

Cómo estudiar un estudio y probar una prueba: lectura crítica de la literatura médica¹

Segunda edición

Richard K. Riegelman y Robert P. Hirsch

PARTE V:

- Capítulo 12. Ejercicios para detectar errores:
ensayos clínicos controlados
- Capítulo 13. Introducción a la prueba de una prueba
- Capítulo 14. Variabilidad de una prueba
- Capítulo 15. El intervalo de lo normal

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: ENSAYOS CLÍNICOS CONTROLADOS

Los siguientes ejercicios están diseñados para poner a prueba su habilidad de aplicar los principios del ensayo clínico controlado. Lea cada ejercicio para detectar errores y luego escriba una crítica señalando los tipos de errores que aparecen en cada componente del marco uniforme.

SANGRE SEGURA. UN NUEVO TRATAMIENTO PARA PREVENIR EL SIDA: EJERCICIO NO. 1

Un investigador creyó que había descubierto un método mejor para evitar el peligro de diseminar el síndrome de la inmunodeficiencia adquirida (SIDA) mediante las transfusiones de sangre, matando el virus en las células transfundidas. Su método exigía tratar a todos los receptores de las transfusiones con un nuevo fármaco llamado "Sangre Segura". En el momento de su descubrimiento, la tasa de transmisión del SIDA debida a transfusiones era de 1 por 100 000 transfusiones.

Una vez conseguida la aprobación del estudio del fármaco en seres humanos, diseñó un ensayo clínico controlado para su uso inicial. En el estudio se preguntó a una muestra aleatoria de todos los receptores de transfusiones en una gran área metropolitana si deseaban recibir el fármaco en las dos semanas siguientes a la transfusión.

El grupo de estudio quedó constituido por 1 000 individuos que aceptaron el tratamiento, y el de control, por 1 000 individuos que lo rehusaron. Los integrantes del grupo de control habían recibido una media de tres transfusiones de sangre por cada 1,5 recibida por los tratados con Sangre Segura. Los investigadores lograron realizar pruebas serológicas para seguimiento del virus de la inmunodeficiencia humana (VIH) en 60% de los que recibieron el fármaco y en 60% de los que lo rechazaron, aproximadamente un mes después de recibir la transfusión.

Los que realizaron dichas pruebas de seguimiento no sabían cuáles pacientes habían recibido o no Sangre Segura. El investigador observó que un individuo del grupo de estudio pasó a ser positivo a la prueba de detección de anticuerpos contra el VIH al mes siguiente de iniciar el tratamiento con Sangre Segura. En el grupo de control dos individuos pasaron a ser positivos a la prueba.

El investigador no detectó ningún efecto indeseable atribuible a Sangre Segura durante el período de seguimiento de un mes. Concluyó, por lo tanto, que el estudio había demostrado que el fármaco era efectivo y seguro, y aconsejó su administración a todos los receptores de transfusiones de sangre.

CRÍTICA: SANGRE SEGURA. UN NUEVO TRATAMIENTO PARA PREVENIR EL SIDA

Diseño del estudio

El investigador intentó realizar un ensayo clínico controlado. Este tipo de estudio es el más indicado para valorar la eficacia de un tratamiento, una vez

que la dosis y el método de administración se han determinado mediante estudios iniciales en seres humanos. Estos estudios no están indicados para realizar investigaciones iniciales en seres humanos.

En el momento del estudio, el riesgo de transmisión del SIDA a través de las transfusiones de sangre era de 1 por 100 000, un riesgo muy bajo. Ya que los ensayos clínicos controlados están destinados a reducir un riesgo ya bajo, es preciso reunir a un número muy elevado de individuos. Se necesitarían millares o incluso millones de individuos para realizar de forma apropiada un ensayo clínico controlado, cuando el riesgo, sin tratamiento, es de 1 por 100 000. Un estudio de esta magnitud no tiene un poder estadístico adecuado. En otras palabras, ese estudio no sería capaz de demostrar una significación estadística para el tratamiento, incluso suponiendo que Sangre Segura fuese capaz de reducir considerablemente la incidencia del SIDA asociado con las transfusiones sanguíneas, por ejemplo, de 1 por 100 000 a 1 por 1 000 000.

Asignación

El investigador identificó una muestra aleatoria de pacientes comparables con los que podrían recibir un tratamiento efectivo. La selección al azar no es un requisito de los ensayos clínicos controlados, pero aumenta la fiabilidad de la extrapolación a los miembros de la población de la que se extrajo la muestra y que no fueron incluidos en el ensayo.

El investigador no asignó al azar a los pacientes a los grupos de estudio y de control. El grupo de control estaba formado por pacientes que rechazaron la administración de Sangre Segura. Este no es un grupo de control ideal, porque los que rehusaron participar podían haber sido diferentes de los que aceptaron, en cuanto a diversos aspectos relacionados con la posibilidad de contraer la infección por el VIH. La asignación al azar, en contraposición a la selección al azar, se considera una característica fundamental de los ensayos clínicos controlados. Por consiguiente, este estudio no fue un verdadero ensayo clínico controlado.

Valoración

Los que valoraron el desenlace de este estudio no sabían cuáles pacientes habían recibido Sangre Segura. Esta valoración a ciegas es objetiva y contribuye a prevenir el sesgo en el proceso de valoración. Sin embargo, la ausencia de enmascaramiento en el proceso de la asignación significa que los pacientes sabían si habían recibido Sangre Segura o no. Esto pudo haber tenido un efecto sobre el desenlace del estudio suponiendo, por ejemplo, que quienes recibieron Sangre Segura creían que estaban protegidos contra el SIDA.

Los investigadores realizaron la prueba de detección de anticuerpos contra el VIH un mes después de que los pacientes recibieran una transfusión. Este período es demasiado corto para valorar adecuadamente si se ha producido o no en un individuo la conversión al estado positivo.

El elevado número de pacientes de los grupos de estudio y de control que se perdieron en el seguimiento constituyó un problema importante al efectuar la valoración, aunque los porcentajes de pérdidas fueron iguales en ambos grupos. Cuando el número de desenlaces adversos es bajo, los sujetos perdidos en el seguimiento son especialmente importantes, porque estos individuos pueden experimentar de forma desproporcionada efectos secundarios o manifestar síntomas a pesar del tratamiento.

Análisis

El investigador no informó sobre pruebas de significación estadística ni intervalos de confianza. Si lo hubiera hecho, no habría sido capaz de demostrar la existencia de una diferencia estadísticamente significativa. Esto no es sorprendente, ya que, con un solo caso más de infección por el VIH, los desenlaces hubieran sido iguales en los grupos de estudio y de control.

En este estudio, el intervalo de confianza hubiera sido muy amplio, indicando que los resultados eran compatibles con la ausencia de diferencias o incluso con una diferencia en la dirección opuesta.

El mayor número de transfusiones de sangre recibidas por los que rehusaron tomar Sangre Segura podría ser una variable de confusión que se debió tener en cuenta mediante un proceso de ajuste en el análisis. El número de transfusiones de sangre es una variable de confusión, dado que es diferente en los dos grupos y está relacionado con el riesgo de desarrollar infecciones por el VIH secundarias a las transfusiones.

Interpretación

Los problemas señalados en el diseño, la asignación, la valoración y el análisis indican que el estudio se debe interpretar con mucha cautela.

El resultado de las pruebas de significación estadística y de los intervalos de confianza implicaría que la diferencia de infecciones por VIH entre los grupos de estudio y de control podrían ser debidas al azar.

El riesgo de desarrollar una infección por el VIH a partir de una transfusión de sangre sin la administración de Sangre Segura es tan pequeño que sería mucho más probable contraer el virus de otra forma. Por lo tanto, ninguna diferencia entre el grupo de estudio y el de control puede atribuirse automáticamente a la administración de Sangre Segura. La diferencia de un grupo a otro en la infección por el VIH puede deberse a diferencias entre otros factores de riesgo del SIDA. No se presentan datos que permitan analizar esos factores, que quizá sean mucho más importantes que las transfusiones de sangre.

Extrapolación

Aunque se demostrara que Sangre Segura es eficaz para prevenir las infecciones por el VIH asociadas con la transfusión, este estudio no permitiría extraer conclusiones acerca de su efectividad o seguridad.

Los ensayos clínicos controlados pueden llevar a conclusiones sobre la eficacia de un tratamiento en las condiciones ideales de una investigación. En cambio, la efectividad implica que el tratamiento ha sido beneficioso en las condiciones habituales de la práctica clínica.

El empleo de Sangre Segura en el medio clínico implicaría su administración a un gran número de individuos. Por eso, los efectos secundarios graves serían importantes por muy raros que fueran. La ausencia de estos efectos entre los que recibieron Sangre Segura no anula la posibilidad de que aparezcan otros efectos poco frecuentes pero graves. De acuerdo con la regla de tres, si se produce un efecto secundario en 1 de cada 1 000 usos, se debe observar a 3 000 individuos para tener una seguridad de 95% de observar al menos un caso de dicho efecto.

Al extrapolar los resultados de un ensayo clínico controlado al uso de un tratamiento en la práctica clínica, es necesario considerar como mínimo lo siguiente:

1. Si el estudio demuestra la eficacia del tratamiento en condiciones ideales.
2. Si los individuos estudiados son similares a los que recibirán el tratamiento.
3. Si los riesgos conocidos o la posibilidad de efectos secundarios raros pero graves no observados en los ensayos clínicos controlados superan los beneficios potenciales.

A veces, también puede ser importante considerar el costo del tratamiento comparado con el de otras opciones.

EJERCICIO NO. 2: VACUNA DE LA GRIPE

Para probar una nueva vacuna contra la gripe, se realizó un ensayo clínico aleatorio. Los participantes del grupo de estudio se eligieron seleccionando al azar a 1 000 familias de una lista de familias de voluntarios para el ensayo. La vacuna se administró a los 4 000 integrantes de las 1 000 familias. Como grupo de control, los investigadores seleccionaron individuos al azar de la guía de teléfonos hasta que 4 000 personas aceptaron la propuesta de recibir una vacuna placebo. Al comparar los grupos, los investigadores observaron que la media de edad del grupo de estudio era de 22 años y la del de control, 35. La posibilidad de que los sujetos poseyeran un termómetro fue dos veces más alta en el grupo de estudio que en el de control, pero, por lo demás, los grupos eran similares cuando se compararon de acuerdo con una larga lista de variables.

Durante la época de gripe del invierno siguiente, se dieron instrucciones a cada individuo de que acudiese a uno de los médicos investigadores siempre que tuviera fiebre, para evaluar clínicamente la posibilidad de que padeciera gripe. Los sujetos del grupo de estudio fueron asignados a un consultorio y los del grupo de control, a otro. Al hacer su evaluación de seguimiento, los investigadores obtuvieron una participación de 95% en el grupo de estudio y de 70% en el de control. Además, observaron que en el grupo de control se produjeron 200 casos de gripe por cada 1 000 personas vacunadas con placebo y seguidas, y 4 casos por 1 000 personas vacunadas en el grupo de estudio. Por último, concluyeron que el nuevo tratamiento reducía el riesgo de gripe a 2% de su tasa previa y recomendaron su administración a la población general para evitar 98% de las muertes causadas por la gripe debida a esta cepa del virus.

CRÍTICA: EJERCICIO NO. 2

Diseño del estudio

Los investigadores no definieron la muestra de estudio que estaban empleando para someter a prueba el nuevo tratamiento. No queda claro si lo estaban probando en familias o en individuos, en sujetos de un grupo de edad o de todas las edades, en voluntarios o en la población general.

Asignación

En los ensayos clínicos aleatorios los investigadores asignan al azar, e idealmente a ciegas, voluntarios al grupo de control y al de estudio. En el caso que nos ocupa, todos los voluntarios fueron asignados al grupo de estudio, se les administró el

tratamiento experimental y se extrajo una muestra separada para constituir el grupo de control. Por lo tanto, la asignación no fue al azar ni a ciegas.

Al vacunar a todos los miembros de una familia se abre la posibilidad de que un tratamiento con éxito parcial parezca casi totalmente efectivo. Si se reduce el riesgo de un miembro concreto de la familia, el riesgo de exposición de los demás miembros de la familia se reduce en gran parte, dado que la exposición familiar es una fuente de transmisión importante. De esta forma, dos factores favorecen la vacunación: la reducción de la exposición y el incremento de la inmunidad.

Los investigadores emplearon una población diferente para seleccionar el grupo de control. Al seleccionar los nombres de la guía telefónica limitaron su muestra a los individuos que aparecían en ella. Con este método no se formó un grupo de control equiparable al de los que recibieron el tratamiento, especialmente en lo que se refiere a los niños. Es probable que estos estén ampliamente representados en una población de familias, pero no en la guía de teléfonos.

Existe la posibilidad de que la distribución de características potencialmente relacionadas con la susceptibilidad a la gripe no sea similar en los grupos de estudio y de control, aunque se obtenga una muestra al azar de la población. Es necesario tener en cuenta las diferencias de edad entre los grupos de estudio y de control (que son de esperar, dado el método de asignación). La diferencia en la frecuencia de posesión de termómetros, aunque sea debida al azar, es importante, porque puede influir en el reconocimiento de la fiebre. Un mayor reconocimiento de la fiebre probablemente resulte en un aumento del número de casos de gripe diagnosticados. Recuerde que la selección al azar de los participantes es un rasgo deseable pero no habitual en un ensayo clínico aleatorio. Sin embargo, la asignación al azar es esencial, porque es la característica distintiva de un ensayo clínico aleatorio.

Valoración

Es probable que el método escogido por los investigadores para diagnosticar la gripe no fuera válido. La gripe es difícil de diagnosticar específicamente y en el estudio no se definieron los criterios estándares para su diagnóstico, como el cultivo de los virus. También, es probable que en los voluntarios el umbral de enfermedad antes de acudir al médico fuera distinto del de los no voluntarios; este hecho, aunado al bajo número de sujetos del grupo de control que poseían termómetros, pudo haber influido en el número de diagnósticos de gripe realizados. El hecho de que los sujetos de los grupos de estudio y de control fueron seguidos en consultorios diferentes sugiere que no se efectuó el enmascaramiento y ello pudo haber influido en la frecuencia de diagnósticos de gripe.

Finalmente, si el seguimiento no incluye a todos los participantes, existe la posibilidad de que los sujetos perdidos en el seguimiento tengan un mejor o peor desenlace que los que pudieron ser seguidos. Una alta proporción de integrantes del grupo de control se perdieron en el seguimiento, lo cual pudo haber influido en la validez de la valoración del desenlace. En general, la valoración del desenlace en este estudio no fue válida.

Análisis

Como la distribución de la edad en el grupo de estudio era diferente de la del grupo de control y la edad es un factor que influye frecuentemente en la susceptibilidad a la infección, es importante ajustar los datos según el efecto de la edad.

Esto se podría haber realizado comparando la tasa de ataque de los individuos de la misma edad para ver si las diferencias entre los vacunados y los no vacunados seguían siendo las mismas. Además, no se aplicó ninguna prueba de significación estadística para determinar cuán probables eran las diferencias observadas, si no existían verdaderas diferencias en las poblaciones de las que se extrajeron las muestras.

Interpretación y extrapolación

Este estudio no consigue cumplir con los estándares mínimos de la asignación, la valoración y el análisis, lo que significa que se debe interpretar con suma cautela. Hay muchas razones, distintas de la de relación causal, que pueden explicar el bajo número de diagnósticos de gripe realizados en el grupo de estudio.

La extrapolación a otras poblaciones requiere una prueba convincente de la existencia de una relación entre los sujetos estudiados. En este estudio faltó la prueba de una relación causal. Aunque se demuestre la existencia de una relación, no se pueden efectuar extrapolaciones de las tasas de ataque de la gripe a las tasas de mortalidad. El estudio no tenía como objetivo estimar la tasa de mortalidad y no proporcionó ninguna prueba de que las tasas de mortalidad de las dos poblaciones fuesen diferentes.

A pesar de que la investigación presenta numerosos problemas de diseño y de que los investigadores extrapolaron mucho más allá de sus datos, es importante darse cuenta de la magnitud del efecto en el grupo de estudio. Una reducción del número de casos de gripe de 98% constituye un efecto muy notable. Este hallazgo exige un análisis riguroso, a pesar de la mala calidad del diseño de la investigación realizada. Si bien es importante, muchas veces es difícil separar la calidad del tratamiento de la calidad de la investigación.

INTRODUCCIÓN A LA PRUEBA DE UNA PRUEBA

El diagnóstico médico puede contemplarse como un intento de tomar las decisiones idóneas manejando información insuficiente. Así, la incertidumbre intrínseca al diagnóstico médico procede de la necesidad de realizar diagnósticos basados en datos inciertos. Los instrumentos diagnósticos empleados en medicina se han considerado tradicionalmente como un medio de reducir la incertidumbre en el diagnóstico. Sin embargo, para utilizar con éxito las pruebas diagnósticas, se debe saber valorar no solo la forma cómo las pruebas reducen la incertidumbre sino también cómo describen y cuantifican la incertidumbre restante.

En épocas pasadas, los instrumentos diagnósticos estaban limitados en gran parte a la historia clínica y al examen físico. Actualmente, estos todavía son potentes instrumentos de diagnóstico. Sin embargo, hoy día se dispone, además de los métodos convencionales, de una tecnología auxiliar con la cual el clínico cuidadoso puede realizar diagnósticos precisos, si la emplea apropiada y selectivamente. La esencia de la práctica de la medicina diagnóstica está constituida por el aprendizaje de cuándo, si es conveniente, debe aplicarse cada elemento de la historia, del examen físico y de la tecnología auxiliar.

El énfasis actual en la calidad con consciencia del costo requiere que los médicos entiendan los principios fundamentales de las pruebas diagnósticas: cuáles son las preguntas que pueden responder y cuáles no, cuáles son las pruebas que aumentan la precisión diagnóstica y cuáles simplemente incrementan el costo.

Para parafrasear a Will Rogers, los médicos tradicionalmente han pensado que nada es cierto, excepto la biopsia y la autopsia. Sin embargo, incluso estos criterios de referencia para el diagnóstico (*gold standards*) pueden errar el objetivo o realizarse demasiado tarde para ser de ayuda alguna. El conocimiento de los principios de las pruebas diagnósticas contribuye a definir el grado de incertidumbre diagnóstica y a aumentar la certeza. Saber cómo vivir con la incertidumbre es una característica central del juicio clínico. El médico habilidoso ha aprendido cuándo debe asumir riesgos para aumentar la certeza y cuándo debe, simplemente, tolerar la incertidumbre.

El principio fundamental de las pruebas diagnósticas reside en la creencia de que los individuos que tienen una enfermedad son distintos de los que no la tienen y que las pruebas diagnósticas permiten distinguir a los dos grupos. Las pruebas diagnósticas, para ser perfectas, requerirían que 1) todos los individuos sin la enfermedad en estudio tuvieran un valor uniforme en la prueba, 2) que todos los individuos con la enfermedad tuvieran un valor uniforme pero distinto en la prueba y 3) que todos los resultados de las pruebas fueran consistentes con los resultados del grupo de los enfermos y del de los sanos (figura 13-1).

Si esta fuera la situación en el mundo real, la prueba perfecta podría distinguir la enfermedad de la salud, y el trabajo del médico consistiría únicamente en solicitar la prueba "adecuada". El mundo real, sea para bien o para mal, no es tan simple. Habitualmente, ninguna de estas condiciones está presente. Existen variaciones en cada uno de los siguientes factores básicos: las pruebas, el grupo de enfermos y el de sanos (figura 13-2).

FIGURA 13-1. Condiciones necesarias de una prueba diagnóstica perfecta

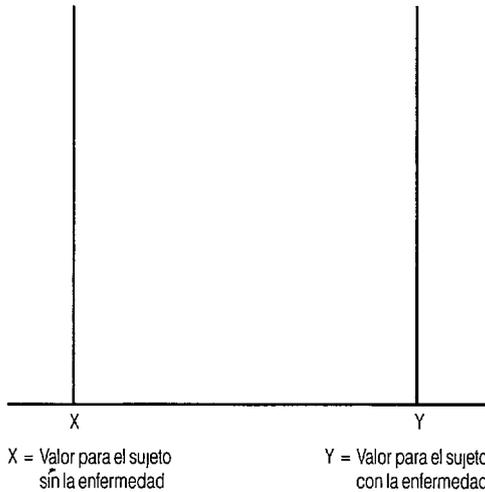
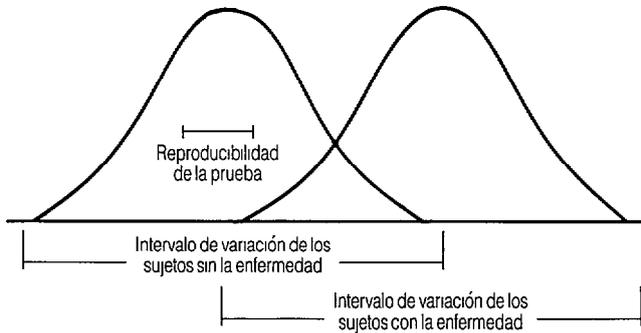


FIGURA 13-2. Los tres tipos de variaciones que afectan a las pruebas diagnósticas

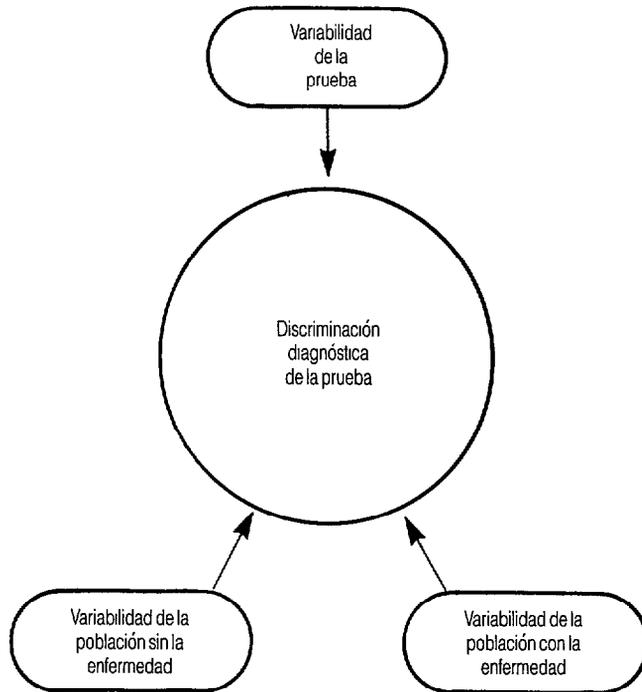


La valoración de las pruebas diagnósticas consiste en gran medida en describir la variabilidad de estos tres factores, y por esa razón se pueden cuantificar las conclusiones, a pesar o a causa de esa variabilidad.

La variabilidad, la reproducibilidad y la exactitud de las pruebas se presentan en el capítulo 14. En el capítulo 15 se revisa y valora la variabilidad de la población de las personas sanas empleando el concepto del intervalo de la normalidad. En los capítulos 16 y 17 se cuantifica la variabilidad de la población de personas enfermas y su relación con la de las sanas por medio de los conceptos de sensibilidad, especificidad y valor predictivo. En esos capítulos se esbozan estos conceptos y se muestran ejemplos de los errores que se cometen al aplicarlos. Seguidamente, se incluyen varios ejercicios para detectar errores, que ofrecen la oportunidad de aplicar esos principios a la evaluación de las pruebas diagnósticas.

Al igual que con el análisis de un estudio, es útil tener una visión panorámica o marco general de la evaluación de una prueba. Este marco se ilustra en la figura 13-3, en la que se representa la variabilidad que existe en las pruebas, en la población sana y en la enferma. También se subraya que estas variaciones deben ser estudiadas e incorporadas en cualquier valoración de la utilidad diagnóstica de una prueba.

FIGURA 13-3. Marco uniforme de la prueba de una prueba



VARIABILIDAD DE UNA PRUEBA

Una prueba perfecta produciría los mismos resultados cada vez que se aplicara en las mismas condiciones. Además, sus mediciones reflejarían exactamente el fenómeno que la prueba intenta medir. En otras palabras, una prueba perfecta sería completamente reproducible y exacta. Definamos estos términos y veamos cómo se utilizan.

La *reproducibilidad* es la capacidad de una prueba para producir resultados consistentes cuando se repite en las mismas condiciones y se interpreta sin conocer sus resultados previos. Sin embargo, en la reproducibilidad de una prueba pueden influir diversos factores.

1. Las condiciones del paciente y del laboratorio bajo las que se realiza la prueba pueden no ser las mismas.
2. La prueba puede estar influida por variaciones de interpretación entre observadores. Este efecto se conoce como *variabilidad interobservador*.
3. La prueba puede verse afectada por variaciones en la interpretación que realiza la misma persona en diferentes momentos. Este efecto se conoce como *variabilidad intraobservador*.

Para valorar el rendimiento de la prueba, los investigadores deben estar seguros de que las condiciones técnicas y biológicas de su realización son idénticas cuando se repite. El no seguir esta precaución produce el error que se muestra en el siguiente ejemplo.

Con el fin de evaluar la reproducibilidad de una prueba para medir la concentración sérica de cortisol se extrajeron dos muestras de sangre de los mismos individuos. La primera muestra se extrajo a las 8 de la mañana y la segunda, al mediodía. Los métodos fueron idénticos y la interpretación se realizó sin conocimiento previo de los resultados de la primera prueba. Este método se aplicó a 100 individuos seleccionados al azar. Los autores observaron que los valores de la segunda prueba eran en promedio el doble de la primera prueba aplicada al mismo individuo. Concluyeron que esta gran variación indicaba que la prueba no era reproducible.

Recuerde que la reproducibilidad es la capacidad de la prueba para producir casi los mismos resultados cuando se realiza en las mismas condiciones. En este ejemplo, los investigadores no repitieron la prueba en las mismas condiciones. La concentración de cortisol tiene un ciclo natural durante el día. Al extraer una muestra de sangre a las 8 de la mañana y otra al mediodía, estaban obteniendo muestras de dos momentos distintos del ciclo. Aunque la prueba fuera perfectamente reproducible cuando se realizase en condiciones de laboratorio idénticas, las distintas situaciones de los pacientes a los que se practicó hubieran producido resultados bastante diferentes.

A menos que una prueba se repita sin conocer el resultado de la primera, la segunda lectura puede estar influida por la primera, como ilustra el siguiente ejemplo.

Un investigador, al estudiar la reproducibilidad de un análisis de orina, solicitó a un técnico de laboratorio experimentado que leyera un sedimento urinario y, sin cambiar de lugar el portaobjetos, lo volviera a leer a los cinco minutos. El investigador observó que la lectura, realizada en las mismas condiciones, produjo resultados perfectamente reproducibles.

En este ejemplo, el técnico conocía los resultados de la primera prueba y era probable que estuviese influido por la primera lectura cuando la repitió a los cinco minutos. Una medida de la reproducibilidad requiere que la segunda lectura se efectúe sin conocimiento previo de la primera. Por esta razón, el técnico no debió haber conocido los resultados de la lectura anterior.

Aunque los observadores no sean conscientes de sus propias lecturas previas o de las de otros, existe la posibilidad de que la variación individual induzca a error cuando se comparan los resultados de la prueba. Siempre que hay que formar un juicio en la interpretación de una prueba, existe la posibilidad de variación inter e intraobservador. Es frecuente que dos radiólogos interpreten la misma placa de rayos X de diferente manera, lo cual se conoce como variación interobservador. Un residente puede interpretar el mismo electrocardiograma de forma diferente por la mañana que en medio de una guardia nocturna. Esto se conoce como variación intraobservador. Por sí mismas, estas variaciones no arruinan la utilidad de la prueba. Sin embargo, es necesario que el médico esté siempre alerta a la posibilidad constante de variaciones en la interpretación de los resultados de la prueba. Las inconsistencias de la técnica o de la interpretación contribuyen en cierta medida a la variabilidad de la mayor parte de las pruebas. Por lo tanto, es preciso aplicar criterios que permitan juzgar cuánta variabilidad puede tolerarse.

En general, es importante que la variabilidad de la prueba sea mucho menor que el intervalo de variabilidad debido a factores biológicos. Por esta razón, el grado de variación de la prueba debe ser pequeño en comparación con el intervalo de normalidad de la prueba (véase el capítulo 15).

Es importante distinguir la reproducibilidad de la exactitud. La *exactitud* de una prueba es la capacidad que tiene de producir resultados cercanos a la verdadera medida del fenómeno anatómico, fisiológico o bioquímico. La exactitud de una prueba requiere que esta sea reproducible y que el resultado no muestre una tendencia sistemática a diferir del verdadero valor en una dirección determinada. Cuando disparamos a un blanco, a veces erramos el tiro porque las balas se dispersan alrededor del centro. También puede haber una tendencia a situar todos los tiros a un solo lado ligeramente apartado del centro. Para ser perfectamente exactos y dar cada vez en el centro del blanco, debe existir reproducibilidad, eliminando de esta forma la dispersión. No debe existir un sesgo o una tendencia a disparar siempre hacia un lado. Por lo tanto, una prueba exacta está exenta de los efectos del azar y de errores sistemáticos o sesgos. Una prueba puede ser muy reproducible y a la vez inexacta, si reproduce valores alejados del valor verdadero. El siguiente caso ejemplifica una prueba con alta reproducibilidad pero poca exactitud, o sea, la diferencia entre reproducibilidad y exactitud.

Se realizó un estudio de 100 pacientes que habían sido operados por fracturas naviculares. De cada paciente, se obtuvieron dos radiografías que se interpretaron de forma independiente. Ambas fueron negativas para una fractura navicular en la semana posterior a la lesión. Los autores concluyeron que los radiólogos habían sido negligentes al no diagnosticar esas fracturas.

Por lo general, las fracturas naviculares no se diagnostican por medio de radiografías tomadas en el momento de la lesión. Desde el punto de vista anatómico,

la fractura existe, pero no suele detectarse en la radiografía hasta que no aparecen signos de reparación. Por consiguiente, no se trata de una negligencia de los radiólogos, sino de la inexactitud de la prueba utilizada. La repetición de las radiografías confirmó la reproducibilidad de los resultados negativos obtenidos por los dos radiólogos: no es posible identificar fracturas naviculares recientes en las radiografías. *Exactitud* significa que la prueba producirá resultados semejantes al verdadero valor anatómico, fisiológico o bioquímico. Dado que las radiografías no son siempre un reflejo exacto de la anatomía, el defecto radicaba en la prueba, no en los radiólogos.

Una prueba puede ser bastante exacta cuando se emplea en un estudio científico, pero puede perder su exactitud cuando se aplica en el medio clínico. Es útil pensar en dos tipos de exactitud: 1) exactitud experimental, es decir, la exactitud de la prueba cuando se utiliza en las condiciones especiales de un estudio, y 2) exactitud clínica, es decir, la exactitud de la prueba cuando se emplea en las situaciones clínicas reales. La diferencia entre ambos conceptos se muestra en el siguiente ejemplo.

En un hospital universitario se realizó un estudio de 500 pacientes que siguieron una dieta baja en grasas durante 3 días. Se mostró que el contenido de grasas en las heces recogidas 72 horas después de finalizar la dieta permitía distinguir bien entre los pacientes con malabsorción y los que no la padecían. Con un protocolo de estudio idéntico aplicado a 500 pacientes ambulatorios no se consiguió demostrar con éxito la presencia de malabsorción. Los autores del estudio ambulatorio concluyeron que los resultados obtenidos con los pacientes hospitalizados eran incorrectos.

El rendimiento de una prueba se valora habitualmente en condiciones experimentales ideales, pero las condiciones reales en las que se aplica muchas veces están lejos de ser ideales. En pacientes ambulatorios puede ser bastante difícil recoger las heces a las 72 horas después de una dieta de tres días de duración baja en grasas. El hecho de que los resultados de los datos de pacientes ambulatorios no concuerden con los de los pacientes hospitalizados puede reflejar simplemente la realidad de las condiciones en la práctica. La exactitud es una propiedad necesaria de una buena prueba. No obstante, la exactitud, por sí sola, no garantiza que la prueba sea válida o útil para el diagnóstico. La *validez* implica que la prueba es una medida apropiada del fenómeno estudiado. Una medida muy reproducible y exacta del tamaño de los pulmones puede proporcionar poca información válida o útil para un diagnóstico. Para establecer la utilidad diagnóstica de una prueba, necesitamos valorar la idoneidad con que la prueba distingue entre las personas sanas y las enfermas.

Antes de tratar de determinar la idoneidad de una prueba para distinguir los que no tienen una enfermedad de los que la tienen, centraremos nuestra atención en analizar cómo medimos la ausencia de enfermedad por medio del concepto del intervalo de lo normal.

EL INTERVALO DE LO NORMAL

Las poblaciones humanas sanas están sujetas a variaciones biológicas intrínsecas. Uno solo necesita pasear por la calle para apreciar las diferencias entre la gente. La altura, el peso y el color de los individuos cubren un espacio que refleja las variaciones grandes, pero no ilimitadas, que pueden existir entre individuos sanos.

En un mundo con información completa sabríamos cuál es el resultado que debe tener un individuo en una prueba determinada. Esto nos permitiría comparar el resultado obtenido en la prueba con el resultado esperado en esa persona. En la realidad, como raramente sabemos cuál debería ser el resultado en un sujeto, estamos obligados a comparar sus resultados con los de otros individuos considerados sanos. Para llevar a cabo esta comparación utilizamos un *intervalo de lo normal* o de la normalidad. El intervalo de lo normal es un mal necesario basado en la suposición de que un individuo concreto debe ser similar a otros individuos.

El concepto del intervalo de la normalidad representa un esfuerzo para medir y cuantificar el intervalo de valores que existen en individuos considerados sanos. Se puede derivar un intervalo de lo normal de cualquier medición en la que existan múltiples posibles valores numéricos para los sujetos sanos. Estos comprenden exámenes de características físicas tales como la tensión arterial, el tamaño del hígado y el pulso o valores de laboratorio como el hematócrito, la velocidad de sedimentación o la creatinina. Aunque el intervalo de medidas normales suele ser amplio, el concepto no incluye a todas las personas que no están enfermas. Este intervalo excluye a propósito a 5% de los individuos considerados sanos, con el fin de crear un intervalo de lo normal suficientemente amplio para describir la mayor parte de las mediciones de las personas sanas, pero no tan amplio como para incluir todos los posibles valores numéricos. Si el intervalo de la normalidad incluyera las mediciones de todos los individuos sin la enfermedad, sería extremadamente amplio, tan amplio que no sería útil para separar a los enfermos de los sanos. El intervalo de lo normal es descriptivo y no diagnóstico; describe a los individuos sanos, no diagnostica la enfermedad. Los valores que se encuentran fuera de ese intervalo podrían ser el resultado de la variación debida al azar, de cambios fisiológicos no asociados con la enfermedad o de cambios patológicos secundarios a la enfermedad.

CONSTRUCCIÓN DEL INTERVALO DE LO NORMAL

Los valores del intervalo de lo normal se pueden construir de la forma que se detalla a continuación:

1. El investigador localiza a un grupo de individuos que se considera que no están enfermos. Este grupo se conoce como *grupo de referencia*. Estos individuos muchas veces son estudiantes de medicina, trabajadores de hospital u otros voluntarios fácilmente accesibles. En general, simplemente se supone que no están enfermos, aunque en algunas circunstancias pueden llevarse a cabo diversas pruebas y un seguimiento para garantizarlo.

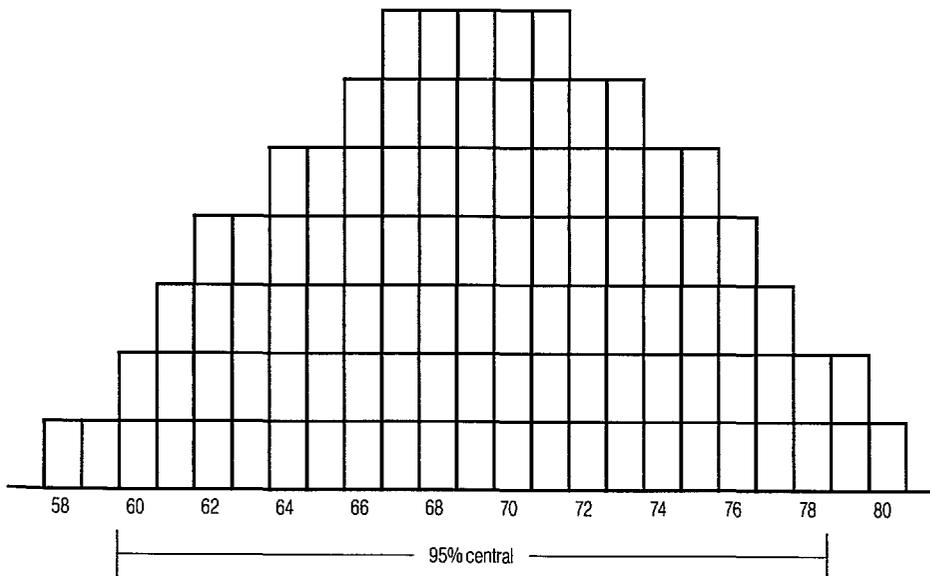
2. El investigador realiza la prueba de interés en todos los individuos del grupo de referencia.
3. Seguidamente, representa gráficamente la distribución de los valores obtenidos mediante la prueba aplicada a los individuos del grupo de referencia.
4. Luego calcula el intervalo de lo normal, que comprende 95% de los valores centrales de la población de referencia. En sentido estricto, el intervalo de la normalidad incluye la media más y menos las mediciones incluidas en dos desviaciones estándar de la media. Si no existe alguna razón para hacerlo de otra forma, el investigador generalmente escoge la parte central del intervalo, de forma que 2,5% de los individuos sanos tengan mediciones mayores y 2,5% de los individuos sanos tengan mediciones menores de los valores del intervalo de la normalidad.

Para ilustrar este método, imagine que los investigadores miden la estatura de 100 estudiantes varones de una facultad de medicina y encuentran valores semejantes a los que aparecen en la figura 15-1. Los investigadores escogerían a continuación un intervalo de la normalidad que incluyera 95 de los 100 estudiantes. Si no tuviesen una razón para hacerlo de otro modo, utilizarían la parte central del intervalo, de forma que el intervalo de la normalidad de este grupo de referencia estaría comprendido entre 60 y 78 pulgadas [152 y 198 cm]. Los individuos que quedarán fuera del intervalo no tendrían necesariamente que tener ninguna enfermedad, simplemente podrían ser individuos sanos excluidos del intervalo de la normalidad.

PRINCIPIOS BÁSICOS

En primer lugar, veamos las implicaciones de los principios del intervalo de lo normal y después los errores que pueden resultar si no se comprenden estas implicaciones.

FIGURA 15-1. Estaturas de 100 estudiantes de medicina varones, utilizadas para construir un intervalo de valores normales



1. Por definición, en cualquier prueba determinada aplicada a un grupo, 5% de los resultados se encontrarán fuera del intervalo de lo normal. Por esta razón, "anormal" y "enfermo" no son sinónimos. Cuantas más pruebas se realicen, más individuos se encontrarán fuera del intervalo de la normalidad, por lo menos en una prueba. Llevando esta proposición a su límite, se puede concluir que una persona "normal" es aquella que no ha sido investigada suficientemente. A pesar de lo absurdo de esta proposición, destaca la importancia de comprender que la definición del intervalo de la normalidad sitúa a propósito a 5% de los individuos que no están enfermos fuera de dicho intervalo. Por este motivo, el término *fuera de los límites de lo normal* no debe entenderse como sinónimo de enfermedad.

2. Los valores que se encuentran en el intervalo de la normalidad no garantizan que los individuos no estén enfermos. La capacidad del intervalo de la normalidad de una prueba para discriminar entre los sanos y los enfermos varía de una prueba a otra. A menos que la prueba sea perfecta para descartar la enfermedad —y pocas pruebas lo son—, las mediciones de algunos individuos que tienen la enfermedad se encontrarán dentro de los límites de la normalidad.

3. Los cambios incluidos en los límites normales pueden ser patológicos. Dado que el intervalo de lo normal incluye un amplio intervalo de valores numéricos, las mediciones de un individuo pueden cambiar considerablemente y todavía encontrarse dentro de los límites de la normalidad. Por ejemplo, el intervalo de la normalidad para la enzima hepática AST varía de 8 a 20 U/L, el del potasio sérico puede variar entre 3,5 y 5,4 mEq/L, y el del ácido úrico, desde 2,5 a 8,0 mg por 100 ml. Es importante considerar no solo si los valores de un individuo se hallan dentro de los límites normales, sino también si han cambiado con el tiempo. El concepto del intervalo de lo normal es más útil cuando no se dispone de datos anteriores para comparar pacientes individuales. Sin embargo, cuando se dispone de esos datos, se deben tener en cuenta.

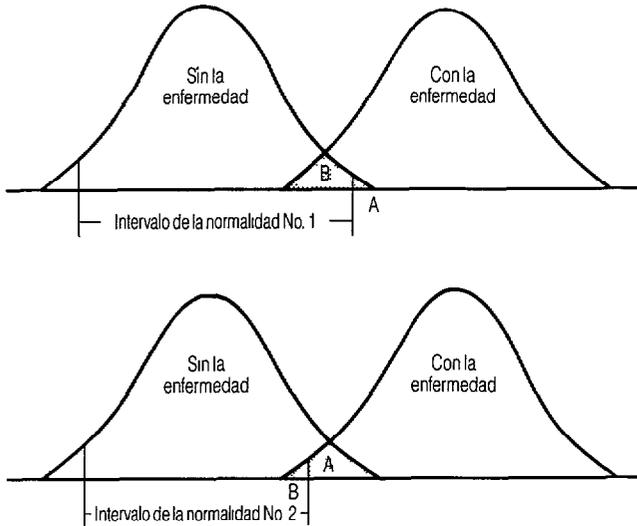
4. El intervalo de lo normal se calcula empleando un grupo concreto de pacientes o una población de referencia considerados sanos. Por consiguiente, cuando se aplica un intervalo concreto a un individuo, se debe averiguar si ese individuo presenta alguna característica que lo diferencia de la población de referencia utilizada para construir el intervalo de la normalidad. Por ejemplo, si para obtener el intervalo de la normalidad del hematócrito se utilizan hombres, este no puede aplicarse necesariamente a las mujeres, que en general tienen hematócritos más bajos.

5. El intervalo de la normalidad no debe confundirse con el intervalo deseable. El intervalo de la normalidad es una medida empírica de cómo son las cosas en un grupo de individuos que se consideran sanos en ese momento. Es posible que amplios sectores de la comunidad tengan resultados de las pruebas más elevados que los ideales y estén predispuestos a desarrollar una enfermedad en el futuro.

6. Los límites superior e inferior del intervalo de lo normal pueden modificarse con fines diagnósticos. El intervalo incluye 95% de los que no presentan una enfermedad o estado concreto. Sin embargo, no es necesario que haya el mismo número de individuos sanos con valores de una prueba por debajo y por encima del intervalo de la normalidad. Existe cierto margen de criterio científico para determinar dónde deben situarse los límites superior e inferior de dicho intervalo. La demarcación de los límites depende del objetivo que persigan el investigador o el clínico al aplicar la prueba.¹ Por

¹ El nivel del intervalo de lo normal fijado definirá la especificidad de la prueba. Este nivel se puede ajustar posteriormente para aumentar (o disminuir) la especificidad. Por esta razón, en la práctica, el intervalo de lo normal siempre se ha ajustado cuando la especificidad no es de 95%

FIGURA 15-2. Modificaciones del intervalo de lo normal. Positivos falsos, A: individuos que no tienen la enfermedad, con valores por encima del intervalo de la normalidad. Negativos falsos, B: individuos que tienen la enfermedad, con valores dentro del intervalo de la normalidad



ejemplo, suponga que la mayoría de los individuos con una enfermedad tienen niveles cercanos al límite superior del intervalo de lo normal en la prueba. Si el investigador está dispuesto a reducir este límite, se puede prever que un mayor número de individuos que tienen la enfermedad tendrán resultados por encima del intervalo de la normalidad. En este caso, el investigador también paga el precio (o acepta el intercambio) de colocar a una mayor proporción de la población sana fuera del intervalo de la normalidad. A veces merece la pena pagar ese precio, sobre todo cuando es importante detectar el mayor número posible de individuos con la enfermedad o cuando las pruebas de seguimiento para clarificar la situación son baratas y prácticas. La figura 15-2 muestra este intercambio.

Al trasladar el intervalo de la normalidad hacia la izquierda, como en el intervalo de normalidad No. 2, observe que el área B se reduce respecto a la del intervalo No. 1 y que el área A aumenta. En otras palabras, una disminución del número de negativos falsos conduce a un aumento de los positivos falsos y viceversa. En esta situación, la prueba identifica como fuera de lo normal a un mayor número de personas con la enfermedad. Al mismo tiempo, clasifica fuera de los límites normales a más individuos sin la enfermedad; por lo tanto, en el intervalo de la normalidad No. 2 estamos aceptando más lecturas positivas falsas a cambio de menos negativas falsas. El número de positivos y negativos falsos que el médico o el sistema sanitario esté dispuesto a tolerar depende de consideraciones éticas, económicas y políticas, así como del conocimiento médico.

Los siguientes ejemplos ilustran los errores que se pueden cometer al aplicar incorrectamente estos principios.

1. En 1 000 exámenes de salud consecutivos de tipo preventivo se realizaron 12 pruebas de laboratorio (SMA-12) en cada paciente, aunque no se encontraron anomalías en la historia médica o en la exploración física. El 5% de las SMA-12 estaban

fuera del límite de la normalidad; es decir, se obtuvo un total de 600 pruebas "anormales". Los autores concluyeron que los resultados obtenidos justificaban la realización de las SMA-12 en todos los exámenes de salud rutinarios.

Veamos el significado de estos resultados. Los valores del intervalo de la normalidad incluyen, por definición, solo 95% de todos aquellos que se consideran exentos de la enfermedad. Si aplicamos esta prueba a 1 000 individuos sin la enfermedad, 5% ó 50 individuos tendrán un resultado fuera del intervalo de la normalidad. Si se aplicaran 12 pruebas a 1 000 individuos sin síntomas ni signos de enfermedad, en promedio, los resultados de 5% de las 12 000 pruebas realizadas se encontrarían fuera del intervalo de lo normal. El 5% de 12 000 pruebas es 600 pruebas. Por eso, aunque los 1 000 individuos no tuvieran ninguna enfermedad, se podría prever que 600 pruebas darían resultados fuera del intervalo de la normalidad. Estos reflejarían simplemente el método de calcular el intervalo de la normalidad. Estos resultados no indican necesariamente la presencia de enfermedad y por sí mismos no justifican el realizar múltiples pruebas de laboratorio en todos los exámenes de salud rutinarios.

Al considerar las implicaciones de los resultados de las pruebas, es importante darse cuenta de que no todos los valores fuera del intervalo de lo normal tienen el mismo significado. Es mucho más probable que los valores bastante alejados del límite sean causados por enfermedad que los valores cercanos a dicho límite y asimismo, que los resultados cercanos a los límites del intervalo se deban a la variabilidad de la prueba o a la variabilidad biológica. Por ejemplo, si el límite superior del hematocrito en un hombre es 52, es más probable que el valor 60 esté asociado con una enfermedad que el valor 53.

2. Se midió la concentración de AST [aspartasa aminotransferasa] en 100 alcohólicos, para valorar su función hepática. En la mayoría, los resultados se encontraban dentro del intervalo de la normalidad. Los autores concluyeron que el hígado de estos alcohólicos funcionaba bien.

Este ejemplo ilustra la diferencia entre el intervalo de la normalidad de las pruebas de laboratorio y el no tener la enfermedad. El hecho de que los resultados de las pruebas de laboratorio de esas personas estuvieran dentro de los límites de la normalidad no es suficiente por sí solo para establecer que su hígado funciona perfectamente, dado que en cualquier prueba algunos resultados correspondientes a individuos enfermos se encontrarán dentro de los límites de lo normal. Cuanto menor sea la capacidad de la prueba para diagnosticar la enfermedad, más elevado será el número de individuos enfermos cuyos resultados se encuentran dentro de los límites de lo normal. Puede que ciertas pruebas no permitan distinguir a las personas enfermas de las sanas. Es posible que en ambos grupos la mayor parte de los resultados estén dentro del intervalo de la normalidad. Esto sucedió con los resultados de la AST. La incapacidad de la prueba para discernir entre enfermos y sanos indica que su capacidad de discriminación diagnóstica es baja y que no es útil para el diagnóstico. Subraya, así, la diferencia entre encontrarse dentro del intervalo de la normalidad y no tener la enfermedad.

Las figuras 15-3 a 15-5 muestran las tres posibles relaciones entre la población sana y la población enferma. La figura 15-3 ilustra una prueba que separa completamente a los que tienen la enfermedad de los que no la tienen. La discriminación diagnóstica de esta prueba es perfecta. La figura 15-4 representa la situación usual, de una prueba que separa parcialmente a los que tienen la enfermedad de los que no la tienen. La figura 15-5 muestra el ejemplo de una prueba que no tiene discriminación diagnóstica. En el caso de la AST, la situación se parece mucho a la de la figura 15-5. A pesar de su utilidad en el diagnóstico de muchas enfermedades hepáticas, la medición

FIGURA 15-3. Prueba que separa completamente los resultados de las poblaciones (con discriminación diagnóstica perfecta)

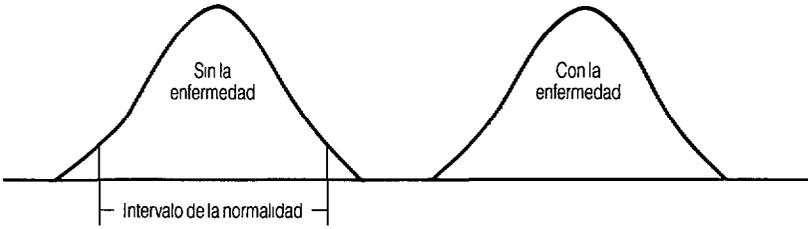


FIGURA 15-4. Prueba que separa parcialmente los resultados de las poblaciones (con discriminación diagnóstica parcial)

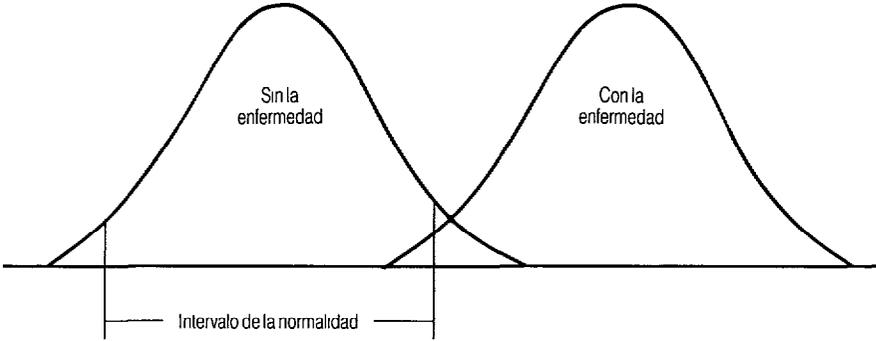
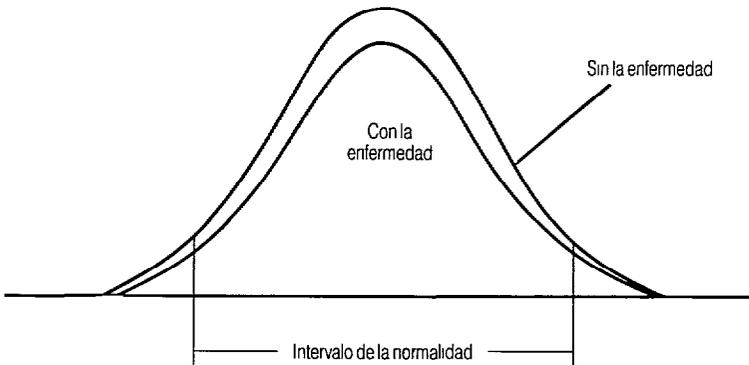


FIGURA 15-5. Prueba que no separa los resultados de las poblaciones (sin discriminación diagnóstica)



de la concentración de AST no es útil para diagnosticar el efecto crónico del alcohol en el hígado. Por eso, a pesar de que se puede calcular el intervalo de la normalidad para cualquier prueba, este intervalo, por sí solo, no indica si la prueba será útil para el diagnóstico. Las mediciones en los individuos con una enfermedad concreta pueden ser idénticas a las de los que no la tienen y viceversa, lo cual indica que la prueba no tiene utilidad diagnóstica para esa enfermedad concreta.

3. Se calculó que el intervalo de la normalidad de la creatinina sérica en 1 000 estadounidenses asintomáticos sin enfermedad renal conocida era de 0,7 a 1,4 mg/dl. Una mujer de 70 años de edad ingresó en el hospital con una concentración de creatinina de 0,8 mg/dl y fue tratada con gentamicina. Cuando fue dada de alta, ese valor era de 1,3 mg/dl. Su médico llegó a la conclusión de que, como la concentración de creatinina se encontraba dentro de los límites de la normalidad, tanto al ingreso como al darla de alta, su paciente no tenía una lesión renal secundaria al tratamiento con gentamicina.

La presencia de un resultado dentro de los límites de la normalidad no garantiza la ausencia de enfermedad. En cada individuo, la medida que indica que no está enfermo puede estar por encima o por debajo de la medida promedio de los demás individuos sin la enfermedad. En este ejemplo, la concentración de creatinina de la paciente aumentó en 60%, si bien todavía se encontraba dentro del intervalo de la normalidad. Dicho cambio sugiere la presencia de un nuevo proceso patológico. Es probable que la gentamicina le haya producido una lesión renal. Cuando se dispone de información previa, es importante considerarla al evaluar el resultado de una prueba. Los cambios, aun dentro del intervalo de la normalidad, *pueden* ser un signo de enfermedad.

4. Se utilizó a un grupo de 100 estudiantes de medicina para calcular el intervalo de valores de la normalidad del recuento de granulocitos. Se escogió un intervalo de forma que incluyera 95 de los 100 recuentos de granulocitos. Los límites del intervalo de la normalidad calculado fueron 2 000 y 5 000. Cuando se preguntó a los autores sobre el recuento de 1 900 en un anciano de raza negra, llegaron a la conclusión de que este se encontraba claramente fuera del intervalo de lo normal y que era preciso realizar más estudios para identificar la causa de ese resultado.

El intervalo de la normalidad depende de la población de referencia sin la enfermedad que se ha seleccionado; esta se define como el intervalo alrededor de un valor promedio que incluye 95% de los individuos de una población de referencia determinada. Sin embargo, la población de referencia sin la enfermedad utilizada para calcular el intervalo de la normalidad puede tener mediciones diferentes de las del grupo de personas en las que queremos usar la prueba.

Es improbable que haya muchos ancianos de raza negra entre el grupo de estudiantes de medicina utilizados para construir el intervalo de la normalidad. De hecho, los hombres de raza negra tienen un intervalo de valores del recuento de granulocitos distinto de los de raza blanca. Por este motivo, el intervalo de lo normal calculado con los estudiantes de medicina puede no reflejar el intervalo de la normalidad aplicable a los ancianos de raza negra. El recuento de granulocitos de este hombre se encontraba probablemente dentro del intervalo de normalidad para su raza, edad y sexo. Como se sabe que los ancianos de raza negra tienen su propio intervalo de la normalidad, este hecho se debe tener en cuenta cuando se interpretan los resultados de la prueba.

5. El intervalo de la normalidad de la concentración de colesterol sérico medida en 100 hombres estadounidenses de raza blanca de 30 a 60 años de edad osciló entre 200 y 300 mg/dl. La concentración de colesterol de un estadounidense de raza blanca de 45

años de edad fue 280 mg/dl. Su médico le dijo que no tenía que preocuparse por el colesterol elevado, dado que su concentración se hallaba dentro de los límites de lo normal.

El intervalo de la normalidad se calcula utilizando los datos recogidos en una población de referencia que en ese momento se considera sana. Es posible que el grupo utilizado esté formado por individuos cuyos resultados en la prueba sean más elevados (o más bajos) que los deseables. Un resultado dentro del límite de la normalidad no garantiza que el individuo se mantendrá sano. Es posible que los hombres estadounidenses, considerados en conjunto, tengan concentraciones de colesterol por encima de los niveles deseables. Si esto es cierto, el paciente con una concentración de colesterol de 280 mg/dl puede muy bien sufrir las consecuencias de la hipercolesterolemia. Cuando existen datos basados en la investigación que sugieren claramente un intervalo deseable de valores para una prueba, es aceptable sustituir el intervalo habitual por el intervalo de la normalidad deseable. Esto se está haciendo cada vez más con el colesterol sérico.

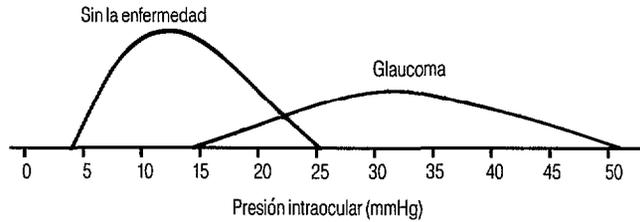
6. En un estudio se demostró que 90% de los que tienen una tensión intraocular mayor de 25 mmHg desarrollarán defectos visuales secundarios al glaucoma en los 10 años siguientes. El 20% de los que tienen tensiones intraoculares de 20 mmHg desarrollarán cambios similares y 1% de los que tienen tensiones intraoculares de 15 mmHg desarrollarán trastornos. Los autores concluyeron que el rendimiento de la prueba se podría mejorar disminuyendo el límite superior de la normalidad de 25 a 15 mmHg, dado que así la prueba podría identificar prácticamente a todos los que se encontraran en riesgo de desarrollar defectos visuales (figura 15-6).

Si el límite superior de la normalidad es de 25 mmHg, casi todo el mundo que no desarrollará glaucoma estará dentro de los límites de la normalidad, aunque un número elevado de los que desarrollarán glaucoma también estarán incluidos en los límites de la normalidad. Por otro lado, si el límite de lo normal se sitúa en 15 mmHg, muy pocos individuos con glaucoma estarán dentro de los límites de la normalidad y un alto número de los que nunca tendrán glaucoma estarán fuera del intervalo de la normalidad.

La capacidad de una prueba para detectar la enfermedad se puede aumentar modificando los límites del intervalo de la normalidad. Si los límites se amplían suficientemente, la prueba incluirá prácticamente a todos los que padezcan la enfermedad. Lamentablemente, esta atractiva solución también sitúa fuera del intervalo de la normalidad a un mayor número de individuos que no tienen ni tendrán la enfermedad. Al aumentar los límites superiores de lo normal, los investigadores incrementan la capacidad de detectar enfermedades futuras, pero solo pagando el precio de seguir a muchos individuos que no tendrán la enfermedad. Al determinar dónde situar los límites superiores, se pueden considerar los factores que se enumeran a continuación.

1. La pérdida de visión por el glaucoma es irreversible en gran parte y puede desarrollarse antes de que el paciente se dé cuenta.
2. El tratamiento suele ser seguro, pero solo parcialmente efectivo para prevenir la pérdida progresiva de la visión.
3. El seguimiento es seguro y acarrea un riesgo bajo, pero el seguimiento de un número elevado de individuos requiere mucho tiempo y es costoso, ya que exige realizar múltiples exámenes repetidos durante largos períodos de tiempo. Además, el seguimiento produce ansiedad en los pacientes.

FIGURA 15-6. Posible distribución de la tensión intraocular de los sujetos con y sin glaucoma



Los factores que hay que sopesar no son todos médicos. Otras consideraciones sociales, psicológicas, económicas o políticas se pueden tomar en cuenta para establecer la línea de demarcación. Es posible que no exista una respuesta correcta. La única salida para esta situación insoluble es que se invente una prueba mejor, en la cual se superponga menos a los que desarrollarán la enfermedad y a los que no lo harán.

El concepto del intervalo de lo normal es un intento para tratar con la variabilidad que existe entre las personas. La comprensión de la utilidad y las limitaciones de este concepto es de capital importancia para comprender las pruebas diagnósticas. El intervalo de la normalidad define los valores numéricos encontrados en 95% de los sujetos considerados sanos de un grupo concreto de referencia. Este intervalo puede no reflejar el nivel deseable y no tiene en cuenta los cambios que se producen respecto de los resultados de pruebas anteriores.

El intervalo de la normalidad *per se* no nos dice nada sobre la utilidad diagnóstica de la prueba. Cada prueba tiene un intervalo de la normalidad que puede o no ayudar a discernir entre los individuos que tienen la enfermedad y los que no la tienen. Para determinar la utilidad de una prueba en el diagnóstico de una enfermedad es necesario examinar los resultados de la prueba en un grupo de individuos con una determinada enfermedad y comparar estos valores con el intervalo de la normalidad de un grupo sin la enfermedad, como haremos en el capítulo 17. Antes de que podamos hacerlo, es preciso que examinemos cómo se puede definir a los individuos que tienen la enfermedad.