = 4,14%		
Tasa de episodios en el grupo experimental (TEE):	TEE = 56 / 2.051 = 2,73%		
Reducción relativa del riesgo (RRR):	RRR = (4,14 - 2,73) / 4,14 = 34,1%		
Reducción del riesgo absoluto (RAR):	RAR = 4,14 - 2,73 = 1,41%		
Número de pacientes que se necesita tratar (NNT):	NNT = 1 / 0,0141 = 71 pacientes		

efecto dañino. En este caso, se ha propuesto cambiar las siglas por las de NNH (*number needed to harm*; número necesario para producir un daño).

Medidas de eficacia de las pruebas de diagnóstico

La utilización de una prueba diagnóstica se asocia siempre con unos índices de eficacia, con unas características medibles que informan de su capacidad de acertar. Estos índices que sirven para evaluar la validez de la prueba (que la prueba mida realmente lo que se quiere medir) se pueden calcular cuando es posible comparar los resultados con los de otra prueba que actúa como prueba de referencia o *gold standard*, y que asegura la existencia o no de la enfermedad buscada.

Cuando se usa una prueba que dicotomiza los resultados (prueba positiva/negativa, enfermos/no enfermos), la comparación entre nuestros resultados y los de una prueba de referencia se puede representar como en la tabla 5, de esta tabla se extraen varias medidas:

- Sensibilidad (S): es la probabilidad de ser positivo en la prueba si realmente se tiene el problema. Una prueba sensible detecta una alta proporción de los casos reales.

$$S = VP / (VP + FN)$$

- Especificidad (E): es la probabilidad de test negativo si se está sano. Un test específico tiene pocos falsos positivos.

$$E = VN / (VN + FP)$$

Para cada prueba y para cada punto de corte (positivo/negativo), la relación entre sensibilidad y especificidad es inversa; al aumentar una disminuye la otra. La frontera entre sano y enfermo no siempre es clara, en muchos casos el punto que los

Tabla 5. Comparación de resultados de dos pruebas diagnósticas

	Resultados del test de referencia	
	Positivo	Negativo
Resultado de la prueba evaluada		
Positivo	VP	FP
Negativo	FN	VN

VP: verdaderos positivos.
FN: falsos negativos.

separa será arbitrario, por lo que si queremos detectar todos los casos (aumentar la sensibilidad) podemos reducir el umbral para que la prueba sea positiva, lo que tendrá la consecuencia de aumentar también los falsos positivos (o sea, reducir la especificidad).

Una forma de integrar la información de sensibilidad y especificidad para elegir el punto de corte más aceptable es calcular los cocientes de probabilidades (*likelihood ratio*). Para cada resultado de una prueba (positivo/negativo) y para cada punto de corte se puede calcular su cociente de probabilidades (CP).

Existe un CP para cada resultado de la prueba. El CP del resultado positivo (CP+) es la probabilidad de ese resultado (+) en pacientes con enfermedad dividida por la probabilidad de que ese mismo resultado se produzca en pacientes sin enfermedad. De forma más breve, es el cociente entre la sensibilidad (S) y el complementario de la especificidad (1 - E).

Del mismo modo, el CP del resultado negativo (CP-) es la probabilidad del resultado negativo en enfermos dividida por la probabilidad del resultado negativo en sanos. En este caso, es lo mismo que el cociente entre el complementario de la sensibilidad y la especificidad.

En un estudio reciente publicado por Silverman et al¹⁹ se analiza en 284 pacientes (tabla 6) cuál es la utilidad de la tomografía de emisión de positrones (PET) para detectar la enfermedad de Alzheimer progresiva en los tres años siguientes. La sensibilidad que consigue la PET para esa detección es del 93% (191/206) y la especificidad del 76% (59/78). Los autores calculan el cociente de probabilidades del resultado negativo al comparar Alzheimer progresivo con no progresivo: $CP- = (1 - S) / E$, en números, $CP- = (1 - 0,93) / 0,76$, que vale 0,1, y

Tabla 6. Valor de la tomografía de emisión de positrones (PET) para detectar Alzheimer progresivo¹⁶

	Resultados de seguimiento de 3 años		
	Alzheimer progresivo	No progresivo	Totales
PET positiva	191	19	210
PET negativa	15	59	74
Totales	206	78	284

concluyen que una prueba PET negativa indica que la progresión de la demencia en los tres años siguientes es improbable.

La sensibilidad y la especificidad son características de la prueba, que se mantienen estables cuando la prueba se aplica en las mismas condiciones. Por tanto, la información que da su cociente de probabilidades también será estable.

Sin embargo, el otro dato fundamental para evaluar la eficacia de una prueba es conocer sobre qué población se está aplicando, en concreto, qué prevalencia de enfermedad existe en esa población. El porcentaje de aciertos que consiga cada resultado de la prueba variará mucho de una población a otra si varía la "cantidad" de enfermedad en los grupos estudiados (puede ser muy diferente si el estudio se hace en grupos de alto riesgo o en población general, por ejemplo).

En la práctica, lo que interesa conocer es, ante la presencia de un resultado, qué probabilidades existen de que sea cierto. Esta información la proporcionan los valores predictivos (VP). El valor predictivo del resultado positivo es la probabilidad de que un caso en el que la prueba sea positiva realmente tenga la enfermedad. Con las siglas de la tabla 5:

$$VP+ = VP / (VP + FP)$$

El valor predictivo del resultado negativo es la probabilidad de que un resultado negativo sea cierto.

$$VP- = VN / (VN + FN)$$

Con una sencilla fórmula en términos de *odds* (probabilidad dividida por su complementario), se relacionan valor predictivo, cociente de probabilidades y prevalencia, lo que se utiliza para calcular el valor predictivo conociendo los otros dos datos. La *odds* posprueba de un resultado es igual al cociente de probabilidades de ese resultado multiplicado por la *odds* preprueba.

$$Odds\ posprueba = CP \times odds\ preprueba$$

siendo la *odds* posprueba lo mismo que el valor predictivo dividido entre su complementario, y la *odds* preprueba la prevalencia de enfermedad (en forma de probabilidad) dividida entre su complementario.

En el caso del ejemplo de Silverman et al¹⁹, el cociente de probabilidades del resultado negativo cuando comparamos el curso no progresivo contra el curso progresivo (cambiando el orden de las columnas de la tabla 6) es 10,5, y la prevalencia de "no progresión del Alzheimer" es de 27,5% (78/284). Con esos datos, a través de la fórmula anterior podemos calcular el valor predictivo negativo:

$$Odds\ posprueba = CP \times odds\ preprueba$$

$$VP- / (1 - VP-) = 10,5 \times 0,275 / (1 - 0,275)$$

$$VP- = 80\%$$

Con esta población y esta prueba, la proporción de pacientes en los que la PET sea negativa y que aun así desarrollará el Alzheimer progresivo en tres años será "relativamente" pequeña, un 20%.

Si variamos los datos de prevalencia o, lo que es lo mismo, si aplicamos la prueba a distintas poblaciones, los datos de rendimiento de la prueba serán diferentes.

El rendimiento de la prueba también se evalúa calculando el error sistemático, que resulta de dividir VP + FP entre VP + FN, y que es importante en el caso de las tasas epidemiológicas donde se persigue acertar con el número total de casos. Es la razón entre el número de casos que da el test y el número de casos de la prueba de referencia. Está bien que esta razón se acerque a la unidad.

Índices de concordancia entre observadores

Las posibles fuentes de error que producen variabilidad en las mediciones se pueden clasificar en: el hecho medido (las personas, los sistemas), el instrumento de medida y el observador consigo mismo o con varios observadores.

La repetibilidad, la fiabilidad de un método de medición, es la probabilidad de que al repetir las medidas sobre el mismo hecho, sobre el mismo dato, obtengamos el mismo resultado. El complementario de esa probabilidad son los errores atribuibles a nuestra manera de realizar la medición.

Una forma de medir la fiabilidad es estudiar cuál es el grado de concordancia entre observadores (dos observadores sobre el mismo dato) o intraobservador (el mismo observador, el mismo dato incluso, en dos momentos distintos).

La repetibilidad en el caso de medidas cuantitativas (como la cifra de la presión arterial o el peso) se puede analizar mediante la desviación estándar obtenida tras mediciones sucesivas. Si se realizan pares de mediciones, se pueden representar en un diagrama de dispersión.

Para los datos medidos de forma cualitativa, los resultados de concordancia entre observadores se pueden analizar en una tabla de contingencia. La concordancia de criterios o grado de acuerdo entre dos evaluadores que observan el mismo fenómeno (p. ej., revisar una historia clínica) se mide con el índice kappa. Este índice trata de informar sobre el grado de acuerdo entre observadores excluyendo del cálculo la proporción de casos en que los observadores coincidirían por efecto del azar. El cálculo del índice kappa relaciona la coincidencia observada no debida al azar con el acuerdo total que podrían tener los observadores. El kappa varía entre 0 y 1 (no acuerdo, máximo acuerdo).

En el entrenamiento de evaluadores para revisar historias clínicas empleando el Appropriateness Evaluation Protocol (AEP), el análisis de la concordancia entre dos revisores que estudiaron por separado la misma muestra de 50 casos dio los resultados de la tabla 7.

El índice kappa en este caso tiene un valor de 0,63, que indica que el grado de acuerdo alcanzado entre los revisores es un 63% superior a lo esperable por el azar. El cálculo de este índice a partir de la tabla de datos se realiza rápidamente en el módulo "Épitable" (opción "comparar proporciones") de la versión 6 de Epi-Info (disponible en: www.cdc.gov), que hace también los cálculos para determinar la significación estadística y el intervalo de confianza. Con otros programas informáticos se calcula el índice kappa a partir de los pares de datos de cada observación.

Tabla 7. Grado de acuerdo entre observadores en la evaluación

	Evaluador 1		Total
	Día adecuado	Inadecuado	
Evaluador 2			
Día adecuado	25	7	32
Inadecuado	2	16	18
Total	27	23	50

Existen otras medidas para evaluar la fiabilidad en estos estudios. Peiró et al²⁰ llevaron a cabo un estudio para evaluar la fiabilidad interobservador y la validez respecto al juicio clínico del AEP en hospitales públicos españoles. Para evaluar la fiabilidad revisaron 614 días de estancia que fueron analizados de forma independiente por tres revisores (dos médicos y una enfermera). Calcularon como índices de resultado el acuerdo observado (casos en que coincidía cada par de revisores entre el total de días revisados), el acuerdo específico (casos en que dos revisores juzgan la estancia como inadecuada entre el total de estancias que han sido inadecuadas para alguno de los revisores) y el índice kappa. Encontraron un alto grado de fiabilidad entre observadores, con cifras de acuerdo específico superiores al 64% y de coeficiente kappa mayores de 0,75.

Las medidas de fiabilidad se calculan en cualquier estudio en el que interviene más de un evaluador. En el estudio de Bates et al²¹ sobre incidencia de sucesos adversos por medicamentos, dos médicos revisaron historias clínicas para clasificar incidentes como sucesos adversos potenciales, graves y prevenibles. El estadístico kappa varió entre 0,81 y 0,98 en el caso de la clasificación de suceso adverso evitable. El valor fue más bajo para el acuerdo sobre si el suceso adverso era grave (kappa entre 0,32 y 0,37).

Otro ejemplo de evaluación de la fiabilidad en un *audit* de calidad lo publica Scinto et al²², que realizaron un estudio de concordancia entre observadores entre 4 revisores de historias clínicas con diagnóstico de neumonía para calcular indicadores de calidad. Los cuatro revisaron 365 casos, y comprobaron el grado de acuerdo entre ellos. El índice kappa presentó valores más bajos para algunos indicadores (0,74 para el uso de antibióticos en las primeras horas del ingreso) y más altos para otros (0,93 para la evaluación de la oxigenación).

El índice kappa depende de la prevalencia, de manera que los valores extremos de prevalencia influyen en la interpretación de este índice. En el último estudio citado, Scinto et al²² realizaron un ajuste por el efecto de la prevalencia en el cálculo de los índices kappa.

Tasas ajustadas

La estandarización o ajuste es una manera de comparar tasas evitando el efecto de una variable que suponemos que distorsiona la comparación.

Cuando se quiere comparar tasas obtenidas de registros poblacionales (datos de mortalidad, registros de cáncer, tuber-

culosis), estandarizar tasas es una estrategia para intentar evitar el efecto de confusión que produce la diferente distribución de variables como la edad, el sexo o la ocupación, en las poblaciones comparadas.

El método directo de estandarización consiste en la comparación de tasas medias ponderadas por edad, por riesgo de mortalidad o por la variable que queramos controlar, tomando los datos de una población como estándar para hacer la ponderación. Esta estrategia se usa cuando se conocen las tasas específicas en cada grupo de la variable por la que queremos ajustar (grupos de edad, grupos de riesgo definido por diagnósticos). El resultado de este ajuste son las razones de tasas estandarizadas, que indican el exceso de enfermedad en una población respecto a otra, una vez ajustado el factor.

Como ejemplo, para comparar la estancia media entre hospitales en la metodología de clasificación de hospitales "Top 20"²³ se elabora el indicador "Estancia media ajustada por casuística y severidad (EMACS)". Este indicador representa la estancia media que habría tenido cada hospital si hubiera atendido a pacientes con la gravedad de los de la población que se ha tomado como estándar (la población del conjunto de hospitales de su tamaño, con la distribución por niveles de gravedad y duración de la estancia en cada nivel).

Con este método, la comparación de los indicadores ajustados permite afirmar en el último cálculo del año 2000 que los hospitales "Top 20" muestran una EMACS inferior en un 8% a la media del resto de hospitales participantes.

Cálculo de casos esperados. Ajuste indirecto

Cuando no se dispone de una población estándar con todos los datos con la que compararse, se recurre a hacer comparaciones a partir de tasas.

Imaginemos que en un hospital (A) queremos averiguar si nuestra tasa de partos por cesárea en el último año (el 32%) está dentro de lo esperado y queremos saber si el factor edad materna influye en que nuestras tasas sean distintas de otras poblaciones. Como las tasas poblacionales de partos por cesárea en cada grupo de edad no son conocidas, recurrimos a unas tasas obtenidas de la bibliografía y analizamos (tabla 8) qué hubiera pasado en nuestra población, qué número de casos hubiéramos esperado, si nuestras tasas fueran las de la bibliografía (de ficción).

Comparados con las tasas estándar, nuestros casos esperados serían 320 en lugar de los 480 observados.

El resultado de la comparación entre la población del estudio y la población estándar es la razón de morbilidad estandarizada (RME o SMR en inglés), que se calcula con el cociente entre el número de casos observado y el número de casos esperado.

$$RME = 480 / 320 = 1,5$$

Y se interpreta que el hospital A tiene un exceso de casos del 50% sobre lo esperado, teniendo en cuenta el efecto de la edad.

Si aplicamos las mismas tasas de referencia a otro hospital (B) con una proporción de cesáreas menor (14,6%) (ta-

Tabla 8. Cálculo de casos esperados. Hospital A

Grupos de edad materna (años)	N.º de partos	Tasas de referencia (%)	Casos esperados	Casos observados
< 25	200	10	20	40
26-35	1.000	20	200	190
36-45	200	30	60	90
> 46	100	40	40	60
	1.500		320	480

bla 9), su RME es 0,75. Es decir, una cuarta parte menos que si el hospital B hubiera tenido las tasas por grupo de edad empleadas como estándar.

Con las fórmulas convenientes se demuestra que la RME es también un cociente entre tasas ajustadas. El numerador es la tasa de la población de estudio, y el denominador es la media ponderada de las tasas específicas por categoría (de edad, de clase social) en la población que comparamos, ponderada en función de la distribución de la propia población de estudio.

La población de estudio cambia cada vez que calculamos una RME, lo que implica que no se deba comparar las RME entre sí. Para calcular cada RME se emplea un estándar distinto (la propia población de estudio), por lo que las tasas de partos por cesárea entre hospitales no pueden ser comparadas entre sí con este método. No se puede decir en nuestro ejemplo que la tasa de cesáreas ajustada por edad en el hospital A sea el doble que en el hospital B.

Una forma más compleja de ajustar es la utilizada para el cálculo del índice de mortalidad ajustado por riesgo (IMAR) y del índice de complicaciones ajustado por riesgo (ICAR) que se usan en la metodología "Top 20". Estos índices se usan para comparar datos entre hospitales teniendo en cuenta el efecto de variables como la gravedad de los pacientes atendidos y los servicios de que dispone el hospital. Estos factores deben ser controlados para poder interpretar las comparaciones de tasas. En una escala de menos a más riesgo, los datos comparados de IMAR dan un índice de riesgo de 0,83 de media para los hospitales "Top 20" y de 1,04 para el resto.

Tabla 9. Cálculo de casos esperados. Hospital B

Grupos de edad materna (años)	N.º de partos	Tasas de referencia (%)	Casos esperados	Casos observados
< 25	400	10	40	35
26-35	1.200	20	240	170
36-45	200	30	60	40
> 46	50	40	20	25
	1.850		360	270

El IMAR y el ICAR son ejemplos de métodos para calcular los casos esperados para un hospital partiendo de un modelo de regresión que incluye una docena de variables. Es una forma más compleja de cálculo pero la esencia es la misma del ajuste de tasas. En este caso, el modelo de regresión hace el mismo papel que las tasas de referencia en el ejemplo de los partos por cesárea.

Cálculo de casos esperados. Distribución de Poisson

Para los sucesos que ocurren con poca frecuencia, como la mortalidad semanal en una sala de hospitalización o las infecciones posquirúrgica de cataratas, se puede, a partir del número de casos promedio observado en los últimos años, asumir que el suceso sigue una distribución de probabilidad de Poisson, y al número de casos que ocurran (o que se espere) asignarle una probabilidad.

Si, por ejemplo, en la UCI del hospital mueren de media tres pacientes cada semana, el que una semana hayan ocurrido 7 o más casos tiene una probabilidad de 0,033. Nosotros podemos poner un límite a esa probabilidad, de manera que a partir de un umbral se considere el dato superior a lo esperado y se inicie una investigación específica.

Errores en la medición. El efecto del azar. Los riesgos alfa y beta

En epidemiología se trabaja habitualmente con muestras que tratan de generalizar los resultados sobre una población. Siempre hay, por tanto, un error aleatorio, un error de muestreo.

Se habla de error aleatorio cuando la equivocación en la medida se da sin un patrón organizado. Lo contrario es la precisión, la fiabilidad, medir siempre lo mismo, lo que se espera, por ejemplo, de un termómetro o del surtidor de una gasolinera. Un revisor de historias clínicas debe ser preciso al buscar los criterios de adecuación de estancias hospitalarias o al buscar los criterios diagnósticos de una neumonía nosocomial.

Cuando elaboramos datos a partir de muestras y con la magnitud de esos datos realizamos inferencias sobre poblaciones, en esa inferencia existe un error aleatorio, que cuantificamos de dos formas: el valor de p y el intervalo de confianza (IC). El valor de p indica la significación estadística. El valor de p depende del tamaño de la desviación respecto de la hipótesis inicial y también del tamaño de la muestra. Por ello, es preferible utilizar intervalos de confianza. Con el IC conocemos, además de si la asociación es estadísticamente significativa, la precisión, es decir, los límites entre los que oscilan los datos obtenidos.

El nivel de decisión $p = 0,05$ o el IC del 95% son una guía que se utiliza para descartar el efecto del azar en las explicaciones de unos resultados. Un valor de p menor de 0,05 dice que es poco probable que una asociación se deba al azar, con lo que podemos atrevernos a afirmar que esa asociación existe.

La principal forma de reducir este error es aumentar el tamaño del estudio, lo que se observa en la fórmula del tamaño muestral, que depende de: *a)* el error alfa: posibilidad de que el efecto se deba al azar, y *b)* el error beta: posibilidad de no detectar un efecto real.

La terminología de riesgo del fabricante (error alfa) y del comprador (error beta) proviene de la industria y son los riesgos que corre cada uno (fabricante y comprador) al aceptar o rechazar lotes de productos basándose en muestreos. En las reglas de decisión para investigar o no los resultados de un sistema de monitorización, el error tipo 1 o alfa se comete al tratar como problemas casos que no lo son, mientras que el error tipo 2 o beta se da al no estudiar problemas que realmente lo son.

Bibliografía

1. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q* 1966;44:166-203 (traducido al castellano en *Rev Calidad Asistencial* 2001;16:S11-S27).
2. Donabedian A. The quality of medical care. *Science* 1978; 200:856-64 (traducido al castellano en *Rev Calidad Asistencial* 2001;16:S29-S38).
3. Morris J. *Uses of epidemiology*. Edimburgo: Livingstone, 1970.
4. Hennekens C. *Epidemiology in medicine*. Boston: Little, Brown and Co., 1988.
5. Sierra Bravo R. *Técnicas de investigación social. Teoría y ejercicios*. Zaragoza: Paraninfo, 1995.
6. Galgano A. *Los siete instrumentos de la calidad total*. Barcelona: Díaz de Santos, 1995.
7. García C, Almenara J, Mérida F, García JJ, Ramos MA. Mortalidad innecesariamente prematura y sanitariamente evitable de la provincia de Cádiz. *Rev Calidad Asistencial* 2001;13: 169-72.
8. Joint Commission on Accreditation of Healthcare Organizations. Sentinel events alert. N.º 25, february 26, 2002. Disponible en: <http://www.jcaho.org>
9. Andrews L, Stocking C, Krizek T, Gottlieb L, Krizek C, Vargish T, et al. An alternative strategy for studying adverse events in medical care. *Lancet* 1997;349:309-13.
10. Monitoring hospital-acquired infections to promote patient safety -United States, 1990-1999. *MMWR* 2000;49:149-53.
11. Sociedad Española de Medicina Preventiva, Salud Pública e Higiene. Evolución de la prevalencia de las infecciones nosocomiales en los hospitales españoles. EPINE 1990-1999. Madrid: IMC. 2001
12. Ebbesen J, Buajordet I, Erikssen J, Brors O, Hilberg T, Svaar H, et al. Drug-related deaths in a department of internal medicine. *Arch Intern Med* 2001;161:2317-23.
13. Kollef M, Sherman G, Ward S, Fraser V. Inadequate antimicrobial treatment or infections. A risk factor for hospital mortality among critically ill patients. *Chest* 1999;115:462-74.
14. García M, Lardelli P, Bueno A, Luna J, Espigares M, Gálvez R. Proportion of hospital deaths associated with adverse events. *J Clin Epidemiol* 1997;50:1319-26.
15. Llorca J, Fariñas C, Delgado M. Fracción atribuible poblacional: cálculo e interpretación. *Gac Sanitaria* 2001;15:61-7.
16. Frick M, Elo O, Haapa K, Heinonen O, Heinsalmi O, Helo P. Helsinki Heart Study: primary prevention trial with gemfibrozil in middle-aged men with dyslipidemia. *N Engl J Med* 1987; 317:1237-45.

17. Meneu R, Peiró S, Márquez S. Influencia de la presentación de resultados de los ensayos clínicos en la intención de prescribir: relativizando el riesgo relativo. *Aten Primaria* 1998;21:446-50.
18. Altman D. Confidence intervals for the number need to treat. *BMJ* 1998;317:1309-12.
19. Silverman D, Small G, Chang C, Lu C, Kung De Aburto M, Che W, et al. Position emission tomography in evaluation of dementia. Regional brain metabolism and long-term outcome. *JAMA* 2001;286:2126-7.
20. Peiró S, Meneu R, Rosello ML, Portella E, Carbonell-Sanchís R, Fernández C, et al. Validez del protocolo de evaluación del uso inadecuado de la hospitalización. *Med Clin (Barc)* 1996; 107:124-9.
21. Bates D, Cullen D, Laird N, Petersen L, Small S, Servi D, et al. Incidence of adverse drug events and potential adverse drug events. Implications for prevention. *JAMA* 1995;274:29-34.
22. Scinto JD, Galusha DH, Krumholz HM, Meehan TP, Qualidigm, Middletown CT. The case for comprehensive quality indicator reliability assessment. *J Clin Epidemiol* 2001;54: 1103-11.
23. García L, Illa C, Arias A, Casas M. Los Top 20 2000: objetivos, ventajas y limitaciones del método. *Rev Calidad Asistencial* 2001;16:107-16.