

## **Cómo estudiar un estudio y probar una prueba: lectura crítica de la literatura médica<sup>1</sup>**

**Segunda edición**

**Richard K. Riegelman y Robert P. Hirsch**

PARTE VII:

- Capítulo 19. Ejercicios para detectar errores:  
la prueba de una prueba
- Capítulo 20. Introducción a las tasas
- Capítulo 21. Muestreo de tasas
- Capítulo 22. Estandarización de tasas

---

<sup>1</sup>Título original: *Studying a Study and Testing a Test. How to Read the Medical Literature*. Second edition. © Richard K. Riegelman, Robert P. Hirsch. Publicado por Little, Brown and Company, Boston, Massachusetts 02108, Estados Unidos de América. Los pedidos del libro en inglés deben dirigirse a esta dirección.

Versión en español autorizada por Little, Brown and Company; se publica simultáneamente en forma de libro (Publicación Científica 531) y como serie en el *Boletín de la Oficina Sanitaria Panamericana* Traducción de José María Borrás, revisada por el Servicio Editorial de la Organización Panamericana de la Salud.

© Little, Brown and Company, 1989. Todos los derechos reservados. Ninguna parte de esta publicación puede ser reproducida ni transmitida en ninguna forma ni por ningún medio de carácter mecánico o electrónico, incluidos fotocopia y grabación, ni tampoco mediante sistemas de almacenamiento y recuperación de información, a menos que se cuente con la autorización por escrito de Little, Brown and Company.

# EJERCICIOS PARA DETECTAR ERRORES: LA PRUEBA DE UNA PRUEBA

Los siguientes ejercicios se han diseñado con el fin de evaluar la capacidad que usted ha adquirido para aplicar los diversos principios usados en *La prueba de una prueba*. Los ejercicios incluyen varios errores que se han ilustrado con ejemplos hipotéticos. Lea cada ejercicio y luego escriba una crítica señalando los tipos de errores cometidos por los investigadores. Compare su crítica con la que se proporciona al final de cada ejercicio.

## EJERCICIO No. 1: VARIABILIDAD DE LAS POBLACIONES SANA Y ENFERMA

Se realizaron dos investigaciones para evaluar la utilidad de una nueva prueba diagnóstica del cáncer de mama. Previamente, se había observado que los resultados de la prueba variaban en menos de 1% cuando se repetían en condiciones iguales y eran leídos por el mismo intérprete.

En el primer estudio, los investigadores escogieron a 100 mujeres con cáncer de mama metastásico y a 100 mujeres sanas sin signos de enfermedad mamaria. En las mujeres sanas, los resultados de la prueba oscilaron entre 30 y 100 mg/dl, y en las pacientes con cáncer, entre 150 y 200 mg/dl. Dado que la prueba diferenciaba perfectamente a un grupo de otro, los investigadores concluyeron que podía considerarse una prueba ideal para el diagnóstico del cáncer de mama y que debía aplicarse inmediatamente al tamizaje de todas las mujeres.

En un segundo estudio en el que se usó la misma prueba, otro investigador comparó a 100 mujeres recién diagnosticadas de cáncer de mama con 100 que padecían una enfermedad benigna de mama. Los resultados de las pacientes de cáncer variaron entre 70 y 200 mg/dl y los de las pacientes con enfermedad benigna, entre 40 y 180 mg/dl. Los autores de este estudio se percataron de la notable superposición entre los dos grupos y concluyeron que la prueba era inútil.

Un lector de ambos estudios, asombrado de que dos investigadores respetados pudieran obtener resultados tan inconsistentes, concluyó que se debían haber cometido errores al notificarlos.

### CRÍTICA: EJERCICIO No. 1

Para revisar estos estudios, es interesante organizar la discusión en torno a los conceptos de variabilidad de la prueba, variabilidad de la población sin la enfermedad y variabilidad de la población con la enfermedad.

#### Variabilidad de la prueba

La reproducibilidad es la medida de la variabilidad de una prueba. Según se afirma en el estudio, cuando la prueba fue realizada en las mismas condiciones por el mismo intérprete, su variabilidad fue de 1%. La medición de la reproducibi-

lidad exige repetir la prueba para demostrar que los resultados de la segunda lectura no están influidos por los de la primera lectura. Como el mismo intérprete repitió las pruebas, es posible que la lectura de los resultados de la segunda prueba estuviera influida por los resultados iniciales. Si esto fuera cierto, la reproducibilidad podría ser menor que la notificada anteriormente. No obstante, en el resto de la discusión supondremos que los autores tenían razón al creer que la prueba era reproducible.

## **Variabilidad del grupo de individuos sin la enfermedad**

En el primer estudio se incluyeron mujeres sanas y los resultados oscilaron entre 30 y 100 mg/dl. En cambio, en el segundo se estudiaron mujeres con enfermedades mamarias benignas y los resultados estuvieron comprendidos entre 70 y 200 mg/dl. Tal vez estos estudios no sean contradictorios y representen dos segmentos diferentes del grupo de mujeres sin la enfermedad. Es posible que la enfermedad mamaria benigna eleve a valores intermedios la concentración del metabolito medida por la prueba.

La medida adecuada del grupo de personas sin la enfermedad es el intervalo de la normalidad, el cual habitualmente incluye a 95% de los individuos del grupo sin cáncer de mama. Sin embargo, aquí los resultados se presentan en intervalos que agrupan a 100% de los individuos; el intervalo total no nos dice nada sobre la forma en que se agrupan los resultados. Estos podrían estar muy concentrados entre 70 y 100. Sin disponer de todos los datos, o al menos del intervalo de la normalidad, es difícil utilizar estos estudios para comparar pacientes con y sin cáncer de mama.

## **Variabilidad de los individuos con la enfermedad**

En el primer estudio se incluyeron pacientes con cáncer metastásico de mama y se estableció un intervalo de resultados comprendido entre 150 y 200 mg/dl. En el segundo, se estudiaron pacientes recién diagnosticadas y los límites del intervalo de los resultados observados fueron 70 y 200 mg/dl. Es posible que esta discrepancia no refleje un error en la notificación de los datos, sino las diferencias entre los grupos estudiados. Las pacientes recién diagnosticadas de cáncer de mama probablemente representan un amplio espectro de la enfermedad, incluidos los estadios iniciales y los metastásicos. Las pacientes con metástasis de cáncer de mama constituirían solo un extremo del espectro de la enfermedad. Por este motivo, el intervalo más amplio de valores encontrado entre las pacientes recién diagnosticadas podría reflejar el grupo más representativo de pacientes con cáncer de mama incluido en el segundo estudio.

## **Discriminación diagnóstica**

Los datos presentados no nos permiten calcular la sensibilidad o la especificidad de la prueba. Puesto que no se menciona la distribución de los resultados individuales, no se puede establecer un intervalo de la normalidad o una separación entre las pruebas positivas y negativas. Por lo tanto, no es aconsejable llegar a conclusiones sobre la utilidad diagnóstica de la prueba.

En el primer estudio, los investigadores incluyeron tanto a individuos que padecían una enfermedad bastante avanzada como a individuos claramente sanos. No es sorprendente que sus resultados parecieran diferenciar correctamente a los grupos. En el segundo estudio, los investigadores incluyeron a pacientes que representaban un espectro más amplio de la enfermedad y también a las que padecían enfermedades benignas. Por lo tanto, tampoco es sorprendente que se produjera una ma-

por superposición de los valores numéricos. Tan incorrecto es inferir que el primer estudio es una prueba perfecta como que el segundo es una prueba inútil. La verdad, que probablemente se encuentra entre ambos extremos, exige valorar la sensibilidad y la especificidad utilizando todos los datos de un amplio espectro de la población enferma y de la sana.

## **EJERCICIO No. 2: EL CONCEPTO DE LA NORMALIDAD**

Un investigador intentó establecer los límites de la normalidad para una nueva prueba diagnóstica de la diabetes, como se describe a continuación.

1. Localizó a 1 000 pacientes hospitalizados por enfermedades distintas de la diabetes.
2. Aplicó la prueba a esos 1 000 pacientes.
3. Trazó la distribución de los valores de la nueva prueba, excluyó 2,5% de los valores del extremo superior y 2,5% del inferior e incluyó el 95% restante en el intervalo de la normalidad.

A continuación, aplicó la nueva prueba en la comunidad y realizó pruebas de tamizaje en voluntarios. A los que tuvieron resultados dentro del intervalo de la normalidad les dijo que no tenían diabetes y a los que tuvieron resultados fuera de esos límites, que padecían diabetes. Un año más tarde aplicó la prueba de nuevo a varios individuos cuyos resultados habían correspondido a la zona inferior del límite de la normalidad y al observar que en esta ocasión los resultados se encontraban en la zona superior de dicho límite, les aseguró que no tenían diabetes.

Un paciente obeso con resultados en la zona superior del intervalo de la normalidad y con una historia familiar de diabetes muy marcada le pidió consejo sobre la forma de evitar el desarrollo de la enfermedad. El investigador le respondió que, como sus resultados se encontraban dentro del intervalo de la normalidad, no debía tener motivos de preocupación.

### **CRÍTICA: EJERCICIO No. 2**

#### **Desarrollo del intervalo de la normalidad**

Al establecer el intervalo de la normalidad, el investigador debe intentar incluir solamente a los individuos que no tengan la enfermedad estudiada. El investigador del estudio anterior llegó a la conclusión de que los individuos hospitalizados con diagnósticos distintos de la diabetes no padecían esta enfermedad. Sin embargo, la diabetes es muy frecuente y los pacientes diabéticos desarrollan una serie de complicaciones que aumentan el riesgo de ser hospitalizados. Por lo tanto, es probable que una proporción de los individuos internados con diagnósticos principales diferentes también tuvieran diabetes y que el investigador no haya establecido un intervalo de la normalidad de pacientes exentos de la enfermedad.

El investigador utilizó como intervalo de la normalidad 95% de los resultados centrales de un grupo de personas que presuntamente no tenían la enfermedad. Aunque este es el procedimiento habitual, puede que no se preste a la máxima discriminación diagnóstica de la prueba. A veces, la modificación de los límites del intervalo de la normalidad puede mejorar la capacidad de la prueba para discriminar entre los que tienen y los que no tienen la enfermedad. No obstante, debe recordarse que cuando cambiamos los límites del intervalo de la normalidad para obtener menos re-

sultados negativos falsos, pagamos el precio de obtener más resultados positivos falsos o viceversa. Si bien puede merecer la pena pagar ese precio, se necesitan más datos antes de poder saber si este es el caso. De cualquier modo, los datos disponibles no proporcionan los medios adecuados para juzgar si la nueva prueba ayuda a discriminar a los diabéticos de los no diabéticos. Lo único que conocemos es el intervalo de la normalidad definido para la prueba.

## **Aplicación del intervalo de la normalidad**

En el caso descrito, no se ha mantenido la distinción entre el concepto del intervalo de la normalidad y el de enfermedad. El autor ha considerado sinónimos el estar fuera del intervalo de la normalidad y tener diabetes, y el estar dentro del intervalo y no tener diabetes. No se han presentado pruebas de que la nueva prueba sea útil para discriminar a los diabéticos de los no diabéticos y es posible que los primeros se encuentren totalmente dentro del intervalo de la normalidad de esta prueba.

Aunque se hubiera demostrado, desde el punto de vista del diagnóstico, que esta prueba es útil para distinguir a los que tienen diabetes de los que no la tienen, es probable que algunos individuos con diabetes tuvieran valores dentro del intervalo de la normalidad y algunos sin la enfermedad, fuera de dicho intervalo. Por definición, el intervalo de la normalidad excluye a 5% de los individuos que no tienen la enfermedad. Por eso, el investigador no puede limitarse simplemente a aplicar la prueba y a identificar a los individuos como diabéticos o no diabéticos.

## **Cambios dentro del intervalo de la normalidad**

El hecho de que los resultados de una prueba aplicada a un individuo cambien, aunque se mantengan dentro del intervalo de la normalidad, puede ser una manifestación de enfermedad. El concepto del intervalo de la normalidad se ha desarrollado principalmente para individuos sobre los que no disponemos de datos basales anteriores. Cuando este es el caso, es preciso comparar, por medio del intervalo de la normalidad, los resultados individuales con los de aquellos que presuntamente están sanos. Si la misma prueba se ha aplicado al individuo con anterioridad, esta información debe tenerse en cuenta.

Un cambio dentro del intervalo de la normalidad puede representar un gran aumento para un individuo determinado; esto se manifiesta especialmente cuando los resultados anteriores de la prueba se encuentran cerca del límite inferior de la normalidad y los posteriores se desplazan hacia su límite superior. Para esos individuos, los cambios que se producen dentro de los límites normales pueden ser manifestaciones precoces de la enfermedad.

## **Grupo de referencia**

El grupo de referencia empleado en este estudio para fijar el intervalo de la normalidad estaba formado en su totalidad por pacientes hospitalizados. Su intervalo de la normalidad podría ser bastante diferente del de otras poblaciones de pacientes jóvenes, ambulatorios y sanos. Por este motivo, al establecer los límites de la normalidad de un grupo y aplicarlo a otro con características diferentes se pudo haber introducido un error.

## Dentro de los límites de la normalidad *versus* lo deseable

Es posible que todos o algunos de los individuos cuyos resultados se encontraban dentro de los límites de la normalidad tuvieran valores más elevados que los deseables. Recuerde que el intervalo de la normalidad refleja cómo son las cosas y no necesariamente cómo deben ser. Posiblemente, una pérdida de peso que en consecuencia disminuya los valores detectados por la prueba prevenga futuros problemas. Esto supone que la prueba discrimina de hecho a los enfermos de los no enfermos, que perder peso influirá en los valores numéricos de los resultados de la prueba y que la reducción de estos últimos mejorará el pronóstico. Sin embargo, lo que interesa en general es que los resultados que caen dentro del intervalo de la normalidad no son necesariamente los deseables.

### EJERCICIO No. 3: DISCRIMINACIÓN DIAGNÓSTICA DE LAS PRUEBAS

Se va a evaluar la utilidad de una nueva prueba para el diagnóstico de la tromboflebitis. La prueba de referencia tradicional ha sido la flebografía y con ella se comparará la nueva prueba. Para valorar la reproducibilidad de la nueva prueba, esta se aplica a 100 pacientes consecutivos con flebografías positivas. Los investigadores observan que 98% de los pacientes diagnosticados de tromboflebitis dan resultados positivos a la prueba. Repiten la prueba en el mismo grupo de pacientes y de nuevo observan que es positiva en 98% de los 100 pacientes. A partir de estos datos concluyen que la reproducibilidad de la nueva prueba es de 100%.

Una vez demostrada la reproducibilidad de la prueba, proceden a estudiar su discriminación diagnóstica, para lo cual deben evaluar el éxito de la prueba en comparación con la flebografía, la prueba de oro o de referencia tradicional. Seguidamente, estudian 1 000 pacientes consecutivos con dolor de piernas unilateral, 500 de los cuales tuvieron flebografías positivas y 500, negativas. Los investigadores clasifican a los individuos como positivos o negativos y presentan los datos del siguiente modo:

PRUEBA NUEVA	FLEBOGRAFÍA POSITIVA	FLEBOGRAFÍA NEGATIVA
Positiva	450	100
Negativa	50	400
	500	500

En este ejemplo los investigadores usaron la definición aceptada de sensibilidad, es decir, la proporción de individuos con resultados positivos en la prueba de referencia que tienen resultados positivos en la nueva prueba. De esta manera,

$$\text{Sensibilidad} = \frac{450}{500} = 0,90 = 90\%$$

También usaron la definición aceptada de especificidad, es decir, la proporción de individuos negativos a la prueba de referencia que tienen resultados

negativos en la nueva prueba. De este modo,

$$\text{Especificidad} = \frac{400}{500} = 0,80 = 80\%$$

Asimismo, calcularon el valor predictivo de una prueba positiva para su grupo de estudio. La definición aceptada del valor predictivo de un resultado positivo es la proporción de individuos con resultados positivos en la nueva prueba que realmente tienen la enfermedad medida con la prueba de oro. De esta manera,

$$\text{Valor predictivo de una prueba positiva} = \frac{450}{550} = 0,818 = 81,8\%$$

Sobre la base de estos resultados, los investigadores llegaron a las siguientes conclusiones:

1. La nueva prueba es totalmente reproducible.
2. La nueva prueba es menos sensible y menos específica que la flebografía y, por eso, es una prueba intrínsecamente inferior.
3. Cuando se aplica a un nuevo grupo de pacientes, por ejemplo, a un grupo con dolor de piernas bilateral, se puede esperar que el valor predictivo de un resultado positivo con la nueva prueba sea igual a 81,8%.

### CRÍTICA: EJERCICIO No. 3

Cuando una prueba se aplica varias veces a los mismos individuos y en las mismas condiciones, el método para valorar su reproducibilidad exige que los resultados de cada individuo sean prácticamente idénticos si la prueba tiene una reproducibilidad de 100%. Los autores declararon que el total de nuevas pruebas positivas fue idéntico cuando se repitieron. Sin embargo, no indicaron si los mismos individuos que fueron positivos cuando se repitió la prueba también fueron positivos la primera vez. Si los mismos individuos no fueron positivos, la prueba no puede considerarse reproducible. Los autores tampoco indicaron si los que realizaron e interpretaron los resultados de las pruebas repetidas conocían los resultados de la primera prueba.

Una prueba de oro es la medida generalmente aceptada de una enfermedad contra la cual se comparan las pruebas nuevas o todavía no probadas, pero, de hecho, no siempre es una medida ideal de la enfermedad para cuyo diagnóstico ha sido diseñada. Una prueba usada como prueba de oro puede considerarse diagnóstica solo por tradición o por aceptación generalizada de su utilidad. No obstante, es posible que una nueva prueba sea una medida más útil de la enfermedad que la aceptada como referencia. Al comparar la sensibilidad y la especificidad de nuevas pruebas con la de oro debemos ser conscientes de que las discrepancias entre las pruebas pueden ser resultado de la imperfección de la prueba de referencia y no de la deficiencia de la nueva prueba.

Cuando los autores concluyeron que la nueva prueba era menos sensible y específica que la flebografía, estaban suponiendo que esta tiene una sensibilidad y especificidad de 100%. Basándose en esta suposición, es imposible que la nueva prueba tenga una sensibilidad y especificidad más altas que la antigua. Si no estamos seguros de que la flebografía siempre es correcta, es prematuro concluir que la nueva prueba es menos útil para diagnosticar la tromboflebitis. Por lo tanto, los autores debie-

ron haber limitado sus conclusiones sobre la sensibilidad y la especificidad de la nueva prueba a una comparación con la flebografía. Si la nueva prueba fuese más segura, barata o práctica que la flebografía, podría llegar a reemplazar a la flebografía en la práctica clínica. A la larga, la experiencia clínica podría demostrar que es lo suficientemente fiable para ser utilizada como prueba de oro. Mientras tanto, lo mejor que se puede esperar de la prueba es que iguale a la prueba de referencia establecida que, por definición, tiene una sensibilidad y especificidad de 100%.

Los autores midieron correctamente la sensibilidad, la especificidad y el valor predictivo de una prueba positiva en su grupo de estudio. Como afirmaron, el valor predictivo de una prueba positiva es la proporción de los positivos a la nueva prueba que realmente tienen la enfermedad medida según el criterio de oro. En este grupo de estudio, la prevalencia de tromboflebitis fue de 50% (500 con tromboflebitis y 500 sin la enfermedad) y, por consiguiente, el valor predictivo de la prueba es 450 dividido por 550, lo que equivale a 81,8%. Sin embargo, el valor predictivo de una prueba es diferente en distintos grupos de pacientes, dependiendo de la prevalencia o probabilidad de la enfermedad anterior a la prueba en el grupo estudiado. No se puede extrapolar directamente un valor predictivo obtenido en un grupo de pacientes a otro grupo en el cual la prevalencia de la enfermedad es distinta. Es de esperar que un grupo de pacientes con dolor de piernas unilateral tenga una prevalencia de tromboflebitis distinta de otro grupo con dolor bilateral.

Los valores predictivos del dolor de piernas bilateral no se pueden estimar basándose únicamente en la sensibilidad y la especificidad que la prueba ha demostrado en los pacientes con dolor unilateral. Sin embargo, si se puede estimar también el porcentaje de individuos con dolor de piernas bilateral que padecen tromboflebitis, entonces es posible estimar los valores predictivos de una prueba en estos pacientes. Supongamos que la prevalencia de tromboflebitis es mucho más baja en los pacientes con dolor de piernas bilateral. Entonces sería de esperar que una nueva prueba positiva tuviera un valor predictivo positivo mucho más bajo que en un grupo de pacientes con dolor unilateral.

Recuerde que, desde el punto de vista clínico, la prevalencia es lo mismo que la probabilidad de padecer la enfermedad antes de realizar la prueba, y que el valor predictivo de una prueba positiva es la probabilidad de padecerla después de obtener un resultado positivo. Dado que la probabilidad de la tromboflebitis en un paciente que presenta dolor bilateral es menos de 50%, la probabilidad de la enfermedad incluso después de un resultado positivo sería mucho menos de 81,8%.

# INTRODUCCIÓN A LAS TASAS

Si usted oye el ruido de cascos, lo más probable es que sea un caballo y no una cebra. Esta metáfora de la clínica señala la obvia pero demasiadas veces olvidada verdad de que las enfermedades comunes se producen frecuentemente y las enfermedades raras, raramente. Cuando los clínicos dicen que una enfermedad es frecuente y otra es rara presuponen una diferencia en las *tasas*.

Todos los clínicos instintivamente utilizan el concepto de tasas. Saben que la enfermedad coronaria es mucho más frecuente en un hombre de mediana edad que en una adolescente. Saben que el cáncer de páncreas es mucho más común en las personas de edad avanzada que en los jóvenes. Saben que la anemia de células falciformes es mucho más probable en una persona de raza negra que en una de raza blanca.

En nuestra discusión anterior sobre las pruebas diagnósticas, señalamos que cuanto más baja es la tasa de prevalencia en una población (o sea, cuanto más rara sea la enfermedad), menor será el valor predictivo de una prueba positiva. Cuando se trata de una enfermedad rara, es menos probable que una prueba positiva indique su presencia. Los clínicos emplean este concepto automática y, quizá, inconscientemente. Saben que es improbable que una mujer joven con cambios en la onda T de un electrocardiograma tenga enfermedad coronaria. Saben que es improbable que un hombre joven con dolor abdominal persistente tenga cáncer de páncreas. Saben que es improbable que una persona joven de raza blanca con dolor articular y anemia tenga anemia de células falciformes. El médico puede apreciar el significado de las tasas sobre la base de su experiencia clínica personal; no obstante, es provechoso que haga uso de los artículos de investigación para mejorar su capacidad de valorar científicamente y objetivamente las tasas de enfermedad. Esta sección tiene como finalidad ayudar al lector a adquirir los conocimientos necesarios para comprender cómo se miden e interpretan científicamente las tasas de enfermedad. Esta comprensión le puede ayudar a escoger el método diagnóstico apropiado y a interpretar los resultados.

Además de facilitar los diagnósticos individuales, la comprensión del significado de las tasas de enfermedad ayuda a valorar los cambios que se producen con el tiempo o como resultado de las intervenciones médicas. Las tasas de enfermedad son un instrumento importante para realizar los tipos de estudios presentados en la *Parte 1, El estudio de un estudio*. La capacidad para reconocer cambios reales y relaciones de causa y efecto verdaderas depende de que se comprendan los principios básicos de la comparación de tasas.

Es posible que no sea evidente la necesidad de estudiar las tasas de las enfermedades. ¿Por qué no comparar simplemente el número de veces que ocurre un suceso? Examinemos el siguiente ejemplo, que muestra los problemas que pueden surgir si se compara únicamente el número de sucesos.

Un panel de revisión hospitalaria evaluó el rendimiento de los médicos del hospital en que usted trabaja. Encontraron que hubo cinco defunciones entre los 1 000 pacientes que usted atendió en el hospital durante el año pasado. El jefe del equipo tuvo solo una defunción entre los 200 pacientes que trató. El panel decidió que

tener cinco veces tantas defunciones como el jefe del equipo indicaba una práctica deficiente de la medicina.

Ahora bien, no es necesario que se prepare para defenderse diciendo: "Puede parecer que está mal, ¡pero realmente lo hago bien!" En lugar de fijarse en el total de defunciones, es más justo considerar cuántas se produjeron en relación con las que podrían haberse producido. Simplemente tiene que señalar que su tasa de mortalidad y la del jefe de su equipo son idénticas: 5 entre 1 000 es lo mismo que 1 entre 200. Las tasas de los sucesos han venido en su ayuda. ¡Vale la pena conocerlas!

Una probabilidad es una proporción en la cual el numerador es el número de veces que ocurre un suceso y el denominador es el número de veces que podría haber ocurrido. Como en todas las proporciones, el numerador está incluido en el denominador. Las tasas realmente son un tipo especial de medida en la que el denominador también incluye una unidad de tiempo.

En medicina, una función importante de las tasas y de las proporciones es la de caracterizar la historia natural de la enfermedad. Habitualmente se usan tres tipos de medidas:

1. Tasa de incidencia: número de casos nuevos que se producen por unidad de tiempo.
2. Prevalencia: probabilidad de tener una enfermedad en un momento dado.
3. Tasa de letalidad: probabilidad de morir de una enfermedad durante un espacio de tiempo a partir de su diagnóstico.

Las tasas de incidencia se definen del siguiente modo:

$$\text{Tasa de incidencia} = \frac{\text{Número de individuos que desarrollan la enfermedad durante un período}}{\text{Total de personas-año}^1 \text{ en riesgo}}$$

Con frecuencia es difícil saber cuántos individuos y durante cuánto tiempo están en riesgo de padecer una enfermedad. Por eso, las tasas de incidencia suelen estimarse mediante la siguiente fórmula:

$$\text{Tasa de incidencia de la enfermedad} = \frac{\text{Número de individuos que desarrollan la enfermedad durante un período}}{\text{Número de individuos en el grupo de riesgo en el punto medio del período de interés} \times \text{la duración de dicho período}}$$

Si, por ejemplo, se quiere conocer la tasa de incidencia de los casos de úlceras duodenales en Nueva York en 1990, esta tasa se calcularía teóricamente de la siguiente manera:

$$\text{Tasa de incidencia de úlceras duodenales en Nueva York en 1990} = \frac{\text{Número de residentes de Nueva York que desarrollaron una úlcera duodenal en 1990}}{\text{Número de residentes de Nueva York en riesgo de desarrollar una úlcera duodenal durante 1990} \times 1 \text{ año}}$$

<sup>1</sup> Una persona-año representa un individuo en riesgo de desarrollar la enfermedad durante un 1 año

Dado que la población de Nueva York fluctúa constantemente, es difícil saber el número real de personas que residieron en la ciudad y por cuánto tiempo vivieron en ella durante 1990. Para calcular la tasa de incidencia aproximada en ese año, se puede usar el censo de Nueva York del 1 de abril de 1990. La tasa aproximada de incidencia de úlceras duodenales en Nueva York en 1990 se calcularía del siguiente modo:

$$\text{Tasa de incidencia de úlceras duodenales en Nueva York en 1990} = \frac{\text{Número de residentes en Nueva York que desarrollaron una úlcera duodenal en 1990}}{\text{Número de residentes de Nueva York en riesgo de desarrollar úlceras duodenales el 1 de abril de 1990 (aproximadamente igual al número de residentes de Nueva York el 1 de abril de 1990)} \times 1 \text{ año}}$$

El tipo de tasa que hemos comentado hasta el momento es una tasa de *incidencia*, que está relacionada con el riesgo de desarrollar una enfermedad durante un espacio de tiempo. El riesgo es el efecto acumulativo de la tasa de incidencia de la enfermedad durante un período específico. Podemos imaginarnos la incidencia como la velocidad a la que uno se desplaza durante un período breve, y el riesgo, como la distancia que uno ha recorrido durante un largo espacio de tiempo, suponiendo que la velocidad es constante.<sup>2</sup> La tasa de incidencia mide los casos nuevos de una enfermedad determinada que se desarrollan por unidad de tiempo, y esto puede ser de ayuda al examinar la causa o etiología de una enfermedad. El riesgo estimado a partir de la probabilidad de desarrollar una enfermedad en un período específico puede contribuir a predecir los sucesos futuros, si se usa con precaución. La enfermedad, una vez desarrollada, puede durar mucho tiempo. Por eso, frecuentemente se usa un segundo tipo de medida que estima la probabilidad de *tener* la enfermedad en un momento determinado. Esta se conoce como *prevalencia* y mide lo frecuente o prevaeciente que es una enfermedad en un momento dado. La prevalencia es muy importante en el diagnóstico, dado que es el punto de partida para estimar la probabilidad anterior a la prueba de que la enfermedad se halle presente. Asimismo, proporciona una estimación de la probabilidad de que la enfermedad esté presente antes de evaluar la historia individual, el examen físico o las pruebas de laboratorio. De esta forma,

$$\text{Prevalencia} = \frac{\text{Número de individuos que tienen la enfermedad en un momento dado}}{\text{Número de individuos que forman parte del grupo en ese momento}}$$

En el ejemplo anterior, la prevalencia de las úlceras duodenales el 1 de abril de 1990 en la Ciudad de Nueva York se calcularía como sigue:

$$\text{Prevalencia} = \frac{\text{Número de residentes de Nueva York con úlceras duodenales el 1 de abril de 1990}}{\text{Número de residentes de Nueva York el 1 de abril de 1990}}$$

<sup>2</sup> Tanto los bioestadísticos como los epidemiólogos establecen una diferencia entre la incidencia acumulativa y las tasas de incidencia. En la incidencia acumulativa, el denominador es el número de individuos en la población al principio de un período determinado. Incidencia acumulativa es sinónimo de riesgo.

Para la mayor parte de las enfermedades, la tasa de incidencia y la de prevalencia se relacionan aproximadamente de la siguiente manera:

$$\text{Prevalencia} = \text{Tasa de incidencia} \times \text{Duración media de la enfermedad}$$

En otras palabras, cuanto más larga sea la duración de la enfermedad, más individuos tendrán la enfermedad en un momento dado y, por lo tanto, más alta será la prevalencia. Las enfermedades crónicas de larga duración, como la diabetes, pueden tener una tasa de incidencia baja, pero una prevalencia elevada en un momento determinado. Las enfermedades agudas de corta duración, como la faringitis estreptocócica, pueden tener una tasa de incidencia elevada, pero una prevalencia baja en un momento dado. Por eso, es importante saber que la prevalencia y la incidencia miden fenómenos distintos. Las tasas de incidencia miden la frecuencia con que se desarrolla un nuevo caso de la enfermedad por unidad de tiempo. La prevalencia mide la probabilidad de *tener* la enfermedad en un momento determinado. Si no se aprecia esta diferencia, se puede cometer el tipo de error que se ilustra con el siguiente ejemplo.

En un estudio sobre la gonorrea asintomática en hombres se tomaron muestras de 1 000 sujetos seleccionados al azar. Se les diagnosticó gonorrea a 10 de ellos. En un segundo estudio, se siguió a un grupo de hombres de la misma población durante un año. Se observó que durante ese lapso de tiempo solo uno de los hombres desarrolló gonorrea asintomática. Al comparar estos estudios, un revisor concluyó que uno de los dos tenía que estar equivocado, ya que las conclusiones eran contradictorias.

Esta aparente incongruencia desaparece, si se distingue entre la tasa de incidencia y la prevalencia. El primer estudio de los casos existentes midió la prevalencia, mientras que el segundo valoró la incidencia. El hecho de que la prevalencia sea mucho más elevada que la incidencia sugiere que la gonorrea asintomática es de larga duración. Esto puede explicarse por el hecho de que, aunque los casos sintomáticos suelen recibir tratamiento, los asintomáticos permanecen en la comunidad sin tratamiento durante un período prolongado.

Además de la tasa de incidencia y de la prevalencia, es necesario definir una tercera medida para caracterizar la historia natural de la enfermedad. Esta medida se conoce como *letalidad*.

$$\text{Letalidad}^3 = \frac{\text{Número de personas fallecidas por una enfermedad durante un período}}{\text{Número de personas diagnosticadas de la enfermedad al inicio del período}}$$

A diferencia de las tasas de incidencia, la letalidad está influida por los éxitos de las intervenciones médicas destinadas a curar las enfermedades. La letalidad es útil para valorar el pronóstico, porque mide la probabilidad de no sobrevivir una vez iniciada la enfermedad. La letalidad durante un período tiene una relación impor-

<sup>3</sup> La letalidad es una proporción que se refiere a la probabilidad de morir de una enfermedad. Cuando esta proporción se multiplica por la tasa de incidencia, se obtiene la tasa de mortalidad.

tante con las tasas de mortalidad de una enfermedad determinada (esto es, el número de defunciones debidas a una enfermedad por persona-año).

$$\text{Tasa de mortalidad} = \text{Tasa de incidencia} \times \text{Letalidad}$$

El no valorar esta relación puede conducir a la confusión que se describe en el siguiente ejemplo.

En un estudio de las tasas de la úlcera duodenal, los autores calcularon correctamente las tasas de mortalidad por úlcera duodenal en los Estados Unidos en 1949 y en 1989. En 1949, la tasa anual de mortalidad fue 5 por 1 000 000 personas-año. Estudios posteriores revelaron que ni la tasa de incidencia ni la prevalencia habían cambiado. Los autores no pudieron interpretar estos datos.

Conociendo la relación que existe entre las tasas de mortalidad y las de incidencia, se entiende que el descenso de las tasas de mortalidad debió ser causado por una reducción de la letalidad. Este descenso de la letalidad puede reflejar el progreso conseguido durante 40 años en el tratamiento de las úlceras duodenales, aunque no se haya progresado en la reducción de la incidencia (casos nuevos) de la enfermedad.

Las tasas de incidencia, la prevalencia y la letalidad miden, respectivamente, la tasa de desarrollo de los casos nuevos de una enfermedad por unidad de tiempo, la probabilidad de tener la enfermedad en un momento dado y la probabilidad de morir por una enfermedad, una vez diagnosticada. Además de estas medidas básicas, en la literatura médica se utiliza con frecuencia una medida conocida como *razón de mortalidad proporcional*, que se define como:

$$\text{Razón de mortalidad proporcional} = \frac{\text{Número de individuos fallecidos por una enfermedad}}{\text{Número de individuos fallecidos por todas las enfermedades}}$$

La razón de mortalidad proporcional mide la probabilidad de que una defunción se deba a una causa determinada. Las razones de mortalidad proporcional son una herramienta útil para determinar cuáles son las causas de muerte más frecuentes. Sin embargo, no nos informan sobre la probabilidad de morir, como muestra el siguiente ejemplo.

Un estudio bien diseñado reveló que los traumatismos fueron la causa de muerte de 4% de las personas mayores de 65 años y que causaron 25% de los fallecimientos entre los menores de 3 años. Los autores llegaron a la conclusión de que los mayores de 65 años tenían una probabilidad mucho menor de morir por traumatismos que los menores de 3 años.

El hecho de que la razón de mortalidad proporcional en los mayores de 65 años sea menor por traumatismos no significa necesariamente que los ancianos tengan una menor probabilidad de morir por esa causa. Dado que entre los mayores de 65 años se producen muchos más fallecimientos, aun 4% de las muertes por traumatismos pueden representar una tasa de mortalidad cercana a la tasa de mortalidad de los menores de 3 años.

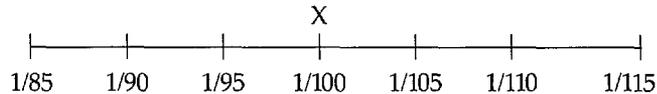
Habiendo ya examinado los tipos de tasas y proporciones que se encuentran con mayor frecuencia en la literatura médica, y distinguido esas medidas de las razones, centraremos nuestra atención en los métodos para calcular las tasas de enfermedad.

# MUESTREO DE TASAS

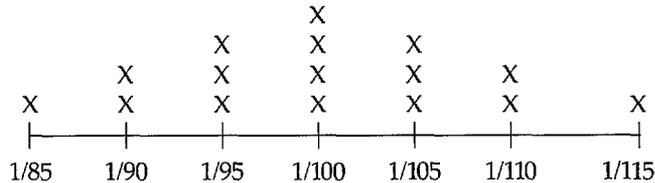
En algunas circunstancias es posible determinar todos los casos de una enfermedad en una población. De ordinario pueden obtenerse tasas de mortalidad para una población, porque los certificados de defunción son documentos legales obligatorios. En consecuencia, ello también permite calcular las tasas de mortalidad de una enfermedad para toda la población. Sin embargo, para la mayor parte de las enfermedades no es factible contar todos los casos en la población y, por esta razón, las técnicas de muestreo son muy útiles. El *muestreo (sampling)* es una técnica mediante la cual el investigador selecciona al azar una porción representativa de la población, estudia esa muestra y luego intenta extrapolar los resultados a toda la población escogida para el estudio.

## ERROR MUESTRAL

El proceso de extracción de una muestra no es perfecto, aunque se realice correctamente. Para valorar dicho proceso y el error intrínseco introducido por el muestreo es preciso comprender el principio que fundamenta esta técnica. Dicho principio afirma que si se extraen muchas muestras al azar, las tasas calculadas con los datos de esas muestras serán iguales, en promedio, a la tasa de la población original. En otras palabras, cada muestra puede tener una tasa mayor o menor que la de la población original. Por ejemplo, la siguiente figura muestra una muestra una tasa de 1 por 100 en la población original,



y si se extraen muestras de esta población, las tasas podrían ser las siguientes:



Observe que, si bien algunas de las tasas obtenidas en determinadas muestras son iguales que las de la población original, otras son mayores o menores. Dado que las muestras solo son exactas en promedio, se dice que una muestra determinada posee un error muestral intrínseco. Si no se aprecia la existencia del error muestral, se puede incurrir en la siguiente interpretación errónea:

Una organización nacional intentó estimar la prevalencia de los portadores de estreptococos realizando cultivos en una muestra aleatoria de 0,1% de todos los niños de la nación. Para verificar los resultados, la misma organización extrajo una segunda muestra aleatoria de 0,1% de los escolares y realizó una segunda encuesta con el mismo protocolo. El primer examen reveló una prevalencia de cultivos positivos de 15 por 1 000, mientras que el segundo examen reveló una prevalencia de 10 por 1 000. Los autores concluyeron que estos resultados inconsistentes eran imposibles, dado que habían usado el mismo método.

Los autores no tuvieron en cuenta el hecho de que el muestreo tiene un error intrínseco. Este error muestral puede explicar las diferencias observadas entre las dos muestras. El ejemplo simplemente señala que dos muestras extraídas de la misma manera pueden producir resultados diferentes solo a causa del azar. Recuerde que un elevado número de muestras proporcionan, en promedio, estimaciones que son idénticas al verdadero valor de la población, pero que puede haber una amplia variación entre dos muestras y entre estas y el verdadero valor de la población.

## EL TAMAÑO DE LA MUESTRA

Un segundo principio importante y necesario para comprender el muestreo afirma que cuantos más individuos formen parte de la muestra, más probable será que la tasa estimada con los datos de la muestra se aproxime a la tasa poblacional. Por esta razón, el tamaño de la muestra condiciona la proximidad de la tasa muestral a la poblacional. Esto no es sorprendente dado que, cuando todo el mundo forma parte de la muestra, está garantizado que la tasa muestral es igual a la poblacional.

Examinemos con más detalle este principio. El factor que más influye en la magnitud del error muestral es el tamaño de la muestra. La relación entre el tamaño muestral y la precisión no es de uno a uno, sino que es una función de la raíz cuadrada. Con muestras pequeñas, el aumento del tamaño muestral aumenta notablemente la precisión del estimador muestral de la tasa poblacional. Sin embargo, a medida que aumenta el tamaño de la muestra, la mejora de la precisión disminuye de forma relativa y los aumentos pequeños o moderados del tamaño muestral añaden poco a la precisión del estimador. Por lo tanto, los investigadores intentan equilibrar la necesidad de precisión con los costos económicos que acarrea el aumento del tamaño muestral. La consecuencia del empleo de muestras pequeñas es que estas pueden variar mucho entre sí y en relación con el verdadero valor de la población. El siguiente ejemplo ilustra la necesidad de tener en cuenta los efectos del tamaño muestral en los resultados del muestreo.

Un investigador extrajo una muestra de 0,01% de los certificados de defunción de la nación y encontró que la tasa de mortalidad por cáncer de páncreas era de 50 por 100 000 personas-año. Un segundo investigador, que extrajo una muestra de 1% de dichos certificados, llegó a la conclusión de que la verdadera tasa de mortalidad era de 80 por 100 000 personas-año. Para resolver esta discrepancia, el segundo investigador identificó todas las defunciones causadas por cáncer de páncreas en el país y obtuvo una tasa de 79 por 100 000 personas-año. Finalmente, llegó a la conclusión de que el primer investigador había realizado su estudio de forma fraudulenta.

El primer estudio empleó una muestra que era la centésima parte de la segunda; por lo tanto, es posible que el error muestral del primer estudio fuera mucho mayor. El hecho de que la segunda muestra fuera más exacta se debe probablemente al aumento de precisión que le confiere su mayor tamaño y no a fraude en el primer estudio.

## EL MUESTREO ALEATORIO

Aunque se han esbozado dos principios importantes del muestreo, todavía queda un aspecto por considerar. Esos dos principios se basan en el supuesto de que la muestra se ha obtenido al azar, lo que significa que todos los individuos de la población tenían la misma probabilidad (o, al menos, una probabilidad conocida) de ser seleccionados para su inclusión en la muestra. Si no se realiza un muestreo al azar, no se puede estimar de forma precisa la proximidad de los resultados muestrales a los de la población. La necesidad de seleccionar las muestras al azar se ilustra en el siguiente ejemplo:

Un investigador de un hospital comarcal estimó que la tasa de infarto de miocardio de su comunidad era de 150 por 100 000 personas-año. Otro investigador de un hospital privado de la misma comunidad estimó que dicha tasa era de 155 por 100 000 personas-año. Dado que sus resultados eran similares, los investigadores concluyeron que la tasa de infarto de miocardio en su comunidad debía situarse entre 150 y 155 por 100 000 personas-año.

En ninguno de los dos estudios se intentó obtener una muestra al azar de la población. Es posible que los pacientes de infarto de miocardio hubiesen escogido selectivamente esos dos hospitales o los hubiesen evitado. Si se hubieran incluido en la muestra todos los hospitales de la zona, la tasa de infarto de miocardio podría haber sido totalmente distinta. Las tasas en este ejemplo se calcularon a partir de datos disponibles, y esto se conoce como *muestreo fortuito* (*chunk sampling*). Este es el tipo de muestreo más simple, dado que los investigadores calculan las tasas a partir de datos fácilmente disponibles. Sin embargo, la falta de representatividad de las muestras fortuitas implica que los resultados obtenidos a partir de las mismas pueden no ser fiables o fácilmente extrapolables a la población.<sup>4</sup>

Cuando se extrae una muestra al azar, los investigadores suelen procurar que todos los individuos de la población tengan la misma probabilidad de ser seleccionados para inclusión en la muestra. Esto se conoce como *muestreo aleatorio simple* (*simple random sampling*). Muchos investigadores han observado que, si se basan exclusivamente en el muestreo aleatorio, no conseguirán incluir suficientes individuos que posean la característica de interés para el estudio. Por ejemplo, si los investigadores estudian las tasas de hipertensión en los Estados Unidos, es posible que estén interesados especialmente en las tasas de hipertensión de las personas de raza negra o de las orientales. Si simplemente extraen una muestra aleatoria, es posible que no incluyan un número suficiente de orientales o de negros. Por lo tanto, los investigadores podrían extraer muestras por separado de los negros, los orientales y el resto de la población. Este procedimiento de obtención por separado de muestras al azar de los diferentes subgrupos o estratos se conoce como *muestreo aleatorio estratificado* (*stratified random sampling*). Esto es permisible, y muchas veces deseable, siempre que el muestreo dentro de cada grupo sea aleatorio. Existen métodos estadísticos distintos para las muestras estratificadas.

Revisemos los principios y los requisitos del muestreo:

<sup>4</sup> No obstante, en ocasiones nos vemos obligados a realizar extrapolaciones a partir de muestras fortuitas. El ejemplo más común ocurre cuando deseamos extrapolar observaciones de investigaciones a los pacientes que vemos más tarde. El paso del tiempo es un aspecto de una población que no se puede muestrear aleatoriamente

1. En promedio, las muestras aleatorias de una población tendrán la misma tasa que la población original. Sin embargo, existe un error muestral intrínseco introducido al incluir solamente una parte de la población.

2. La magnitud del error muestral está influido por el tamaño de la muestra obtenida. El aumento del tamaño muestral reduce la magnitud del error muestral, pero el aumento de la precisión desciende a medida que incrementamos el tamaño de la muestra.

3. Los principios del muestreo se basan en el supuesto de que las muestras se obtienen al azar. Mediante el muestreo estratificado es posible garantizar un número suficiente de casos en cada categoría de interés. No obstante, el muestreo debe ser aleatorio en cada categoría o estrato. Si no se realiza un muestreo al azar, no existe ningún método que relacione con precisión la tasa obtenida en la muestra con la verdadera tasa de la población de la que se ha extraído.

# ESTANDARIZACIÓN DE TASAS

En el capítulo anterior esbozamos los requisitos para calcular tasas con exactitud mediante las técnicas de muestreo aleatorio. Ahora trataremos de comparar las tasas calculadas. Supondremos que las tasas se han calculado correctamente y mostraremos las precauciones que deben tomarse al compararlas, incluso las que se han calculado mediante técnicas de muestreo aleatorio adecuadas.

Compararemos las tasas de muestras extraídas de dos grupos diferentes. Estos grupos pueden ser dos hospitales, dos condados, dos países, dos fábricas, o el mismo hospital, ciudad o fábrica comparado en dos momentos distintos en el tiempo. Compararemos las tasas para determinar la magnitud de las diferencias entre las tasas de las poblaciones o el grado de cambio de las tasas con el tiempo. Estas comparaciones son importantes para las tasas que se calculan a partir de datos obtenidos por muestreo, así como para las calculadas a partir de toda la población.

Cuando se utilizan tasas para comparar probabilidades o riesgos de enfermedad es importante considerar si las poblaciones difieren en algún factor que se sabe que influye en el riesgo de contraer la enfermedad. Esta consideración corresponde al ajuste según las variables de confusión discutido anteriormente.

Es posible que al realizar un estudio, el investigador ya sepa que factores como la edad, el sexo o la raza influyen en el riesgo de desarrollar una enfermedad determinada. En este caso, el investigador ajustaría o "estandarizaría" las tasas según esos factores. La importancia de la estandarización se puede apreciar considerando las tasas del cáncer de pulmón. Dado que la edad es un factor de riesgo conocido para el cáncer de pulmón, se gana poco descubriendo que una comunidad de jubilados tiene una tasa de cáncer de pulmón más elevada que el resto de la comunidad. De forma similar, si una fábrica tiene una fuerza laboral más joven que otra, es erróneo comparar directamente las tasas de cáncer de pulmón de las dos fábricas, sobre todo si uno desea extraer conclusiones sobre la seguridad de las condiciones de trabajo.

Para evitar este problema, las tasas de enfermedad se pueden ajustar para tener en cuenta los factores que ya se sabe que influyen notablemente en el riesgo. Este proceso de ajuste se denomina *estandarización* (*standardization*). La edad es el factor que requiere estandarización con más frecuencia, pero podemos ajustar según cualquier factor del que sepamos que produce un efecto. Por ejemplo, al usar las técnicas de estandarización para comparar tasas de hipertensión de dos muestras, con objeto de estudiar la importancia de las diferencias en el suministro de agua, se puede ajustar según la raza, ya que se sabe que la tasa de hipertensión de las personas de raza negra es más elevada.

El principio utilizado en la estandarización de las tasas es el mismo que se emplea para ajustar según las diferencias entre los grupos de estudio, tal como se comentó en la *Parte 1, El estudio de un estudio*. Los investigadores comparan las tasas entre individuos que son similares en cuanto a la edad o a otros factores según los cuales se ajustarán los datos. Antes de mostrar un ejemplo del método, veamos cuán erróneos pueden ser los resultados si no se realiza una estandarización.

En un estudio se comparó la incidencia de cáncer de páncreas en los Estados Unidos con la de México. La tasa por 100 000 personas-año en los Estados Unidos fue tres veces mayor que la de México. Los autores llegaron a la conclusión de que los estadounidenses tenían un riesgo tres veces más elevado de padecer cáncer de páncreas que los mexicanos, suponiendo que el diagnóstico fuera igualmente exacto en ambos países.

La interpretación de este estudio es superficialmente correcta; si los datos son fiables, el riesgo de cáncer de páncreas es mayor en los Estados Unidos. Sin embargo, se sabe que el cáncer de páncreas se produce con mayor frecuencia en las personas ancianas. Es posible que el hecho de que la población mexicana sea más joven explique la diferencia entre las tasas de cáncer de páncreas. Esta puede ser una cuestión importante si estamos examinando la causa de esta enfermedad. Si la distribución de edad no explica esta diferencia, los autores habrán detectado una diferencia importante e inesperada que exige otra explicación. Por este motivo, los autores deben estandarizar sus datos según la edad y observar si persisten las diferencias.

La estandarización de las tasas frecuentemente se realiza comparando una muestra especial que se está estudiando con la población general. Para realizar este tipo de estandarización empleamos a menudo el denominado *método indirecto* (*indirect method*). Mediante este método se compara el número de sucesos observado, como las defunciones, en la muestra de interés con el número de sucesos que serían de esperar si la muestra estudiada tuviese la misma distribución de edad que la población general. Cuando la muerte es el desenlace de interés, el método indirecto permite calcular una razón conocida como *razón de mortalidad estandarizada* (*standardized mortality ratio*).

$$\text{Razón de mortalidad estandarizada} = \frac{\text{Número observado de muertes}}{\text{Número esperado de muertes}}$$

La razón de mortalidad estandarizada es un instrumento útil para comparar una muestra extraída de una población de interés con la población general. Sin embargo, cuando se interpreta esta razón es importante recordar que a menudo no se espera que una población especial en estudio tenga la misma tasa de mortalidad que la población general.

Por ejemplo, cuando se compara un grupo de empleados con la población general, se debe recordar que, en muchos casos, el estar empleado requiere estar sano o, al menos, no estar incapacitado. La necesidad de tener en cuenta este efecto del empleo se ilustra en el ejemplo que figura a continuación.

En un estudio sobre nuevos empleados de una industria química, la razón estandarizada de mortalidad por todas las causas de muerte fue 1. El investigador concluyó que, como la razón de mortalidad estandarizada era 1, la industria química no presentaba riesgos para la salud de sus trabajadores.

Para interpretar este estudio, es importante recordar que los nuevos empleados por lo común son más sanos que las personas de la población general. Por esta razón, sería de esperar que tuvieran una tasa de mortalidad más baja que la de la población general.<sup>1</sup> Por consiguiente, esta razón de mortalidad estandarizada de 1 ó 100% puede no reflejar los riesgos a que están expuestos estos trabajadores sanos.

<sup>1</sup> N del E. Este hecho se denomina *efecto del trabajador sano* (*healthy worker effect*).

**CUADRO 22-1. Comparación de tasas de cáncer de vejiga urinaria**

Grupo de edad (años)	Número de sujetos	Número de casos de cáncer de vejiga urinaria	Tasa de cáncer de vejiga urinaria en cada grupo de edad <sup>a</sup>
FÁBRICA A			
20-30	20 000	0	0 por 100 000
30-40	20 000	10	50 por 100 000
40-50	30 000	20	67 por 100 000
50-60	20 000	80	400 por 100 000
60-70	10 000	90	900 por 100 000
Total	100 000	200	200 por 100 000
FÁBRICA B			
20-30	10 000	0	0 por 100 000
30-40	10 000	4	40 por 100 000
40-50	20 000	6	30 por 100 000
50-60	50 000	140	280 por 100 000
60-70	10 000	50	500 por 100 000
Total	100 000	200	200 por 100 000

<sup>a</sup> La tasa se obtiene a partir del número de casos y el número de sujetos en el grupo de edad. Las tasas no se pueden sumar a lo largo de la columna.

Cuando se estudian dos grupos de una población o cuando se valoran los *cambios temporales* de una sola población, es preferible utilizar el método directo de estandarización. El método directo funciona de la siguiente manera: Supongamos que los investigadores desean comparar las tasas de cáncer de vejiga urinaria en dos grandes industrias. Los datos del cáncer de vejiga en las dos industrias se presentan en el cuadro 22-1. Observe que las tasas globales de ambas muestras son de 200 por 100 000 trabajadores. Note también que las tasas de cada grupo de edad son tan elevadas o más en la fábrica A que en la B. A pesar de que las tasas de la fábrica B son más bajas, en un primer momento puede parecer sorprendente que las tasas globales sean las mismas. No obstante, examinando el número de individuos en cada grupo de edad, es obvio que la fábrica A tiene una fuerza de trabajo más joven que la B. La fábrica B tiene 60 000 trabajadores de 50 a 70 años de edad, mientras que la A solo tiene 30 000 en esos grupos de edad. Como se sabe que el cáncer de vejiga aumenta con la edad, la juventud de la fuerza laboral de la fábrica A reduce la tasa global de esa empresa. Por este motivo, es erróneo examinar solo las tasas globales, ya que la de la industria B se eleva como consecuencia de una estructura de edad más madura. Esto es especialmente aplicable cuando la seguridad del medio ambiente de la fábrica es la cuestión de mayor interés.

Para evitar este problema, los autores deben estandarizar las tasas para ajustarlas según las diferencias de la edad y, de este modo, hacer una comparación más correcta. Para realizar la estandarización, se subdivide cada muestra para indicar el número de individuos, el de casos de la enfermedad y la tasa de incidencia en cada grupo de edad. Cuando se dividen los datos en grupos según una característica como la edad, cada división se denomina un *estrato (strata)*. El resultado de esta división se muestra en el cuadro 22-1.

**CUADRO 22-2. Método de estandarización según la edad**

Grupo de edad (años)	Tasa de cáncer de vejiga en la fábrica A	Número de sujetos en la fábrica B	Número de casos que se producirían en la fábrica A si la distribución de edades fuera igual a la de la fábrica B <sup>a</sup>	Número de casos de cáncer de vejiga que realmente se produjeron en la fábrica B
20-30	0/100 000	10 000	0	0
30-40	50/100 000	10 000	5	4
40-50	67/100 000	20 000	13	6
50-60	400/100 000	50 000	200	140
60-70	900/100 000	10 000	90	50
Total		100 000	308	200

<sup>a</sup> Los valores de esta columna se calculan multiplicando los de las columnas precedentes

A continuación, los autores han de determinar cuántos casos de cáncer de vejiga urinaria hubieran aparecido en la fábrica A si su estructura de edad fuese igual a la de la fábrica B. Seguidamente se detallan las etapas de este proceso.<sup>2</sup>

1. Empezando con el grupo de edad de 20 a 30 años, los autores multiplican la tasa de cáncer de ese grupo en la fábrica A por el número de individuos en el grupo de edad correspondiente de la fábrica B. Este producto es el número de casos que se hubieran producido en la fábrica A si hubiera tenido el mismo número de individuos en ese grupo de edad que la B.

2. A continuación, los autores efectúan este cálculo para cada grupo de edad y suman los totales de casos de los diferentes grupos. Esto produce el total de casos que habrían aparecido si la fábrica A hubiera tenido la misma distribución de edad que la B.

3. Los autores han estandarizado las tasas según la edad y ahora pueden comparar directamente el número de casos que han aparecido en la fábrica B con el de los que se habrían producido en la A si hubiera tenido la misma distribución de edad que la B. Los autores ahora ya han ajustado la fábrica A según la edad a la distribución de edades de la fábrica B.<sup>3</sup>

Vamos a aplicar este procedimiento a los datos del cáncer de vejiga urinaria, tal como se muestra en el cuadro 22-2.

En la fábrica A se habrían producido 308 casos de cáncer de vejiga urinaria si su distribución de edad hubiera sido la misma que en la fábrica B, pero en realidad en la B solo se produjeron 200. Estos resultados constituyen mejores medidas para comparar el riesgo de los trabajadores de desarrollar cáncer de vejiga en cada industria que las frecuencias no ajustadas. Las cifras ajustadas acentúan el hecho de que, a pesar de la igualdad de las tasas globales, la fábrica A tiene una tasa igual o más alta en cada grupo de edad. Por lo tanto, para realizar comparaciones equitativas entre po-

<sup>2</sup> El método que se presenta no es necesariamente el único o el mejor para realizar la estandarización. Por razones estadísticas, es habitual ponderar los estratos según la inversa de la varianza del estimador en cada estrato, como se realiza cuando se utiliza el método de Mantel-Haenszel.

<sup>3</sup> También es posible ajustar según la edad en la dirección opuesta; es decir, ajustar según la edad la fábrica B a la distribución de edad de la fábrica A. Aunque la conclusión general hubiese sido la misma, las estimaciones habrían sido distintas.

**CUADRO 22-3. Comparación de las tasas de mortalidad por fibrosis quística, 1969 y 1989**

Grupo de edad (años)	Tasa de mortalidad	Población	Número de defunciones
1969			
0-10	5/100 000	1 000 000	50
10-20	10/100 000	1 000 000	100
20-40	1/100 000	<u>2 000 000</u>	<u>20</u>
		4 000 000	170
1989			
0-10	3/100 000	1 000 000	30
10-20	6/100 000	1 000 000	60
20-40	4/100 000	<u>2 000 000</u>	<u>80</u>
		4 000 000	170

blaciones que difieren en su estructura de edad y en las que se sabe que la edad influye en el riesgo de padecer la enfermedad, es preciso estandarizar los resultados según esta variable. Si se conocen otros factores que influyen en las tasas, se puede aplicar el mismo proceso para estandarizar o ajustar según esos factores.

La estandarización produce medidas resumidas de grandes cantidades de datos, razón por la cual es tentador estandarizar los datos siempre que se comparan dos grupos.

No obstante, observe que, al estandarizar, los cálculos dan un peso mayor o "ponderan" a los subgrupos más numerosos. Por eso, cuando se produce un cambio importante en un solo subgrupo, especialmente si es pequeño, este efecto puede ser ocultado por el proceso de estandarización. Además, a veces puede que se haya logrado progresar en la disminución de la mortalidad por una causa en los grupos más jóvenes, con aumento en los de más edad. Como se puede observar en el siguiente ejemplo, cuando la mortalidad por una causa determinada se desplaza hacia los grupos de edad más avanzada, el efecto puede quedar encubierto por la estandarización.

Se realizó un estudio sobre las tasas de mortalidad por fibrosis quística, para determinar si los progresos en su tratamiento durante la niñez se reflejaban en las tasas de mortalidad de un estado de gran tamaño con una distribución de población estable. En el cuadro 22-3 se presentan los datos de 1969 y 1989.

Observe que en este ejemplo la mortalidad por fibrosis quística en los grupos de edad de 0 a 10 años y de 10 a 20 años descendió entre 1969 y 1989. Sin embargo, la mortalidad por esta enfermedad aumentó en el grupo de 20 a 40 años durante el mismo período. Este aumento se contrapesa con el descenso entre los más jóvenes, de forma que las tasas de mortalidad global en ambos años fueron 170 por 4 000 000 ó 4,25 por 100 000 personas-año. Resulta tentador intentar estandarizar esos datos y obtener una medida ajustada de la tasa de mortalidad. Sin embargo, si estandarizamos aplicando la distribución de edad de 1989 a las tasas de mortalidad de 1969 (o viceversa), los resultados después del ajuste no serían distintos de los anteriores al ajuste. Esto sucede porque las distribuciones de edad de las poblaciones de 1969 y de 1989 son iguales.

Lamentablemente, la estandarización no nos ayuda a apreciar lo que ocurre en este caso. Tanto las tasas brutas o no ajustadas como las estandarizadas

oscurecen el hecho de que se ha producido un descenso importante de las tasas de mortalidad en los grupos de 0 a 10 y de 10 a 20 años de edad. Para darse cuenta de este cambio, es preciso examinar directamente los datos reales de cada grupo de edad.<sup>4</sup>

Es importante darse cuenta de que tanto las tasas brutas como las ajustadas pueden no revelar diferencias o cambios que solo se producen en uno o en pocos grupos de edad. Sobre todo, es probable que los cambios en un grupo pasen desapercibidos cuando los otros grupos de edad cambian en la dirección opuesta.

Una situación en la que se observan con frecuencia cambios en dirección opuesta es el retraso de la muerte hasta una edad más avanzada sin que se logre la curación. Es importante entender este principio para valorar el error cometido en el siguiente estudio:

Un investigador estudió un nuevo tratamiento para el cáncer de mama, que en promedio prolongaba 5 años la supervivencia en el estadio 2 de la enfermedad. Con toda confianza predijo que, si su tratamiento se aplicaba ampliamente, la tasa global de cáncer de mama descendería sobremanera en los siguientes 20 años.

Este investigador no se dio cuenta de que, cuando se prolonga la vida pero la muerte se retrasa a edades más avanzadas, las tasas de mortalidad globales ajustadas según la edad no mejoran necesariamente. A pesar del éxito de este nuevo tratamiento, los autores no están afirmando que curará la enfermedad. Cuando solo se prolonga la vida, los enfermos pueden morir de la enfermedad a una edad más avanzada. En este caso, las tasas de mortalidad por cáncer de mama pueden descender en los grupos de edad más jóvenes y aumentar en los de edad más avanzada. Por esta razón, las tasas de mortalidad ajustadas según la edad a veces no revelan los progresos realizados.

---

<sup>4</sup> Además, es posible aplicar métodos estadísticos con los que se obtienen medidas conocidas como la *esperanza de vida (life expectancy)*. La esperanza de vida tiene en cuenta el impacto del aumento de años de vida vividos incluso en ausencia de una curación. Esta medida se obtiene a partir de tablas de vida transversales que valoran la probabilidad hipotética de supervivencia del individuo promedio que alcanza una edad determinada. Suponiendo una experiencia de mortalidad estable, la esperanza de vida se basa en la experiencia de la mortalidad de una población concreta, como la de los Estados Unidos, en un año determinado.