



Investigación clínica: en busca de la certidumbre

La importancia del diálogo entre la práctica clínica y la investigación científica

Alberto Lifshitz

Práctica médica e investigación clínica:
claves para generar conocimiento
y mejorar la atención

Carla Martínez Castuera-Gómez *et al.*

Estudios de proceso (prueba diagnóstica)

Juan O. Talavera *et al.*

Cómo localizar artículos para resolver
una pregunta clínica

Rodolfo Rivas-Ruiz *et al.*

Presentación

Investigación clínica

Juan O. Talavera

Esta serie de artículos intenta ofrecer al médico en áreas de atención clínica un instrumento que le permita interpretar las lecturas que realiza en forma habitual para resolver los problemas de salud que presentan nuestros pacientes. De esta forma, no dependerá de la interpretación de otros y podrá identificar, además, los errores de interpretación involuntarios o voluntarios que se observan en las publicaciones científicas. La serie da comienzo con una descripción de distintos abordajes, de los que hay que resaltar dos: el abordaje arquitectónico, que se basa en el juicio clínico para describir el fenómeno de causalidad y los estudios de proceso (diagnóstico); y el abordaje jerárquico, que tiene como eje la calidad de la información, y en el que se muestran cuatro diseños base: el ensayo clínico, la cohorte, el diseño de casos y controles y la encuesta transversal. Además, se hace referencia a una estrategia que nos permite entender el porqué de la prueba estadística y del tamaño de muestra, seguido de la diferencia que existe entre la significación estadística y la relevancia clínica; esta última es la que determina la utilidad de la manobra. Luego, se describe el procedimiento de búsqueda sistemática, estrategia encaminada a encontrar en forma ordenada y rápida los artículos que pueden responder a los cuestionamientos que se generan durante el día a día de la atención clínica. El suplemento cierra con un par de ejemplos: el primero, que integra lo que se propone indispensable para una revisión estructurada de la literatura y el segundo, que muestra la combinación del modelo arquitectónico y el modelo jerárquico.

Prefacio

La importancia del diálogo entre la práctica clínica y la investigación científica

Alberto Lifshitz^a

^aConsejero emérito de educación, investigación y políticas de salud, Instituto Mexicano del Seguro Social, Distrito Federal, México

Comunicación con: Alberto Lifshitz
Teléfonos: (55) 5623 2421, 5623 2300,
extensión 43038
Correo electrónico: alifshitz@liceaga.facmed.unam.mx

Aunque la práctica clínica se nutre de los resultados de la investigación científica y esta última se alimenta de las necesidades de la práctica clínica, lo cierto es que en los últimos tiempos estos dos mundos se han distanciado de una manera inconveniente. Uno parece ser el mundo de la ciencia y otro el de la clínica. Aun en las estructuras curriculares para la formación de médicos, se distinguen dos etapas muy claramente definidas: la de las ciencias básicas y la de las disciplinas clínicas, al grado de que parecen dos carreras distintas. Todos los planes curriculares tienen que recurrir a actividades de integración porque se suelen ver como compartimientos separados. A mayor abundamiento, en muchas escuelas ya los maestros de ciencias básicas no son clínicos sino biólogos o químicos, de tal modo que no tienen la perspectiva de la práctica profesional del médico, y muchos profesores de clínica han olvidado las ciencias básicas si no es que las menosprecian o les temen. Ahora se han agregado nuevas ciencias básicas como la epidemiología, la estadística, la comunicación y la informática, y se percibe más bien una tendencia a salirse de la dicotomía básico-clínica y aventurarse en la fundamental-aplicada (Bandiera G, Boucher A, Neville A, Kuper A, Hodges B. Integration and timing of basic and clinical sciences education. *Med Teach.* 2013;35(5):381-7. doi: 10.3109/0142159X.2013769674. Epub 2013 Feb 27). Pero además, el ejercicio clínico tiene el riesgo de convertirse en una actividad empírica, refleja, estereotipada cuando se aleja de la ciencia, aún de la llamada ciencia clínica.

El movimiento de la epidemiología clínica significó un cambio en la forma de ver la actividad arquetípica de los médicos al incorporar métodos propios de la ciencia ya no a la indagación de los aspectos básicos de la medicina sino a la práctica clínica misma, y no solo como una estrategia para crear conocimiento sino para atender más adecuadamente a los pacientes. De esta propuesta surgieron muchos avances metodológicos, varios de los cuales se agruparon dentro de la medicina basada en evidencias. Uno de los logros más importantes para la atención de los pacientes ha sido precisamente la aplicación de estos métodos a la búsqueda de mejores soluciones para los enfermos. Este suplemento es una aportación en este sentido y no necesariamente para formar investigadores sino para formar mejores médicos que integren la investigación a su práctica cotidiana. Al fin y al cabo, la atención de los pacientes es un espacio apropiado para esta integración de visiones complementarias: de allí surgen las necesidades de investigación y allí llegan los resultados como mejores soluciones que las previas.

Por supuesto que la formación tradicional de los médicos no abarca suficientemente esta capacidad de identificar en lo cotidiano los problemas que tendrían que abordarse mediante la ciencia, ni la de acechar la aparición de las soluciones para aplicarlas oportunamente y menos la de juzgar la validez y la confiabilidad de cuanto se publica y difunde. Lamentablemente, el exceso de información está plagado de pseudociencia, ya sea publicidad que aparenta ser científica o resultados bien intencionados pero con defectos metodológicos. Quienes atienden pacientes por lo menos tendrían que discernir lo valioso de lo superfluo, lo promocional de lo científico, lo aplicable de lo teórico, lo confiable de lo cuestionable, lo válido de lo que no lo es. El insumo básico para la atención médica es, ciertamente, la información y por ello esta tiene que ser de calidad.

Pero la clínica es también un espacio apropiado para la creación de conocimiento. Lo que ocurre es que no está suficientemente generalizada la motivación, la disciplina, la curiosidad ni la metodología para hacer efectiva esta potencialidad. Este suplemento es, entonces, una valiosa herramienta para despertar al científico que los clínicos tienen dentro y verter esta capacidad en beneficio de sus pacientes y del progreso

de la profesión. Mucho se ha cuestionado si la clínica es o no una ciencia. Lo que se puede afirmar es que es un espacio en el que se pueden poner a prueba los conocimientos generados por la ciencia, un territorio en el que surgen necesidades de investigación científica, una actividad que sigue un método de indagación semejante al de la ciencia y un ámbito en el que ciertamente se pueden desarrollar investigaciones centradas en los pacientes.

Es verdad que hay muchos y muy buenos textos de metodología de la investigación y de análisis crítico de la literatura científica, pero este suplemento tiene la ventaja de que va dirigido a quienes tienen la responsabilidad de atender pacientes en una institución como es el Instituto Mexicano del Seguro Social; está escrito por quienes tienen esta experiencia, adicional a la formación metodológica que también fue orientada hacia la investigación clínica. La potencialidad de encontrar preguntas que se puedan abordar mediante la investigación y de acechar los resultados de las investigaciones para aplicarlos oportunamente a los pacientes de todos los días ha sido insuficientemente aprovechada. Este suplemento de la Revista Médica del IMSS es una herramienta para avanzar en ese camino.

Editorial

Práctica médica e investigación clínica: claves para generar conocimiento y mejorar la atención

Carla Martínez Castuera-Gómez, Juan O. Talavera

El conocimiento médico que integra la investigación clínica a la práctica médica cotidiana incrementa la calidad de la atención de manera inmediata. En el proceso de la generación de conocimiento médico se distinguen cuatro pasos: hacer un cuestionamiento sobre la práctica médica, analizar el conocimiento publicado en la literatura especializada, desarrollar un protocolo de investigación y publicar los resultados. Es indispensable propiciar decisiones fundadas en la investigación clínica para favorecer el desarrollo de estrategias que mejoren la calidad de la atención.

Palabras clave: investigación clínica, calidad de la atención de salud, publicaciones periódicas

Investigar, crear y compartir el conocimiento son de las actividades más nobles que el ser humano puede desempeñar, dado que invariablemente su objetivo es mejorar la condición de la vida en general. Esta nobleza es evidente en el campo de la medicina: los resultados de la investigación determinan la calidad de vida que tendrán las personas sanas y las que padezcan alguna enfermedad. Por lo tanto, la trascendencia de la investigación en el área médica estriba en su inherente responsabilidad social.

En consideración a lo anterior, la presente reflexión busca contribuir a la idea de que es posible atender esta responsabilidad cuando el personal de salud mantiene una relación simbiótica entre la práctica médica, el quehacer de la investigación clínica y la publicación del conocimiento médico.

De la práctica a la generación de conocimiento

El proceso de generación de conocimiento médico puede incidir en mejorar la calidad de la atención médica cuando se inicia en la práctica médica, se enriquece con la investigación clínica y culmina con la publicación.

La práctica médica se puede definir como la estrategia que el médico sigue de manera cotidiana al elegir las mejores alternativas de atención —dentro de su bagaje de conocimiento y recursos— para tratar determinada condición de salud. Cuando el médico se encuentra con situaciones que no puede resolver de la forma como regularmente lo hace, llega al momento de comenzar a generar conocimiento médico.

El primer paso de ese proceso se da cuando el médico se plantea una pregunta para tratar de resolver un problema nacido de su práctica profesional, ya sea que trate de formular un diagnóstico, estimar el pronóstico o decidir la causa del problema o un mejor tratamiento. El cuestionamiento es una destreza que el médico desarrolla de manera casi natural. Actividades rutinarias como la exploración física, la elaboración o revisión del historial clínico, la prescripción de un fármaco diferente ante complicaciones o persistencia de enfermedades, entre otras, implican un cuestionamiento. A este cuestionamiento le siguen la búsqueda de causas, la comparación de casos y la identificación de condiciones irregulares para tomar decisiones sobre el tratamiento de determinada condición de salud. Cuestionarse, responder y decidir son tareas inherentes a la profesión médica y la creación de conocimiento. Cuando el médico se invo-

lucra en actividades académicas y de investigación a la par que ejerce su práctica profesional, se afinan y agudizan las destrezas de cuestionamiento y la toma de decisiones asertivas.

En consecuencia, el médico que no se involucra en la investigación está desperdiciando la oportunidad de desarrollar sus habilidades profesionales y está faltando a su responsabilidad social por no utilizar su conocimiento y sus capacidades para mejorar la calidad de vida de las personas. Más aún, el desarrollo de la investigación clínica debe ser una condición contenida en el diseño de los sistemas de salud, de modo tal que deben existir tareas administrativas y médicas que faciliten su realización.

El siguiente paso en la generación de conocimiento médico es buscar respuestas mediante la consulta y el análisis crítico de la literatura especializada. La importancia de este paso es que reduce el riesgo de invertir tiempo, recursos humanos, financieros y bienes en buscar respuestas a preguntas ya planteadas o, peor aún, llegar a respuestas inconclusas o que ya han sido propuestas. Además, la revisión exhaustiva y crítica de la literatura es crucial, porque asegura que el manuscrito sea original y novedoso y tenga sustento científico adecuado y una alta estimación de factibilidad. Cuando dichos factores están presentes en un manuscrito, este tiene más posibilidades de haber resuelto acertadamente el cuestionamiento planteado y de lograr convertirse en material de publicación por la relevancia del conocimiento generado.

Este paso aparentemente plantea dos retos: el acceso a fuentes de información y la búsqueda selectiva. En realidad, el reto es uno: saber buscar. Internet y PubMed son poderosas fuentes de información de fácil acceso para todos los médicos, pero si no se sabe utilizar parámetros de búsqueda se convierten en un depósito inacabable de información de baja calidad que desalienta la investigación. Ante este reto se plantea una solución sencilla: enseñar estrategias de búsqueda selectiva y ponerlas constantemente en práctica. Esta propuesta es un aspecto en el que el personal médico y el administrativo pueden influir para mantener la simbiosis *práctica médica-investigación clínica-publicación*.

El tercer paso del proceso de generación del conocimiento médico es diseñar y ejecutar el protocolo de investigación clínica. La elaboración, el contenido, las características y la ejecución de un protocolo son temas ampliamente tratados y están fuera del alcance de la presente reflexión, cuyo interés central es subrayar que el conocimiento médico se genera cuando la investigación clínica logra proponer una respuesta al cuestionamiento que surge de la práctica médica. Pese a ello, es importante hacer énfasis en que la investigación clínica y el desarrollo del protocolo deben seguir

estrategias de control de calidad que salvaguarden el rigor metodológico y a los pacientes participantes. Lo anterior se logra con la inclusión y observancia de principios éticos mínimos. La participación de los comités de ética, el registro internacional de ensayos clínicos, la revisión entre pares y las opiniones de los consejos editoriales, entre otros, son mecanismos para supervisar el apego a principios éticos que garantizan el desarrollo de la investigación de calidad.

La ejecución del protocolo de investigación genera una respuesta al cuestionamiento. Si bien la respuesta puede ser diferente de la que se había inferido o de la que se esperaba, se tiene la certeza de haber llegado a ella reuniendo y probando evidencia. Independientemente de la respuesta que se encuentre, se abre camino para el cuarto paso del proceso y se llega al momento de elegir la revista para publicar la información obtenida.

En la actualidad se tiende a seleccionar la revista considerando principalmente su factor de impacto: “today, too many of our postdocs believe that getting a paper into a prestigious journal is more important to their career than doing the science itself”.¹ Sin embargo, esta decisión debería fundamentarse en la audiencia a quien va dirigida la información, la facilidad de acceso que la revista ofrece al público médico, los requisitos editoriales y, por último, el factor de impacto. Este orden de prioridades de selección es el idóneo si el objetivo principal de publicar es que los resultados de la investigación clínica sean accesibles y los médicos los integren a su práctica cotidiana para mejorar la atención que ofrecen.

Por otro lado, este orden de prioridades combate la presión que impone buscar la publicación en una revista con factor de impacto y la consecuente frustración cuando no se logra esto. Si bien los sistemas académicos se valen de parámetros como el factor de impacto para evaluar la productividad científica, en el ámbito local es posible crear mecanismos de evaluación e incentivos que promuevan la publicación de conocimiento médico en revistas serias, de fácil acceso y amplia distribución entre el público médico, independientemente del factor de impacto. En nuestro país, y particularmente en nuestra Institución, la *Revista Médica del Instituto Mexicano del Seguro Social* es un espacio único y privilegiado que debe ser considerado para motivar la publicación de conocimiento médico.

De acuerdo con un artículo editorial publicado en *Proceedings of the National Academy of Sciences*, numerosos estudiantes de posdoctorado afirman que elegirían publicar sus trabajos académicos en sus revistas favoritas, aquellas en las que encuentran los escritos que disfrutan leer, si no fueran evaluados con base en el factor de impacto.¹ Más aún, si —como

se ha argumentado— el conocimiento médico publicado permite compartir y promover las mejores prácticas, entonces la selección de la revista para publicar no puede estar definida únicamente por el factor de impacto.²

Tomando en cuenta lo anterior, parecería conveniente promover la publicación del conocimiento derivado de la investigación en la práctica clínica en revistas de fácil acceso, dado que esta característica favorecerá su aplicación en el área médica. Por ejemplo, cuando se publica en revistas locales se aumentan las posibilidades de que el lector conozca al autor y viceversa. Esto puede ser un estímulo importante para que más médicos que se ven en el mismo nivel que los autores se sientan atraídos a crear y compartir su conocimiento por medio de la generación de conocimiento. Además, los médicos que leen el conocimiento publicado por colegas que conocen pueden ser más propensos a integrarlo en su práctica si el autor es una persona a quien respetan, en parte porque los lectores tienen la posibilidad de dialogar con el autor y porque tienen la certeza de que el autor conoce las condiciones de su servicio médico o, por lo menos, su circunstancia local o nacional. Ese conocimiento es visto con autoridad y no como una receta importada que no puede aplicarse a las propias circunstancias. Al seleccionar este tipo de revistas se minimiza la tentación de falsear los resultados o la información para lograr la publicación, impulso que es mayor cuando se trata de una revista con alto factor de impacto.¹

Finalmente, si recordamos que hacer investigación es un acto de responsabilidad social, la selección de la revista para publicar no debería realizarse en función del prestigio sino de la posibilidad de compartir el conocimiento. Por lo tanto, promover la mejora de la práctica médica se relaciona directamente con promover la publicación del conocimiento médico basado en la investigación clínica. Mientras más se integre el quehacer médico a la investigación clínica y como resultado se publique el conocimiento que se genere, mayores serán las posibilidades de incidir en mejorar la atención médica, cerrando así el círculo virtuoso de la generación de conocimiento.

Hasta aquí hemos intentado sustentar el argumento de que la relación *práctica médica-investigación clínica-publicación* incide en la calidad de la atención médica. Al igual que otros autores, creemos que la investigación clínica por sí misma tiene tres efectos positivos:³⁻⁶

1. Los pacientes que participan en un proyecto de investigación reciben mejor calidad de atención.
2. Aumenta la motivación del médico y su satisfacción en el trabajo.

3. Los sistemas de salud se benefician de la eficacia y la eficiencia que muestran tanto los médicos en su práctica como los pacientes en su tratamiento.

Sin embargo, es con la publicación y divulgación del conocimiento derivado de la investigación clínica como se asegura que estos beneficios se extiendan y reproduzcan a través de la relación *práctica médica-investigación clínica-publicación*. El recorrido descrito es el ideal para mantener esa simbiosis e incidir en la mejora de la atención médica. No obstante, por infortunio no es el camino que siempre se sigue. Es posible, y más frecuente de lo deseable, encontrar publicaciones de investigación clínica poco originales o poco sustentadas e inconclusas, con muy bajo control de calidad y, en ocasiones, sin respeto a los principios éticos correspondientes. Las consecuencias no han sido menores: credibilidad erosionada de algunas revistas, desinterés en publicar el conocimiento que se genera con la investigación clínica y en realizar investigación, desactualización de los médicos y tendencia a reducir la efectividad de su práctica y mínima o nula creación de conocimiento aplicable sobre las dolencias del paciente.

Por el contrario, cuando el proceso para generar conocimiento originado de la práctica médica y la investigación clínica se sigue de manera ordenada, se genera un ambiente virtuoso que estimula la simbiosis *práctica médica-investigación clínica-publicación*. El médico que se dedica a la atención médica, realiza investigación clínica y concreta el proceso con la publicación en revistas accesibles para sus colegas, se convierte en autoridad y modelo a seguir. Quien logra resolver las necesidades de la práctica médica a través de la investigación clínica desarrolla buenos hábitos de atención y facilita que esta actitud se reproduzca entre el personal de salud con el que trabaja. En resumen, se estima que hay mejora inmediata en la atención de los pacientes.

Conclusiones

La imposibilidad de un médico para atender parte de su responsabilidad social, por no involucrarse en actividades académicas y de investigación, podría considerarse abrumadora. Sin embargo, no hay razón para tal interpretación cuando se entiende que la responsabilidad de ese profesional es la generación del conocimiento médico y su aplicación en la mejora de la atención al paciente. Es responsabilidad del personal administrativo y de los diseñadores de los sistemas de salud promover ambientes propicios para que los médicos realicen investigación clínica y publiquen sus resultados. Teniendo en mente lo anterior se

presentan cuatro aspectos sobre los cuales se invita a pensar:

- No toda la práctica médica debe convertirse en material de investigación, pero toda investigación debe convertirse en material para tomar decisiones en la práctica clínica.
- El adiestramiento en técnicas para la búsqueda de información y el análisis adecuado de la literatura son alternativas sencillas y económicas que ayudarán a los médicos a refinar sus habilidades de cuestionamiento y de toma de decisiones en favor de la mejor atención a los pacientes. Evidentemente se requiere un adiestramiento básico que permita evaluar la calidad de la información y evitar que esta se acepte sin que medie una reflexión crítica.
- El apoyo a la publicación y divulgación en las revistas médicas locales puede ser un mecanismo para estimular la simbiosis *práctica médica-investigación clínica-publicación*.
- La creación de un ambiente propicio para que los médicos hagan investigación clínica es una oportunidad que tienen los administradores y responsables de los sistemas de salud para que faciliten la generación de conocimiento médico que incida en la calidad de la atención.

Por lo anterior, se sugiere estimular las actividades académicas y de investigación en sesiones de discu-

sión entre los médicos adscritos y los residentes, en las que se transmitan las herramientas de búsqueda de la literatura y análisis crítico de la misma, en función de resolver cuestionamientos que surjan de la práctica médica. Debido a que numerosos centros de salud son también centros de aprendizaje, esta tarea solo implicaría la organización de tiempos, la habilitación de un aula o sala de juntas con equipo informático, acceso a internet y a sistemas de comunicación interactiva que permitan consultar en tiempo real la literatura médica disponible y promover la comunicación entre médicos de distintos centros de atención.

Finalmente, la promoción y el apoyo a las revistas locales se puede lograr si los médicos piden que existan esos espacios para publicar y desde el nivel administrativo se facilita su producción y distribución.

El conocimiento que se genera y no se comparte es conocimiento inútil, pues no existe posibilidad de aplicarlo, reproducirlo y mejorarlo. La publicación es el mecanismo más poderoso para compartir el conocimiento porque, por un lado, obliga a los generadores del mismo a estructurarlo y ordenarlo de tal manera que sea accesible y, por el otro, porque la publicación materializa el conocimiento para su memoria y consulta. La publicación del conocimiento médico, sustentado en la práctica médica y la investigación clínica, es conocimiento útil que permitirá mejorar la calidad de la atención médica y cumplir con la responsabilidad social inherente a la medicina.

Referencias

1. Marder E, Kettenmann H, Grillner S. Impacting our young. *Proc Natl Acad Sci USA*. 2010;107(50):21233. doi: 10.1073/pnas.1016516107
2. McIntyre E, Eckermann SL, Keane M, et al. Publishing in peer review journals. Criteria for success. *Aust Fam Physician*. 2007;36(7):561-2.
3. Jowett SM, Macleod J, Wilson S, et al. Research in primary care: extent of involvement and perceived determinants among practitioners from one English region. *Br J Gen Pract*. 2000;50:387-9. Texto libre en <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1313704/pdf/10897537.pdf>
4. Starfield BL, Macinko J. Contribution of primary care to health systems and health. *Milbank Quarterly*. 2005;83:457-502. doi: 10.1111/j.1468-0009.2005.00409.x.
5. Sullivan F, Butler C, Cupples M, et al. Primary care research networks in the United Kingdom. *BMJ*. 2007;334:1093-4. doi: 10.1136/bmj.39190.648785.80.
6. Soler-González J, Ruiz C, Serna C, et al. The profile of general practitioners (GPs) who publish in selected family practice journals. *BMC Res Notes*. 2011 May 26;4:164. doi: 10.1186/1756-0500-4-164. Texto libre en <http://www.biomedcentral.com/1756-0500/4/164>

I. Diseños de investigación

Juan O. Talavera

La investigación clínica se encarga principalmente del estudio de grupos de sujetos enfermos con el fin de establecer un diagnóstico, estimar un pronóstico e iniciar un tratamiento. Con este objetivo utiliza el método científico desde distintos puntos de vista: el arquitectónico, que se divide en estudios de causa-efecto y de proceso; el metodológico, en el que se incluyen los estudios de ensayo clínico, cohorte —casos y controles— y encuesta; y por objetivos, en el que se ubican los estudios de prueba diagnóstica, pronóstico, tratamiento y estudios de factores de riesgo o agente etiológico. Estos diseños se consideran primarios, es decir, utilizan información obtenida directamente del sujeto en estudio; sin embargo, existen otros que utilizan la información de los estudios primarios, a los cuales se les denomina secundarios o diseños de integración.

Palabras clave

investigación
proyectos de investigación
ensayo clínico

Este artículo fue publicado originalmente en Rev Med Inst Mex Seguro Soc 2011; 49 (1): 53-58 y ha sido revisado para la presente edición.

Introducción

La investigación clínica, conocida como epidemiología clínica —término que bajo el concepto actual fue acotado por Alvan R. Feinstein (previamente había sido utilizado por John R. Paul, para referirse a lo que ahora conocemos como epidemiología social y medicina comunitaria)— se dedica al estudio de grupos de personas con el fin de obtener evidencias para tomar decisiones en el cuidado del paciente; es decir, se encarga del estudio de la estructura y la función de la investigación realizada en sujetos enfermos. Sin embargo, en ocasiones se superpone con la epidemiología clásica y estudia al sujeto antes de que desarrolle la enfermedad. Por otro lado, los conocimientos obtenidos en la epidemiología clínica se aplican en el paciente como ente individual, mientras que en la mayoría de las ocasiones los conocimientos obtenidos en la epidemiología clásica se aplican a un grupo de sujetos.

El método de investigación en la epidemiología clínica es único y corresponde al método científico. No obstante, con fines didácticos se han realizado clasificaciones desde distintos puntos de vista, de los cuales tres son los más frecuentes.

El primero, denominado *arquitectónico*, se basa en la descripción más apegada del suceso real e incluye los estudios de causa-efecto y de proceso. El segundo, conocido como *metodológico*, se caracteriza por jerarquizar la calidad de la información obtenida de los grupos en estudio; abarca el ensayo clínico, el de cohorte —casos y controles— y la encuesta. El tercero utiliza el propósito que conlleva en la práctica clínica cotidiana y lo conocemos como *enfoque por objetivos*; se divide en estudios de diagnóstico, pronóstico, tratamiento y de factores de riesgo o agente causal (causalidad).

Los estudios en los que no se considera la imposición de una maniobra por el investigador, y que por lo tanto no son un experimento sino una observación, respetan los principios del método científico y sustituyen la maniobra experimental por una maniobra natural o impuesta con fines ajenos a la investigación.

Enfoque arquitectónico

Cuando hablamos de estudios de causa-efecto nos referimos al cambio sufrido en el estado basal del sujeto al recibir una maniobra, por ejemplo: al estimar en un paciente previamente sano (estado basal) que sufre un traumatismo craneoencefálico (maniobra observacional), la probabilidad de fallecer o quedar con secuelas (desenlace); o al evaluar en un paciente con cefalea (estado basal) si un analgésico que se le prescribe (maniobra) disminuyó el dolor (desenlace).

Lo anterior significa que los estudios de causa-efecto no solo incluyen la búsqueda de un agente etiológico o factor de riesgo, sino también de factores pronósticos e, incluso, de acciones terapéuticas. Por otro lado, en los estudios de proceso se evalúa la calidad de los procedimientos, ya sea al comparar el procedimiento por analizar con un estándar o con otra ejecución del mismo, por ejemplo: para estimar la sensibilidad y la especificidad del ultrasonido de cuello (procedimiento en estudio), este es comparado en pacientes con obstrucción de la carótida (contra arteriografía de la carótida). En los casos sin estándar de oro, el procedimiento en estudio se contrasta con otra ejecución del mismo estudio en la valoración de la misma lesión por dos radiólogos, con el fin de evaluar la coincidencia más allá de lo esperado por el azar (figuras 1 y 2).

Enfoque metodológico

Con base en la calidad de la información obtenida, el enfoque metodológico intenta jerarquizar los distintos diseños de tal manera que pueda decidirse qué estudio sobre un mismo tema es más confiable al tener una menor probabilidad de cursar con sesgos y, por lo tanto, en cuál deben fundamentarse las decisiones en relación con los pacientes.

Es importante considerar que diseños en niveles jerárquicos inferiores llevados a cabo en forma adecuada pueden superar a otros de niveles superiores pero mal estructurados; aunado a esto, estudios en niveles jerárquicos inferiores pueden ser suficientes para dar respuesta a una pregunta de investigación; además, en no raras ocasiones son los únicos que pueden efectuarse.

En la descripción de los diseños es necesario tener en cuenta cuatro características básicas y la medición de la ocurrencia del desenlace.

Características básicas

1. Imposición o no de una maniobra con fines de investigación. Un estudio se considera experimental si la maniobra fue impuesta por el investigador y observacional cuando dicha maniobra es natural (ejemplo, la presencia de alguna enfermedad) o impuesta con fines ajenos a la investigación (tabaquismo, alcoholismo, etcétera).
2. Seguimiento o no del paciente a través del tiempo. Se considera que un estudio es longitudinal cuando el paciente es evaluado en alguna de sus características de interés a través del tiempo (en más de una ocasión); en la mayoría de las ocasiones se hace referencia al cambio del estado basal al de resultado o desenlace, por ejemplo: el seguimiento de un grupo de médicos sin antecedentes de cardiopatía isquémica (estado basal) durante cinco años y la medición durante dicho periodo de la aparición de enfermedad coronaria (desenlace). La investigación es transversal cuando el paciente es evaluado en forma estacionaria (solo en una ocasión), por ejemplo: la medición de hipertensión en un grupo de pacientes diabéticos en quienes se busca asociar el descontrol metabólico con la hipertensión. Mientras los estudios longitudinales permiten evaluar distintos factores como causantes del cambio del estado basal al subsecuente con la certeza de la temporalidad de exposición a ellos, frecuentemente en los transversales no existe la certeza de la relación temporal, aun cuando se establecen asociaciones entre variables para las que artificialmente se asume cuál es la maniobra y el desenlace.
3. Direccionalidad en la obtención de la información. Un estudio es prolectivo cuando la obtención de la información corresponde tanto al estado basal, la maniobra y el desenlace. Se realiza en tiempo real

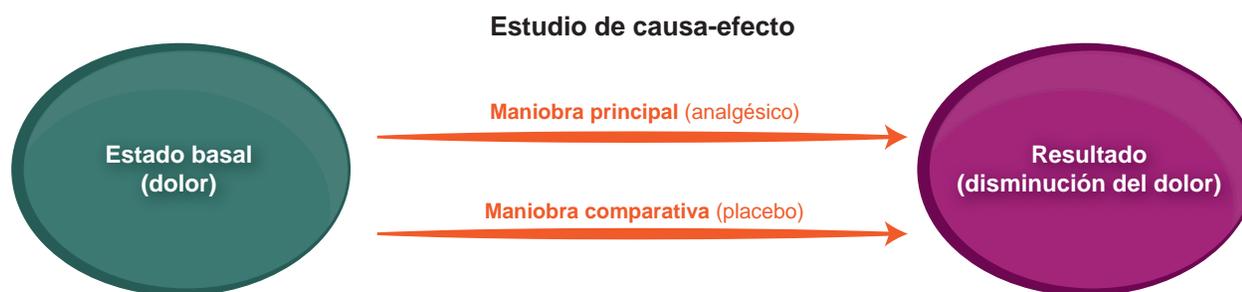


Figura 1 En un estudio de causa-efecto se busca establecer la asociación entre la maniobra y el cambio en el estado basal del sujeto, lo que genera un resultado. Deben considerarse tres componentes: **el estado basal del sujeto**, **la maniobra principal** y **el desenlace o resultado**; de acuerdo con la pregunta, puede ser necesaria o no la maniobra comparativa

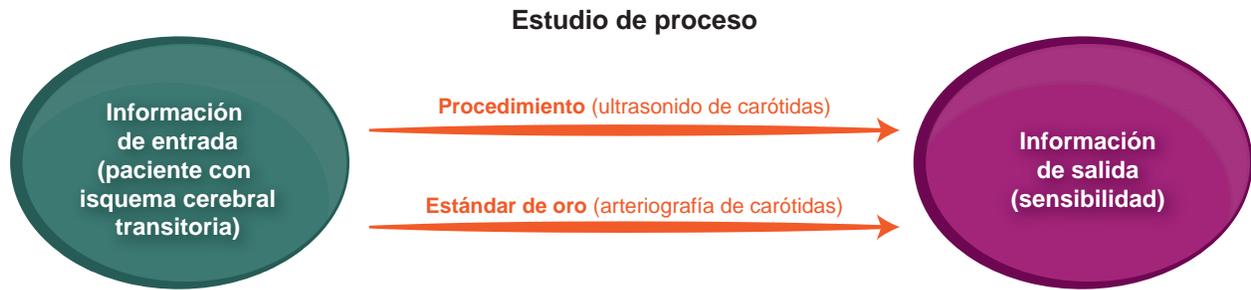


Figura 2 En estudios de proceso se busca evaluar la confiabilidad del procedimiento, para lo cual es necesaria información de entrada (sustrato), la ejecución de un procedimiento a comparar con el estándar de oro o con otra ejecución del procedimiento, lo que da como resultado información de salida

con fines de investigación, es decir, simultáneamente a la exposición de la maniobra y a la ocurrencia del desenlace. Es retrolectivo, cuando la obtención de la información es realizada una vez que la exposición a la maniobra y el desenlace han ocurrido. Existe la posibilidad de que un estudio sea retroproyectivo cuando al momento de obtener la información la maniobra ya ocurrió pero el resultado aún no, y, por lo tanto, su medición se realiza en el momento en que este ocurre (figura 3).

4. Búsqueda o no de asociación entre dos variables. Un estudio se considera descriptivo cuando el objetivo consiste en mostrar el espectro de características del grupo en estudio. Con frecuencia los resultados de estudios descriptivos son utilizados con fines comparativos, por ejemplo: cuando la prevalencia de cierta enfermedad en una población determinada es comparada con la prevalencia de la misma enfermedad en una población analizada previamente. Por otro lado, se trata de un estudio comparativo cuando se busca asociación entre la maniobra y el desenlace, o entre un estándar y la calidad de un producto o procedimiento (cuando se trata de un estudio diagnóstico). Un ejemplo de estudio comparativo es la búsqueda de asociación entre obesidad (maniobra natural) y la resistencia a la insulina (desenlace), o al comparar el diagnóstico ultrasonográfico de colecistitis aguda (procedimiento) con los hallazgos quirúrgicos (estándar de oro).

Medición de la ocurrencia del desenlace

La medición de la frecuencia del desenlace se puede realizar de dos formas de acuerdo con el diseño metodológico:

1. *La incidencia (incidencia acumulada)* se refiere al número de casos nuevos que se presentan en un periodo y en una población determinados; es propia de los estudios con seguimiento, es decir, de las cohortes (ya sea observacionales o experimentales).

Puede tener distintas denominaciones: cuando lo que se estudia es la mortalidad y no la ocurrencia de una enfermedad se le conoce como tasa de mortalidad.

2. *La prevalencia*, o número de casos existentes en un momento determinado en una población determinada, es propia de los estudios transversales, con excepción del estudio de casos y controles.

La proporción caso-control no es una forma de medir la ocurrencia del desenlace sino una simple relación caso-control creada artificialmente.

Diseños básicos

El orden jerárquico, designado por la calidad de la información obtenida, sitúa al *ensayo clínico* en primer lugar, ya que permite obtener la información en forma directa y con control sobre la maniobra, consecuentemente con el menor número de errores. Le sigue la cohorte, luego el estudio de casos y controles y, finalmente, la encuesta.

El *ensayo clínico* se caracteriza por ser un estudio prolectivo y longitudinal, en el que se planea la aplicación de la maniobra (experimental) a la cual se desea atribuir el cambio en el estado basal (comparativo); el ensayo clínico es experimental cuando cuenta con un grupo comparativo, con asignación aleatoria de la maniobra y evaluación cegada del resultado. Sin embargo, en ocasiones no se dispone de un grupo comparativo y es el estado basal la característica que hay que comparar con el resultado (*estudio de antes y después*), o no es posible realizar una asignación aleatoria de la maniobra o una evaluación cegada de la misma, lo que define al ensayo clínico como *casi experimental*. El ensayo clínico puede ser definido como una cohorte experimental, dado que cuenta con todas las características de una cohorte con la asignación de la maniobra. Como se trata de un estudio longitudinal, permite estimar la incidencia como medida de la ocurrencia de la enfermedad.

Direccionalidad en la obtención de la información

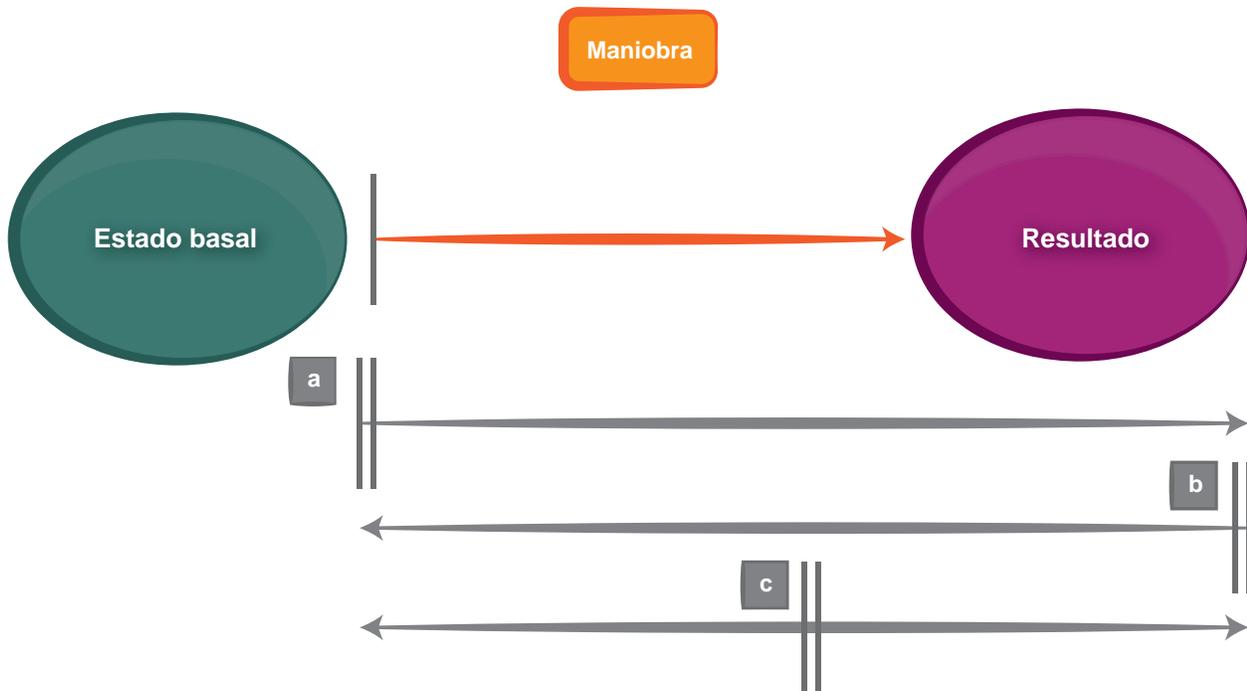


Figura 3 Cuando la captura de la información se inicia en el estado basal antes de la maniobra y el resultado, el estudio se considera prospectivo (a); cuando la captura se realiza una vez acontecidos la maniobra y el resultado se considera retrospectivo (b), y cuando la captura se realiza una vez que ocurrió la maniobra pero antes del resultado se trata de un estudio retroprospectivo (c)

La *cohorte* es el diseño ideal entre los estudios observacionales. Se caracteriza por disponer de un grupo de sujetos elegidos de acuerdo con características comunes en un tiempo determinado y que son seguidos a través del tiempo en alguna de sus características (*longitudinal*), donde la obtención de la información (*proyectiva, retrolectiva o retroproyectiva*) puede o no coincidir con la ocurrencia de la maniobra o el resultado, y en el cual siempre se busca la asociación entre la maniobra y el resultado (*comparativo*). Aun cuando el diseño puede ser retrospectivo, situación en la que se le denomina cohorte histórica, la orientación va de la causa (maniobra) al efecto (resultado). Por ejemplo: puede realizarse un estudio de pronóstico donde interesa saber qué pacientes con enfermedad vascular cerebral van a fallecer durante los primeros días posteriores al evento, para lo cual se revisa la información en los expedientes de todos los pacientes que ingresaron al hospital durante el año previo al estudio; dado que ya ocurrieron la maniobra (características presentes en las primeras horas de la enfermedad vascular cerebral, conocidas como indicadores pronósticos) y el resultado o desenlace (muerte en los primeros siete días del evento), se trata de un estudio retrospectivo; sin embargo, el análisis y la captura de datos deberán hacerse con todos los pacientes, empezando con las manifestaciones clínicas presentes al ingreso y pos-

teriormente midiendo el desenlace. A diferencia del estudio de casos y controles que puede cubrir estas mismas características, en la cohorte se dispone de la información de todos los pacientes que sufrieron la enfermedad vascular cerebral durante el año y, por lo tanto, se dispone de la incidencia del desenlace, mientras que en los estudios de casos y controles no se cuenta con toda la población sino con una proporción artificial de casos-controles, tal como se describe a continuación.

El diseño de *casos y controles* se caracteriza, a la inversa de los diseños mencionados, por ir del efecto a la causa. Se empieza con un grupo de sujetos con el desenlace de interés (*resultado*) que corresponde a los casos y se selecciona un grupo testigo que no sufrió el desenlace (*controles*), posteriormente se busca la asociación entre la maniobra y el desenlace (*comparativo*). Por lo tanto, se trata de un estudio *retrospectivo y observacional*. Existen controversias en cuanto al seguimiento o no de las variables, por lo cual algunos autores lo consideran un estudio transversal, ya que se obtiene toda la información en un solo tiempo, mientras que para otros es *longitudinal* debido a que es factible hacer una recapitulación de la temporalidad de la maniobra hasta el momento del desenlace. En este diseño no existe una medición de la ocurrencia del desenlace, simplemente hay una relación caso-control creada artificialmente.

Cuadro I Diseños de acuerdo con el enfoque metodológico

Diseño	EXP/OBS	LONG/TRANS	PROL/RETROL	COMP/DESC	MEDICIÓN
Ensayo clínico	Experimental	Observacional	Prolectivo	Comparativo	Incidencia
Cohorte	Observacional	Longitudinal	Prol/Retrol/RP	Comparativo	Incidencia
Casos y controles	Observacional	Long/Trans	Retrolectivo	Comparativo	Prop. C-C
Encuesta	Observacional	Transversal	Retrolectivo	C/D	Prevalencia

El enfoque metodológico considera cuatro características: 1. La imposición o no de la maniobra con fines de investigación: estudio experimental (EXP) u observacional (OBS), respectivamente. 2. Seguimiento (LONG) o no (TRANS) del paciente a través del tiempo. 3. Direccionalidad en la obtención de la información: prolectivo (PROL), retrolectivo (RETROL) y retroprolectivo (RP). 4. La búsqueda o no de asociación de dos o más variables: comparativo (C) y descriptivo (D), respectivamente. La medición de la ocurrencia del desenlace (MEDICIÓN), ya sea a través de incidencia, prevalencia o simplemente la proporción caso-control (prop. C-C)

La *encuesta* es entre los diseños observacionales el más sencillo pero también el más limitado en sus aseveraciones; se lleva a cabo en una muestra representativa de la población en estudio y el objetivo más común es delinear las características de la población (*descriptivo*); sin embargo, también puede utilizarse para establecer una asociación entre dos o más variables (*comparativo*). Frecuentemente es imposible determinar si la maniobra precede al resultado, ya que la obtención de la información ocurre una vez que se han sucedido la maniobra y el resultado (*retrolectivo*) y a un solo tiempo (*transversal*). A diferencia del estudio de casos y controles no existe una proporción predeterminada del número de casos y controles, de hecho no existe una selección de la población a partir del desenlace, sino que una vez seleccionada la población (por cualesquiera que sean los criterios) se mide la exposición a la maniobra, en este caso observacional, y el desenlace. Por lo tanto, lo que se obtiene es la prevalencia del desenlace.

En el cuadro I se hace un resumen de las características propias de cada diseño. Cabe mencionar que existen combinaciones de estos diseños y en ocasiones es difícil definirlos.

Enfoque por objetivos

El enfoque basado en la práctica clínica es con el que estamos más habituados; además, en él es posible distinguir la mayor diferencia entre la epidemiología clínica y la epidemiología clásica. En la epidemiología

clínica, que estudia grupos de pacientes, el objetivo principal es resolver un problema ya existente en un grupo de personas, para lo cual debe establecerse un diagnóstico (estudio diagnóstico), estimar un pronóstico (estudios de pronóstico) e iniciar una maniobra terapéutica (ensayo clínico experimental o casi experimental). Sin embargo, como se mencionó en párrafos anteriores, es común que la epidemiología clínica se sobreponga a la clásica y aborde problemas de factores de riesgo, como los cardiovasculares (estudio de factores de riesgo o agente etiológico, este último cuando el agente es único).

Estudios complementarios

Hasta el momento hemos mencionado solo estudios que ocupan información primaria, sin embargo existe un grupo denominado “estudios de integración”, caracterizado por conjuntar los datos obtenidos en los estudios primarios. Estos están formados por cuatro diseños: estudios de revisión (metaanálisis y revisiones sistemáticas), lineamientos de práctica clínica, análisis de decisión y análisis económicos.

Agradecimientos

A los doctores Niels H. Wachter-Rodarte, Susana Castañón-Robles, Rodolfo Rivas-Ruiz y Jorge Salmerón-Castro por sus sugerencias, que permitieron mejorar en lo posible el presente artículo.

Bibliografía

1. Cañedo DL. Investigación clínica. México: Interamericana; 1987.
 2. Feinstein AR. Clinical epidemiology. The architecture of clinical research. Philadelphia: WB Sanders; 1985.
 3. Feinstein AR. Directionality and scientific inference. *J Clin Epidemiol*. 1989;42:829-33.
 4. Feinstein AR. Scientific standards in epidemiologic studies of the menace of daily life. *Science*. 1988;242:1257-63.
 5. Hernández-Ávila M, Garrido-Latorre F, López-Moreno S. Diseño de estudios epidemiológicos. *Salud Publica Mex*. 2000;42:144-54.
 6. Hughes M, Williams P. Challenges in using observational studies to evaluate adverse effects of treatment. *NEJM*. 2007;356:1705-7.
 7. Hulley S, Cummings S. Designing clinical research. Philadelphia, USA: Lippincott Williams & Wilkins; 1988.
 8. Kelsey J, Whittemore A, Evans A, Thompson W. Methods in observational epidemiology. Second edition. New York, USA: Oxford University Press; 1996.
 9. Meinert C. Clinical trials: design, conduct and analysis. New York, USA: Oxford University Press; 1986.
 10. Moreno A, Valle C, Romero G. Epidemiología clínica. Segunda edición. México: Interamericana-McGraw-Hill; 1994.
 11. Petitti D. Meta-analysis, decision analysis and cost-effectiveness analysis: methods for quantitative synthesis in medicine. Second edition. New York, USA: Oxford University Press; 2000.
 12. Schlesselman J. Case control studies: design, conduct, analysis. New York, USA: Oxford University Press; 1982.
 13. Wachter N, Lifshitz A. Qué es la epidemiología clínica y para qué le sirve al clínico. *Rev Med IMSS*. 1989;27:171-4.
 14. Walker AM. Observation and inference. An introduction to the methods of epidemiology. Chestnut Hill, MA: Epidemiology Resources Inc.; 1991.
 15. Weiss NS. Scientific standards in epidemiologic studies. *Epidemiology*. 1990;1:85-6.
- Lecturas recomendadas de ejemplos**
- Casos y controles
16. Cruz-Anguiano V, Talavera J, Vázquez L, Antonio A, Castellanos A, Lezana M, *et al*. The importance of quality of care in perinatal mortality: a population-based case-control study in Chiapas, Mexico. *Arch Med Res*. 2004;35:554-62.
- Cohorte
17. Brea-Andrés E, Aburto-Gudiño E, Vázquez-Estupiñán F, Nellen-Humel H, Talavera-Piña JO, Wachter-Rodarte N, *et al*. Incidencia de delirium y morbilidad asociada en medicina interna. *Acta Psiquiátrica y Psicológica de América Latina* 2000;46:359-62.
- Diagnóstico
18. Talavera J, Wachter N, Laredo F, López A, Martínez V, González J, *et al*. A rating system for prompt clinical diagnosis of ischemic stroke. *Arch Med Res*. 2000;31: 576-84.
- Encuesta
19. Gómez-Díaz R, Martínez-Hernández A, Aguilar-Salinas C, Violante R, Alarcón A, *et al*. Percentile distribution of the waist circumference among Mexican pre-adolescents of a primary school in México City. *Diabetes Obes Metab*. 2005;7:716-21.
- Ensayo clínico controlado
20. González-Ortiz M, Guerrero-Romero JF, Violante-Ortiz R, Wachter-Rodarte N, Martínez-Abundis E, Aguilar-Salinas C, *et al*. Efficacy of glimepiride/metformin combination versus glibenclamide/metformin in patients with uncontrolled type 2 diabetes mellitus. *J Diabetes Complications*. 2009;23: 376-9.
- Estudios de proceso
21. Gómez R, Aguilar-Salinas CA, Morán-Villota S, Barradas-González R, Herrera-Márquez R, Cruz M, *et al*. Lack of agreement between the revised criteria of impaired fasting glucose and impaired glucose tolerance in children with excess body weight. *Diabetes Care*. 2004;27:2229-33.
 22. Pérez-Cuevas R, Reyes-Morales H, Flores-Hernández S, Wachter-Rodarte N. Efecto de una guía de práctica clínica para el manejo de la diabetes tipo 2. *Rev Med Inst Mex Seguro Soc*. 2007;45(4):353-60.
- Pronóstico
23. Cruz M, Maldonado-Bernal, C, Mondragón-González R, Sánchez-Barrera, Wachter N, Carvajal-Sandoval, *et al*. Glycine treatment decreases proinflammatory cytokines and increases interferon-g in patients with type 2 diabetes. *J Endocrinol Invest*. 2008;31:694-9.
- Riesgo
24. Cruz M, García-Macedo I, García-Valerio Y, Gutiérrez M, Medina-Navarro R, Durán G, *et al*. Low adiponectin levels predict type 2 diabetes in Mexican children. *Diabetes Care*. 2004;27:1451-3.
- Tratamiento
25. Nellen H, Flores G, Wachter N. Treatment of human immunodeficiency virus enteropathie with a gluten-free diet. *Arch Intern Med*. 2000;160:244.

II. Estudios de proceso (prueba diagnóstica)

Juan O. Talavera, Niels H. Wachter-Rodarte, Rodolfo Rivas-Ruiz

Una prueba diagnóstica tiene como fin establecer la presencia de salud o enfermedad, incluso, graduar el grado de afección. Las pruebas diagnósticas suelen evaluarse matemáticamente. Así, se estiman la sensibilidad y la especificidad, una vez conocida la existencia o no de enfermedad; en la práctica clínica suele actuarse en forma inversa: de la positividad o negatividad de una prueba hacia la presencia o no de la enfermedad y, por lo tanto, se utilizan los valores predictivos positivos y negativos. Las estrategias matemáticas permiten cuantificar lo observado, pero se requiere juicio clínico para determinar la calidad de esa observación; en consecuencia se deben considerar algunas características: a) la selección bajo los mismos criterios para casos y testigos; b) la inclusión de todo el espectro de severidad de la enfermedad (procurando que todos los estratos cuenten con un número importante de sujetos); c) la interpretación del estándar de oro y de la prueba en estudio debe ser a ciegas y por expertos; d) la interpretación de los resultados debe mostrar la aplicación de prueba en la práctica cotidiana; e) se debe comprobar la reproducibilidad de la prueba. No debe olvidarse que habitualmente se atiende a un paciente a la vez; por lo tanto, es indispensable un conocimiento pleno del desempeño de la prueba diagnóstica y considerar la parte clínica para su correcta aplicación.

Palabras clave

investigación
proyectos de investigación
técnicas y procedimientos diagnósticos

Este artículo fue publicado originalmente en Rev Med Inst Mex Seguro Soc 2011; 49 (2): 163-170 y ha sido revisado para la presente edición.

Introducción

En la parte 1 de esta serie [Rev Med Inst Seguro Soc 2011; 49(1):53-58] se mencionaron los distintos enfoques desde los cuales se abordan los problemas clínicos: *enfoque arquitectónico*, basado en el fenómeno natural; *enfoque metodológico*, basado en la jerarquía de la información; *enfoque clínico*, basado en los objetivos de la práctica médica. Se desglosaron las características principales del enfoque metodológico y, además, se mencionaron los estudios de integración.

Sin embargo, en la práctica clínica, las preguntas suelen estar relacionadas con la necesidad de establecer un diagnóstico o atribuir causalidad, ya sea a través de un estudio pronóstico, de tratamiento o al intentar identificar lo que provocó determinada alteración o enfermedad. Es aquí donde el enfoque arquitectónico se empata con el enfoque por objetivos.

Entre los estudios de proceso, de acuerdo con el enfoque arquitectónico se encuentra la prueba diagnóstica (enfoque por objetivos). Asimismo, dentro de los estudios de causalidad se sitúan los de pronóstico, de tratamiento y de factores de riesgo o agente causal (enfoque por objetivos). En el presente artículo describiremos las herramientas más comúnmente utilizadas en las pruebas diagnósticas.

En la práctica clínica, una prueba diagnóstica tiene como objetivo identificar la condición de salud o de enfermedad del sujeto en estudio. Frecuentemente, ante una enfermedad permite establecer la severidad del padecimiento, por ejemplo: en un paciente con déficit neurológico súbito la tomografía permite definir el diagnóstico (accidente vascular cerebral isquémico); mientras que si ya se cuenta con el diagnóstico, la tomografía hace posible conocer la extensión de la lesión.

El uso de las matemáticas durante el proceso diagnóstico tiene como fin estimar el grado de eficacia y certidumbre de las pruebas en la práctica clínica. A continuación se describen las principales características de toda prueba diagnóstica, tanto con el uso de datos clínicos como de hallazgos de laboratorio y de gabinete.

Características de una prueba diagnóstica

La forma de evaluar la eficacia de una prueba diagnóstica depende del tipo de dato (variable) que se va a utilizar. Por lo tanto, resulta importante identificar el tipo de variable. Las variables básicas son las que conocemos como *cualitativas de tipo nominal o dicotómicas*, y se refieren a aquellas en las que solo nos percatamos de su presencia o en las que solo existen dos opciones (por ejemplo nacionalidad, presencia o no de enferme-

dad, hombre o mujer). Las variables *cualitativas ordinales* son aquellas en las que solo se puede identificar qué lugar ocupan las características evaluadas en el grupo pero no conocemos de qué tamaño es la diferencia entre unas y otras (por ejemplo: el grado de afección de una enfermedad —leve, moderada o grave—, o la intensidad de un dato clínico identificado con cruces, en cuyo caso, aun cuando se reconoce que + es menor que ++ y en consecuencia menor que +++, no puede afirmarse que ++ sea el doble de +). Y, finalmente, las *variables cuantitativas*, aquellas en las que se conoce la distancia entre una intensidad y otra; y esta distancia entre valores siempre es equidistante. Se denominan *discretas o discontinuas* cuando no pueden fraccionarse (por ejemplo: cuántos hijos tiene una familia [0, 1, 2, 3]), y *continuas* cuando pueden identificarse fracciones entre uno y otro valor (por ejemplo: peso de 52.0 kg, 52.2 kg o 52.250 kg).

La sensibilidad y la especificidad son características propias de toda prueba diagnóstica e indican su eficacia. La *sensibilidad* se refiere a la proporción de individuos enfermos que tienen la prueba positiva. La *especificidad* se refiere a la proporción de individuos no enfermos que tienen la prueba negativa.

El cálculo de sensibilidad y especificidad utiliza *datos nominales o dicotómicos* y se basa en el uso de una tabla de 2 x 2, en la que se contrasta el dato a prueba contra el diagnóstico definitivo obtenido a través de un parámetro ideal llamado *estándar de oro*,

que representa la prueba con mayor confiabilidad para demostrar una enfermedad, por ejemplo: el resultado histopatológico (seminoma de testículo), los hallazgos quirúrgicos (colecistitis) o la interpretación de los estudios de imagen (accidente vascular cerebral por tomografía o resonancia magnética) o los estudios de imagen intervencionista (tipo de cardiopatía congénita por cateterismo cardiaco) o de laboratorio (insuficiencia renal por depuración de creatinina).

En la figura 1 se muestra el cálculo de la sensibilidad y la especificidad de la rigidez de nuca para el diagnóstico de hemorragia subaracnoidea, entre pacientes con déficit neurológico de inicio súbito de probable causa vascular. Se observa una sensibilidad de 59 % con una especificidad de 94 %, lo que significa que 59 % de los pacientes con hemorragia subaracnoidea pueden presentar rigidez de nuca y entre los que no tienen hemorragia subaracnoidea, 94 % no presenta rigidez de nuca.

Los cálculos de sensibilidad y especificidad se dirigen a partir de la presencia o ausencia de una determinada enfermedad, hacia la probabilidad de presentar o no cierto dato. Sin embargo, en la clínica, el acercamiento suele ser a la inversa: se parte del resultado de la prueba positiva o negativa hacia la probabilidad de tener o no determinada enfermedad. Este tipo de orientación corresponde a lo que conocemos como valores predictivos. El *valor predictivo positivo* repre-

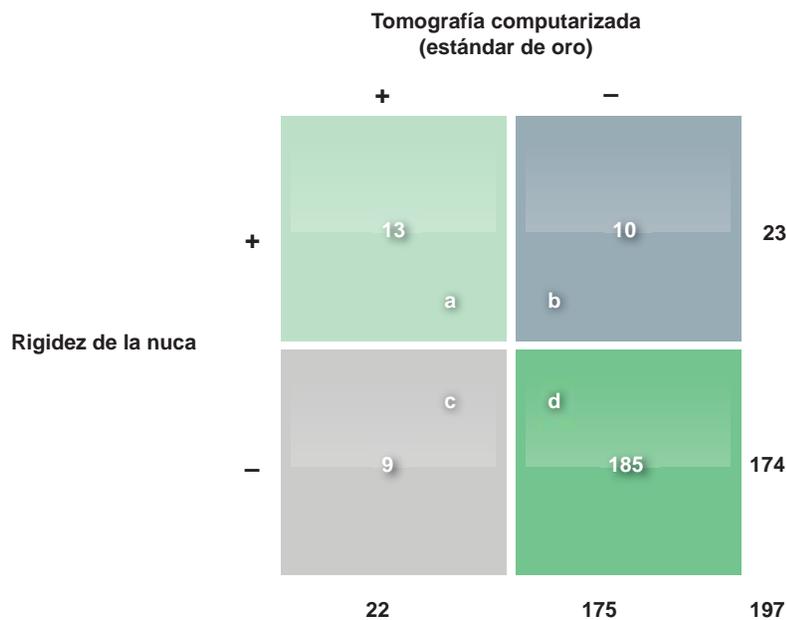


Figura 1 Estimación de sensibilidad y especificidad de la rigidez de nuca en el diagnóstico de la hemorragia subaracnoidea

Sensibilidad	$a/a + c = 0.59$ (59 %)	Especificidad	$d/b + d = 0.94$ (94 %)
Falsos positivos	$b/b + d = 0.6$ (6 %)	Falsos negativos	$c/a + c = 0.41$ (41 %)
Valor predictivo positivo	$a/a + b = 0.57$ (57 %)	Valor predictivo negativo	$d/c + d = 0.95$ (95 %)
Prevalencia	$a + c / a + b + c + d = 0.11$ (11 %)	Certeza diagnóstica	$a + d / a + b + c + d = 0.90$ (90 %)

senta la probabilidad que tiene un paciente con cierta prueba positiva (signo, síntoma, resultado de laboratorio, de gabinete o algún índice) de cursar con una enfermedad determinada; el *valor predictivo negativo* representa la probabilidad que tiene un paciente con cierta prueba negativa de estar libre de una enfermedad determinada.

En la figura 1 se observa un valor predictivo positivo de 57 % y un valor predictivo negativo de 95 %; esto significa que entre los pacientes con cuadro clínico de accidente vascular cerebral, un sujeto con rigidez de nuca tiene una probabilidad de 57 % de presentar hemorragia subaracnoidea, mientras que un paciente sin rigidez de nuca tiene una probabilidad de 95 % de no tener hemorragia subaracnoidea.

Mientras los valores de sensibilidad y especificidad se consideran constantes, lo cual no es cierto como mencionaremos más adelante, los valores predictivos se ven afectados por la prevalencia de la enfermedad. Por ejemplo, en la figura 2, donde solo incrementamos la prevalencia de la enfermedad de 11 a 56 %, respetando la proporción de sujetos enfermos con la prueba positiva y con la prueba negativa, se puede ver cómo la sensibilidad y la especificidad se conservan, mientras los valores predictivos se modifican: el valor predictivo positivo es de 93 % y el valor predictivo negativo de 65 %. De tal forma que un incremento en la prevalencia provoca un incremento en el valor predictivo positivo, con una disminución del valor predictivo negativo (una prueba positiva en una población con alta prevalencia de la enfermedad prácticamente

hace el diagnóstico, sin embargo, una prueba negativa no lo descarta); al contrario, una disminución de la prevalencia provoca un incremento en el valor predictivo negativo y una disminución del valor predictivo positivo (una prueba negativa en una población con una baja prevalencia de la enfermedad prácticamente descarta la enfermedad).

Si la prevalencia de la enfermedad en la población de la cual se obtuvieron los valores predictivos de la prueba diagnóstica es diferente a la prevalencia de la enfermedad en nuestra población, no es posible hacer uso de dichos valores predictivos. Sin embargo, el teorema de Bayes permite estimar los valores predictivos con el uso de la sensibilidad y especificidad de la prueba, y la prevalencia de la entidad en estudio en nuestra población. En el cuadro I se muestra cómo el incremento de la prevalencia de 11 a 56 % provoca un incremento en el valor predictivo positivo de 57 a 94 %. En este ejemplo se observa claramente cómo una prueba positiva en una población con prevalencia baja (11 %) tiene una probabilidad aproximada de 50 % de tener la enfermedad, mientras que con una prevalencia alta (56 %) prácticamente hace el diagnóstico.

Otra estrategia práctica para estimar la probabilidad de la enfermedad ante una prueba positiva, pero a distintos valores de prevalencia, es el uso del nomograma de Fagan y la razón de verosimilitud (RV, *likelihood ratio*). La RV positiva (RVP) se obtiene del cociente sensibilidad/1-especificidad. Por su parte, la RV negativa (RVN) se obtiene del

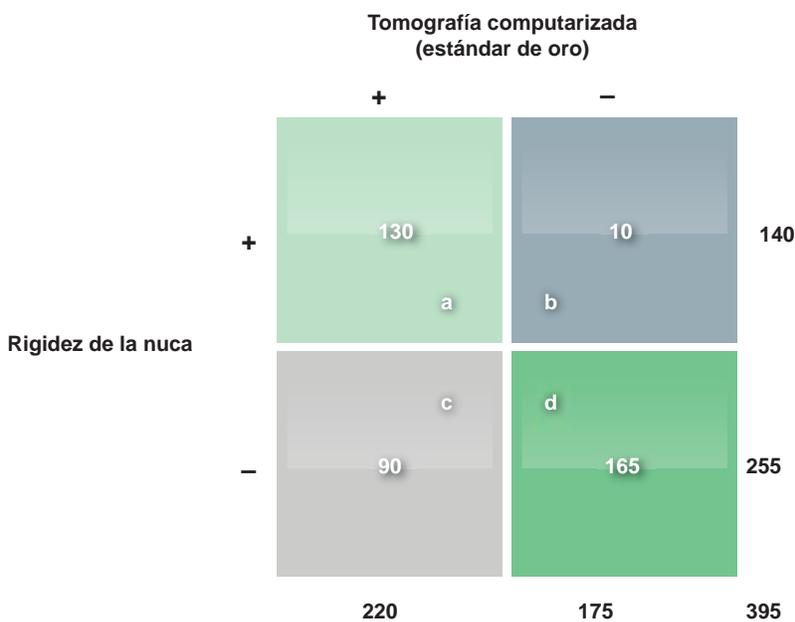


Figura 2 Modificación de los valores predictivos de la rigidez de nuca en el diagnóstico de hemorragia subaracnoidea con el cambio en la prevalencia

Sensibilidad = 59 %

Especificidad = 94 %

Valor predictivo positivo = 93 %

Valor predictivo negativo = 65 %

Prevalencia = 56 %

cociente de 1-sensibilidad/especificidad. El nomograma de Fagan se divide en tres partes. En la primera columna está la posibilidad preprueba (prevalencia). En medio se encuentran los valores de la RV y en la última columna la probabilidad posprueba. La probabilidad posprueba para una RVP se refiere a la probabilidad de obtener un resultado positivo cuando la prueba es positiva y corresponde al VPP; la probabilidad posprueba para una RVN se refiere a la probabilidad de obtener un resultado positivo cuando la prueba es negativa, equivalente a 1-VPN. En la figura 3 se encuentran los ejemplos para una prevalencia de 11 y 56 %.

Se dijo que la sensibilidad y especificidad de una prueba no son dependientes de la prevalencia de la enfermedad, sin embargo, los valores varían de acuerdo con el grado de severidad predominante de la enfermedad en el grupo en estudio. Por ejemplo, diagnosticar un cáncer pulmonar en un estadio avanzado con una radiografía de tórax es sencillo y rara vez pasará desapercibido, o sea, rara vez existirán falsos negativos y la sensibilidad será elevada, sin embargo, difícilmente se detectará si intentamos diagnosticarlo en sujetos asintomáticos, en estadio temprano, lo cual provocará un alto porcentaje de falsos negativos y una baja sensibilidad. Por lo tanto, considerar que la sensibilidad obtenida de una prueba en una población es aplicable a otra población implica que la distribución de la severidad de la enfermedad es igual en ambos lugares, ya que si en el primero predomina la proporción de sujetos en estadios avanzados, la sensibilidad será alta, y si en el segundo predomina un estadio inicial, la sensibilidad será baja. Contar con los mismos criterios de inclusión entre diferentes estudios de poblaciones distintas no garantiza que la distribución de los sujetos conserve una proporción similar de sujetos en cada estadio de la enfermedad y, consecuentemente, la sensibilidad puede ser distinta.

Uso de datos ordinales y cuantitativos

A diferencia de los datos nominales, cuando la prueba en estudio corresponde a datos ordinales o cuantitativos (con más de un punto de corte), es necesaria la elaboración de una curva ROC (*receiver operator characteristic*), la cual permite determinar en cuál de los puntos de corte se obtiene la certeza diagnóstica más elevada.

En la figura 4 se observan los diferentes rangos de valores de fosfoquinasa de creatina en líquido cefalorraquídeo expresados en $\mu\text{U/mL}$, con sus respectivas frecuencias, y se esquematiza el cálculo de sensibilidad y especificidad de acuerdo con los distintos puntos de corte a través de la elaboración de tablas de 2×2 .

En estas tablas se construyen intervalos con los distintos valores de la prueba en estudio y se tabulan en dos columnas; la primera muestra las frecuencias de sujetos con la enfermedad en cada uno de los intervalos y la segunda muestra la frecuencia de sujetos sin la enfermedad dentro de esos mismos intervalos. Los valores más alterados se sitúan arriba (primeros intervalos) y los menos alterados abajo. Se calcula el porcentaje acumulado hacia arriba y hacia abajo de cada punto de corte, en las dos columnas. En la columna de enfermos a partir del punto de corte hacia arriba se estima la sensibilidad y en la columna de los controles el porcentaje de falsos positivos (1-especificidad).

Con los valores de sensibilidad y el porcentaje de falsos positivos se grafican los resultados: los valores de la sensibilidad en el eje de la ordenada (Y), y la proporción de falsos positivos (1-especificidad) en el eje de la abscisa (X); un valor de especificidad de 90 % corresponde a 10 % de falsos positivos (figura 5). El mejor punto de corte corresponde dentro de la curva ROC al punto más cercano del ángulo superior izquierdo de

Cuadro I Teorema de Bayes

$p(E+/P+) = \frac{p(P+/E+) p(E+)}{p(P+/E+) p(E+) + p(P+/E-) p(E-)}$	
$p(E+/P+) =$	probabilidad <i>a posteriori</i> de tener cierta enfermedad ante una prueba positiva; corresponde al <i>valor predictivo positivo</i> (VPP).
$p(P+/E+) =$	probabilidad de resultado positivo de la prueba cuando el paciente padece la enfermedad; corresponde a la <i>sensibilidad</i> .
$p(E+) =$	probabilidad <i>a priori</i> de tener la enfermedad de acuerdo con la población a la que se pertenece; corresponde a <i>prevalencia</i> .
$p(P+/E-) =$	probabilidad de un resultado positivo de la prueba cuando el paciente no padece la enfermedad; equivale a los <i>falsos positivo</i> o <i>1-especificidad</i> .
$p(E-) =$	probabilidad <i>a priori</i> de no tener la enfermedad y corresponde a <i>1-prevalencia</i> [$1 - p(E+)$].
Prevalencia	11 % 56 %
Sensibilidad	59 % 59 %
Especificidad	94 % 94 %
VPP	57 % 94 %
VPN	95 % 64 %

La estimación del valor predictivo negativo se realiza de la misma forma invirtiendo los signos de la fórmula [por ejemplo: $p(E+/P+)$ cambia a $p(E-/P-)$]

la curva, o al punto dentro de la tabla que contiene el menor valor de $b + c$ (valores que pertenecen a la suma de falsos positivos y falsos negativos) o el mayor valor de $a + d$ (valores que pertenecen a la suma de verdaderos positivos y verdaderos negativos). En este caso, el punto de corte es de $\geq 16 \mu\text{U/mL}$, el cual permite clasificar correctamente como enfermo o sano a 79.6 % de los pacientes, con sensibilidad de 61.5 % y especificidad de 96.5. Sin embargo, de acuerdo con la utilidad que se le dé a la prueba puede seleccionarse más de un punto: donde se privilegia a la sensibilidad o a la especificidad (mayor valor predictivo negativo o positivo).

Existen casos en que no solo la prueba en estudio contiene más de dos estratos, sino incluso el estándar de oro. En estos casos se puede estimar el porcentaje de aciertos y de errores. En la figura 6 se contrasta el diagnóstico clínico de tromboembolia pulmonar considerando como estándar de oro el diagnóstico por gammagrama ventilatorio perfusorio; el porcentaje de aciertos corresponde a las casillas donde coinciden

tanto el diagnóstico clínico como el estándar de oro, es decir, en las celdas a, e, i ($40 + 90 + 70$), siendo este de 66.66 %, y nuestro porcentaje de errores que sobrestiman el diagnóstico corresponde a las celdas b, c, f ($30 + 20 + 10$), siendo este de 20 %; finalmente, el porcentaje de error que subestima el diagnóstico lo forman las celdas d, g, h ($7 + 30 + 3$), siendo éste de 13.33 %. Sin embargo, existe la posibilidad de que se deseara manejar el desenlace con solo dos posibilidades; en tal caso se podrían agrupar los gammagramas con baja y moderada probabilidad y hablar de un gammagrama con alta probabilidad de embolia pulmonar o sin alta probabilidad, o agrupar los de alta y mediana probabilidad y dejar en un solo grupo los de baja probabilidad. Este mismo procedimiento puede realizarse con la escala clínica, de tal forma que al tener solo cuatro celdas se puedan utilizar los estimadores tradicionales de la utilidad de una prueba diagnóstica, o conservar los tres estratos de nuestra prueba en estudio y calcular una curva ROC.

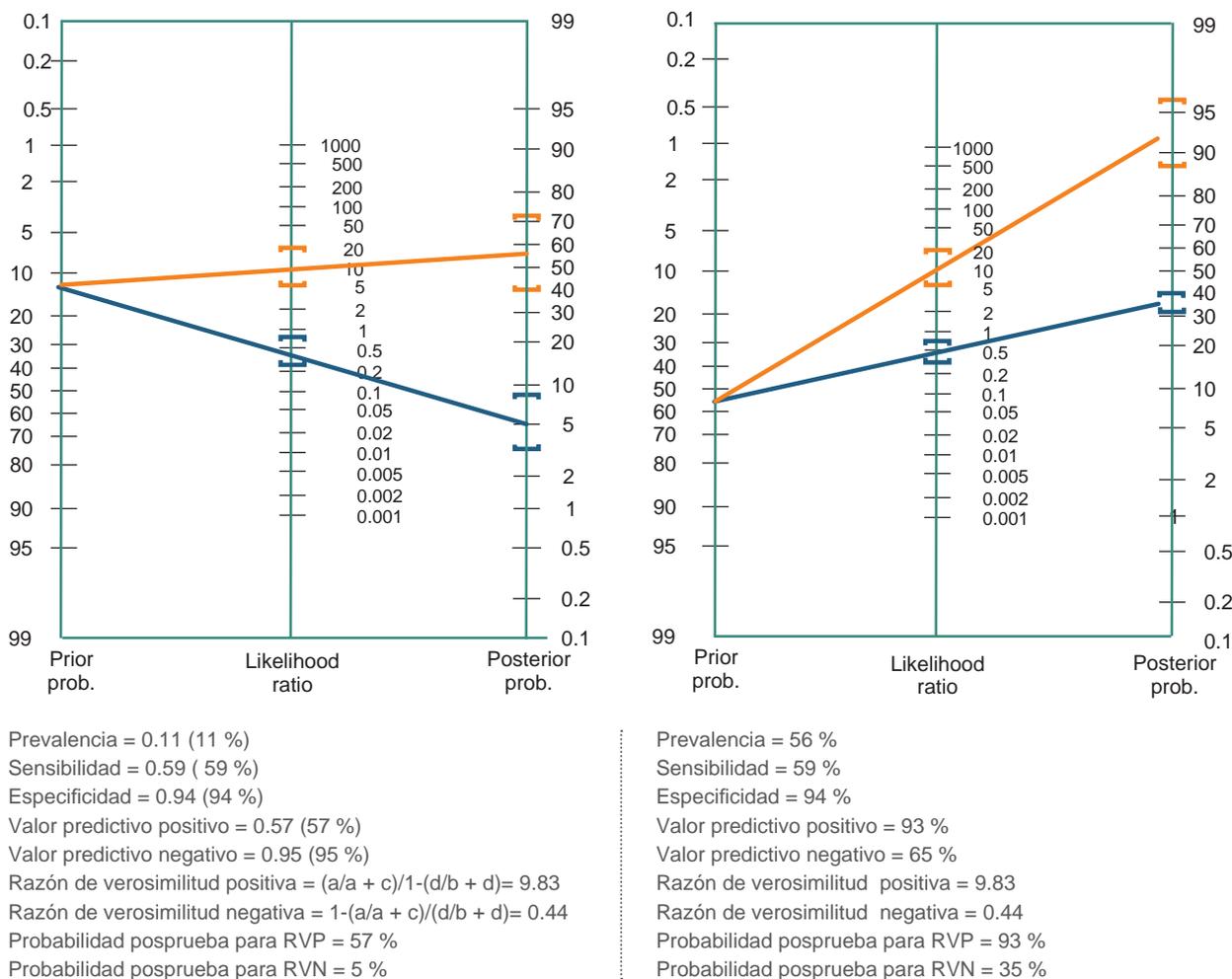
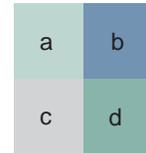


Figura 3 Uso del nomograma de Fagan y razones de verosimilitud

Creatinfosfocinasa (μU/mL)	Daño		≥ 76	≥ 61	≥ 46	≥ 31	≥ 16	≥ 0		
	Orgánico	Metabólico								
≥ 76	2	0	2	0						
61-75	1	0	25	27	3	0				
46-60	2	0		24	27	5	0			
31-45	6	0			22	27	11	0		
16-30	6	1				16	27	17	1	
≤ 15	10	26					10	26	27	27
								0	0	



Sensibilidad (a/a + c)	%	7	11	19	41	63	100
Especificidad (d/b + d)	%	100	100	100	100	96	0
Falsos positivos	%	0	0	0	0	4	100
Errores		25	24	22	16	11	27

Figura 4 Estimación de sensibilidad y especificidad en distintos puntos de corte para identificar daño orgánico en pacientes con coma

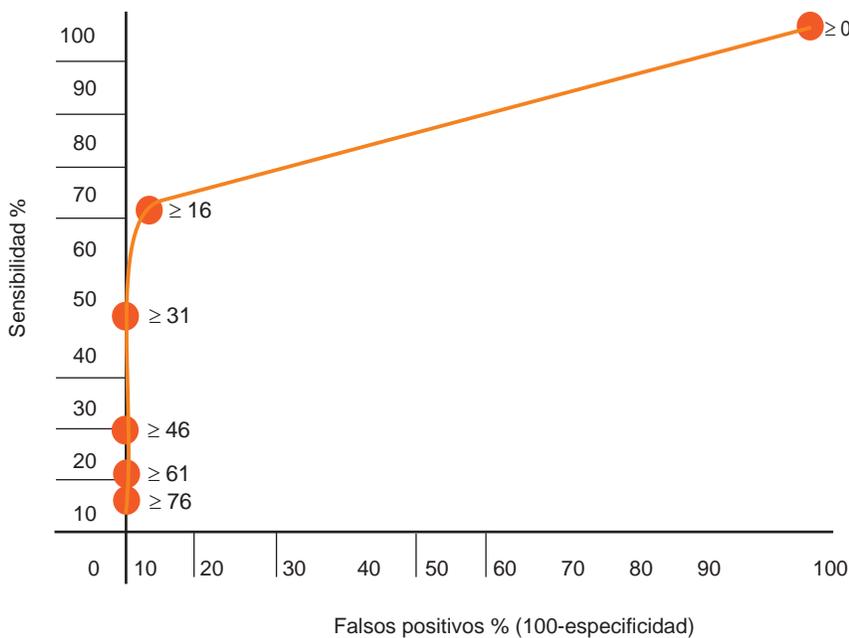


Figura 5 Curva ROC

**Gammagrama ventilatorio perfusorio
(estándar de oro)**

		Alta	Moderada	Baja	
Diagnóstico clínico	Alta	40	30	20	90
		a	b	c	
		7	90	10	
	Moderada	d	e	f	
		3	30	70	103
		Baja	g	h	
	50		150	100	

Figura 6 Valoración de la eficacia del diagnóstico clínico para identificar trombo-embolia pulmonar

Celdas a, e, i = aciertos, en este caso de 66.66 %
 Celdas b, c, f = errores que sobrestiman el diagnóstico, en este caso de 20 %
 Celdas d, g, h = errores que subestiman el diagnóstico, en este caso de 13.33 %

Aplicaciones de la prueba diagnóstica

Debe quedar claro que la aplicación de una prueba puede tener distintos objetivos:

1. Si se desea una prueba de escrutinio se debe utilizar una prueba con alta sensibilidad, aun cuando tenga baja especificidad (por ejemplo: tiras reactivas para medir glucosa en sangre, para buscar sospechosos de diabetes mellitus).
2. Si se desea descartar determinada enfermedad, se utiliza una prueba con alta sensibilidad y de ser posible alta especificidad (valor predictivo negativo elevado, por ejemplo: ELISA para VIH), pues si bien cuando es positiva no hace diagnóstico, cuando es negativa sí lo descarta.
3. Si se desea confirmar un diagnóstico en un paciente con sospecha de que cursa con una determinada enfermedad, se utiliza una prueba con alta especificidad y de ser posible alta sensibilidad (valor predictivo positivo elevado, por ejemplo: Western-Blot para VIH), pues aunque no siempre descarta la enfermedad cuando es negativa, si es positiva hace el diagnóstico.

Solicitar exámenes en exceso, justificados o no, genera resultados anormales incluso en personas normales, lo que a su vez desencadena una cascada de exámenes de mayor costo y riesgo, además de angustia para el paciente.

Errores comunes al elaborar una prueba diagnóstica

Ya se mencionó la forma de estimar la eficacia de una prueba diagnóstica y cómo hacer uso de ella, sin embargo, se deben vigilar posibles causas de errores sistemáticos entre los que destacan dos principalmente:

1. La selección inadecuada de pacientes.
2. La interpretación inadecuada tanto de la prueba en estudio como del estándar de oro.

La selección de un espectro inadecuado de pacientes puede suceder desde el punto de vista clínico o patológico. Por ejemplo: no es igual la eficacia del estudio citológico de esputo para detectar cán-

cer pulmonar en un paciente con tabaquismo intenso y prolongado, pérdida de peso, tos con expectoración hemoptoica y disnea, que en un paciente que solo tiene tos y expectoración blanquecina; tampoco es igual la eficacia de la medición del antígeno carcinoembrionario para detectar cáncer de colon en un paciente con estadio A de Dukes, en comparación con un paciente con estadio D. Es indispensable que toda prueba diagnóstica se realice con la participación de pacientes que incluyan todo el espectro de la enfermedad y que, además, se mencione la proporción de pacientes en cada estrato, de tal manera que se pueda determinar su utilidad en otras poblaciones. Por otro lado, deben considerarse las enfermedades concomitantes y las terapias utilizadas que puedan modificar la eficacia de la prueba en estudio. El grupo control debe haberse seleccionado bajo los mismos criterios que el grupo problema, es decir, bajo la misma puerta de entrada, de tal forma que la comparación tenga significación clínico-metodológica.

En lo que se refiere a los errores más frecuentes durante la elaboración de una prueba diagnóstica, es común que al evaluar la prueba en estudio se tenga conocimiento previo del resultado del estándar de oro; esto genera un sesgo en la interpretación debido a que el evaluador espera un cierto resultado. Ocasionalmente, la realización y la evaluación de la prueba en estudio precede al estándar de oro e influye sobre la selección de los pacientes que se someten a este último, o sobre su interpretación cuando tiene un componente subjetivo y, no rara vez, la prueba en estudio es parte del estándar de oro con el cual se contrasta. Todas estas desviaciones sobrestiman la utilidad de la prueba.

Estos dos grandes errores pueden evitarse durante la aplicación de una prueba diagnóstica si se consideran

los valores de sensibilidad y especificidad solamente cuando:

- a) El espectro de la enfermedad en la población donde se va a aplicar sea igual al espectro de la enfermedad con la cual se elaboró el estudio.
- b) La evaluación de la prueba en estudio y del estándar de oro se haya realizado en forma ciega e independiente en todos los pacientes.

Finalmente, debe hacerse énfasis en que si la evaluación de la calidad de una prueba diagnóstica depende en parte de estrategias matemáticas, es más relevante el juicio clínico del que deriva. Y si bien la estimación de sensibilidad y especificidad se inician a partir de la presencia o no de la enfermedad, en la práctica clínica el estudio del paciente ocurre a partir de la presencia o ausencia del síntoma o signo (clínico o paraclínico).

Además, en todos los casos deberá evaluarse la reproducibilidad de la prueba, siempre y cuando los grupos en estudio sean comparables; esto significa que, además de seleccionar ambas poblaciones bajo los mismos criterios, la distribución de los sujetos dentro de los distintos grados de severidad de la enfermedad debe ser similar. No debe olvidarse que en forma cotidiana se trata a un paciente a la vez y que, por lo tanto, tener un conocimiento pleno de la severidad de la enfermedad en el grupo en estudio es indispensable para su aplicación posterior, de tal forma que el paciente pueda ser evaluado y tratado de acuerdo con la severidad de su padecimiento y no de acuerdo con el promedio de la severidad de la enfermedad en el grupo en que se evaluó la prueba diagnóstica o el tratamiento.

Bibliografía

1. Altman DG, Bland JM. Diagnostic tests 1: sensitivity and specificity. *BMJ*. 1994;308:1552.
2. Altman DG, Bland JM. Diagnostic tests 2: predictive values. *BMJ*. 1994;309:102.
3. Fagan TJ. Nomogram for Bayes's theorem. *N Engl J Med*. 1975;293:257.
4. Feinstein AR. *Clinical epidemiology. The architecture of clinical research*. Philadelphia: W. B. Saunders Company; 1985.
5. Grund B, Sabin C. Analysis of biomarker data: logs, odds ratios, and receiver operating characteristic curves. *Curr Opin HIV AIDS*. 2010;5(6):473-9.
6. Jaeschke R, Guyatt G, Lijmer J. Diagnostic tests. En: Guyatt G, Rennie D, editors. *Users' guides to the medical literature*. Chicago: AMA Press; 2002: p. 121-140.
7. Sackett DL, Straus S, Richardson WS, Rosenberg W, Haynes RB. *Evidence-based medicine. How to practice and teach EBM*. Second edition. Edinburgh: Churchill Living-stone; 2000. p. 67-93.
8. Sackett DL, Haynes RB. The architecture of diagnostic research. *BMJ*. 2002;324:7336-56.
9. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical epidemiology. A basic science for clinical medicine*. Third edition. US: Little Brown; 2009.
10. Soreide K, Korner H, Soreide JA. Diagnostic accuracy and receiver-operating characteristic curve analysis in surgical research and decision making. *Ann Surg*. 2011; 253(1):27-34.
11. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Diagnostic methods 2: receiver operating characteristic (ROC) curves. *Kidney Int*. 2009;76(3):252-6.

III. Estudios de causalidad

Juan O. Talavera, Niels H. Wachter-Rodarte, Rodolfo Rivas-Ruiz

Si bien la necesidad de resolver un problema clínico lleva a establecer un punto de partida para abordarlo (estudio de riesgo, pronóstico o tratamiento), en todos los casos se busca atribuir causalidad. El razonamiento clínico desglosado en el libro *Clinical Epidemiology. The architecture of clinical research* ofrece una guía sencilla para entender este fenómeno y utiliza tres componentes: estado basal, maniobra y desenlace. En este modelo se describen distintos errores sistemáticos (sesgos) que pueden ocurrir al soslayar características de esos tres componentes básicos. Las omisiones de características en el estado basal provocan un *ensamble inadecuado* de la población y el *sesgo de susceptibilidad*; en la aplicación o evaluación de la maniobra, el *sesgo de ejecución*; y en la evaluación del desenlace, el *sesgo de detección* y el *sesgo de transferencia*. Así, es importante hacer énfasis en que si esta forma de razonamiento facilita el entendimiento del fenómeno causal, las variables por seleccionar en los estudios a las que se les atribuirá o no causalidad requieren razonamientos clínicos adicionales que evalúen su pertinencia.

Palabras clave
investigación
causalidad

Este artículo fue publicado originalmente en Rev Med Inst Mex Seguro Soc 2011; 49 (3): 289-294 y ha sido revisado para la presente edición.

Introducción

Cuando se trata de predecir un evento futuro, el médico tiene que distinguir dos procesos: uno que ocurre antes de que aparezca la enfermedad y otro que se desarrolla una vez que la enfermedad se ha presentado. Al primero se le conoce como riesgo y se caracteriza por la asociación entre una serie de factores presentes en el sujeto sano (conocidos como factores de riesgo) y el desarrollo de la enfermedad; al segundo se le conoce como pronóstico y se caracteriza por la asociación entre una serie de características presentes al principio de la enfermedad (conocidas como indicadores pronósticos) y su desenlace.

A estos dos acontecimientos se agregan múltiples intervenciones, ya sean preventivas o terapéuticas; las primeras tienen el objetivo de evitar la aparición de la enfermedad y las segundas, revertir o reducir el daño causado por esta. El evento por el cual una condición basal (salud o enfermedad) se ve modificada por una maniobra (factores de riesgo, indicadores pronósticos o tratamiento), lo cual provoca una nueva condición conocida como desenlace (prevención o aparición de enfermedad y progresión o resolución del daño), corresponde a un suceso de causalidad. Es decir, en estos tres casos —ya sea que nuestro objetivo consista en identificar factores de riesgo, un agente etiológico, indicadores pronósticos, o evaluar un tratamiento— se busca atribuir causalidad.

Si bien la necesidad de resolver un problema clínico nos lleva a establecer un punto de partida para poder abordarlo —estudio de riesgo, pronóstico o tratamiento—, en la realidad existe una fuerte asociación entre sus componentes. Por esta razón, resulta indispensable que cuando se valore cualquiera de ellos se considere la pertinencia de los otros dos dentro de la evaluación. Esta acción se lleva a cabo frecuentemente bajo el término de *control de factores de confusión*.

De esta manera, el estudio de causalidad para evaluar un tratamiento no se limita solo a la evaluación de la terapia sino que obliga a estimar la contribución de todos los indicadores pronósticos existentes en el estado basal que participan en la enfermedad correspondiente.

De igual forma, cuando se quiere evitar con alguna maniobra la aparición de una enfermedad estamos obligados a evaluar los distintos factores de riesgo asociados específicamente a dicha enfermedad. Esta obligación de medir el impacto de los distintos factores de riesgo o indicadores pronóstico cuando se evalúa una terapia se corresponde con la obligación de evaluar los distintos procedimientos terapéuticos, cuando lo que se desea evaluar son los factores de riesgo o indicadores pronóstico.

Razonamiento clínico en estudios de causalidad

El razonamiento clínico desglosado en el libro *Clinical Epidemiology. The architecture of clinical research* nos ofrece un enfoque sencillo para entender el fenómeno de la causalidad. En la figura 1 se muestra el modelo base integrado por el estado basal, la maniobra y el desenlace. En este modelo se describen distintos errores sistemáticos (sesgos) que se pueden favorecer al omitir características de los tres componentes básicos.

Errores en el estado basal

Los dos primeros errores se asocian con omisiones de características del estado basal y son el ensamble inadecuado y el sesgo de susceptibilidad.

Ensamble inadecuado se refiere a la selección de una población no susceptible de presentar el desenlace buscado con una maniobra propuesta; por ejemplo, resulta poco práctico probar una vacuna en una población con baja incidencia de la enfermedad que pretendemos prevenir, porque el tamaño de la muestra tendría que ser enorme; también resulta inconveniente querer evaluar el efecto protector sobre el riñón de un inhibidor de la ECA en una población de pacientes diabéticos de reciente diagnóstico, porque el seguimiento tendría que ser muy prolongado.

El *sesgo de susceptibilidad* hace alusión a la probabilidad premaniobra que tiene el sujeto de presentar cierto desenlace; por ejemplo, la presencia de sobrepeso u obesidad incrementa la probabilidad de un infarto en el paciente diabético, independientemente del mal control metabólico con el que curse.

Las características propias que deben describir al estado basal para evitar estos errores se muestran en las figuras 2a y 2b, es decir, la forma de selección de la población, la demarcación diagnóstica y la estratificación pronóstica.

Dentro de la estratificación pronóstica, la anatomohistología ha sido utilizada como el principal indicador, sobre todo en oncología, seguida por el aspecto funcional. En la práctica clínica es común el uso de múltiples indicadores pronósticos con el fin de estadificar la enfermedad de acuerdo con la condición del paciente. Las siguientes agrupaciones de estratificación son las más comunes:

Primarias

- **Estratificación por estatus:** incluye el estado funcional, el nutricional y el mental del paciente. El estado funcional ha sido evaluado con escalas como la Karnovsky o la ECOG, basadas en la capacidad del paciente para realizar sus actividades cotidianas, de tal forma que un paciente que no es autosuficiente está más afectado que aquel que puede realizar sus labores. El estado nutricional tiene impacto en la respuesta inmunológica y en la estabilidad hemodinámica. Se ha observado que pacientes con niveles bajos de albúmina presentan un incremento importante en la mortalidad en comparación con aquellos con niveles mayores. Otras formas de evaluar el estado nutricional pudieran ser el índice de masa corporal y el índice cintura-cadera, cuando se busca evaluar el impacto del sobrepeso o la distribución de la grasa corporal; asimismo, dos de las características más importantes para evaluar el estado mental son la presencia de depresión y ansiedad, entre muchas otras condiciones.
- **Estratificación morfológica:** se refiere a la localización y tipo de daño propio de la patología. Un ejemplo es la estirpe histológica de las tumoraciones y los marcadores citogenéticos o inmunofenotípicos (por ejemplo, dos tumores con la misma extensión de la enfermedad pueden tener distinto pronóstico de acuerdo con la estirpe histológica, la presencia de marcadores tumorales o las alte-

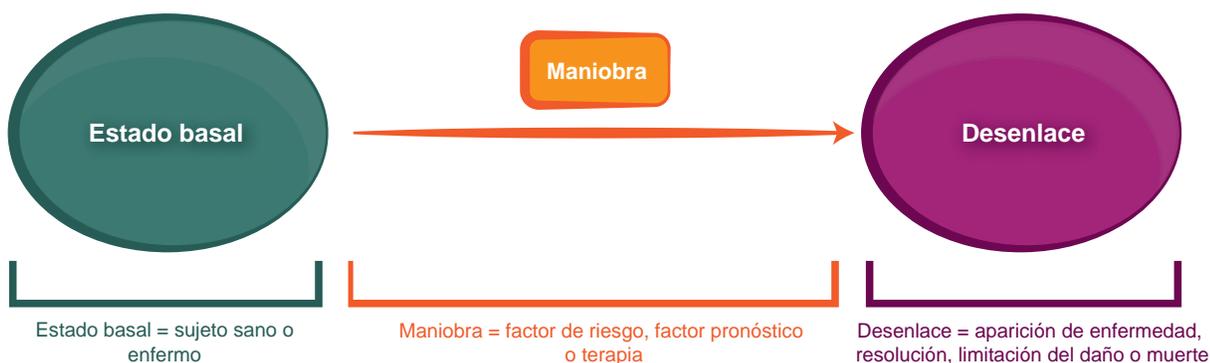


Figura 1 Modelo base del fenómeno de causalidad

raciones en el cariotipo; también, un paciente con falla cardíaca puede tener distinto pronóstico de acuerdo con el grado y tipo de daño valvular).

- Estratificación clínica: considera la severidad de la enfermedad, por ejemplo, el paciente con insuficiencia cardíaca grado IV (edema agudo pulmonar) no tiene la misma probabilidad de muerte que el paciente con grado II (disnea a esfuerzos moderados), aun cuando la afección anatómica en ambos casos sea una estenosis mitral con el mismo diámetro de apertura valvular.
- Estratificación cronométrica: considera dos componentes, la edad del paciente y la duración de la enfermedad. En cuanto a la primera, muchas enfer-

midades tienen mayor impacto en los extremos de la vida y se asocian con mayor susceptibilidad a un desenlace desfavorable; asimismo, el sujeto mayor tiene una esperanza de vida menor. Respecto a la duración de la enfermedad, si dos pacientes tienen el mismo daño pero en uno la enfermedad es de inicio reciente mientras en el otro de larga evolución, el pronóstico será mejor en el último porque ya están seleccionados aquellos pacientes con una menor agresividad de la enfermedad.

- Estratificación por comorbilidad: se refiere a la coexistencia de algún otro proceso patológico que pueda alterar el resultado de interés. Los distintos padecimientos ejercen diferente impacto sobre el

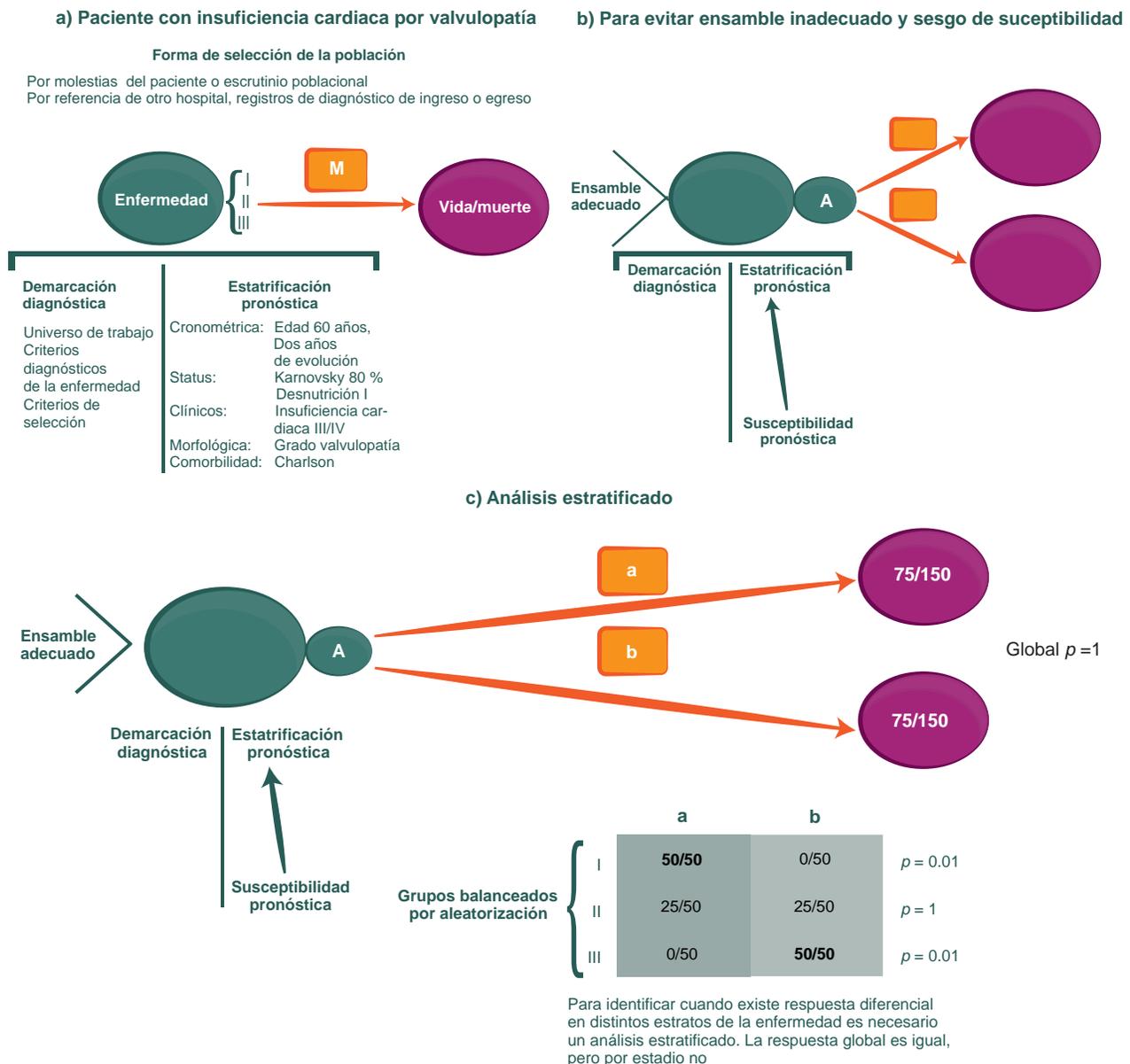


Figura 2 Características por considerar en el estado basal

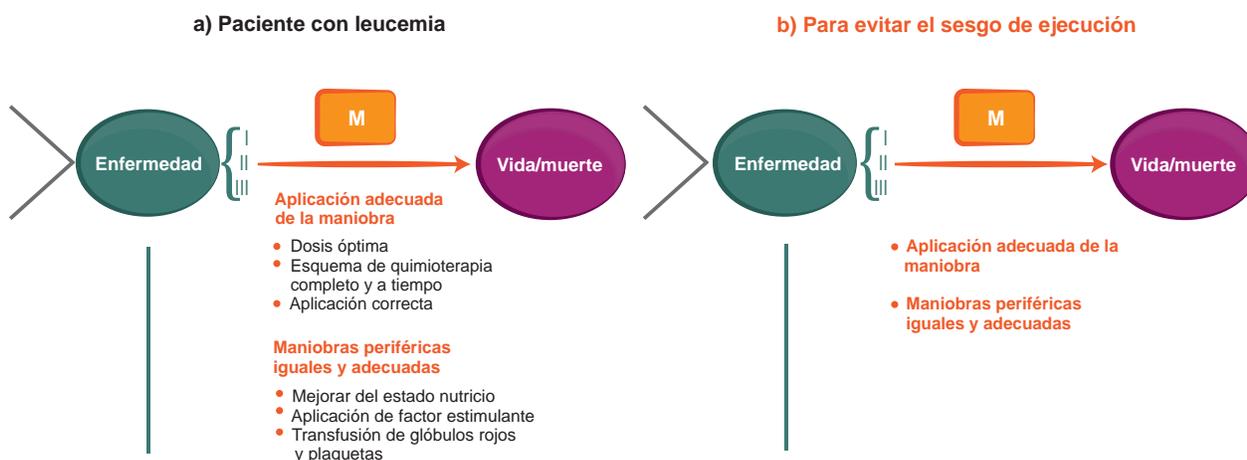


Figura 3 Características por considerar en la maniobra

desenlace, e incluso en un mismo padecimiento el impacto está generalmente en relación con el grado de afección, por ejemplo, en un paciente con infarto agudo del miocardio el pronóstico es mejor cuando la comorbilidad es artritis reumatoide que cuando es diabetes mellitus.

- Estratificación por maniobra previa: aquí pueden identificarse dos apartados: el primero y más utilizado es la respuesta temprana a una maniobra ya sea preventiva o terapéutica, es decir, se espera un mejor pronóstico ante una respuesta favorable en forma temprana. El segundo se refiere al impacto adverso de una maniobra. Se conoce que prácticamente cualquier maniobra conlleva un riesgo; sin embargo, no en todas es de la misma magnitud. De tal forma, para cualquier terapia habrá que considerar la seguridad como un indicador pronóstico.
- Estratificación por herencia: se ha identificado el impacto que tiene la carga genética como factor de riesgo para varias enfermedades y con la mayor agresividad de estas o mayor riesgo de daño a órgano blanco, como en la diabetes.

Secundarias

- Las condiciones sociales, económicas y culturales, así como la forma de afrontar la enfermedad suelen tener un impacto menor que los componentes biológicos dentro del pronóstico; sin embargo, en ocasiones son determinantes, como tener acceso a servicios de salud en eventos urgentes o el cambio de estilo de vida en algunas enfermedades crónicas.

Una estrategia propia de los ensayos clínicos para evitar el sesgo de susceptibilidad es la asignación aleatoria de los sujetos al brazo de tratamiento, con la que se busca, entre otras cosas, que los factores conocidos y no conocidos relacionados potencialmente con el

desenlace queden distribuidos de forma equilibrada entre los grupos por comparar. Otro beneficio consiste en evitar que quien asigna se vea tentado a situar a un sujeto con mejor pronóstico en algún brazo en particular, pues esto facilita el cegamiento de los tratamientos y busca distribuir homogéneamente a los sujetos con distinta probabilidad de apego al tratamiento y distinta probabilidad de abandono del estudio. Debe quedar claro que si bien con la asignación aleatoria se busca que los grupos por comparar queden distribuidos en su estado basal en forma homogénea, no muestra el efecto de las maniobras en los distintos estratos (figura 2c).

Errores en la maniobra

El tercer error sistemático, conocido como *sesgo de ejecución*, se asocia con omisiones en la aplicación o evaluación de la maniobra, y se refiere a las diferencias generadas por la distinta calidad entre las maniobras por comparar o por la utilización desigual de maniobras adicionales entre los grupos (también conocidas como maniobras periféricas); por ejemplo, no es lo mismo una cirugía realizada por un cirujano recién graduado que la realizada por el médico con amplia experiencia, ni resultan comparables dos cirugías cuando en una de ellas los pacientes son nutridos o llevados a hemoglobina normal, mientras que en el otro grupo no. Las características que deben considerarse en las maniobras para evitar estos errores se muestran en las figuras 3a y 3b, las cuales consisten en la aplicación adecuada de la maniobra y la aplicación por igual de maniobras periféricas.

En los ensayos clínicos existe una estrategia que busca manejar el error generado por la aplicación inadecuada de la maniobra, y es la forma de analizar los datos, ya sea a través de un análisis o por intención

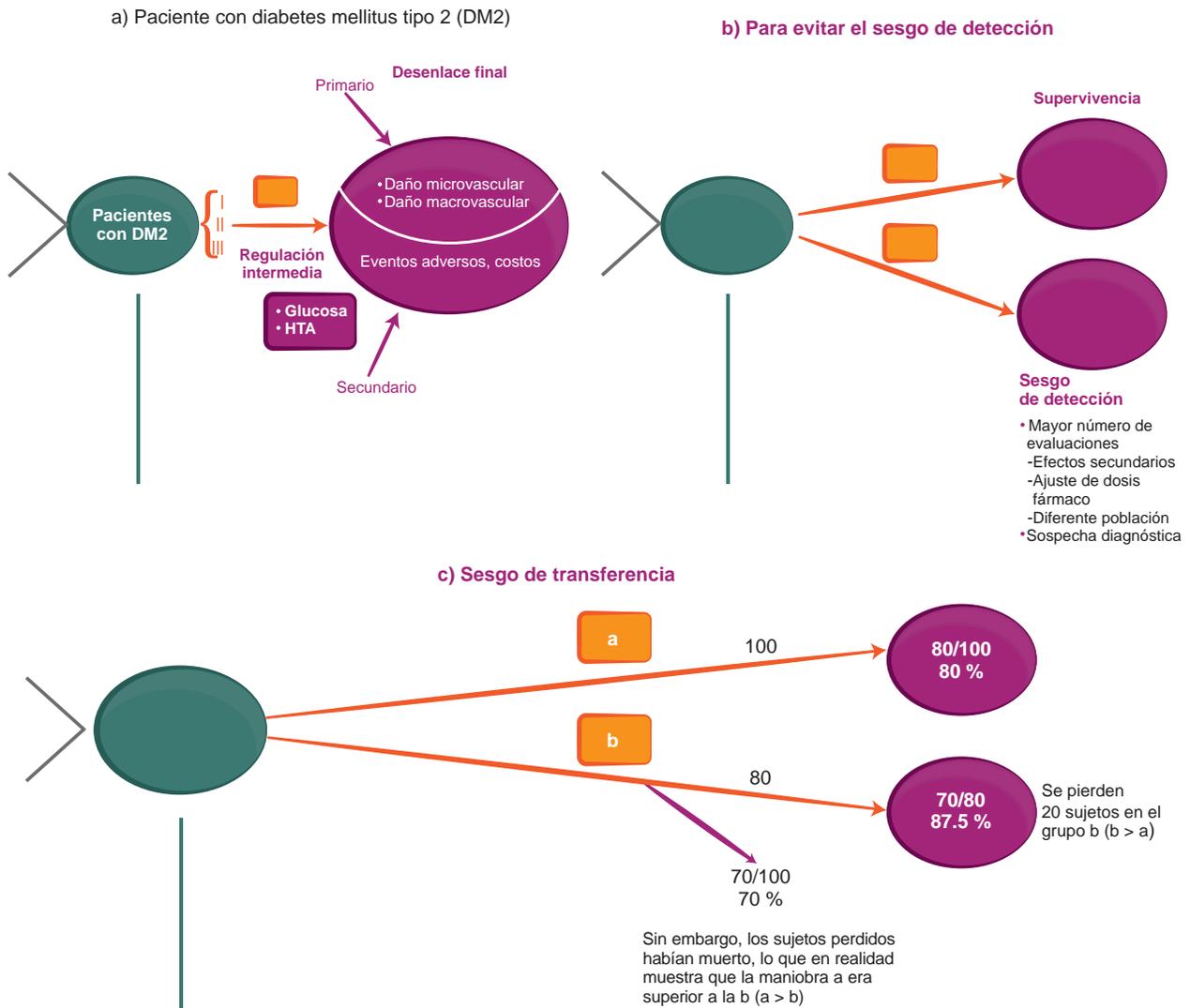


Figura 4 Principales características por considerar al evaluar el desenlace

al tratamiento o del análisis por protocolo. El *análisis por intención a tratamiento* consiste en analizar a los sujetos en el grupo al cual fueron asignados al inicio del estudio, independientemente de que hayan cumplido o no el protocolo terapéutico. El análisis por protocolo consiste en analizar sólo a quienes cumplieron el protocolo terapéutico. En los estudios observacionales, dado que no existe la asignación aleatoria de la maniobra, se gradúa esta al interior de los grupos, con lo que se pueden comparar los distintos grados de calidad en su aplicación.

Errores en el desenlace

Durante la evaluación del desenlace se da el sesgo de detección, relativo a la detección desigual del desenlace entre los grupos, el cual se presenta principalmente por dos razones:

- Un mayor número de evaluaciones en algún grupo, debido principalmente a más efectos secundarios, un ajuste continuo de la dosis o a la comparación de poblaciones con accesibilidad médica distinta.
- La presencia de sospecha diagnóstica.

En la evaluación del desenlace es importante identificar si se trata de un desenlace final o una regulación intermedia; por ejemplo, en el paciente diabético el desenlace final es evitar daño a órganos blanco; sin embargo, una regulación intermedia es el control de la glucosa; esta última puede considerarse un desenlace final si en el paciente descontrolado se busca disminuir la sintomatología.

Otro aspecto importante en la evaluación del desenlace es la identificación y diferenciación entre el desenlace primario y el secundario. Este punto resulta relevante, ya que tanto los criterios de selección, la

estratificación pronóstica, la maniobra y la estimación del tamaño de muestra se realizan sobre el desenlace primario y no sobre el secundario. Por lo tanto, los resultados obtenidos en la mayoría de los estudios son solo exploratorios para desenlaces secundarios (figuras 4a y 4b).

El último sesgo tiene que ver también con el desenlace, se genera por la pérdida de los sujetos en estudio y se conoce como *sesgo de transferencia* (figura 4c). Si bien en estudios longitudinales se incrementa el tamaño de muestra 20 % si se consideran las posibles pérdidas, es importante hacer énfasis en que este incremento no resuelve el sesgo de transferencia, sino más bien mantiene la estabilidad de los datos.

Consideraciones finales

En los estudios longitudinales resulta fácil aplicar esta guía para estudiar el fenómeno de causalidad; en los transversales continúa siendo aplicable, pero consti-

tuye un mayor reto que se traduce en la creación de un modelo artificial en cuanto al establecimiento temporal de sus componentes. Se recomienda tomar en cuenta los elementos aquí descritos no solo para la lectura de un estudio de causalidad sino también para la creación de una propuesta de investigación.

Es importante insistir en que si esta forma de razonamiento facilita el entendimiento del fenómeno causal, lo apropiado en las variables, por seleccionar a las que se les atribuirá o no causalidad, es tomar en cuenta las consideraciones clínicas adicionales que evalúan su pertinencia. Los principios básicos fueron descritos en 1965 por sir Austin Bradford Hill y fueron actualizados en el año 2000 por Kaufman y Poole; seguramente con el paso del tiempo habrá de aumentar el número de factores por considerar cuando se juzgue una posible relación causal.

Esperamos que el enfoque de causalidad aquí descrito que desglosa la base de la práctica clínica facilite la interpretación de la literatura médica y sirva de guía para planear propuestas de investigación e incrementar la calidad de la atención médica.

Bibliografía

1. Charlson ME, Frederic LS. The therapeutic efficacy of critical care units from two perspectives; a traditional cohort approach vs. a new case control methodology. *J Chron Dis.* 1987;40:31-9.
2. Feinstein AR. *Clinical epidemiology. The architecture of clinical research.* Philadelphia: WB Saunders; 1985.
3. Feinstein AR. Directionality and scientific Inference. *J Clin Epidemiol.* 1989;42:829-33.
4. Fletcher R, Fletcher S, Wagner E. *Clinical epidemiology: the essentials.* 2nd. ed. Baltimore: Williams & Wilkins; 1988.
5. Rothman KJ, Greenland S, Lash TL. *Modern epidemiology.* Third edition. Baltimore: Williams & Wilkins; 2008.
6. Sackett D, Haynes R, Tugwell P. *Epidemiología clínica una ciencia básica para la medicina clínica.* Madrid: Ediciones Díaz de Santos; 1989.
7. Talavera JO. Pronóstico. En: Ramiro M, Lifshitz A, Halabe J, Frati A, editores. *El internista. Medicina interna para internistas.* 3a. ed. México: Nieto Editores; 2008. p. 1893-1898.

IV. Pertinencia de la prueba estadística

Juan O. Talavera, Rodolfo Rivas-Ruiz

Cuando observamos la diferencia entre dos terapias o la asociación de un factor de riesgo o indicador pronóstico con su desenlace, tenemos que evaluar la certeza del resultado. Esta evaluación se basa en un juicio que utiliza información relacionada con el diseño del estudio y el manejo estadístico de la información. En este artículo se menciona específicamente la pertinencia de la prueba estadística seleccionada. Las pruebas estadísticas se eligen a partir de dos características: el objetivo del estudio y el tipo de variables. El objetivo se puede dividir en tres grupos de pruebas: *a)* en las que se desea mostrar diferencias entre grupos, o de un mismo grupo antes y después de una maniobra; *b)* en las que se busca mostrar la relación entre las variables; *c)* en las que se pretende predecir un desenlace. En cuanto a los tipos de variables tenemos dos: las cuantitativas (continuas y discontinuas) y las cualitativas (ordinales y dicotómicas). Por ejemplo, si se busca demostrar diferencias en edad (variable cuantitativa), entre pacientes con lupus eritematoso sistémico, con y sin afección neurológica (dos grupos), la prueba apropiada es la *t* de Student para muestras independientes; pero si lo que se compara entre esos mismos grupos es la frecuencia de mujeres (variable binomial), entonces la prueba estadística pertinente es la χ^2 .

Palabras clave

investigación biomédica
proyectos de investigación
estadística y datos cuantitativos

Este artículo fue publicado originalmente en Rev Med Inst Mex Seguro Soc 2011; 49 (3): 289-294 y ha sido revisado para la presente edición.

Introducción

Cuando observamos la diferencia entre dos terapias o la asociación de un factor de riesgo o indicador pronóstico con su desenlace, surge un cuestionamiento: ¿el resultado es real? Decidir si es real requiere dos juicios que se complementan:

1. La planeación y el desarrollo del proceso que documentan dicha diferencia o asociación están libres de errores, o cuando menos estos son de una magnitud menor que no modifica la orientación de la diferencia o de la asociación (es decir, diseño apropiado y ejecución adecuada).
2. El tamaño de la muestra es suficiente para mantener la estabilidad de los datos y la prueba estadística es apropiada para el objetivo.

La planeación y el desarrollo del proceso se han mencionado en los tres capítulos previos de esta serie. Por otra parte, la estabilidad de los datos será discutida a detalle en un artículo posterior al hablar de tamaño de muestra y valor de *p*.

En este artículo discutiremos la pertinencia de la prueba estadística seleccionada. Sin duda, este conocimiento nos permitirá comprender con mayor precisión los resultados obtenidos en estudios de investigación clínica y, por supuesto, incrementará nuestra capacidad para hacer un uso adecuado de los mismos.

Objetivo del estudio y tipo de variable

Las pruebas estadísticas se eligen a partir de dos características: el objetivo del estudio y el tipo de variables. Dentro de los objetivos del estudio podemos identificar tres:

1. Demostrar diferencias entre grupos o diferencias en un mismo grupo antes y después de una maniobra (por ejemplo, el tratamiento con el fármaco *A* reduce en mayor proporción la presión sanguínea que el tratamiento con el fármaco *B*).
2. Mostrar relación (correlación) entre variables (por ejemplo, la creatinina sérica se eleva conforme disminuye la función renal).
3. Predecir un desenlace (por ejemplo, la probabilidad de que el sujeto con vida sedentaria y sobrepeso desarrolle diabetes mellitus tipo 2).

Frecuentemente se superponen los modelos, y es así que en ocasiones para demostrar diferencias entre dos grupos se hace uso de modelos inicialmente identificados para predecir un desenlace; esto ocurre sobre todo cuando se requiere ajustar la maniobra principal

(fármaco A versus fármaco B) por múltiples factores (edad, sexo, índice de masa corporal, etcétera). Pero también acontece el fenómeno opuesto, cuando se busca predecir un suceso que ocurrirá a futuro pero solo se cuenta con una o dos variables predictivas; en este caso se utiliza una prueba para demostrar diferencias.

Es importante aclarar que la correlación básicamente sirve para ver la magnitud de la asociación entre variables, aunque debe quedar claro que no determina causalidad. De hecho ninguna prueba estadística puede hacerlo. Para ello, se requiere cubrir una serie de principios descritos por sir Austin Bradford Hill.

Definir el tipo de variable resulta relevante debido a que es el eje para seleccionar la prueba apropiada dependiendo del objetivo buscado. Dentro del tipo de variable existen dos grupos:

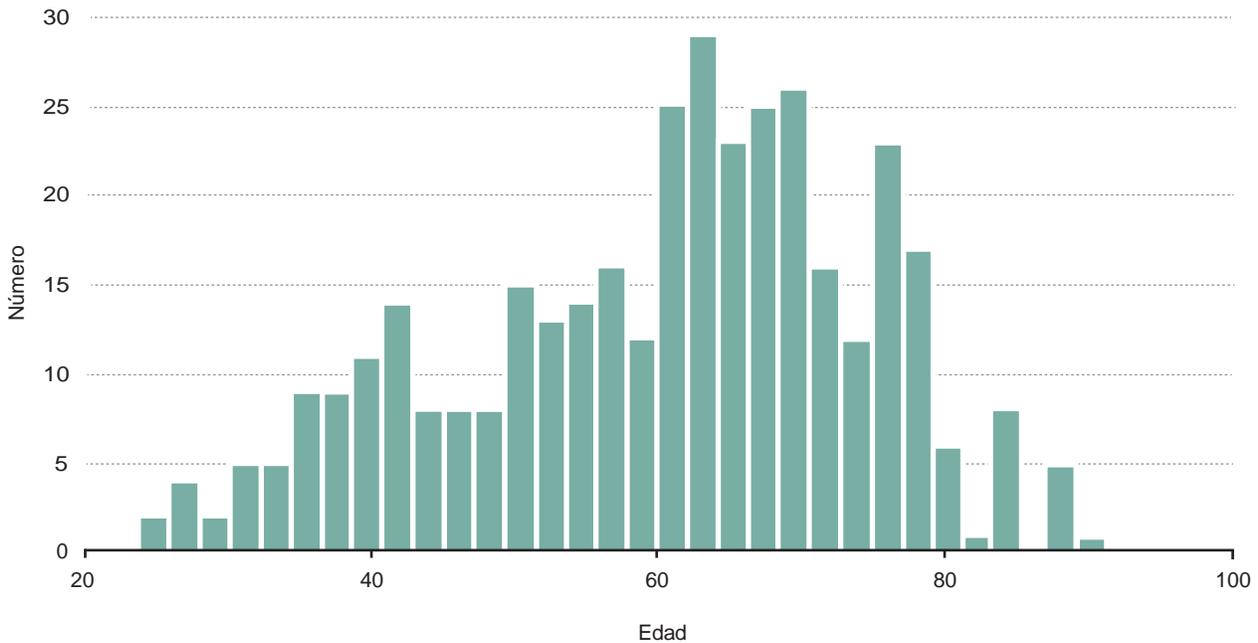
1. Cuantitativas: continuas y discontinuas o discretas. Las primeras se caracterizan porque pueden tomar cualquier valor a lo largo de un continuo (por ejemplo, talla 1.75 m). Por su lado, las discontinuas o discretas utilizan solo números enteros (número de partos, 1, 2, 3...). En ambas, la distancia entre una unidad y otra a través de toda su escala es equidistante.
2. Cualitativas: dentro de estas se encuentran la ordinal y la dicotómica. La ordinal permite dar un orden

Cuadro I Peso de sujetos en estudio bajo dos esquemas terapéuticos

Grupo A			Grupo B		
77			65		
78			69		
80			77		
82			78		
85	83.5	Promedio	85	83.5	Promedio
85	85.0	Mediana	85	85.0	Mediana
85	85.0	Moda	89	85.0	Moda
86			93		
88			96		
89			98		

Las medidas de tendencia central son iguales, pero la dispersión de los datos es distinta

a la característica en estudio y, a diferencia de como pasa con las cuantitativas, la distancia entre dos categorías no es equidistante (por ejemplo, falla cardíaca, grados I a IV). Las dicotómicas, como su nombre lo dice, son aquellas en las que solo hay dos categorías, que pueden ser binomiales (es una opción u otra, por ejemplo, hombre o mujer) o nominales (se refiere a la presencia o no de la característica, por ejemplo, vivo a seis meses, sí o no).



Media 59.79
 Desviación estándar de 13.882. Dos desviaciones estándar a cada lado de la media reflejan 95 % de la población
 Promedio de 59.79, con IC 95 % = 32.03-87.55

Figura 1 Histograma

Es importante mencionar el manejo que sufrirá el tipo de variable durante el proceso de análisis, empezando por la recolección de datos en forma “cruda”, lo que significa que solo se trata de una compilación de información de un grupo de sujetos. Para que estos datos tengan un significado de utilidad, se necesita organizarlos y resumirlos. La forma más simple de organización son las tablas de distribución de frecuencia; sin embargo, en ocasiones es más fácil entender su representación gráfica a través de un histograma o polígono de frecuencias. No obstante la utilidad de esta información, se necesita información cuantitativa de los datos recolectados, es decir, se requieren *índices numéricos* que reflejen las distintas distribuciones de probabilidad, cuya función primordial es *modelar* el comportamiento de una gran variedad de fenómenos biológicos. Dentro de estos índices numéricos encontramos las medidas de tendencia central y las medidas de dispersión.

1. Medidas de tendencia central (cuadro I y figura 1).

a) *Media*: es la suma de un conjunto de datos divididos por el número total de ellos. El símbolo para representar la media de una población es la letra griega mu (μ), y la media de una muestra está representada por \bar{x} . Es la medida de resumen más utilizada para variables cuantitativas.

b) *Mediana*: es el valor que se encuentra exactamente a la mitad del total de los datos. La mediana divide una distribución de datos ordenados en exactamente dos partes iguales. La ventaja de la mediana como medida de tendencia central es que no se ve afectada por el valor de datos extremos, fenómeno que sí ocurre con la media. Es el tipo de medida de resumen más utilizada para las variables cuantitativas que no siguen una distribución normal y para variables ordinales.

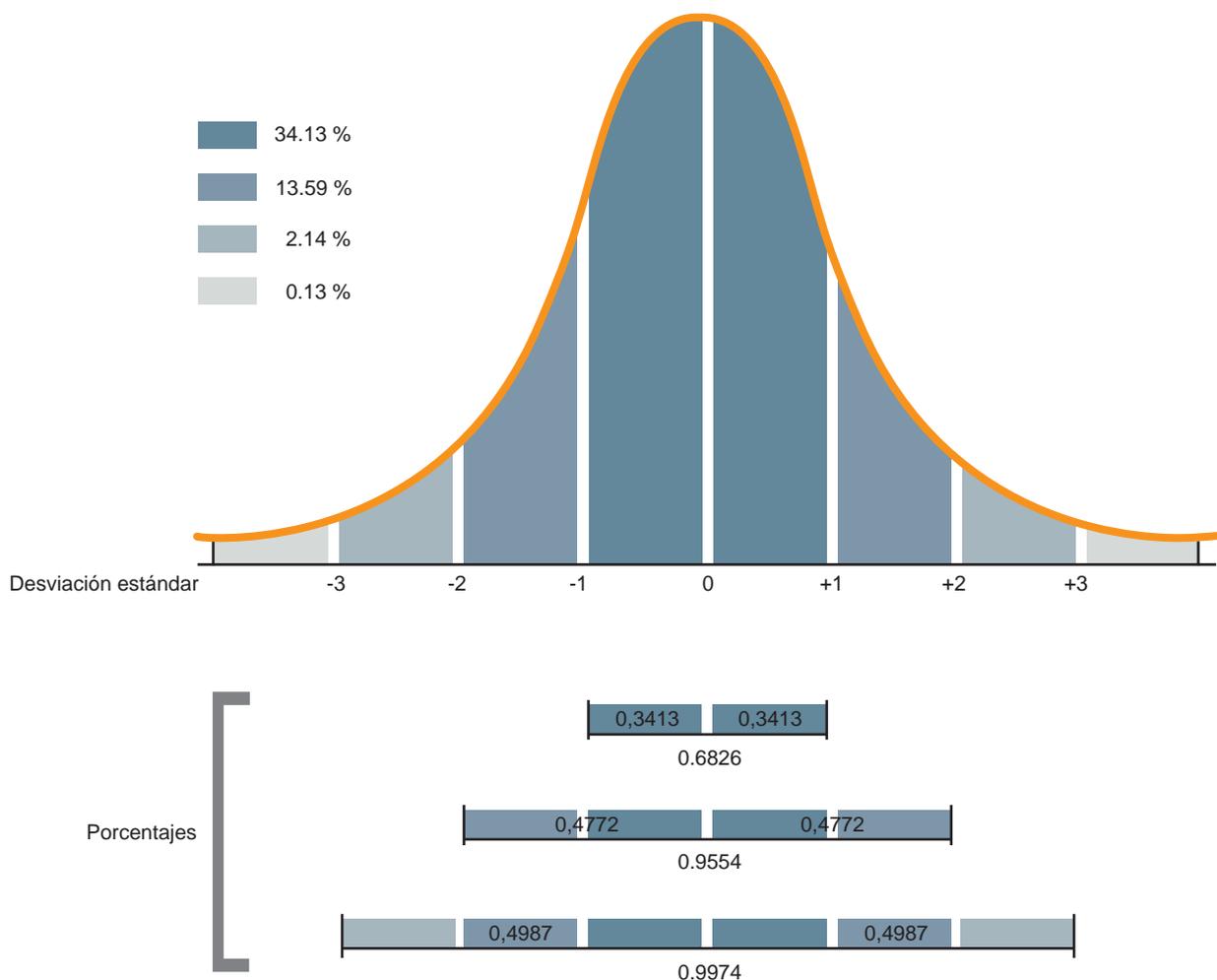


Figura 2 Curva de distribución normal

c) *Moda*: se refiere al valor que más se repite en una distribución. Esta medida es de poco uso en medicina

2. Medidas de dispersión más comunes.

a) *Desviación estándar*: refleja la variación entre todo el conjunto de datos y se utiliza cuando estos siguen una distribución normal.

b) *Percentil*: describe la posición de un valor de la distribución. Se utiliza para variables cuantitativas que no siguen una distribución normal y para variables de tipo ordinal.

c) *Rango*: diferencia entre el valor más alto y el valor más bajo de la distribución.

d) *Rangos intercuartílicos*: se refieren a los valores del primer y tercer cuartil

En investigación clínica, como en muchos otros fenómenos de la vida real, los datos más frecuentemente analizados son los cuantitativos, que en la mayoría de los casos muestran una distribución gaussiana, también conocida como distribución normal, que se caracteriza por tener forma acampanada, ser simétrica respecto a su media, tener valores de frecuencia decrecientes conforme se alejan de la media y que nunca llegan a cero (asintótica). La moda y la mediana son iguales a la media, aproximadamente 68 % de los datos se encuentran en ± 1 desviación

estándar respecto a su media y 95 % en ± 2 desviaciones estándar (figura 2). Es así que si la agrupación de datos es de tipo cuantitativo con distribución normal, su medida de resumen será la media y su medida de dispersión, la desviación estándar. Sin embargo, si su distribución no es gaussiana, al igual que para una variable de tipo ordinal, su medida de resumen será la mediana y su medida de dispersión, el percentil o rango. En general, estas variables no tienen medidas de dispersión y cuando se utilizan se prefieren los intervalos de confianza del 95%

Pertinencia de la prueba estadística

Una vez que conocemos nuestro objetivo y las características de nuestros datos (tipo de variable) podemos considerar la pertinencia de la prueba estadística (cuadro II). Sin embargo, existen dos consideraciones más cuando el objetivo sea demostrar diferencia:

1. Si se trata de un estudio donde se compara el valor de un dato antes y después de una maniobra, ya sea observacional o experimental, se le conoce como de muestras relacionadas, pero si se trata de la comparación de datos entre distintos grupos, se le denomina de muestras no relacionadas.
2. Si se trata de una comparación entre distintos grupos, es necesario determinar si será entre dos o más.

Cuadro II Selección de la prueba estadística conforme al objetivo y al tipo de variable

Tipo de variable	Tipo de muestra	Demostrar diferencia		Mostrar relación ^{&}	Predecir 1 variable [‡]
		Dos grupos	Tres grupos	Dos variables	Variable desenlace
Cuantitativa (distribución normal)	NR	t de Student*	Anova 1 factor	Pearson	Regresión lineal
	R	t de Student**	Anova 1 factor		
Cualitativa ordinal (libre distribución)	NR	U Mann-Whitney	Kruskal-Wallis	Spearman	
	R	Wilcoxon	Friedman		
Cualitativa dicotómica	NR	χ^2 (o prueba exacta de Fisher)	χ^2 (de tendencia lineal)	Coeficiente phi	Regresión logística/
	R	McNemar			

NR = no relacionada; R = relacionada; R = medición de la variable en el mismo sujeto en dos tiempos distintos

* t de Student para muestras independientes

** t de Student para muestras relacionadas

& Para la correlación entre 2 variables se utiliza la prueba de aquella en la escala inferior (en realidad ninguna escala es inferior, sin embargo, se han ordenado de las cuantitativas continuas a las dicotómicas, pasando por las cuantitativas discontinuas y las ordinales).

‡ La variable predictiva puede ser cuantitativa, dicotómica u ordinal (estas últimas transformadas a variables tipo *dummy*)

Ya con la información completa, con el cuadro II podemos verificar si la selección de la prueba estadística fue la adecuada conforme a la variable y al objetivo. Por ejemplo, si se compara la edad (variable cuantitativa con distribución normal en este caso) entre pacientes con lupus eritematoso sistémico, con y sin afección neurológica (dos grupos), la prueba apropiada es la *t* de Student para muestras independientes. Pero si lo que se compara entre esos mismos pacientes es la frecuencia de mujeres (variable binomial), entonces la prueba estadística pertinente es la χ^2 . Si lo que se compara entre los dos grupos es su grado de actividad lúpica (escala ordinal), la prueba estadística pertinente es la *U* de Mann-Whitney (escala ordinal o cuantitativa de libre distribución). Por otro lado, si lo que nos muestran es la magnitud de la asociación (relación) entre edad (variable cuantitativa con distribución normal) y el grado de actividad lúpica (variable ordinal), la prueba pertinente es la *r* de Spearman. Finalmente, si lo que se busca es predecir el peso de un niño (variable cuantitativa) a partir de la edad (variable

cuantitativa), tipo de alimentación (variable ordinal: buena, regular o mala) y el sexo (dicotómica), la prueba pertinente es la regresión lineal. Pero si lo que se quiere es predecir la probabilidad de infarto (dicotómica nominal) en los siguientes 10 años a partir de la edad (cuantitativa), el riesgo aterogénico (ordinal, bajo, medio y alto) y el sexo (dicotómica binomial), la prueba pertinente es la regresión logística múltiple.

Por último, esperamos que el presente artículo permita entender la razón de la selección de las pruebas estadísticas más empleadas en la investigación en salud y que, de igual manera, sirva de guía a quienes se inician en la estadística. No es suficiente para definir si los resultados obtenidos son reales, será necesario tener en consideración el diseño y la ejecución del estudio y la estabilidad de los datos, pero este último aspecto merece ser analizado en otro apartado. En los siguientes capítulos de esta serie se abordarán a fondo las pruebas *t* de Student, *U* de Mann Whitney (en la que se abordará cómo seleccionar el tipo de distribución de las variables cuantitativas) y las pruebas de χ^2 .

Bibliografía

1. Armitage P, Berry G, Matthews JNS. Statistical methods in medical research. 4th ed. Massachusetts, MA: Blackwell Publishing; 2002.
2. Bland M. Introduction to medical statistics. 3rd ed. Oxford: Oxford University Press; 2003.
3. Feinstein AR. Clinical epidemiology. The architecture of clinical research. Philadelphia, PA: W.B. Saunders; 1985.
4. Feinstein AR. Multivariable analysis: an introduction. New Haven, CT: Yale University Press; 1996.
5. Feinstein AR. Principles of medical statistics. New York, NY: Chapman and Hall/CRC; 2002.
6. Le Chap T. Introductory biostatistics. Hoboken, NJ: New Jersey: John Wiley and Sons; 2003.
7. Peat J, Barton B. Medical statistics. A guide to data analysis and critical appraisal. Malden, MA: Blackwell Publishing; 2005.
8. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. 3rd ed. Saddle River, NJ: Pearson/Prentice Hall; 2009.

Toda la información disponible en línea



<http://revistamedica.imss.gob.mx/>

V. Tamaño de muestra

Juan O. Talavera, Rodolfo Rivas-Ruiz, Laura Paola Bernal-Rosales,
Lino Palacios-Cruz

En investigación clínica resulta imposible e ineficiente estudiar a todos los pacientes con una patología específica, por lo que es necesario abocarse a una muestra. La estimación del tamaño de muestra garantiza la estabilidad de los resultados y permite prever la factibilidad del estudio dependiendo del costo y de la disponibilidad de los pacientes. La estructura básica de estimación del tamaño de muestra parte de la premisa que busca demostrar —entre otras cosas— que la diferencia entre dos o más maniobras en el estado subsecuente es real. Para ello se requiere conocer el valor de la diferencia esperada (δ) y la medida de dispersión de los datos que le dieron origen (desviación estándar), que suelen obtenerse de estudios previos. Posteriormente se consideran otros componentes: α , porcentaje de error tipo I aceptado en la aseveración de que la diferencia entre las medias es real, generalmente de 5 %; y β , porcentaje de error tipo II aceptado en la aseveración de que la no diferencia entre las medias es real, generalmente de 15 a 20 %. Estos valores se sustituyen en la fórmula o en algún programa electrónico de estimación del tamaño de muestra. Si bien las medidas de resumen y dispersión pueden variar conforme la medición del resultado y consecuentemente la fórmula, el principio es el mismo.

Palabras clave

tamaño de muestra
intervalo de confianza

Este artículo fue publicado originalmente en Rev Med Inst Mex Seguro Soc 2011; 49 (3): 289-294 y ha sido revisado para la presente edición.

Introducción

En investigación clínica es imposible e ineficiente estudiar a todos los sujetos enfermos con una patología específica; en consecuencia, cuando leemos un artículo los resultados que nos muestran corresponden a una muestra del total de la población. El número de sujetos incluidos en un estudio está determinado por una serie de características que se mencionarán más adelante, pero cuyo objetivo principal es responder a una pregunta con la certeza de que el resultado encontrado es real. Aunado a esto, la estimación del tamaño de muestra antes de iniciar un estudio permite considerar su factibilidad dependiendo de la disponibilidad de pacientes y del costo. La ausencia del cálculo en el tamaño de muestra puede provocar una dispensa innecesaria de recursos tanto económicos como humanos. Es posible que se incrementen los gastos del estudio en forma innecesaria por el número excedente de sujetos que se incluyan en él, o que la inversión realizada sea infructuosa al incluir un número de sujetos insuficiente para contestar la pregunta de investigación.

La estructura básica de estimación del tamaño de muestra parte de la premisa que busca demostrar que la diferencia observada entre las mediciones realizadas antes y después de una maniobra, o entre dos maniobras en el estado subsecuente, es real y no se debe a efectos del azar. Esta estructura es la misma independientemente del tipo de variables necesarias para contestar la pregunta de investigación. En otros casos no se desea demostrar la veracidad de una diferencia sino obtener el valor promedio de una determinada característica dentro de una población con una precisión marcada por el límite superior e inferior del intervalo de confianza (IC), que en la mayoría de los casos se solicita de 95 o 99 %.

Estimación para dos grupos

Este propósito queda ejemplificado cuando se desea demostrar que las cifras de presión sanguínea son diferentes con cierto fármaco versus otro y que esa diferencia no se debe al azar. Para estimar el tamaño de muestra, lo primero que se necesita en este ejercicio es el promedio (\bar{X}) de las cifras de tensión arterial diastólica (TAD) de los pacientes que tomaron un fármaco (grupo A) u otro (grupo B): en el supuesto de que el promedio de TAD en el grupo A sea de 90 mm Hg y en el grupo B de 85 mm Hg, entonces la diferencia entre las medias será de 5 mm Hg, valor que representa el primer componente, el cual se identifica como *delta* (δ).

En segundo término, será necesario disponer de alguna medida de la variación de los valores al interior de cada grupo, dado que habrá pacientes con presiones mucho menores y mucho mayores del promedio; por

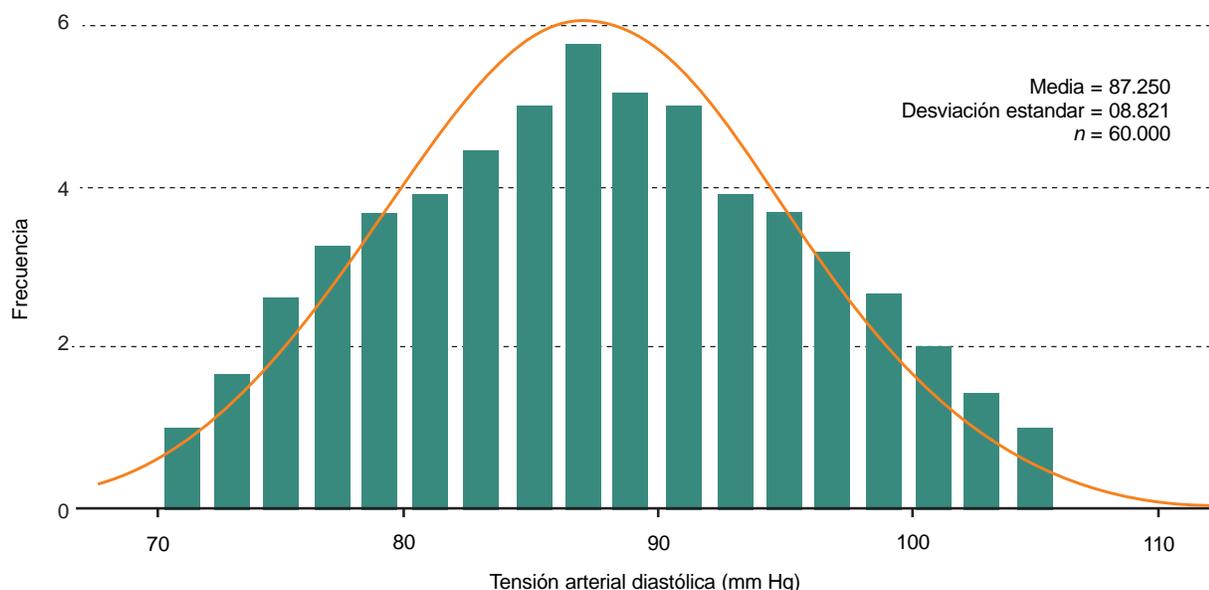


Figura 1 Grupo total de pacientes hipertensos bajo tratamiento farmacológico

ejemplo, de 60 a 112 mm Hg. Con este valor se podrá observar la variación al interior de cada grupo y, al mismo tiempo, conocer si los valores entre grupos se sobrepone en demasía en relación con la diferencia de promedio. En una variable cuantitativa, como en el modelo descrito, la medida de dispersión se conoce como desviación estándar (DE).

Como se observa en la figura 1, el promedio de TAD para la población total es de 87 mm Hg, con una desviación estándar de 9 mm Hg; mientras que en la figura 2a, el promedio de TAD del grupo A es 90 ± 9 mm Hg ($\bar{x} \pm DE$) y el promedio de TAD (figura 2b) del grupo B es 85 ± 8 mm Hg ($\bar{x} \pm DE$). Esto significa que la población general tiene un promedio de 87 mm Hg pero que sus valores a dos desviaciones estándar van de 69 a 105 mm Hg ($\bar{x} \pm 2 DE$). En el grupo A con promedio de 90 mm Hg sus valores van de 72 a 108 mm Hg ($\bar{x} \pm 2 DE$), y en el B con promedio de 85 mm Hg sus valores van de 69 a 101 mm Hg ($\bar{x} \pm 2 DE$). Los valores promedio y de dispersión de la variable de interés suelen obtenerse de información existente en estudios previos ya publicados o preliminares.

Una vez que tenemos una medida de resumen (promedio) y su correspondiente medida de dispersión (DE), hay que considerar:

1. ¿Con qué certeza queremos demostrar que la diferencia de TAD entre grupos es real? Cuando este punto no se toma en cuenta podemos incurrir en lo que se conoce como error tipo I: aceptar que la diferencia es real sin que lo sea.
2. ¿Con qué certeza queremos demostrar que la no-diferencia es real? Cuando este punto no se toma

en cuenta podemos caer en lo que se conoce como error tipo II: aceptar que la no diferencia es real.

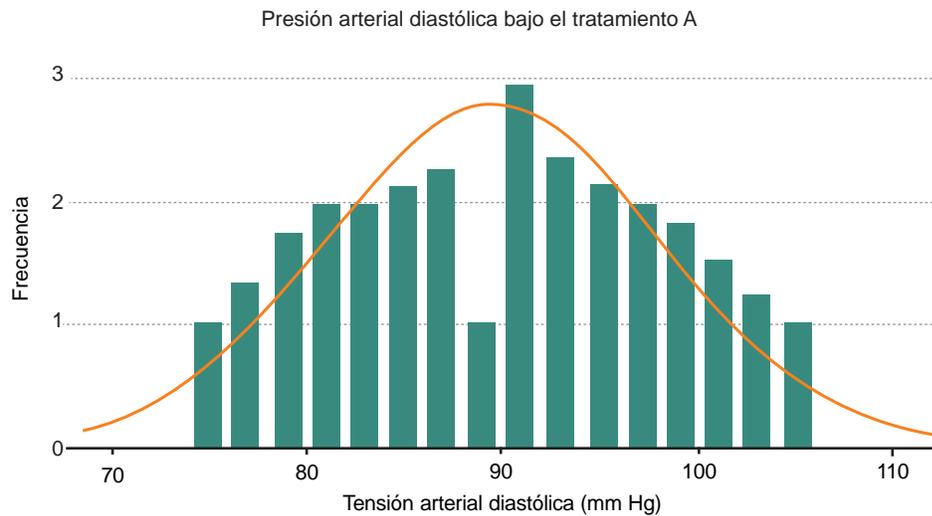
La certeza con la que habitualmente se acepta que una diferencia es real está en 95 % y corresponde a un valor de alfa (α) de 0.05, lo cual indica que al determinar que hay diferencia en la cifra de TAD entre los grupos existe 95 % de seguridad de que esa diferencia sea real y solo se acepta un error de 5 %.

Para aceptar que la no diferencia encontrada es real, se debe iniciar con una capacidad preestablecida de encontrar significación cuando hay una diferencia, lo que se conoce como poder y se representa por la diferencia de $1 - \beta$ (β). El valor del poder aceptado puede variar de 80 a 95 % y corresponde a una β del 20 al 5 %, respectivamente.

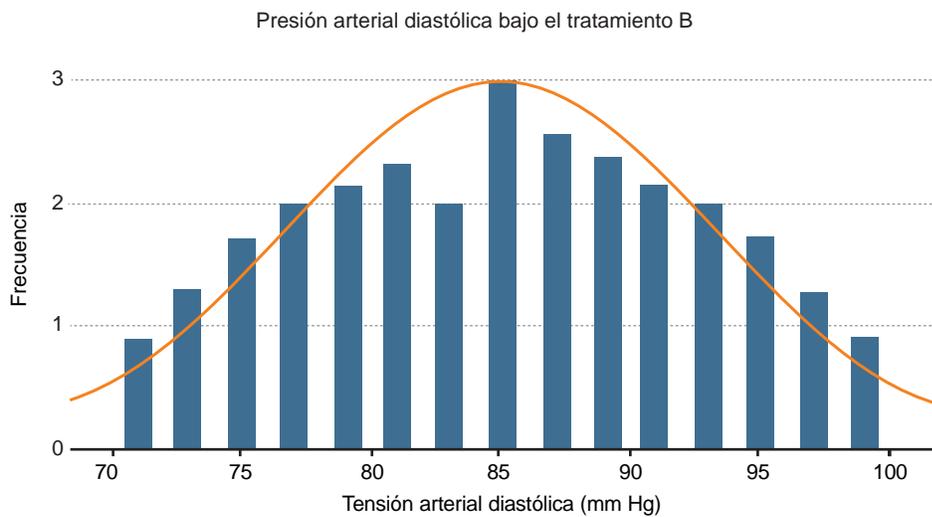
Para este momento se dispone ya de todos los componentes necesarios para estimar el tamaño de muestra:

- δ : diferencia entre las medidas de resumen (en el ejemplo es la diferencia entre las medias).
- DE: medida de dispersión, que en el ejemplo es desviación estándar.
- Error tipo I o α : error aceptado en la aseveración de que la diferencia entre las medias es real, generalmente de 5 % (0.05).
- Error tipo II o β : error aceptado en la aseveración de que la no diferencia entre las medias es real, generalmente va de 5 a 20 %.

Ignorar estos distintos componentes comúnmente provoca que al final del estudio el tamaño de muestra sea insuficiente, con lo que a pesar de contar con una



Media = 89.530
Desviación estándar = 08.803
n = 30



Media = 84.970
Desviación estándar = 08.369
n = 30

Figura 2 Pacientes hipertensos bajo tratamientos A y B, respectivamente

diferencia clínica significativa ($\geq 10\%$) no se encuentre una diferencia estadística ($p < 0.05$), lo que significa un poder bajo ($< 80\%$) y por lo tanto un error tipo II.

Z_β = valor de z relacionado con un $\beta = 0.20$ (poder de 80%).

DE= desviación estándar

μ_1 = media de grupo A

μ_2 = media de grupo B

Diferencia de medias

Con los componentes anteriores se estima el tamaño de muestra mediante la fórmula para diferencia de medias:

$$n = 2 \left[\frac{(Z_\alpha - Z_\beta) DE}{\mu_1 - \mu_2} \right]^2$$

Donde:

Z_α = valor de z relacionado con $\alpha = 0.05$ (se extrae de tablas de referencia)

De acuerdo con el ejemplo, la sustitución de los valores sería de la siguiente forma:

$Z_\alpha = 1.96$

$Z_\beta = -0.84$

DE= 9 mm Hg

$\mu_1 = 90$ mm Hg

$\mu_2 = 85$ mm Hg

Y sustituyendo en la fórmula:

$$n = 2 \left[\frac{(1.96 - (-0.84))9}{90 - 85} \right]^2 50.80 \approx 51$$

Por lo tanto, es necesario incluir 51 pacientes en cada grupo si se desea obtener 80 % de posibilidad (80 % de poder) para detectar una diferencia de medias de 5 mm Hg o más entre los dos grupos de tratamiento.

Diferencia de proporciones

Se utiliza cuando el resultado de interés está expresado en términos de proporciones. Ejemplo: comparación de dos grupos de pacientes con sobrepeso. El primer grupo de pacientes recibe fármaco y el segundo, orientación alimentaria. Si el evento de resultado se evalúa después de seis meses y se mide como la proporción de pacientes que logran normalizar su peso (índice de masa corporal menor a 25), ¿qué se necesita?

$$\alpha = 0.05$$

$$\beta = 0.10$$

$\pi_1 - \pi_2$ = (diferencia de proporciones) proporción del grupo 1 menos la proporción del grupo 2, que sea clínicamente significativa

DE = la fórmula para su determinación es 1 – proporción de grupo, que queda incluida dentro de la fórmula global

La fórmula para la determinación del tamaño de muestra para diferencia de proporciones es:

$$n = \left[\frac{Z_\alpha \sqrt{2\pi_1(1-\pi_1)} - Z_\beta \sqrt{\pi_1(1-\pi_1) + \pi_2(1-\pi_2)}}{\pi_1 - \pi_2} \right]^2$$

Donde:

$$Z_\alpha = (\alpha = 0.05) 1.96$$

$$Z_\beta = (\beta \approx 0.10-0.20) \approx -1.645, -0.84$$

$$\pi_1 = \text{proporción de grupo 1}$$

$$\pi_2 = \text{proporción de grupo 2}$$

$\pi_1 - \pi_2$ = diferencia entre proporción de grupo 1 – proporción de grupo 2, que sea clínicamente significativa

Suponiendo que para el problema de estudio se esperara que a seis meses el grupo que recibe tratamiento farmacológico tenga éxito en 70 % de los casos, mientras que el grupo con orientación alimentaria tenga éxito en 50 % de los casos, se sustituirán los valores en la fórmula de la siguiente manera:

$$n = \left[\frac{1.96 \sqrt{2 \times 0.70 \times 0.30} - (-1.645) \sqrt{(0.70 \times 0.30) + (0.50 \times 0.50)}}{0.70 - 0.50} \right]^2$$

$$n = \left[\frac{2.435}{0.20} \right]^2 = 12.18^2 = 148.35 \text{ sujetos por grupo}$$

El resultado anterior se tiene que redondear al dígito superior. De tal forma, la muestra debe incluir 149 sujetos en cada grupo de estudio si se quiere tener 90 % de posibilidad (90 % de poder) para detectar al menos una diferencia de 20 % en el porcentaje de éxito de pérdida de peso entre los dos grupos de tratamiento que se ejemplifican.

Estimación para un grupo

Por otro lado, cuando el objetivo es obtener el valor promedio de una determinada característica dentro de una población, lo que se requiere para estimar el tamaño de muestra es el valor promedio (proporción o media) y su límite superior e inferior marcado por el IC, que en la mayoría de las ocasiones se solicita de 95 o 99 %.

Para una proporción

Para estimar el tamaño de muestra de la prevalencia o proporción de un evento o característica se deben identificar distintos componentes, empezando por la medida de resumen (p_0), que corresponde a la proporción esperada, y la precisión de esta (d), que equivale a la mitad de la amplitud del IC. Si entendemos este apartado, podemos despegar la fórmula de tamaño de muestra a partir de la fórmula de la precisión, que a su vez proviene de la estimación de la desviación estándar de una proporción:

$$d = Z_\alpha \sqrt{\frac{p_0 \times q_0}{n}}$$

Al despejar n se obtiene:

$$n = \frac{Z_\alpha^2 \times p_0 \times q_0}{d^2}$$

En este caso, $q_0 = (1 - p_0)$; por lo tanto, si queremos buscar una prevalencia (p_0) de 20 %, el valor de q_0 sería $1 - 0.2 = 0.8$. Por tanto, para hacer el cálculo del tamaño de muestra para una proporción debe considerarse lo siguiente:

- La precisión (d , igual a la 1/2 de la amplitud del IC), cuyo valor en general es conferido por el investigador y se corresponde con el grado de error que pudiera ser tolerado hacia cada lado de la

- media; por ejemplo, para un error de 8 % a partir de la media, su d sería de $0.0064 \cdot 0.082 = 0.0064$.
- La confianza, también llamada Z_α , corresponde a $1 - \alpha$.
 - El valor de p_0 que se pretende estimar.

Ejemplo: ¿cuántos prematuros será necesario estudiar para verificar si la prevalencia estimada de enfermedad ósea metabólica en una población de la unidad de cuidados intensivos neonatales es de 20 % considerando una precisión de 8 % y un α de 0.05 %?

Con un nivel de confianza de 95 % ($\alpha = 0.05$; $Z_\alpha = 1.96$), $Z_\alpha^2 = 3.8416$, que al despejar:

$$N = (3.8416 \times 0.2 \times 0.8) / 0.0064$$

$$N = 96.04$$

Por lo tanto, el tamaño de muestra necesario será de 97 niños para una prevalencia esperada de 20 %, con un IC que va de 12 a 28 %.

Como podemos observar, el tamaño de la muestra dependerá de la precisión esperada del error a partir de la media, de modo que para un IC más estrecho se requiere una d menor, en general se usan los valores de 0.08 y de 0.04, y este último es el de mayor precisión (o menor error); por lo tanto, se necesitará un mayor tamaño de la muestra. De la misma manera, si se desea cambiar el nivel de confianza de 95 a 99 %, como se exige en los estudios de determinantes genéticos, el tamaño de muestra se incrementará. En el cuadro I se muestran algunos ejemplos de variación de acuerdo con esos parámetros.

Para una media

Si se entiende lo anterior, resulta sencillo comprender los componentes para estimar el tamaño de muestra

de una media. De igual forma, la base es la fórmula del IC de la media:

$$IC \text{ de } 95 \% = \bar{x} \pm Z_\alpha = \frac{DE}{\sqrt{n}}$$

En este caso, la precisión (d) se calcula de la siguiente manera:

$$d = Z_\alpha \frac{DE}{\sqrt{n}}$$

Por lo tanto, la fórmula para el cálculo de tamaño de muestra para estimar una media es:

$$n = \frac{Z_\alpha^2 \times DE^2}{d^2}$$

Esta fórmula requiere del conocimiento de Z_α , DE y la d deseada. De esta manera, el tamaño de muestra para una media esperada depende de Z_α (1.96 para $\alpha = 0.05$), de la desviación estándar que ha sido observada en los estudios previos, así como de la precisión deseada.

Consideraciones finales

Debe quedar clara la cuestión de que los anteriores supuestos no son los únicos para estimar el tamaño de una muestra, de tal suerte que si queremos estimarlo para demostrar diferencias en tasas de incidencia acumuladas (*Hazard risk ratio*) o en unidades obtenidas en modelos como las curvas de supervivencia de riesgos proporcionales de Cox, la estimación es más compleja por considerar el desenlace a lo largo del tiempo; sin embargo, el concepto base es el mismo.

Por otro lado, si la intención es controlar múltiples variables de confusión o explorar múltiples factores de riesgo mediante de un modelo de regresión logística múltiple, entonces será necesario utilizar un número de eventos por variable, para lo cual se requerirán entre 10 y 20 sujetos por cada una en el menor de los grupos de desenlace (de tal suerte si la mortalidad es de 30 %, este es el menor de los grupos, ya que sobrevivirá el restante 70 %).

Cuadro I Diferentes tamaños de muestra según distintos valores de nivel de confianza (α), de la prevalencia (p) y la precisión (d)

a (Z_α)	p	d	n
0.05(1.960)	0.2	0.08	97
0.05(1.960)	0.2	0.04	385
0.01(2.576)	0.2	0.08	166
0.01(2.576)	0.2	0.04	664

Lecturas recomendadas

Cohen J. Statistical power analysis for the behavioural sciences. Second edition. New Jersey: Lawrence Earlbaum; 1988.

Dawson B, Trapp R. Basic and clinical biostatistics. Fourth edition. Lange Medical Books-McGraw-Hill; New York, USA; 2004.

- Hulley SP, Cummings SR. Diseño de la investigación clínica. Un enfoque epidemiológico. Barcelona: Doyma; 1993.
- Portney LG, Watkins MP. Foundations of clinical research. Applications to practice. Appleton and Lange; Chicago, IL: 1993.
- Feinstein AR. Principles of medical statistics. London, UK: Chapman and Hall-CRC; 2002.

Para el cálculo del tamaño de muestra

- Brixton Health. [Sitio web]. Epicalc 2000. Disponible en [http:// www.brixton-health.com/epicalc.html](http://www.brixton-health.com/epicalc.html)
- EpiInfo 2000. Disponible en <http://huespedes.cica.es/huespedes/epiinfo/espanol.htm>
- Department of Biostatistics, Vanderbilt University. [Sitio web]. PS: Power and Sample Size Calculation. Disponible en <http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize>

VI. Relevancia clínica

Juan O. Talavera, Rodolfo Rivas-Ruiz, Marcela Pérez-Rodríguez

En la clínica suele seleccionarse la maniobra que logra un desenlace cuando menos de 10 % de superioridad directa o cuando el número necesario a tratar es ≈ 10 . Si bien estos parámetros sirven para estimar la magnitud de una asociación, estamos obligados a diferenciar las medidas de impacto (riesgo atribuible, fracción prevenible), de asociación (riesgo relativo, razón de momios, razón de riesgo) y de frecuencia (incidencia y prevalencia) que son aplicables cuando el desenlace es nominal. También debemos identificar la forma de medir la fuerza de la asociación y la magnitud de la asociación cuando la variable de desenlace es cuantitativa. No es raro interpretar las medidas de asociación como si fueran de impacto, por ejemplo, para un riesgo relativo de 0.68 se asume una reducción de 32 % del desenlace sin considerar que esta es una reducción relativa que igual proviene de una relación de 0.4/0.6, 0.04/0.06 o 0.00004/0.00006; sin embargo, la reducción directa es de 20 % (60-40 %), de 2 % y de 2 por cada 100 000, respectivamente. Por lo tanto, para estimar el impacto de una maniobra es importante contar con la diferencia directa o el número necesario a tratar.

Palabras clave

medidas de asociación
exposición
riesgo o desenlace
riesgo relativo
número necesario a tratar

Este artículo fue publicado originalmente en Rev Med Inst Mex Seguro Soc 2011; 49 (6): 631-635 y ha sido revisado para la presente edición.

Introducción

Aun teniendo un estudio bien diseñado, con un análisis estadístico y tamaño de muestra adecuados, en el que se muestra significación estadística en la asociación entre una maniobra y un desenlace (trátase de la asociación entre un factor de riesgo o maniobra preventiva y la ocurrencia de una enfermedad, o entre un factor pronóstico o maniobra terapéutica y la evolución de la enfermedad), el clínico tiene la necesidad de identificar la magnitud de dicha asociación —impacto de la maniobra— para considerar su utilidad en la práctica clínica cotidiana, donde las más de las veces se considera el beneficio de una maniobra terapéutica y suele seleccionarse la que logra un desenlace favorable con al menos 10 % de superioridad directa sobre otras. Esto significa que, por ejemplo, si el desenlace es supervivencia y la maniobra seleccionada es la *A*, se espera que esta sea 10 % superior a la maniobra estándar *B* (70 % de supervivencia a dos años con la maniobra *A* contra 60 % de la maniobra *B*), o si el desenlace es el nivel de glucosa, entonces se espera una reducción mínima de 10 % (de 140 a 126 mg/dL). Y si el desenlace es insuficiencia cardíaca, se espera una reducción mínima de 10 % del grado de falla cardíaca (que en su conjunto cuando menos exista 10 % más de pacientes que mejoraron su grado de falla cardíaca). Hay que hacer notar que la sustracción de una proporción a partir de otra se realizó en forma directa, mientras que para datos cuantitativos se estima 10 % a partir del valor de referencia.

En salud pública o medicina preventiva, las diferencias directas menores a 10 %, incluso de 4 a 7 %, resultan de suma relevancia, dado que las poblaciones susceptibles pueden ser de millones de sujetos. Lo mismo sucede en la atención clínica en la que el desenlace no deseado es de alrededor de 10 %, para lo cual cualquier reducción esperada será inferior a este y cuya relevancia dependerá de la gravedad y el costo de la alteración. Por otro lado, en caso de eventos adversos resultan significativas las diferencias incluso menores a 10 %, sobre todo dependiendo de la intensidad del evento. No obstante, en la clínica se considera deseable una ganancia mínima de 10 % para la mayoría de las situaciones.

Si bien para los médicos clínicos resulta común y entendible una diferencia de porcentaje para estimar el impacto de una asociación, en la literatura existe una serie de cálculos conocidos como *medidas de impacto*, que, si bien resultan discretamente más elaborados, acaban siendo una asociación entre proporciones. En el proceso de obtener las medidas de impacto se estiman las *medidas de asociación*, indicadores que evalúan la fuerza con la que una variable o característica se asocia con otra, mismas que no tendrían sentido si

Cuadro I Tabla de doble entrada para medidas de frecuencia relativa (ejemplo), asociación e impacto

	Desenlace +	Desenlace –	Total
Expuestos (tratados)	a 5	b 95	a + b = 100
No expuestos (placebo)	c 15	d 85	c + d = 100
Total	a + c = 20	b + d = 180	

Ensayo clínico y cohorte	Fórmula	Ejemplo	Interpretación
Incidencia de expuestos (Ie)	$Ie = a/a + b$	$5/100 = 0.05$	5 casos nuevos en 100 sujetos o 5 %
Incidencia de observados o no expuestos (Io)	$Io = c/c + d$	$15/100 = 0.15$	15 casos nuevos en 100 sujetos o 15 %
Riesgo relativo (RR)	$RR = Io - Ie$	$0.05/0.15 = 0.33$	Existe una protección. Reducción relativa o del riesgo. El riesgo está por debajo de la unidad
Reducción absoluta del RR (RAR) (riesgo atribuible [RA])	$RR = Io - Ie$	$0.15 - 0.05 = 0.1$	La reducción directa del riesgo atribuida al tratamiento es de 10 %
Número necesario a tratar (NNT)	$NNT = 1/RAR$	$NNT = 1/0.1 = 10$	Se tienen que exponer 10 personas para ver el efecto benéfico en una
Fracción atribuible (FA) (para $RR > 1$)	$Ie - Io/Ie$	Como el RR en este ejemplo es > 1 , no se calcula la FA	Se interpreta como la proporción de casos expuestos que se deben al factor de riesgo
Reducción del riesgo relativo (RRR) (para $RR < 1$, fracción prevenible)	$RRR = 1 - RR \times 100$	$1 - 0.33 \times 100 = 67 \%$	El 67 % de los casos se evitaron debido al factor de exposición

Casos y controles, y encuesta transversal

Prevalencia de los expuestos (Pe) (solo en encuesta transversal)	$Pe = a/a + b$		Número de eventos en el grupo expuesto (se usa en los estudios transversales)
Prevalencia de los no expuestos (Po) (solo en encuesta transversal)	$Po = c/c + d$		Número de eventos en el grupo no expuesto o control (se usa en los estudios transversales)
Prevalencia del factor de exposición en casos	$PfrCa = a/a + c$	$5/20 = 0.25$	25 % de los casos estuvo expuesto al factor de exposición
Prevalencia del factor de exposición en controles	$PfrCo = b/b + d$	$95/180 = 0.527$	52.7% de los controles estuvo expuesto al factor de exposición
Razón de momios (RM)	$a \times d/b \times c$	$RM = 5 \times 85/15 \times 95$ $RM = 425/1.425$ $RM = 0.29$	El grupo expuesto está protegido. El riesgo está por debajo de la unidad

La incidencia y la prevalencia son medidas de frecuencia; el riesgo relativo y la razón de momios se consideran medidas de asociación; y la reducción absoluta del riesgo y la reducción del riesgo relativo, medidas de impacto. Otra medida de asociación es la razón de riesgo, obtenida en el análisis de supervivencia de riesgos proporcionales de Cox (Hazard risk ratio, HRR). También se puede calcular el riesgo atribuible y la fracción prevenible a partir de RM (en vez de lo usar Pe y en vez de lo usar Po).

no se acompañarán de la certeza de que dicha asociación es real y no debida al azar, para lo cual se estima la *significación estadística* (una asociación es real cuando el valor de $p < 0.05$). Antes que estos dos tipos de medidas, durante el proceso de manejo de datos,

tenemos que hacer uso de lo que se conoce como *medidas de frecuencia*, las cuales miden el número absoluto de eventos. Debe resaltarse que en la mayoría de las ocasiones lo que observamos en los artículos son *medidas de frecuencia relativa*, en las que se relaciona el

Cuadro II Ejemplos de RR e intervalos de confianza de 95 %

Ejemplos de estudios	A		B		RR (IC 95 %)	RR (IC 95 %)
	Eventos	Total	Eventos	Total		
Aspirina (A) contra placebo	65	5000	95	5000	0.68 (0.50, 0.94)	
Consumo de café (A) contra placebo (B)	25	5003	24	5000	1.04 (0.60, 1.82)	
Con dislipidemia (A) contra sanos (B)	205	5000	115	5000	1.78 (1.42, 2.23)	

RR = riesgo relativo; IC 95 % = intervalo de confianza de 95 %; RRR = reducción del riesgo relativo
 Las aspirinas tienen una RRR de 32 %, estadísticamente significativa; la dislipidemia tiene un incremento del RR de 78 %, estadísticamente significativo. El consumo de café tiene un incremento relativo de 4 %, no significativo estadísticamente

número de eventos con el número total de individuos en la población o muestra en estudio, de tal forma que en una etapa posterior se pueden realizar comparaciones entre grupos con *n* distinta (cuadro I).

En la clínica, las mediciones de la asociación entre dos variables (maniobra y desenlace) a través de riesgo relativo (RR), razón de momios (OR, RM) y razón de riesgo (*Hazzard risk ratio*, HR) son comunes y se interpretan en forma similar; variables con un valor inferior a 1 se consideran de protección mientras

que aquellas con valor por arriba de 1, de riesgo. De esta forma tenemos que el riesgo común de la población o muestra de sufrir o tener el evento de interés sin identificar ningún factor ya sea de protección o de riesgo es de 1 (que corresponde a la incidencia o prevalencia del evento en el total de la muestra o población en estudio). Pero si identificamos un factor de riesgo vemos que la incidencia en este subgrupo se incrementa y en el que no la tiene se disminuye en relación con el riesgo de la población o muestra total.

Cuadro III Medidas de asociación y equivalentes para variables cuantitativas

Variable dependiente cualitativa (nominal)			Variable dependiente cuantitativa	
Medidas de frecuencia	Medidas de asociación	Medidas de impacto	Fuerza de asociación	Magnitud de la asociación
Incidencia • Tasa de incidencia • Incidencia acumulada	RR (razón de incidencia acumulada)	Riesgo atribuible (fracción etiológica, RAR y NNT)	r^2	% de diferencia de las medias Coeficiente <i>b</i>
	HR (<i>Hazzard risk ratio</i>)	RRR, FP (fracción prevenible)	R^2	% de diferencia de las medias a través de la ecuación de la regresión
Prevalencia • Prevalencia puntual • Prevalencia de periodo	RM (razón de momios de prevalencia o productos cruzados)			$(\hat{y} = a + b1X1)$
		<i>r</i>	% de diferencias de proporciones a través de la ecuación de probabilidad	
		Coeficiente <i>b</i>		
		R^2		$\hat{y} = 1/1 + e^{-(a + b1X1...)}$

$\hat{y} = 1/1 + e^{-(a + b1X1...)} = \text{probabilidad del evento}$

RRR = reducción del riesgo relativo

El NNT (número necesario a tratar) es una forma relativamente nueva de estimar la magnitud de la asociación

Por ejemplo, si consideramos el uso de aspirina para prevenir infarto del miocardio en población en la que la incidencia a un año es de 1.6 %, la incidencia en el grupo expuesto a aspirina será de 1.3 %, mientras que en el grupo control de 1.9 % con un riesgo relativo de 0.68 (0.013/0.019), lo que significa que hay una reducción del riesgo relativo de 32 %. Hasta aquí parece haber una asociación entre el uso de la aspirina y la reducción de infarto, pero habrá que revisar el intervalo de confianza de 95 % para dicho riesgo relativo: si el intervalo en sus límites (inferior y superior) está por debajo de la unidad se considera estadísticamente significativo, pero si el valor superior rebasa la unidad (1) entonces no es estadísticamente significativo y, en consecuencia, no puede descartarse que el valor puntual observado de 0.68 se deba al azar. De igual manera, cuando hablamos de un factor de riesgo es de esperar que el límite inferior del intervalo de confianza de 95 % se encuentre por arriba de la unidad (1) para que sea estadísticamente significativo (cuadro II).

Las medidas de frecuencia, asociación e impacto están basadas en la presencia o no de un evento o desenlace y, por lo tanto, se trata de una variable nominal, pero en la clínica existen numerosas variables de desenlace que se miden a través del cambio en el valor de una variable cuantitativa, en la cual de igual forma interesa conocer la fuerza y la magnitud de la asociación, por lo que resulta relevante contar con un equivalente.

En el cuadro III se muestran en un contexto global las medidas de frecuencia relativa, medidas de asociación y medidas de impacto, descritas básicamente para una variable dependiente nominal. Se agregan otras medidas también aplicables que pueden definir la fuerza de asociación —medidas de asociación—:

- El coeficiente de determinación r^2 , que mide el porcentaje de explicación de una variable a partir de la otra y que es el cuadrado de la r obtenida en una correlación, para este caso coeficiente phi.
- El coeficiente beta, valor obtenido durante un modelo de regresión (en este caso logística), que corresponde al logaritmo de la razón de momios.
- La R^2 similar a la r^2 , cuyo resultado se obtiene del modelo de regresión.

En cuanto a la magnitud de la asociación, del resultado de un modelo de regresión se puede obtener la probabilidad estimada de la ocurrencia de un fenómeno ($y = 1/1 + e^{-(a + bIXI...)}$), que en la base de la ecuación para su cálculo suma los coeficientes beta de las distintas variables, y, finalmente, calcula su RM global. Con esta ecuación si se comparan dos tratamientos se puede estimar la diferencia de

dicha probabilidad (diferencia de proporciones), aun ajustada por múltiples variables de interés; de igual forma, se pueden comparar distintas probabilidades de ocurrencia de un fenómeno por exposición a distintos valores de una variable cuantitativa.

En el mismo cuadro III se muestra cuando la variable dependiente es cuantitativa: las unidades para medir la fuerza de la asociación se limitan a la r^2 de Pearson, al coeficiente b y a la R^2 , estas dos últimas como resultado del modelo de regresión lineal.

Finalmente, para evaluar la magnitud de la asociación de una variable cuantitativa se utilizan las diferencias de las medias, más específicamente la proporción de la diferencia de las medias, ya sea estimada en forma directa o como resultado de la ecuación de regresión (en la regresión lineal el valor de la variable dependiente se obtiene en forma directa).

Una medida para la magnitud de la asociación que se ha popularizado en la actualidad es el *número necesario a tratar* (NNT = 1/RAR), el cual se refiere a la cantidad de sujetos que se requiere tratar para obtener el beneficio en uno cuando se compara con placebo; cuando este número resulta negativo, se le conoce como *número necesario a dañar*. Por lo tanto, para definir si una maniobra es clínicamente significativa puede seguir usándose una diferencia directa de 10 % o el número necesario a tratar (NNT), en el que si bien no hay un parámetro preestablecido se considera ideal un valor alrededor de 10, con el que correspondería tratar 10 sujetos para obtener en uno el beneficio deseado (equivalente a 10 %). Cabe mencionar que, por lo general, en los ensayos clínicos rara vez se utiliza placebo como grupo de comparación, por lo que este número puede estar subestimado cuando se compara con otra maniobra activa.

Comentarios

El uso apropiado de las medidas de frecuencia, asociación o impacto y sus equivalentes resulta indispensable a fin de evitar errores comunes que se cometen en la práctica clínica. No es raro interpretar las medidas de asociación como si fueran medidas de impacto; por ejemplo, si la OR o RM, el RR o el HR de una maniobra es de 0.68, se asume una reducción de 32 % del desenlace. Sin embargo, hay que tener en cuenta que esta es una reducción relativa que lo mismo proviene de una relación de 0.4/0.6, que de 0.04/0.06 o de 0.00004/0.00006 (RR=0.66); empero, en el primer caso, el NNT es de 5, en el segundo es de 50 y en el tercero es de 50 000. Por lo tanto, para estimar el impacto de una maniobra es importante contar con la diferencia directa o el NNT (RAR).

Bibliografía

1. Cordell WH. Number needed to treat (NNT). *Ann Emerg Med* 1999;33:433-436.
2. Feinstein AR. Principles of medical statistics. New York, NY: Chapman and Hall/CRC; 2002.
3. Guyatt GH, Sackett DL, Cook DJ. Users guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help in caring for my patients? Evidence Based Medicine Working Group. *JAMA* 1994;271:59-63.

Para el cálculo en línea

4. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BMJ*. 1995; 310:452-4.
5. KT Clearing House. [Sitio web]. Odds ratio to NNT converter. Disponible en <http://ktclearinghouse.ca/cebm/practise/ca/calculators/ortonnt>
6. Sociedad Española de Hipertensión/Liga Española para la Lucha contra la Hipertensión Arterial. [Sitio web]. Odds ratio, riesgo relativo y número necesario a tratar. Disponible en <http://www.seh-lelha.org/odd-ratio.htm>
7. University of British Columbia. [Sitio web]. UBC clinical significance calculator. Disponible en <http://spph.ubc.ca/sites/healthcare/files/calc/clinsig.html>

Sistema bibliotecario

información científica **a tu alcance**



Salas de lectura (SL)

Centros de Documentación en Salud (CDS)

Salas de consulta electrónica (SCE)

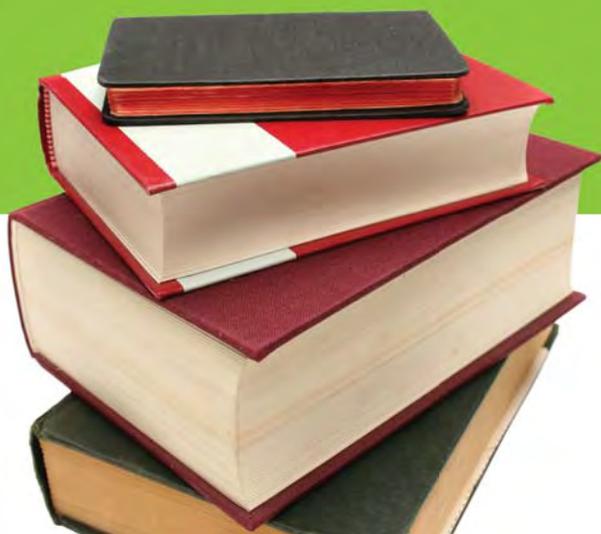
Centro Nacional de Investigación Documental en Salud (CENAIDS)

Material documental

Revistas, libros impresos
y electrónicos

- Préstamo en sala, a domicilio e interbibliotecario
- Educación de usuarios, visitas guiadas y talleres
- Consulta a bases de datos por medio de Internet
- Servicio de recuperación de documentos nacionales y extranjeros
- Difusión y promoción, alertas bibliográficas y disseminación selectiva de información

www.edumed.imss.gob.mx



Coordinación de Educación en Salud
Centro Médico Nacional Siglo XXI
5627 6900 extensión 21152, 21153 y 21256
División de Innovación Educativa
Av. Cuauhtémoc 330 Col. Doctores, México, D.F.



VII. Búsqueda sistemática: cómo localizar artículos para resolver una pregunta clínica

Rodolfo Rivas-Ruiz, Juan O. Talavera

En el proceso de responder a las dudas generadas durante la atención médica, la cantidad de artículos que aparecen durante la búsqueda resulta tan vasta que debe considerarse una estrategia para acotarla. El presente artículo describe el proceso para buscar y seleccionar la información que nos ayude a responder a las necesidades de nuestros pacientes. El juicio de la calidad y pertinencia de la respuesta dependerá de cada lector. Hay que buscar en sitios donde existe arbitraje médico para las publicaciones, razón por la que recomendamos PubMed, e iniciar la búsqueda una vez desglosado el acrónimo PICO, donde *P* = pacientes, *I* = intervención, *C* = comparador y *O* = *outcome* o resultado. El acrónimo PICO comparte los componentes con el modelo clásico de la arquitectura de la investigación descrito por el doctor Alvan R. Feinstein. Una buena búsqueda debe participar en la respuesta a nuestra pregunta en los primeros 20 artículos; de no suceder así, se tendrá que ser más específico con el uso de filtros.

Palabras clave

PubMed
MeSH
investigación clínica

Este artículo fue publicado originalmente en Rev Med Inst Mex Seguro Soc 2012; 50 (1): 53-58 y ha sido revisado para la presente edición.

Introducción

En el proceso de responder a las dudas generadas en el transcurso de la atención médica, la cantidad de artículos que aparecen durante la búsqueda resulta tan vasta que hay que considerar una estrategia que en poco tiempo nos permita encontrar los que respondan a nuestras necesidades como médicos, de tal forma que no resultemos sofocados por un mar de información. El presente artículo describe el proceso para buscar de forma sistemática documentos que nos ayuden a responder a las necesidades de nuestros pacientes, si bien el juicio de la calidad y la pertinencia dependerá de cada lector.

La accesibilidad a la información médica ha cambiado con Internet y los medios electrónicos. En el mundo se estima que existen 20 000 revistas del área, las cuales aportan aproximadamente 2 millones de artículos por año. Esta cantidad de artículos, que representa nuevo conocimiento, genera una gran dificultad para estar actualizado en todos los aspectos de la medicina.

El problema se agrava con las publicaciones en Internet de temas médicos sin arbitraje, que dependen de la buena voluntad de quien las edita y en ocasiones responden a fines diferentes a los científicos. Lamentablemente, los metabuscadores como Google o Yahoo las identifican con facilidad, lo que provoca que sean muy consultadas por pacientes y algunos médicos.

Por estos motivos, la búsqueda de la literatura médica debe realizarse en sitios en los que las publicaciones tienen un arbitraje y de acuerdo con un sistema que evite la omisión de artículos relevantes y la inclusión de artículos poco específicos para resolver nuestro cuestionamiento. Es así que la búsqueda sistemática ofrece un protocolo transparente, reproducible y auditable.

El buscador que recomendamos es PubMed, debido a que es sencillo, gratuito y, lo más relevante, los artículos que aparecen cuentan con arbitraje por expertos. Además, recientemente ha incluido opciones para hacer búsquedas en dispositivos móviles. Este sistema se encarga de difundir el acervo de la base de datos Medline de la National Library of Medicine de Estados Unidos, la cual cuenta con más de 21 millones de artículos¹ (en áreas como genética, medicina, enfermería, psicología, veterinaria, entre otras), de los cuales 90 % tiene resumen en inglés; algunas revistas vinculan al artículo en texto completo desde esta página. Esta biblioteca médica se ostenta como la más grande del mundo y ha comenzado a integrar artículos en extenso, si bien aún son pocas las revistas con libre acceso.

Ahora bien, el primer paso para resolver una pregunta es estructurarla adecuadamente a partir de los tres apartados del enfoque arquitectónico expuesto en

los capítulos anteriores: *estado basal, maniobra y desenlace*.² Para una búsqueda electrónica se ha propuesto una adaptación del modelo arquitectónico del doctor Alvan R. Feinstein, en la que se forma el acrónimo PICO, donde la *P* es pacientes, con la especificación en su caso de la enfermedad; la *I*, intervención o maniobra, tratamiento, factor de riesgo, indicador pronóstico e incluso un procedimiento diagnóstico; la *C* es comparador, que puede ser un grupo placebo, otro tratamiento o una maniobra observacional; y, finalmente, la *O*, de *outcome*, corresponde al resultado o desenlace³—este acrónimo puede tener algunas variaciones como PEO (pacientes, exposición, *outcome*) o PICOST, donde la *S* y la *T* corresponden a tipo de estudio y tiempo de seguimiento—.⁴ Trasladémoslo a un ejemplo en el que

un clínico se pregunta si el uso de albumina reduce la mortalidad en pacientes con choque hipovolémico, comparado con el uso de solución salina. Con este planteamiento se formaría el siguiente acrónimo:

P = pacientes con hipovolemia

I = tratamiento con albumina

C = solución salina

O = mortalidad

Con este acrónimo la pregunta quedaría así:

¿Reducirá el uso de albumina (al compararlo con solución salina) la mortalidad en los pacientes con hipovolemia?

1

Una herramienta que complementa este método es el acrónimo MeSH (Medical Subject Headings), vocabulario controlado de la National Library of Medicine de Estados Unidos mediante el cual se indizan y organizan los artículos en PubMed. Estas palabras permiten tener la definición de lo que se busca. A su catálogo se ingresa desde la pantalla principal de PubMed seleccionando tres opciones: el tipo de catálogo (MeSH) (1), la palabra por buscar (2) y el botón *Search* (3), como se muestra en la figura 1.

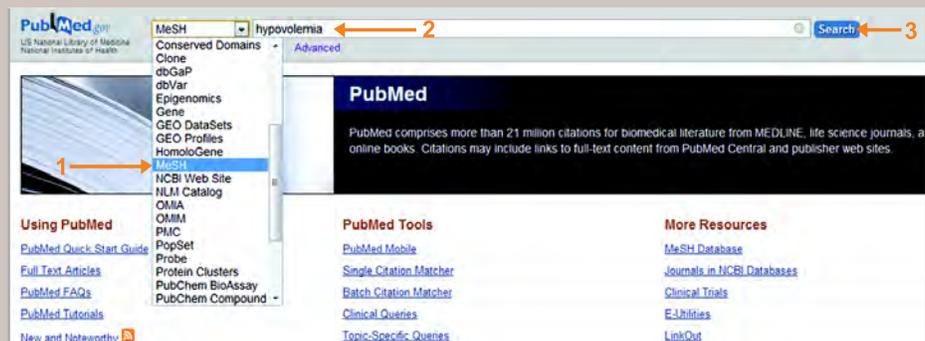


Figura 1 Opciones en PubMed para consultar el catálogo de palabras MeSH

Variantes en el proceso

Para términos novedosos, no consignados en el catálogo MeSH o si se desconoce bajo qué nomenclatura se encuentra registrado un concepto, se pueden usar *text words* o palabras libres, que serán identificadas en cualquier lugar de los artículos: título, resumen o cuerpo del artículo. La ventaja es una búsqueda amplia, con el riesgo o inconveniente de que puede arrojar artículos no relacionados directamente con el tema. Otro inconveniente es que las *text words* se deben escribir directamente en la caja de búsqueda junto con el operador booleano

```
PubMed search builder
((("Hypovolemia"[Mesh]) AND
"Albumina"[Mesh] AND (saline
solution) AND ("Mortality"[Mesh]
OR "mortality" [Subheading] OR
"Hospital Mortality"[Mesh] )
```

En nuestro ejemplo, solución salina (saline solution) no está registrada como palabra MeSH; la utilizamos por considerarla de uso muy difundido. Fue ingresada como *text word* (manualmente junto con su operador booleano)

2

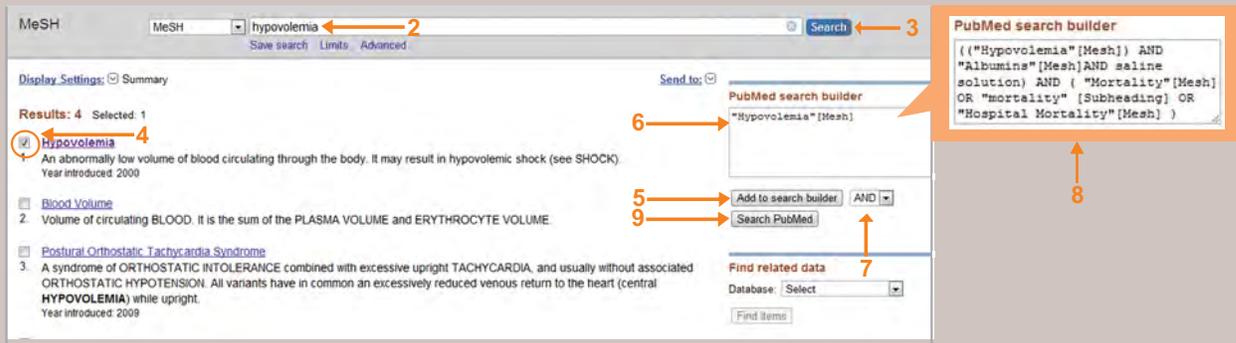


Figure 2 The MeSH words browser offers additional advantages

Con el primer término del acrónimo PICO ingresado (2) (en nuestro ejemplo hipovolemia/hypovolemia), como se muestra en la figura 2, será necesario activar el recuadro de selección (4) y oprimir la opción *Add to search builder* (5) para ingresar el término a la caja de texto (6). Los pasos 2 a 6 se repetirán para ingresar las demás palabras del acrónimo PICO, las cuales se unirán entre sí con ligandos, que corresponden a operadores booleanos (7):

- *and* (y) para unir uno o más criterios, lo que permite realizar búsquedas más específicas
- *or* (o) se ocupa para incluir uno u otro término, haciendo la búsqueda más amplia.
- *not* (no), que se utiliza para hacer una exclusión total del término que le sigue.

Veamos cómo se combinarían las palabras de nuestro acrónimo si agregamos operadores booleanos:

¿Reducirá el uso de albumina al compararlo con (AND) solución salina (AND) la mortalidad en (AND) los pacientes con hipovolemia?

Como se observa en la misma figura 2, en la caja de búsqueda (8) se irán agregando las palabras del acrónimo PICO, los operadores booleanos y, automáticamente, paréntesis, de tal manera que al terminar de

incluir todos los términos en el sistema, la búsqueda será registrada de la siguiente forma:

((“Hypovolemia”[Mesh]) AND “Albumins”[Mesh] AND saline solution) AND (“Mortality”[Mesh] OR “mortality” [Subheading] OR “Hospital Mortality”[Mesh])

Lo anterior se debe a que al combinar los términos, el sistema PubMed incluye paréntesis para realizar la búsqueda siguiendo una lógica semejante a la de las notaciones algebraicas, es decir, resuelve primero la búsqueda de los paréntesis interiores y sus resultados son combinados con los exteriores.

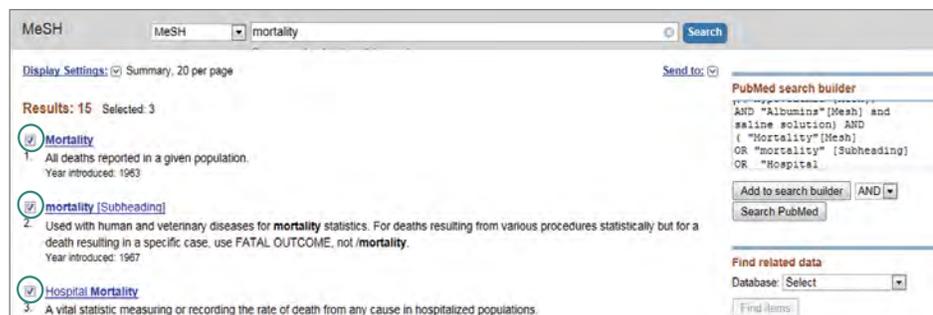
Una vez que se han ingresado todos los términos del acrónimo PICO a la caja de búsqueda, lo que resta es oprimir el botón *Search PubMed* (9).

La importancia de construir con anterioridad la pregunta PICO estriba en que el orden de ingreso de los términos será respetado, lo que permitirá una búsqueda orientada específicamente a encontrar información relacionada con nuestro cuestionamiento.

Una buena búsqueda debe lograr encontrar la solución a nuestra pregunta en los primeros 20 artículos (cuando existen estudios). Cuando no se logre obtener algún artículo al buscar enfermedades muy raras, la búsqueda debe hacerse solo con dos o tres términos o hay que ampliarla con el operador booleano *or*.

Cuando para dar cabida a un mayor número de artículos se seleccionan dos o más términos MeSH en un mismo paso, las palabras son unidas auto-máticamente con OR

En nuestro ejemplo, se seleccionaron tres opciones de mortality



3

Como se muestra en la figura 3, PubMed cuenta además con otros recursos para mejorar las pesquisas. Uno de ellos es Related citations (10), el cual genera un mecanismo de identificación que despliega los artículos más parecidos al artículo que dentro de nuestro listado seleccionemos como el ideal, con lo que se amplía la gama de documentos que podemos consultar. Como se observa, en el ejemplo se encontraron 10 artículos (11); al emplear Related citations se obtuvieron 130 (12). Otro recurso de PubMed son los filtros o límites (13), a los que es posible acceder a partir del buscador principal.

Los límites o filtros constituyen un sistema muy útil para acotar la búsqueda a fechas (14), tipo de artículo (ensayo clínico, cohorte, metaanálisis, guía de práctica clínica) (15), especie (humanos y animales) (16), lenguaje (17), sexo (18) y otros parámetros. Con estos límites se obtienen resultados más específicos, cuestión que es indispensable cuando el número de artículos que han sido identificados es abundante (figura 4).

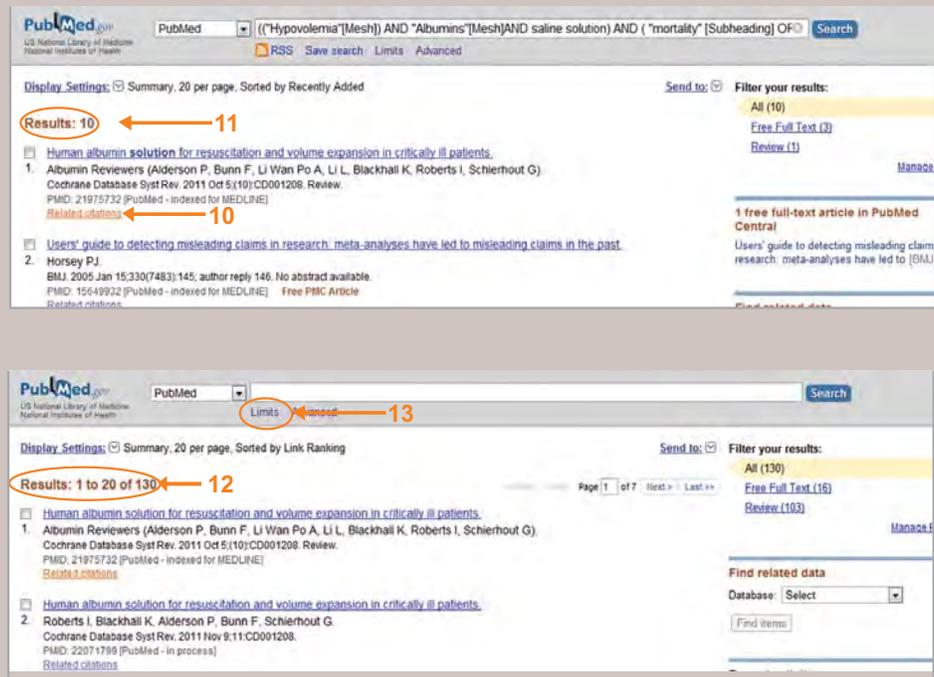


Figura 3 Otros recursos para las búsquedas

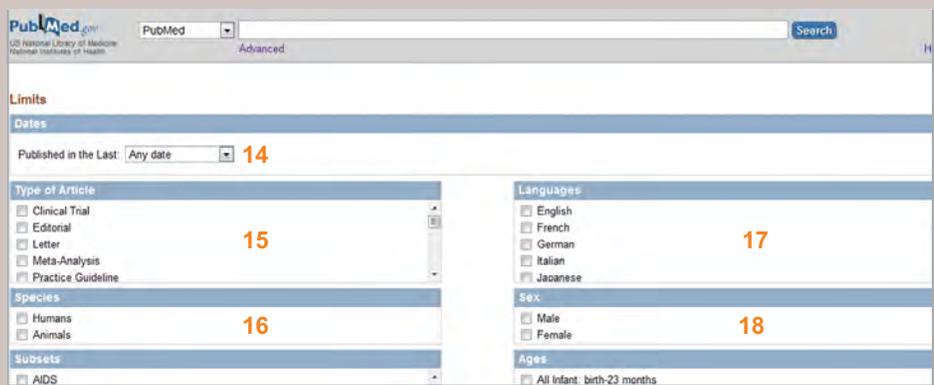
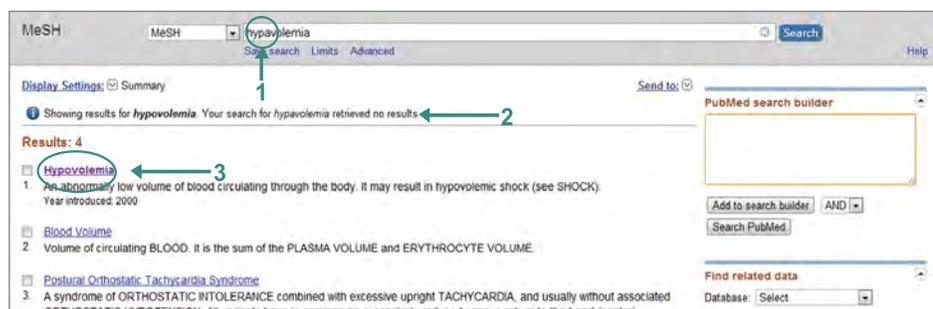


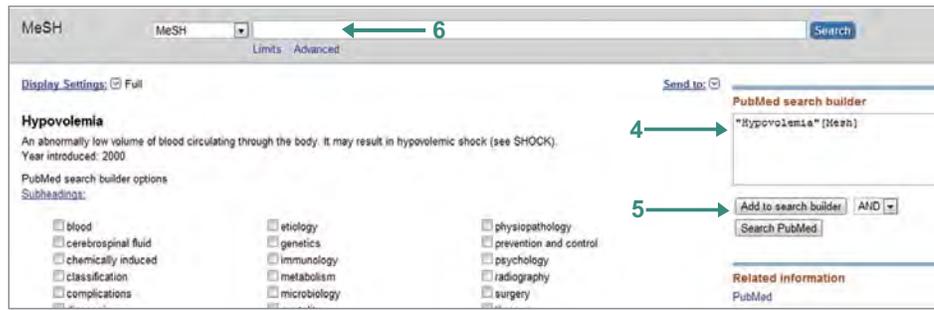
Figura 4 Detalle de la pantalla de límites

Si el usuario comete un error de escritura (1), el sistema lleva a una pantalla en la que hay una advertencia y se despliegan términos que pueden sustituir o están relacionados con el deseado (2). Si el usuario activa el término MeSH (3), se hará visible otra pantalla





La nueva pantalla despliega la definición del término y conceptos relacionados para que el usuario verifique si es el deseado. Podrá agregarlo en la caja de búsqueda (4) con la opción Add to search builder (5). Para continuar, solo deberá ingresar el siguiente término PICO (6)



4

En los metabuscadores como Google o Yahoo se puede usar este mismo sistema del acrónimo PICO. Lo único que hay que hacer es teclear las palabras en inglés y unirlas con sus términos booleanos, como puede observarse en la figura 5. En Google es posible encontrar más artículos que en PubMed y algunos que pueden estar patrocinados o no estar avalados por pares. Sin embargo, cuando se respeta el orden de las palabras PICO y uno se limita solo a ellas, el resultado suele ser similar al encontrado en PubMed en casos complementarios. En este ejemplo podemos ver resultados similares a los que se obtuvieron en PubMed, con la conveniencia de que la mayoría de las ocasiones se cuenta con el texto completo.

Esta estrategia electrónica comparte los componentes del modelo clásico de la arquitectura de la investigación descrita por el doctor Alvan R. Feinstein en su libro *Clinical Epidemiology*.⁶ Este modelo fue citado recientemente por Julian P. T. Higgins y Sally Green en el capítulo 6 de *Cochrane Handbook for Systematic Reviews of Interventions*,⁵ y fue empleado por *The Cochrane Collaboration* para la elaboración de revisiones sistemáticas.⁷⁻⁸ Este acrónimo ha sido utilizado recientemente por el modelo GRADE como mecanismo de búsqueda para realizar guías de práctica clínica

Es importante mencionar que para búsquedas más amplias, como las revisiones sistemáticas, se deben consultar otras fuentes además de PubMed, como

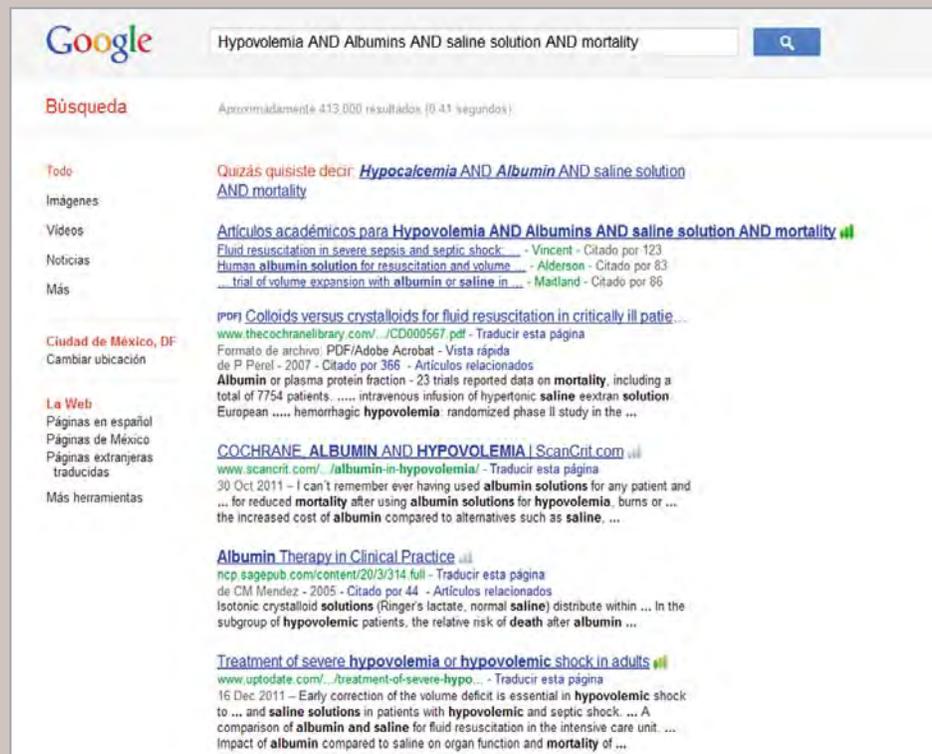


Figura 5 Los metabuscadores responden al acrónimo PICO con la ventaja de incluir no solo artículos de PubMed, sino otras publicaciones locales. Tienen la “desventaja” de identificar un gran número de resultados, lo que a veces impide una consulta completa

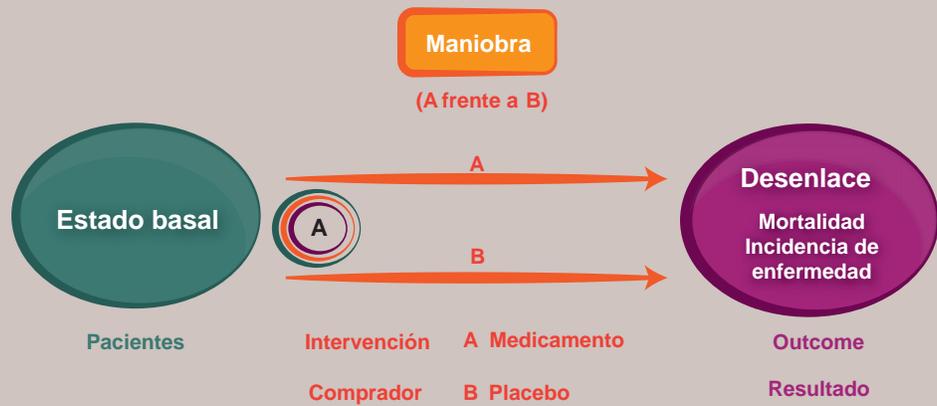


Figura 6 Utilidad del modelo clásico de la arquitectura de la investigación propuesto por el doctor Alvan R. Feinstein

EMBASE, LILACS, Imbiomed, resúmenes de congreso e, incluso, metabuscadores como Google y Yahoo.

Consideramos que este mecanismo de búsqueda y de formulación de preguntas clínicas, basado en el modelo arquitectónico y reducido al acrónimo PICO, es uno de los más útiles en la práctica clínica actual, debido a que es sumamente sensible para los motores de búsqueda electrónicos disponibles, incluso en dispositivos portátiles.

La ventaja del esquema tradicional (figura 6) es que permite identificar las partes de un estudio, los posibles sesgos, el análisis estadístico, la factibilidad

del estudio o la significación clínica, y constituye la base de la búsqueda electrónica.^{2,9-12}

Difundir y fomentar estos mecanismos de búsqueda en los hospitales podría ayudar considerablemente a resolver preguntas clínicas de un modo más rápido —con práctica calculamos que no más de 10 minutos— e incrementar la certeza en la prescripción, en la selección de una prueba diagnóstica o en la emisión de un pronóstico, facilitando así la enseñanza de la medicina, la discusión entre pares y en general el trabajo del clínico. Como complemento de una adecuada lectura y una comprensión de los artículos, este abordaje podría mejorar la calidad de la atención.

Referencias

1. US National Library of Medicine, National Institutes of Health. Pubmed. Consultado en <http://www.ncbi.nlm.nih.gov/pubmed/>
2. Talavera JO, Wachter-Rodarte NH, Rivas-Ruiz R. Clinical research III. The causality studies. *Rev Med Inst Mex Seguro Soc.* 2011;49(3):289-94.
3. Stone PW. Popping the (PICO) question in research and evidence based practice. *Appl Nurs Res.* 2002;15(3):197e-198e.
4. Tricco A, Tetzlaff J, Moher D. The art and science of knowledge synthesis. *J. Clin epidemiol.* 2011;64:11-20
5. Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0. (Actualizado en marzo de 2011). The Cochrane Collaboration; 2011.
6. Feinstein AR. *Clinical epidemiology. The architecture of clinical research.* US: WB Saunders; 1985.
7. Egger M, Smith GD, Altman D. *Systematic reviews in health care: meta-analysis in context.* Second edition. London: BMJ; 2001.
8. Khan K, Kunz R, Kleijnen J, Antes G. *Systematic reviews to support evidence-based medicine.* Second edition. London: Royal Society of Medicine; 2011.
9. Talavera JO. Clinical research I. The importance of research design. *Rev Med Inst Mex Seguro Soc.* 2011;49(1): 53-8.
10. Talavera JO, Rivas-Ruiz R. Clinical research IV. Relevancy of the statistical test chosen. *Rev Med Inst Mex Seguro Soc.* 2011;49(4):401-5.
11. Talavera JO, Rivas-Ruiz R, Bernal-Rosales LP. Clinical research V. Sample size. *Rev Med Inst Mex Seguro Soc.* 2011;49(5):517-22.
12. Talavera JO, Rivas-Ruiz R. Clinical research VI. Clinical relevance. *Rev Med Inst Mex Seguro Soc.* 2011;49(6):631-5.

VIII. Revisión estructurada de un artículo

Juan O. Talavera, Rodolfo Rivas-Ruiz

Se han intentado varias estrategias para seleccionar un artículo bajo los supuestos de pertinencia y buena calidad, que en su mayoría dependen de la presencia o no de una serie de características y en otras ocasiones del juicio de quienes clasifican el artículo. Sin embargo, estas estrategias no nos permiten conocer la magnitud del error. Debido a que no hay artículo perfecto, resulta relevante identificar la magnitud del error y su impacto en el resultado final; para ello, es necesario desarrollar destrezas que nos permitan revisar un artículo, identificar los posibles errores y generar una idea de su impacto sobre el resultado. De acuerdo con la información contenida en las partes I a VII de esta serie de artículos sobre investigación clínica, hemos intentado mostrar su aplicación en una revisión estructurada de un artículo de causalidad, empezando por la revisión del estado basal, la maniobra y el resultado, con los errores sistemáticos (sesgos) que se generan en cada apartado, seguidos por la pertinencia de la prueba, la oportunidad del tamaño de muestra y, finalmente, la relevancia clínica.

Palabras clave

artículo de revista
causalidad
estadística y datos numéricos
sesgos
tamaño de la muestra
medidas de asociación, exposición, riesgo o desenlace

Este artículo fue publicado originalmente en Rev Med Inst Mex Seguro Soc 2012; 50 (2): 163-166 y ha sido revisado para la presente edición.

Introducción

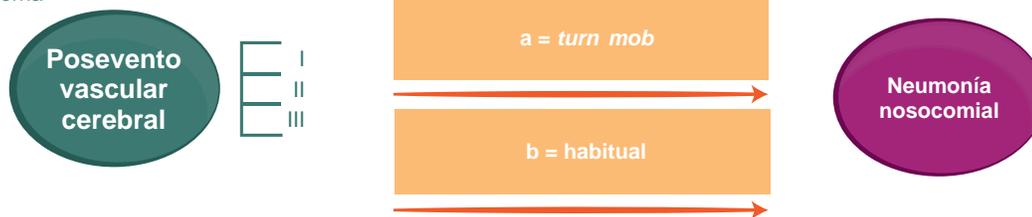
Se han intentado varias estrategias para seleccionar un artículo bajo los supuestos de pertinencia y buena calidad, que en su mayoría dependen de la presencia o no de una serie de características y en otras ocasiones del juicio de quienes clasifican el artículo. Esto conlleva a una clasificación de “adecuado” o “inadecuado”, o en el mejor de los casos a una graduación de mayor a menor calidad o pertinencia. Sin embargo, estas estrategias no nos permiten en realidad conocer la magnitud del error. Y debido a que no hay artículo perfecto, resulta relevante identificar la magnitud del error y el impacto que este pudo haber tenido sobre el resultado final; para ello, se vuelve necesario desarrollar destrezas que nos permitan revisar un artículo en forma estructurada, identificar los posibles errores y generar una idea del impacto de los mismos sobre el resultado. Es decir, no podemos depender de una clasificación ni del juicio de otros para decidir qué leer y qué no leer, o qué considerar adecuado o no adecuado. Tendremos que aprender la estructura base mínima que nos permita evaluar por nosotros mismos la pertinencia de cada artículo, sus errores y sus resultados.

De las partes I, III a VI, de la presente serie sobre investigación clínica, hemos intentado mostrar las características que consideramos básicas para realizar una lectura y una revisión estructurada de un artículo de causalidad (factor de riesgo o agente etiológico, pronóstico o tratamiento), una vez que se ha identificado el artículo a partir de una búsqueda sistemática (tema abordado en la parte VII). Comenzamos con un modelo integrado por el estado basal, la maniobra y el resultado (descritos en el artículo I), con los errores sistemáticos (sesgos) que se generan al definir y operar cada uno de estos apartados (artículo III). Y continuamos con la pertinencia de la prueba (parte IV), la estimación del tamaño de muestra (parte V) y, finalmente, la relevancia clínica (parte VI).

A continuación haremos un ejercicio sobre el uso de dicha información bajo una propuesta de revisión estructurada; para ello, utilizaremos un artículo de nuestra propia autoría: “Reduction in the incidence of poststroke nosocomial pneumonia by using the ‘Turn-Mob’ Program”, publicado en Journal of Stroke and Cerebrovascular Diseases 2010;19:23-28. El objetivo del estudio fue demostrar la eficacia de un programa de movilización en cama llamado *turn-mob* para disminuir la incidencia de neumonía nosocomial en pacientes con evento vascular cerebral isquémico.

En la figura 1 encontramos las características del estado basal como la forma de selección de la prueba y la demarcación pronóstica; se puede apreciar que la aleatorización logró balancear las características de los grupos, con excepción de la enfermedad pulmonar

Forma de selección de la población
 Paciente con déficit neurológico agudo,
 duración > 12 horas, proveniente de urgencias
 o medicina interna



Demarcación diagnóstica	Estratificación pronóstica: grupo a frente a grupo b	
< 48 horas evolución	<i>Cronométrica</i>	Edad 72 y 74 años
Sin requerimiento de apoyo ventilatorio	<i>Estatus IMC</i>	Normal 18 frente a 17 %; sobrepeso 69.4 frente a 70.5 %; Obesidad 12.6 frente a 12.5 %
Primer evento vascular	<i>Clínica</i>	Déficit motor, hemiparesia 66.7 frente a 75.9 % Hemiplegía 33.3 frente a 24.1 %; afasia 50.5 frente a 40.2 % Déficit sensitivo: 56.8 frente a 40.2; reflejo nauseoso 82 frente a 79.5 % Glasgow puntuación 15, 40.5 frente a 32.1 % NIHSS puntuación 2-7, 30.6 frente a 32.1 % 8-13, 41.4 frente a 43.8 % 14-18, 16.2 frente a 17.9 % 19-23, 11.7 frente a 6.3 %
Sin evidencia clínica de IVR superiores inferiores	<i>Morfológica</i>	Subtipo de enfermedad cerebrovascular Infarto parcial de circulación anterior 88.3 frente a 90.2 %
Sin agitación psicomotora	<i>Comorbilidad</i>	DM 50.5 frente a 42 %; HTA 83 frente a 84 %; EPOC 7 frente a 14 %; ECV 39 frente a 40 %
Diagnóstico tomográfico de EVC isquémico	<i>Tratamiento previo</i>	Corticosteroides, antibiótico
Se excluyeron aquellos que presentaron IVR en las primeras 48 horas		

IVR = infección de vías respiratorias; EVC = evento vascular cerebral; IMC = índice de masa corporal; DM = diabetes mellitus; HTA = hipertensión arterial; EPOC = enfermedad pulmonar obstructiva crónica

Figura 1 Características del estado basal: demarcación diagnóstica (criterios de selección) y estratificación (demarcación) pronóstica (variables que impactan en el desenlace independiente de la maniobra)

obstructiva crónica, discretamente mayor en el grupo *b* (14 % frente a 7 %, $p = 0.088$), que pudo impactar en el resultado final. Dado que no se realizó un análisis estratificado, no es posible observar el efecto de cada maniobra según distintos factores de riesgo, con lo que el resultado se puede atribuir principalmente a las características promedio de la población estudiada.

En la figura 2 hay que considerar la calidad de la aplicación de la maniobra principal (programa *turn-mob* contra cambios de posición habitual) y verificar que se apliquen las maniobras periféricas en forma similar en ambos grupos.

Si bien no existieron diferencias en las maniobras periféricas, la aplicación del programa *turn-mob* fue estandarizada inicialmente y verificada día a día; por su parte, nunca se estandarizó ni se verificó la aplicación del tratamiento habitual, por lo cual no hay garantía de que se llevara a cabo; más aún, al egresar el paciente del hospital, no contó en casa con el apoyo de enfermería. Esto pudo representar más que una superioridad del programa *turn-mob* sobre el tra-

tamiento habitual: el resultado de la aplicación del programa *turn-mob* contra nada.

En lo referente al desenlace, no existió la posibilidad de haber detectado en forma diferencial la presencia de neumonía nosocomial, dado que a todos los pacientes se les realizaron rayos X de tórax al egreso o ante la mínima sospecha clínica. De igual forma, no existió problema por la pérdida de pacientes (sesgo de transferencia), ya que tan solo se excluyeron dos pacientes de un total de 225 por la presencia de neumonía en las primeras 48 horas del ingreso hospitalario (figura 3).

Comentarios generales

Como comentario global del diseño metodológico y el desarrollo del proyecto, podemos decir que la selección de la población fue adecuada (ensamble adecuado), al considerar sujetos con alta probabilidad de desarrollar una neumonía nosocomial y en quienes era factible

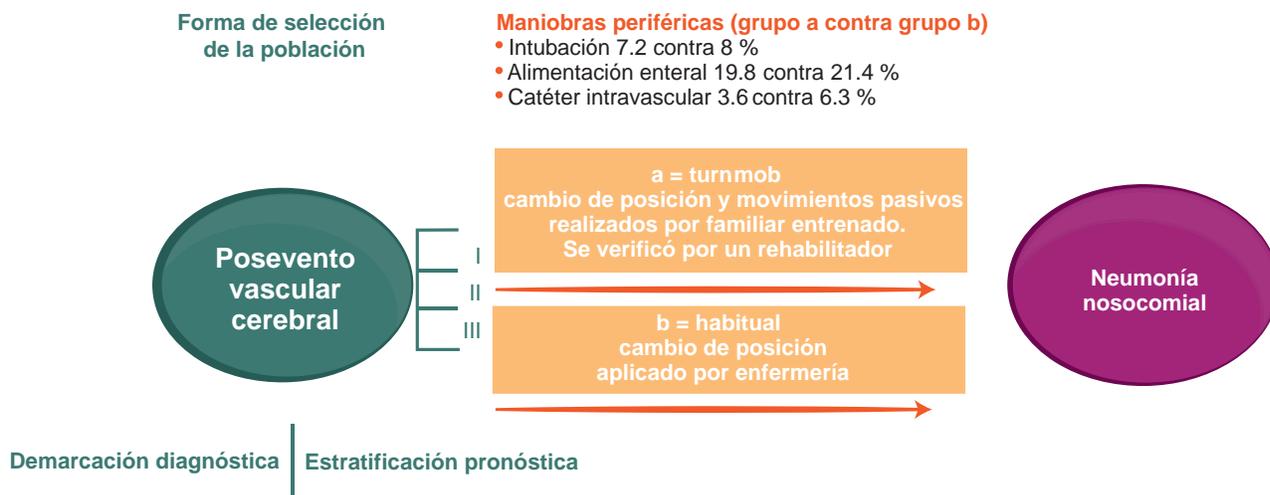


Figura 2 Características que hay que considerar durante la aplicación de la maniobra

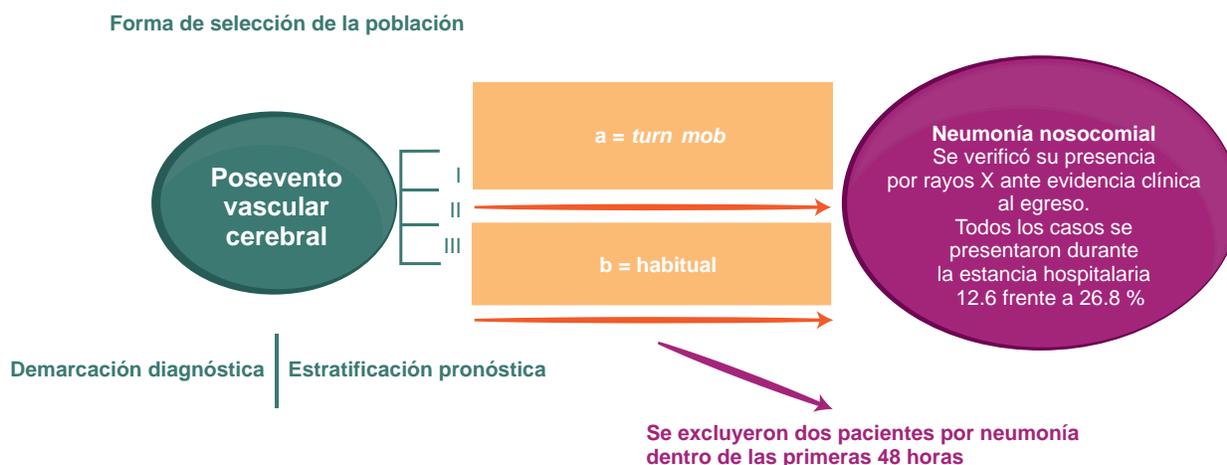


Figura 3 Características que se deben considerar en el resultado

la aplicación del programa *turn-mob*. Se demostró que la distribución de distintos factores pronósticos fue similar entre los grupos, con lo que se evitó parcialmente el sesgo de susceptibilidad, dado que no se realizó un análisis estratificado que permitiera evaluar la maniobra en distintos grupos de riesgo (susceptibilidad pronóstica). En cuanto a la maniobra, no se cuidó la adecuada ejecución de la maniobra habitual, con lo que no podemos garantizar que no hubiese existido un sesgo de ejecución. La medición del desenlace fue igual para ambos grupos, lo que evitó el sesgo de detección. Finalmente, no observamos pérdidas que pudiesen haber invertido la diferencia observada del desenlace entre los grupos (no hubo sesgo de transferencia).

En relación con la prueba utilizada (tema desarrollado en la parte IV de esta serie sobre investigación clínica),

la χ^2 muestra la comparación entre dos grupos de una variable de desenlace nominal, como la presencia o no de neumonía nosocomial.

Por otro lado, si bien se demostró que no había diferencia entre la presencia de las distintas características y el grupo de tratamiento (prueba χ^2), hubiese resultado atractivo un ajuste multivariado del efecto del programa *turn-mob*, por las múltiples características del estado basal y de las comaniobras que pudiesen haber impactado en el desenlace. En este caso, la prueba de regresión logística múltiple hubiese resultado adecuada, debido a que el desenlace es nominal.

En cuanto al tamaño de muestra (abordado en la parte V), no se menciona cómo se calculó; sin embargo, recordemos que este cálculo se lleva a cabo para obtener el número necesario de pacientes para demostrar que una diferencia esperada entre dos gru-

pos es real y no al azar. En este caso, la diferencia clínica observada de 12.6 % frente a 26.8 % resultó estadísticamente significativa, con lo cual podemos asumir que es real, dado que la probabilidad de que se deba al azar es menor a 5 % ($p < 0.05$). Y si bien no se describen los cálculos, con la incidencia mencionada en la introducción de 2 a 23 % podemos estimar que se utilizó el valor mayor y se consideró una reducción directa de alrededor de 15 %, con lo que da un tamaño de muestra entre 90 y 103 sujetos por grupo (fórmula Fleiss-Kelsey) y si a esto agregamos 20 %, obtenemos un valor de alrededor de los 225 sujetos incluidos en el estudio (estimación de tamaño de muestra para diferencia de proporciones).

Finalmente, en general se consideraron como clínicamente relevantes las diferencias directas mayores a 10 % o un NNT \approx 10 (IC-VI). En este caso la diferencia fue de 14.2 y el NNT de 7.04 pacientes (que redondeado equivale a 8) para ver el beneficio en uno. Con estos resultados claramente podemos concluir que es clínicamente relevante.

Conclusiones

No podemos descartar la presencia de un sesgo de ejecución donde el tratamiento habitual no se hubiese realizado, en cuyo caso la conclusión no sería que el programa *turn-mob* es mejor que la movilización habitual realizada por enfermería, sino que se tendría que concluir que el programa *turn-mob* en un paciente con posevento vascular cerebral isquémico es mejor que la no rotación ni movilización. Por otro lado, no podemos identificar si el programa *turn-mob* conserva su beneficio en distintos estratos de severidad del paciente, dado que no se realizó un análisis estratificado, ni se ajustó a través de un análisis multivariado; es probable que no se hayan realizado dichos análisis por el tamaño de muestra, ya que 44 casos de neumonía nosocomial resultan insuficientes al estratificar o ajustar. Como podemos ver, todo estudio tiene errores, no obstante existe información valiosa; sin embargo, para ponderarla resulta indispensable tener un conocimiento mínimo en investigación clínica.

Bibliografía

- Talavera JO. Clinical research I. The importance of the research design. *Rev Med Inst Mex Seguro Soc.* 2011;49(1): 53-8.
- Talavera JO, Wachter-Rodarte NH, Rivas-Ruiz R. Clinical research II. Studying the process (the diagnosis test). *Rev Med Inst Mex Seguro Soc.* 2011;49(2):163-70.
- Talavera JO, Wachter-Rodarte NH, Rivas-Ruiz R. Clinical research III. The causality studies. *Rev Med Inst Mex Seguro Soc.* 2011;49(3):289-94.
- Talavera JO, Rivas-Ruiz R. Clinical research IV. Relevancy of the statistical test chosen. *Rev Med Inst Mex Seguro Soc.* 2011;49(4):401-5.
- Talavera JO, Rivas-Ruiz R. Clinical research V. Sample size. *Rev Med Inst Mex Seguro Soc.* 2011;49(5):517-22.
- Talavera JO, Rivas-Ruiz R. Clinical research VI. Clinical relevance. *Rev Med Inst Mex Seguro Soc.* 2011;49(6):631-5.

IX. Del juicio clínico al ensayo clínico

Juan O. Talavera, Rodolfo Rivas-Ruiz

Se describen dos estrategias para entender la causalidad y documentarla con la mejor evidencia: el juicio clínico y el ensayo clínico. En el primero se identifican el estado basal, la maniobra y el desenlace, cada uno con características que muestran la complejidad del fenómeno de causalidad, cuyo control permite evitar errores sistemáticos: en el estado basal, el ensamble inadecuado y el sesgo de susceptibilidad; durante la aplicación de la maniobra, el sesgo de ejecución; en la medición del desenlace, los sesgos de detección y transferencia. En el ensayo clínico se mencionan las tácticas que intentan aislar el efecto de la maniobra principal del de otros componentes del fenómeno de causalidad previamente descritos en el juicio clínico. Para ello, se aprovechan la oportunidad de manipulación de la maniobra y la temporalidad en la relación causal. Entre sus características, están la asignación y el cegamiento de la maniobra, la factibilidad de la detención anticipada de esta, el análisis de acuerdo con la adherencia de la maniobra, los grupos por comparar, la temporalidad de la maniobra comparativa y el consentimiento informado. Cuando el médico aplica consciente y estructuradamente este conocimiento con su paciente, mejora su eficiencia y acerca la práctica médica a la investigación clínica.

Palabras clave

ensayo clínico
sesgo

Este artículo fue publicado originalmente en Rev Med Inst Mex Seguro Soc 2012; 50 (3): 267-272 y ha sido revisado para la presente edición.

En “Investigación clínica III” de esta misma serie se abordó el razonamiento clínico (juicio clínico) como modelo lógico para explicar el fenómeno de causalidad, previamente descrito por el doctor Alvan Feinstein en sus libros *Clinical Biostatistic* y *Clinical Epidemiology. The Architecture of Clinical Research*. Según Feinstein, todo médico sensato debe conocer ese razonamiento. Nos atrevemos a decir que no tan solo todo médico sensato lo conoce, sino que lo aplica durante su práctica clínica. Sin embargo, en ocasiones le falta hacerlo en forma consciente y, en consecuencia, estructurada. De igual forma, en el número I de esta misma serie se mencionaron los diseños de investigación como una estrategia para obtener la evidencia de dicha causalidad. Dentro de ellos, el ensayo clínico aporta la evidencia de mayor calidad.

El presente artículo muestra estas dos estrategias para explicar y documentar el fenómeno de causalidad, e intenta mostrarlas en paralelo, de tal manera que a partir de una se entienda con facilidad la razón de la otra:

- El juicio clínico, o razonamiento clínico/arquitectura de la investigación clínica, como descripción fenomenológica de la investigación clínica.
- El ensayo clínico, como diseño que ofrece la mayor calidad de la información durante el proceso de investigación clínica, al intentar controlar o cuando menos documentar la participación de todo componente dentro del fenómeno de causalidad.

Juicio clínico

Para explicar el fenómeno de causalidad en el razonamiento clínico, se describen estado basal, maniobra y resultado (y sus características), así como cinco fuentes de error sistemático en las que se puede caer al omitirlas: dos en el estado basal, una durante la aplicación o medición de la maniobra y dos en el desenlace.

Fuentes de error en el estado basal (figura 1)

a) *Ensamble inadecuado*. Generalmente se presenta cuando se omiten componentes de la demarcación diagnóstica. Está definido por el lugar de donde proviene la población, los criterios diagnósticos y los criterios de selección.

b) *Sesgo de susceptibilidad pronóstica*. Generalmente se observa cuando se omite la estratificación pronóstica. En él se deben considerar todos los factores presentes en el estado basal que pueden impactar en el desenlace, independientemente del efecto de la maniobra.



Figura 1 Características que hay que considerar en el estado basal para evitar un ensamble inadecuado y un sesgo de susceptibilidad

Fuentes de error en la maniobra (figura 2)

a) *Sesgo de ejecución.* Suele presentarse cuando no se consideran los distintos componentes para tener una potencia óptima de la maniobra, por lo que la calidad de la maniobra resulta deficiente; también se presenta cuando no se consideran aquellas acciones que la acompañan antes, durante o después, y que se conocen como comaniobras o maniobras periféricas. Además, se debe especificar la comparabilidad de la maniobra (eficacia, efectividad y eficiencia), así como la multiplicidad de maniobras y la concurrencia temporal de la maniobra comparativa.

Fuentes de error en el desenlace (figura 3)

a) *Sesgo de detección.* Identificación desigual del desenlace, ya sea por sospecha diagnóstica o por desigual número de evaluaciones del desenlace entre los grupos.

b) *Sesgo de transferencia.* Pérdida de seguimiento de pacientes cuya causa no es el azar. El incremento de 20 % del tamaño de muestra no resuelve el problema cuando las pérdidas están asociadas con la maniobra, simplemente mantiene la estabilidad de los datos para que durante el análisis estadístico se conserve el poder de la prueba.

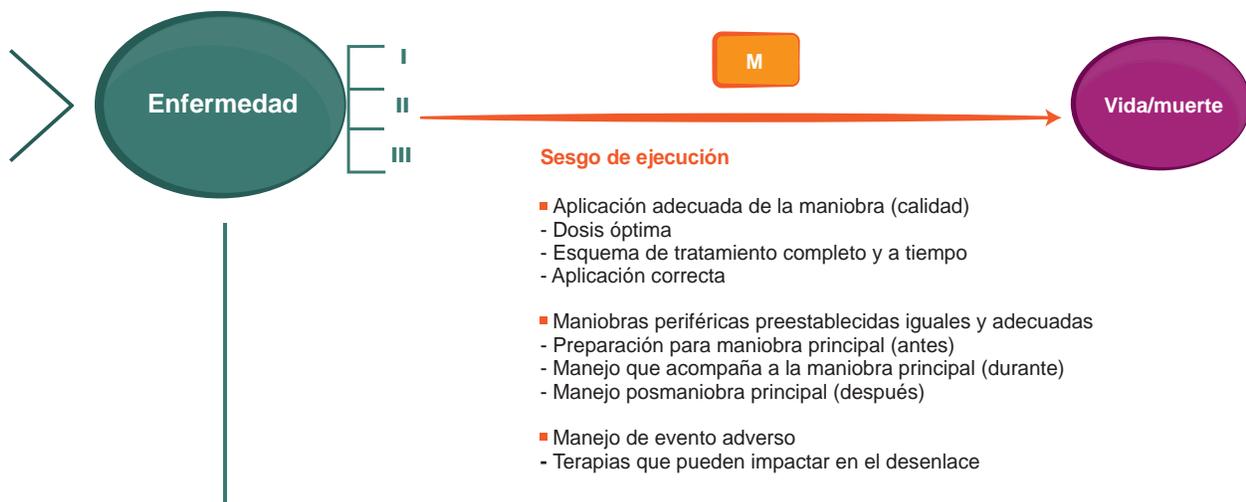


Figura 2 Características que se deben considerar en la maniobra para evitar el sesgo de ejecución

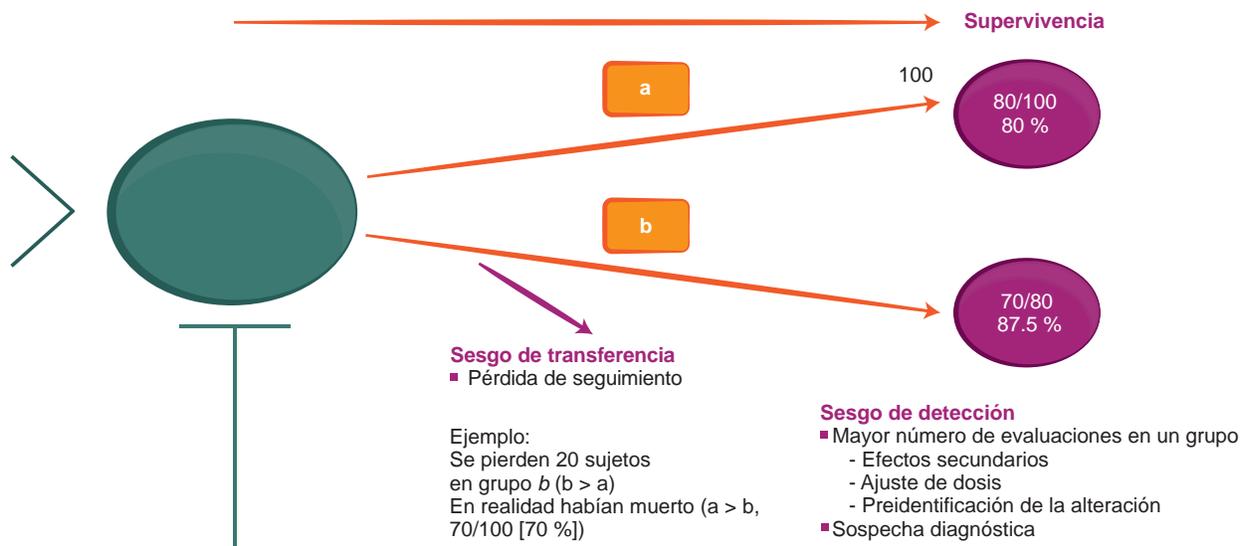


Figura 3 Características que se deben considerar en el desenlace para evitar el sesgo de detección diagnóstica y de transferencia

Ensayo clínico

El ensayo clínico permite obtener la información con tal calidad que intenta aislar el resultado que provoca la maniobra principal sobre el estado basal y controla los componentes que pueden participar en el desenlace o provocar una evaluación sesgada del mismo.

El ensayo clínico, a diferencia de los estudios observacionales, permite manipular la maniobra, lo que le confiere características propias.

Dentro de las características que acompañan a la maniobra, ya sea en un periodo anterior inmediato, durante o en un periodo posterior, las siguientes son exclusivas del ensayo clínico (figura 4):

Ensayo clínico — Experimental — Longitudinal — Prolectivo — Comparativo

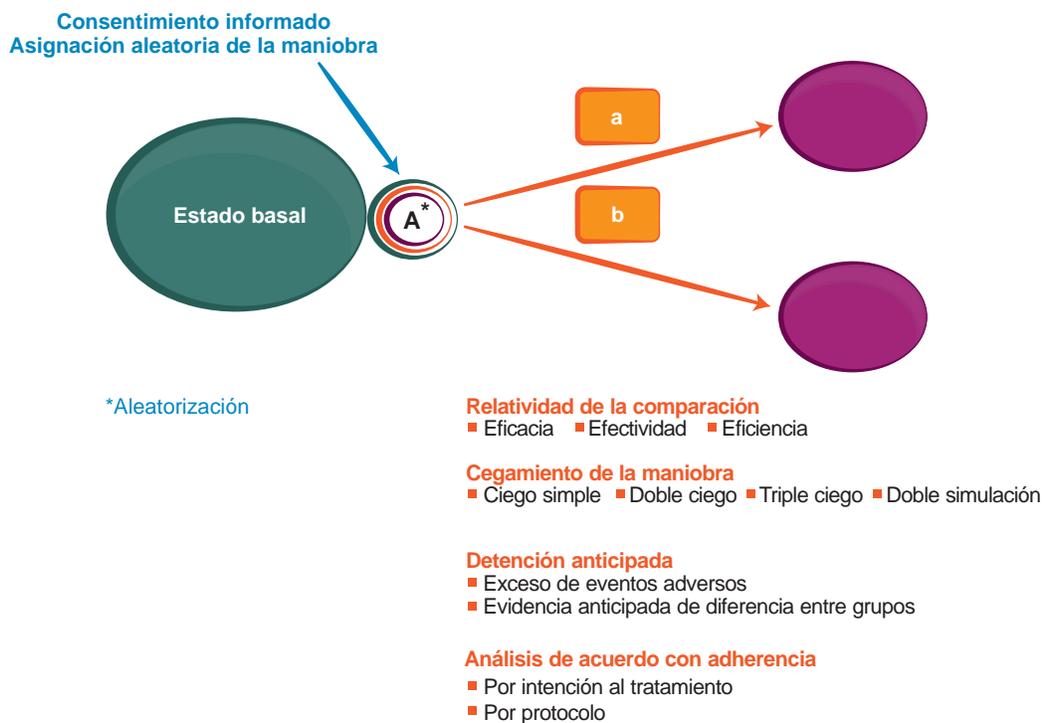


Figura 4 Características del ensayo clínico

- La asignación de la maniobra: es la característica distintiva entre el ensayo clínico y otros diseños, ya que solo el ensayo clínico da la oportunidad de asignar la maniobra. La asignación aleatoria de la maniobra busca generar grupos de condiciones basales similares entre las distintas maniobras (para evitar sesgo de susceptibilidad pronóstica), evitando de esta forma discrepancias que pudiesen posteriormente ser las causantes de diferencias en el desenlace. Si bien es una estrategia altamente popular, no evita la presencia del fenómeno de “transestratificación” ni especifica el impacto de la maniobra en distintos estratos de pronóstico (véase “Investigación clínica III”). Este fenómeno se puede prever si se realiza una asignación aleatoria por estratos, siempre y cuando el análisis de los resultados sea al interior de cada estrato y no solo en forma global. De igual forma, la asignación aleatoria tiene otras funciones como el respeto al principio ético de proveer a cada individuo la misma probabilidad de recibir la maniobra experimental, y la posibilidad de que queden asignados a cada brazo de tratamiento sujetos con similares probabilidades de apego a la maniobra (para evitar sesgo de ejecución) y similar probabilidad de abandono del estudio (lo que disminuye el sesgo de transferencia). Por último, cabe mencionar que la asignación aleatoria facilita el cegamiento de la maniobra. Es así como la maniobra disminuye la probabilidad de sesgos propios del estado basal, de la maniobra y del desenlace.

- Cegamiento de la maniobra: esta estrategia busca principalmente evitar que la subjetividad participe en la evaluación del desenlace (para evitar sesgo de detección). Se subdivide en tres categorías dependientes de quienes dentro del proceso de investigación desconocen la maniobra de tratamiento:

a) *Ciego simple*: se considera cuando el paciente ignora qué tratamiento está recibiendo, es decir, desconoce a qué maniobra fue asignado.

b) *Doble ciego*: cuando el paciente y el investigador desconocen el brazo de tratamiento.

c) *Triple ciego*: cuando el paciente, el investigador y quien analiza los datos desconocen el brazo de tratamiento.

Aunado a esto, cuando la forma de aplicar un medicamento es diferente (por ejemplo: el fármaco *a* se aplica dos veces al día y el fármaco *b* tres veces al día; o fármaco *a* se administra vía oral y el fármaco *b* en forma intramuscular), o cuando la imagen física del fármaco es distinta (fármaco *a*, pastilla azul; fármaco *b*, amarilla), se utiliza la doble simulación (*doble dummy*); por ejemplo, si el paciente recibe el fármaco

a solo dos veces al día y el *b* tres veces, al *a* habrá que agregarle tres placebos del *b*, que deberá tomar de igual forma tres veces al día y al *b*, dos placebos del *a*, que habrá de tomar dos veces al día.

- Detención anticipada: los ensayos clínicos pueden ser interrumpidos por dos razones inherentes al tratamiento: diferencia anticipada entre los grupos en el desenlace principal, siempre que no exista probabilidad de que dichas diferencias se pierdan al tener la muestra o el seguimiento completo; y por la presencia de eventos adversos, por arriba del intervalo de confianza superior de 95 %, estimado de acuerdo con el tamaño de muestra o periodo de seguimiento correspondiente.
- Análisis de acuerdo con la adherencia a la maniobra: difícilmente un ensayo clínico con periodo de seguimiento que supere unos cuantos días llega al final con un apego de todos los participantes cuando menos a 80 % de la maniobra (por ejemplo, el consumo del fármaco en 80 % de las dosis). En general se espera que entre los grupos de tratamiento los pacientes que no se apegaron sean similares en número y características —basales y en maniobras periféricas—; de igual forma, que los sujetos con falta de apego sean de características similares a quienes llegaron al final del estudio con un apego adecuado. Es así que asumiendo falta de apego al azar entre los grupos, se analizan los datos a través de dos estrategias:

a) *Análisis por intención de tratamiento (intention to treat analysis)*, la cual se caracteriza por incluir por igual en la evaluación del desenlace a los sujetos que cumplieron con un apego adecuado ($\geq 80\%$) y a quienes no (apego $< 80\%$).

b) *Análisis por protocolo*, cuando la decisión estriba en solo incluir en el análisis los datos provenientes de los sujetos con un apego $\geq 80\%$.

En el análisis por intención al tratamiento suele observarse una disminución de las diferencias entre los grupos de tratamiento, mientras que en el análisis por protocolo suele conservarse lo que podría ser la diferencia real entre las maniobras, siempre y cuando las pérdidas hayan sido al azar; de otra manera, podría estarse favoreciendo a uno de los grupos (imáginese que no se apegan los sujetos con más eventos adversos y que estos son diferenciales entre las maniobras, o que no se apegan los sujetos con mejor o peor respuesta al tratamiento y que esta fuera de igual forma diferencial entre los grupos; de ocurrir, se presentaría el sesgo de ejecución).

Entre otras características de los ensayos clínicos, no exclusivas, dado que pueden ser consideradas en

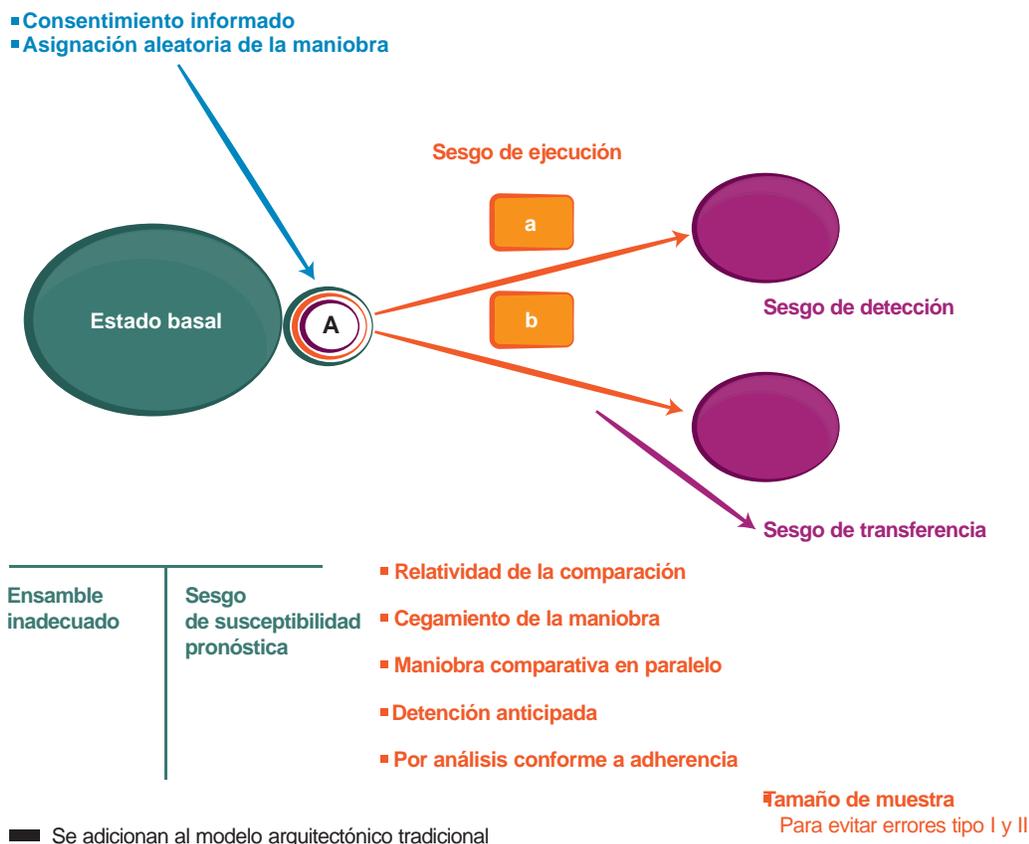


Figura 5 Características del ensayo clínico en paralelo al razonamiento clínico

estudios observacionales, se encuentran las siguientes (figura 4):

- Grupos por comparar. Es importante evaluar cuál es el comparador de la maniobra principal, ya que dependiendo de este se ha clasificado a los ensayos clínicos en estudios de eficacia, efectividad y eficiencia:

a) *Eficacia*: cuando se compara la maniobra activa contra placebo o contra nada. Esta comparación pretende demostrar que la maniobra activa funciona mejor que no hacer nada o que tan solo dar un placebo.

b) *Efectividad*: representa la comparación de la maniobra activa con un tratamiento estándar; por lo tanto, busca demostrar la superioridad de una maniobra contra otra. Este estudio debe ser ponderado con cuidado, ya que no encontrar diferencias significativas no significa que las maniobras sean iguales o equivalentes. Si lo que se busca es demostrar equivalencia, habrá que estimar el tamaño de la muestra para una diferencia máxima de alrededor de 3%. Si lo que se busca es que no haya inferioridad, habrá que considerar una diferencia máxima de 9%.

c) *Eficiencia*: se refiere no a una comparación, sino al impacto de la maniobra una vez que se aplica en la comunidad.

- Temporalidad de la maniobra comparativa. En la mayoría de las ocasiones, los ensayos clínicos que comparan dos o más maniobras tienen la virtud de hacerlo dentro de un tiempo calendario y, en consecuencia, con aplicación de la maniobra en forma simultánea (en paralelo). Otra modalidad de comparación distinta son los estudios cruzados, en los que las maniobras por comparar se llevan a cabo en tiempos sucesivos y en forma alterna en cada uno de los sujetos de estudio; la gran ventaja es que los sujetos por comparar son los mismos y, por lo tanto, el resto de variables fuera de la maniobra principal son idénticas; no obstante, estos estudios tienen algunos problemas como: 1) el efecto de arrastre (*carry over*), en el que al introducir la segunda maniobra las condiciones basales del sujeto han cambiado por acción de la primera, o 2) cuando la enfermedad ha cambiado por sí misma durante el periodo de tiempo de aplicación entre la primera y la segunda maniobra. Por otro lado, este tipo de diseño es propio de patologías estables con mínimos cambios esperados en el periodo programado

de estudio (en el que al retirar la primera maniobra verdaderamente se regresa al estado basal previo), y en patologías cíclicas (cuyo comportamiento es prácticamente el mismo en cada ciclo).

Al comparar distintas maniobras al mismo tiempo o en periodos muy cercanos, se prevé que las condiciones de diagnóstico de la patología estudiada sean similares y que la posibilidad de acceso a las maniobras periféricas sea semejante; de esta forma, se evita que las diferencias entre las terapias no sean dadas por diferencias en el diagnóstico (sesgo de susceptibilidad) o en la accesibilidad a maniobras periféricas (sesgo de ejecución), o en los criterios diagnósticos (ensamble inadecuado), o en los criterios de evaluación del desenlace (sesgo de detección). Finalmente, debemos mencionar que en un ensayo clínico las condiciones basales y el tiempo de seguimiento de los sujetos incluidos y aleatorizados a una u otra terapia es el mismo.

- Consentimiento informado. Dado que en todos los casos la maniobra será asignada aun cuando sea de riesgo mínimo, se requiere cuidar los principios

éticos de investigación en seres humanos. (Por ello, se resaltarán los principios que se deben considerar para salvaguardar los derechos y el bienestar de los pacientes que participan en proyectos de investigación.)

Conclusiones

Identificar y organizar mentalmente los pormenores del fenómeno de causalidad durante el curso clínico de una enfermedad, y conocer las razones de las características propias del ensayo clínico, permite comprender la unión de la práctica clínica con la investigación clínica y, en consecuencia, facilita la explotación razonada y estructurada de ambas en forma bidireccional para el beneficio de los pacientes. Es relevante resaltar, tal y como lo mencionó el doctor Alvar Feinstein, que la gente más habituada al manejo de la causalidad es el médico clínico, quien cada vez que asigna una maniobra a su paciente está aplicando este conocimiento y destrezas, y que hacerlo en forma consciente y estructurada sin duda alguna mejorará su eficiencia y acercará la práctica médica a la investigación clínica (figura 5).

Lecturas recomendadas

- Feinstein AR. *Clinical Biostatistics*. Saint Louis: The CU Mosby Co; 1977.
- Feinstein AR. *Clinical epidemiology. The architecture of clinical research*. Philadelphia: WB Saunders; 1985.
- Feinstein AR. Directionality and scientific inference. *J Clin Epidemiol*. 1989;42(9):829-33.
- Portney LG, Watkins MP. *Foundations of clinical research: applications to practice*. Third edition. New Jersey: Pearson-Prentice Hall; 2009.
- Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. Baltimore: Williams & Wilkins; 2008.
- Sackett D, Haynes R, Tugwell P. *Epidemiología clínica una ciencia básica para la medicina clínica*. Madrid: Ediciones Díaz de Santos; 1989.
- Talavera JO. Clinical research I. The importance of the research design. *Rev Med Inst Mex Seguro Soc*. 2011; 49(1):53-8.
- Talavera JO, Wachter-Rodarte NH, Rivas-Ruiz R. Clinical research III. The causality studies. *Rev Med Inst Mex Seguro Soc*. 2011; 49(3): 289-94.

X. Del juicio clínico al diseño de cohorte

Juan O. Talavera, Rodolfo Rivas-Ruiz

Después del ensayo clínico, el segundo diseño de investigación con mayor calidad de información es la cohorte. Si bien no se cuenta con la posibilidad de asignación de la maniobra, se tiene la oportunidad de contar con el seguimiento de los sujetos a lo largo del tiempo. Cualquier investigación que intenta explicar el fenómeno de causalidad corre el riesgo de caer en sesgos; sin embargo, las particularidades de los estudios de cohorte intentan evitarlos. Sus principales características son: 1 ser observacional, aspecto en que el investigador solo mide la presencia de la maniobra, característica que divide a los sujetos en expuestos y no expuestos; 2 ser longitudinal, lo que ofrece la oportunidad de seguir al sujeto a lo largo del tiempo, documentando la secuencia temporal de aparición de los componentes del fenómeno de causalidad; 3 tener direccionalidad en las mediciones, lo que genera la existencia de cohortes prolectivas, retrolectivas y retroprolectivas (las primeras son las de mayor calidad por contar con la medición en tiempo real de las variables de interés); 4 ser comparativo.

Palabras clave

estudios de cohortes
estudios de seguimiento
estudios longitudinales
estudios prospectivos
estudios retrospectivos

Este artículo fue publicado originalmente en Rev Med Inst Mex Seguro Soc 2012; 50 (4): 383-388 y ha sido revisado para la presente edición.

El estudio de cohorte se caracteriza por el seguimiento de un grupo de sujetos con características similares a lo largo del tiempo. Después del ensayo clínico, es el segundo diseño de investigación con mayor calidad en la obtención de información. Si bien no se cuenta con la posibilidad de asignación de la maniobra que es propia del ensayo clínico, se tiene la oportunidad de disponer del seguimiento de los sujetos a lo largo del tiempo y, en consecuencia, con la congruencia de medir la maniobra antes de la aparición del desenlace (maniobra observacional, dado que no es asignada por el investigador —también es conocida como “medir la exposición”—).

Es importante mencionar que cualquier estudio de investigación que intenta explicar el fenómeno de causalidad corre el riesgo de caer en sesgos, ya sea al definir el estado basal (por ensamble inadecuado y sesgo de susceptibilidad), durante la maniobra (sesgo de ejecución) o al medir el desenlace (sesgo de detección y sesgo de transferencia), según se muestra en las figuras 1a, 1b y 1c, descritas previamente en “Investigación clínica III” e “Investigación clínica IX” de esta misma serie. Sin embargo, las características de los estudios de cohorte intentan evitarlos.

Características principales (cuadro I)

Exposición a la maniobra

Se trata de un estudio observacional, por lo que el investigador solo puede medir la exposición a la maniobra, a diferencia del ensayo clínico, en el que el investigador la asigna. Hay que mencionar que si bien el ensayo clínico es el diseño ideal para evaluar una maniobra terapéutica, en la actualidad se acepta la evaluación de esta por medio de estudios observacionales como el de cohorte (se puede evaluar el efecto de un fármaco indicado por alguien ajeno al investigador, por ejemplo, los estudios fase IV). Incluso resulta el modelo ideal cuando se busca evaluar una maniobra que no puede ser asignada por el investigador por aspectos éticos.

Es importante mencionar que la maniobra divide a la cohorte en los grupos por comparar; en su estado basal, los sujetos conforman la cohorte como un solo grupo que comparte características similares, y con la maniobra principal se distribuyen en *expuestos* y *no expuestos*. Se deberá estimar el efecto de la variable principal sobre el estado basal para generar el desenlace, siempre ajustando por las variables de confusión que se pueden presentar en el estado basal (ensamble inadecuado y sesgo de susceptibilidad) o durante la acción de la maniobra principal (sesgo de ejecución). En un ensayo clínico se busca controlar las variables



Figura 1a Características que hay que considerar para evitar un ensamble inadecuado y el sesgo de susceptibilidad

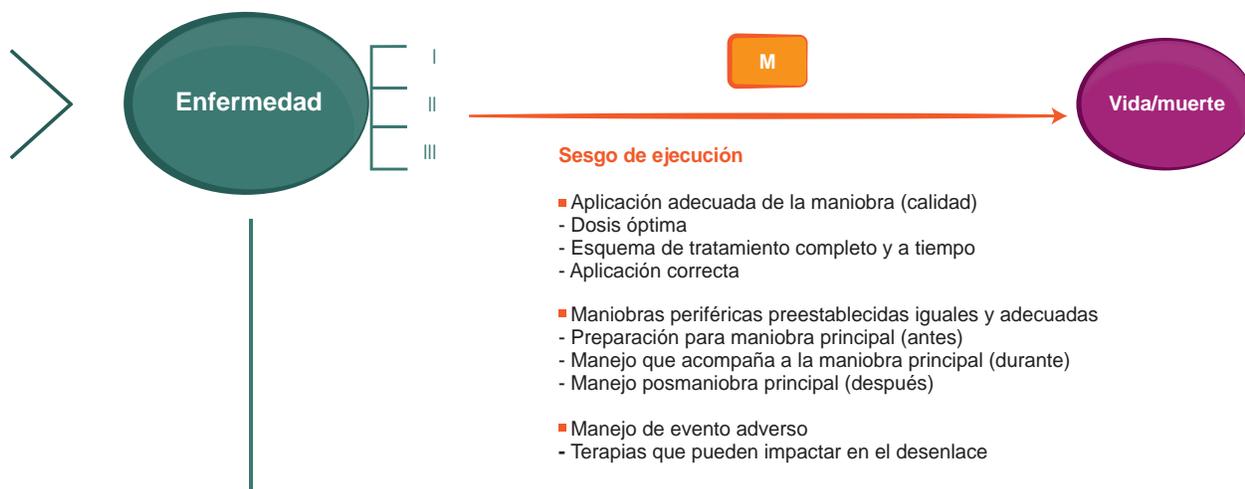


Figura 1b Características que se deben considerar para evitar el sesgo de ejecución

que confunden mediante la asignación aleatoria de la maniobra, posibilidad que no existe en el diseño de cohorte, razón por la que se debe ser acucioso en la medición de posibles variables confusoras.

Seguimiento del sujeto

La segunda y más importante característica de este diseño es su carácter *longitudinal*, es decir, existe seguimiento del sujeto en estudio, con la medición de la o las variables de interés a lo largo del tiempo, de tal forma que puede documentarse el cambio (por ejemplo, de valores de glucosa) o la aparición de la varia-

ble de interés (por ejemplo, infarto, muerte, evento adverso).

Durante el seguimiento de la cohorte, existe la posibilidad de incluir sujetos en un momento similar dentro del curso clínico de su enfermedad —generalmente al inicio, lo que se conoce como una *cohorte de inicio o incipiente*—, y de seguirlos en forma homogénea durante un tiempo previamente establecido, ya sea hasta el fin del periodo de seguimiento o hasta el desenlace. En estos casos el estudio se conoce como *cohorte cerrada*, caracterizado por contar con periodos similares de seguimiento (figura 2a). En contraparte, existe la *cohorte abierta o dinámica*, cuando se

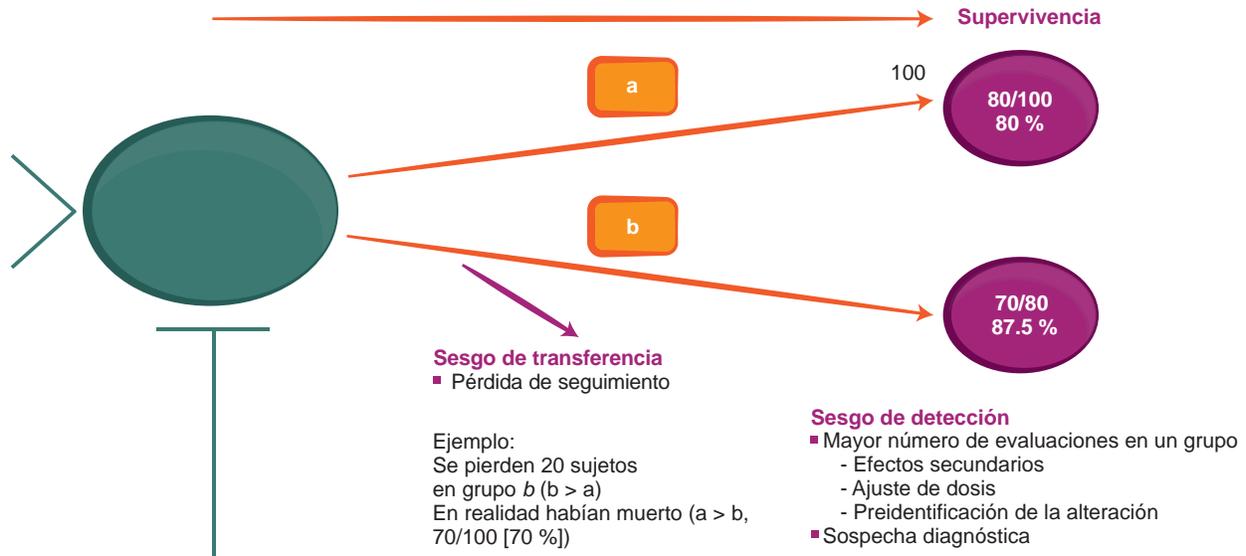


Figura 1c Características que hay que considerar para evitar el sesgo de detección y transferencia

acepta la inclusión y salida de los sujetos de estudio en distintos momentos durante el curso clínico de la enfermedad, en cuyo caso se cuenta con periodos de seguimiento heterogéneos (figura 2b).

Dado el seguimiento de los sujetos de estudio, existe la probabilidad de que se presente el sesgo de ejecución si no se dispone de una maniobra homogénea y constante al interior de cada grupo y ante maniobras periféricas heterogéneas entre los grupos. Además, al ser un diseño que implica el seguimiento de sujetos en el tiempo, la posibilidad de que ocurran pérdidas de ellos es elevada, lo que provoca un sesgo de transferencia. Finalmente, debe mencionarse que particularmente en las cohortes dinámicas se puede propiciar el ensamble inadecuado o el sesgo de susceptibilidad, al incluir a sujetos con menor o mayor probabilidad de sufrir el desenlace, por ejemplo, al incluir en periodos posteriores al estado basal solo a sobrevivientes (*cohorte de sobrevivientes*).

Direccionalidad en las mediciones

La tercera característica del diseño de cohorte es la direccionalidad en la medición de la información, que da lugar a lo que conocemos como estudio de cohorte prolectiva (prospectiva), a la cohorte histórica o cohorte retrolectiva (retrospectiva) y a la cohorte ambispectiva o retroprolectiva (retroprospectiva) (figura 3).

La *cohorte prolectiva o prospectiva* se caracteriza por la medición de las características basales, de seguimiento y del desenlace en tiempo real y bajo estándares previamente establecidos, lo que proporciona gran calidad a la obtención de dicha información, de tal manera que la evaluación del impacto de la maniobra principal sobre el estado basal para generar un desenlace es de gran precisión.

En la medición de la maniobra principal y demás variables participantes en el fenómeno de causalidad

Cuadro I Características del diseño de cohorte

Diseño	Exp/Obs	Long/Trans	Prol/Retrol	Comp/Desc	Medición
Cohorte	Observacional	Longitudinal	Prol/Retrol/Rp	Comparativo	Incidencia

El enfoque metodológico considera cuatro características: 1. La imposición o no de la maniobra con fines de investigación: estudio experimental (Exp) u Observacional (Obs), respectivamente. 2. Seguimiento (Long) o no (Trans) del paciente a lo largo del tiempo. 3. Direccionalidad en la obtención de la información: prolectivo (Prol), retrolectivo (Retrol) y retroprolectivo (Rp). 4. La búsqueda o no de asociación de dos o más variables: comparativo (Comp) y descriptivo (Desc), respectivamente. La medición de la ocurrencia del desenlace (Medición), ya sea a través de incidencia, prevalencia o simplemente la proporción caso-control

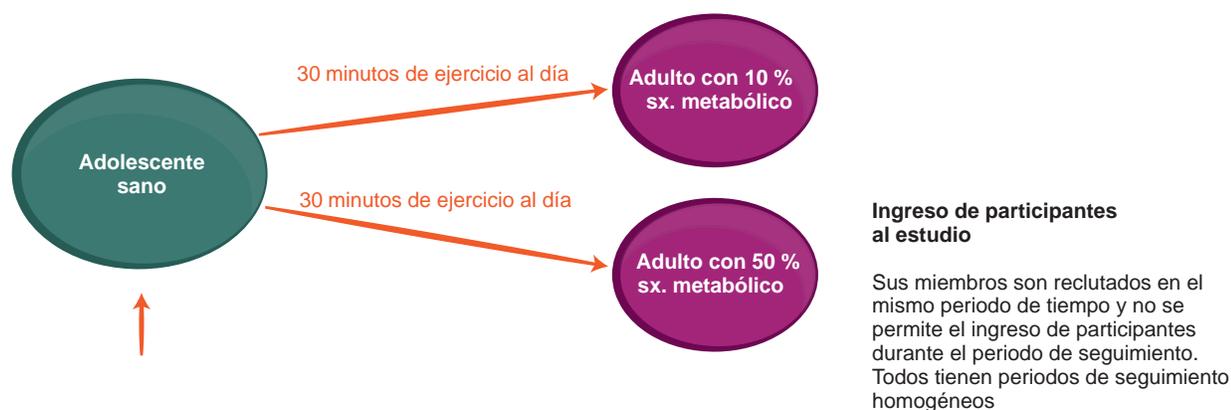


Figura 2a Diseño de cohorte cerrada

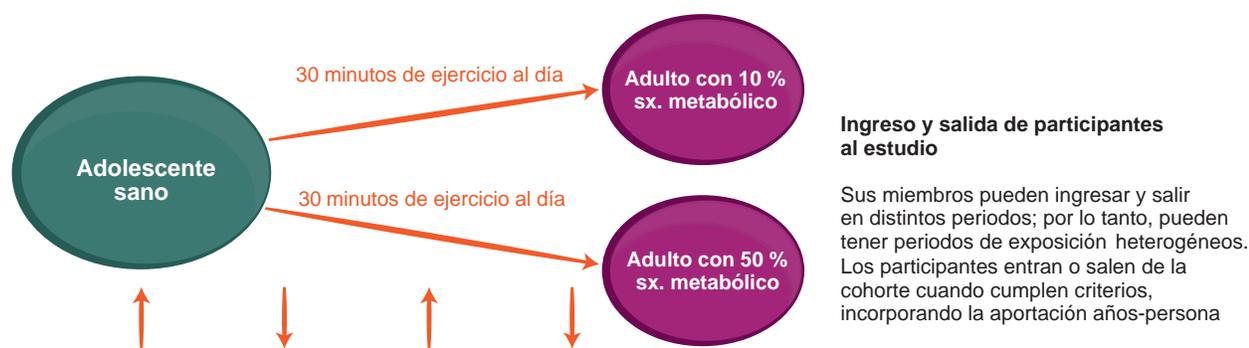


Figura 2b Diseño de cohorte abierta o dinámica

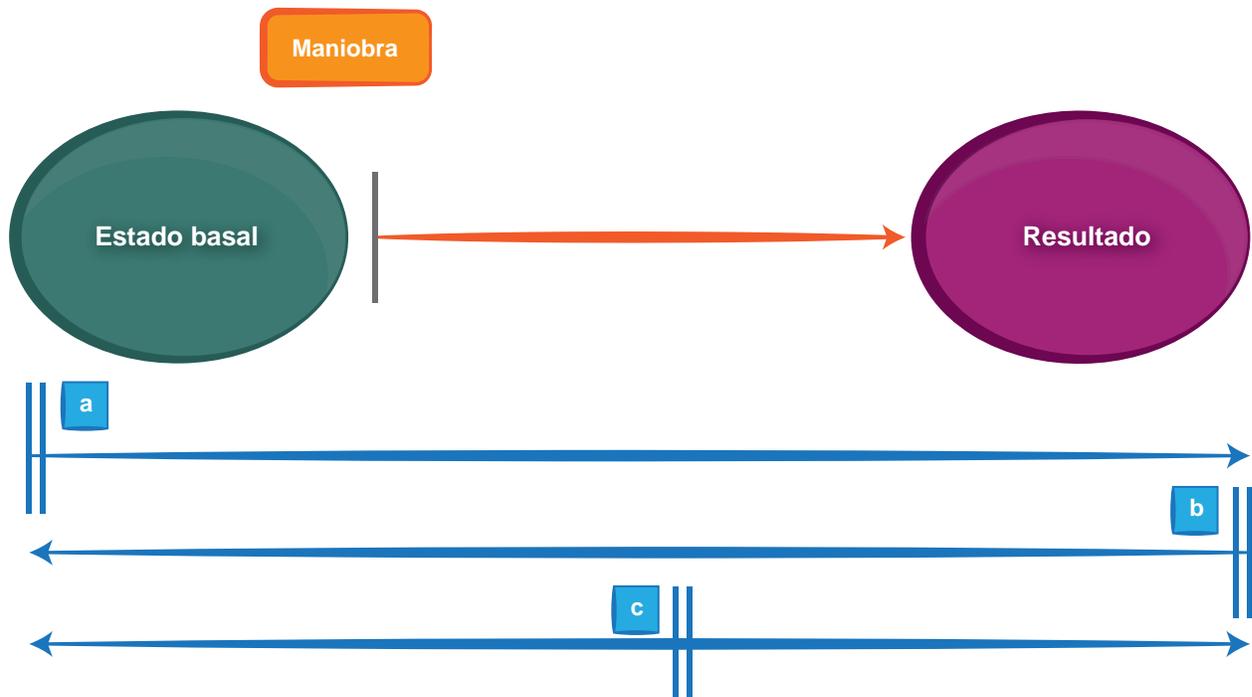
(variables de confusión), son múltiples las posibilidades que pueden generarse, como la medición con criterios tan específicos como se desee o medir el grado de exposición a esta, ya sea en el estado basal o durante el seguimiento —simulando el apego en el caso de la maniobra principal (lo que evita el sesgo de ejecución)—. La previsión y la medición de las posibles maniobras que den lugar a confusión permiten realizar ajustes, ya sea en el estado basal (con lo que se evade el sesgo de susceptibilidad) o durante la ejecución de las maniobras periféricas (para evitar el sesgo de ejecución). Por último, la medición objetiva, específica y homogénea de la ausencia del desenlace en el estado basal, y de la ocurrencia del mismo durante el seguimiento o al término del estudio, impide un ensamble inadecuado al principio (cuando ya estaba en forma incipiente el desenlace al inicio del estudio) y posteriormente el sesgo de detección.

Para simular la ceguera de la maniobra propia solo del ensayo clínico, en el estudio de cohorte se espera que la medición de las variables en el estado basal la haya realizado personal independiente a quien evalúe la exposición a la maniobra y, a su vez que ambos sean

independientes de quien evalúe el desenlace. Las ventajas que ofrece la planeación anticipada a los sucesos dentro del fenómeno de causalidad solo son propias de los estudios de cohorte prolectiva y del ensayo clínico. Es así que entre los estudios observacionales, la cohorte prolectiva resulta el modelo con mayor calidad en la recolección de información ideal para evaluar causalidad.

La *cohorte histórica* o *retrolectiva* no permite medir el impacto de la maniobra con la misma precisión que la cohorte prolectiva, dado que no se mide en tiempo real ninguna variable en ninguno de los componentes descritos en el diseño arquitectónico —razonamiento o juicio clínico—. En la cohorte histórica, la población que se decide evaluar ya estuvo expuesta a la variable de interés y ya sufrió o no el desenlace, con lo cual ha acabado el periodo de seguimiento. Sin embargo, si bien no se puede medir ningún componente en tiempo real, deben existir criterios específicos para cada variable por medir, pero propios y esperables en un registro clínico habitual. Durante la planeación del estudio, los investigadores debieron especificar criterios para cada variable por medir y estrategias para mejorar la calidad

Direccionalidad en la obtención de la información



a = Cohorte prospectiva: todas las variables, ya sean del estado basal, de la exposición de la maniobra o del desenlace, se miden en tiempo real. b = Cohorte retrospectiva: la medición se realiza cuando el tiempo de seguimiento ha terminado y el desenlace ha sucedido; en consecuencia, la exposición a la maniobra, las condiciones basales y el desenlace no se miden en tiempo real. c = Cohorte retroprospectiva, es una combinación: las condiciones basales ya acontecieron, la exposición a la maniobra ha ocurrido en forma total o por un tiempo parcial, pero el desenlace aún no y, por lo tanto, se mide en tiempo real

Figura 3 Tipo de cohorte de acuerdo con la direccionalidad de las variables

de la información. Una de ellas consiste en fragmentar el registro clínico en tres apartados: el que corresponde al estado basal, el de la exposición de la maniobra y el de la medición del desenlace, de tal manera que cada bloque de información pueda ser revisado en forma independiente (similar a lo descrito en la cohorte prospectiva). Si bien cuenta con la gran desventaja de que parte de la información puede no encontrarse en el registro clínico o su calidad puede ser dudosa, la cohorte histórica muestra lo que ocurre en la práctica real, con lo cual, al evaluar una maniobra terapéutica, el resultado es más cercano a lo que ocurrirá una vez que esta se aplique en la población, a diferencia de lo que sucede con el ensayo clínico o con la cohorte prospectiva, sin el efecto de la vigilancia ni la acuciosidad en las mediciones ni en el seguimiento del sujeto.

Búsqueda de asociación

La cuarta característica del diseño de cohorte es la búsqueda de asociación. En realidad, son pocos los estudios descriptivos en la actualidad; sin embargo, todo estudio describe

las características de su población en el primer párrafo de resultados. La cohorte es un estudio comparativo, ya sea porque compara la exposición de los sujetos en estudio con distintas maniobras o con el cambio o la aparición de alguna característica a lo largo del tiempo.

Comentarios

Es importante resaltar en qué momento ocurre el ensamble de la población en el diseño de cohorte, puesto que es una de las características que claramente lo diferencia de los otros diseños observacionales. En la cohorte, la población se integra en el estado basal, sin importar la direccionalidad de las mediciones. Por ejemplo, si se trata de una cohorte prospectiva de pacientes con diabetes mellitus tipo 2 y se quiere seguir por 10 años, todo paciente que recién se diagnostique con la enfermedad en una población específica, y que cubra los criterios de selección, podrá ingresar y será seguido durante 10 años, con medición de las variables en tiempo real. Pero si se trata de una cohorte

retrolectiva (histórica), todo paciente perteneciente a la población de interés que 10 o más años atrás se haya diagnosticado con diabetes mellitus tipo 2 y que en aquel momento contara con los criterios de selección, podrá ingresar y será seguido en sus registros de aquel momento hasta cubrir el tiempo de seguimiento o la aparición del desenlace; evidentemente en ese caso las variables no serán medidas en tiempo real.

Lecturas recomendadas

- Feinstein AR. Clinical biostatistics. Saint Louis: Mosby; 1977.
- Feinstein AR. Clinical epidemiology. The architecture of clinical research. Philadelphia: WB Sanders; 1985.
- Feinstein AR. Directionality and scientific inference. *J Clin Epidemiol.* 1989;42:829-33.
- Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Third edition. New Jersey: Pearson-Prentice Hall; 2009.
- Rothman KJ, Greenland S, Lash TL. Modern epidemiology. Third edition. Baltimore: Williams & Wilkins; 2008.
- Talavera JO, Wachter-Rodarte NH, Rivas-Ruiz R. Clinical research III. The causality studies. *Rev Med Inst Mex Seguro Soc.* 2011;49(3):289-94.
- Talavera JO. Clinical research I. The importance of research design. *Rev Med Inst Mex Seguro Soc.* 2011;49 (1):53-8.

XI. Del juicio clínico al diseño de casos y controles

Juan O. Talavera, Rodolfo Rivas-Ruiz

El diseño de casos y controles, al igual que la cohorte histórica, carga con una serie de posibles sesgos derivados de reconstruir los hechos una vez que ha sucedido el desenlace, además de los sesgos generados por la selección del grupo control. Se caracteriza por contar con una serie de casos, para los cuales se identifica un grupo comparativo (controles). Es decir, se parte del desenlace hacia la causa y, en consecuencia, se obliga a reconstruir los hechos en el sentido opuesto a como sucede el fenómeno de causalidad. No obstante, habrá que tener en mente el diseño arquitectónico y considerar en cada apartado —estado basal, maniobra y desenlace— aquellas características necesarias para demostrar el efecto de la maniobra, evitando un ensamble inadecuado y los sesgos de susceptibilidad, de ejecución y de detección. El sesgo de transferencia solo podrá ser controlado cuando se cuente con una población definida, ya sea de base poblacional o anidada en una cohorte. Cuando no se cuente con una población definida, este diseño solo se recomienda para enfermedades raras.

Palabras clave

estudios de casos y controles
ensayo clínico

Este artículo fue publicado originalmente en *Rev Med Inst Mex Seguro Soc* 2012; 50 (5): 505-510 y ha sido revisado para la presente edición.

Si bien el estudio de casos y controles aparentemente es un diseño sencillo para resolver cuestionamientos, sin duda alguna es el más complejo. Al igual que la cohorte histórica, carga con una serie de posibles sesgos derivados de la reconstrucción de los hechos que anteceden al desenlace, además de los sesgos en la selección del grupo control. Por lo anterior, este diseño deberá considerarse solo en las ocasiones en que no sea posible responder la pregunta clínica por medio de un ensayo clínico o de un estudio de cohorte.

La recolección de la información necesaria para documentar el fenómeno de causalidad —descrito bajo el concepto de arquitectura de la investigación o juicio clínico (figuras 1, 2 y 3)— se lleva a cabo bajo condiciones ideales por medio del ensayo clínico, cuya característica más importante es la asignación de la maniobra (experimental). Cuando este diseño no es posible se recurre a la cohorte, que conserva la oportunidad de seguir a la población en estudio a lo largo del tiempo, con la posibilidad de documentar la maniobra antes de que se presente el desenlace (longitudinal). Sin embargo, habrá que considerar el diseño de casos y controles si la rareza del fenómeno que se analiza, la dificultad para completar el tamaño de la muestra o el uso pertinente de los recursos, obligan a ello.

Este diseño se caracteriza por contar con una serie de casos para los cuales se identifica un grupo control (grupo comparativo). A diferencia del ensayo clínico y del estudio de cohorte —en el que se asigna (experimental) o se identifica la maniobra antes del desenlace (observacional) y se le da seguimiento hasta la evaluación de este (longitudinal)—, en el estudio de casos y controles se trata de reconstruir el efecto de la maniobra una vez que ha ocurrido el desenlace (para los casos) o se ha documentado su ausencia (grupo control) (figura 4). Es decir, se parte del desenlace y se reconstruye la información en dirección hacia la probable causa (figura 5); este diseño obliga a reconstruir los hechos en el sentido contrario a como ocurre el fenómeno de causalidad.

Características principales

El diseño de casos y controles tiene limitaciones, en la documentación de la información, que son similares a las del estudio de cohorte histórica (cuadro I) y, en consecuencia, los sesgos son semejantes.

Exposición a la maniobra

Se trata de un estudio observacional en el que solo se mide la exposición a la maniobra. A diferencia de los estudios de cohorte, aquí la maniobra no divide a los



Figura 1 Características que hay que considerar para evitar un ensamble inadecuado y el sesgo de susceptibilidad

sujetos en dos grupos (en la cohorte, en expuestos y no expuestos), sino que la identificación de la exposición parte del hecho de ser caso o control, lo que provoca que al interior de cada uno de estos grupos (casos o controles) se genere un subgrupo de sujetos expuestos y no expuestos (figura 5). Documentar el efecto de la maniobra principal en el estudio de casos y controles —a diferencia de lo que ocurre en el ensayo clínico, en el que las condiciones basales y las comaniobras están controladas y la maniobra principal se asigna en forma aleatoria—, implica registrar todas las posibles variables de confusión presentes en el estado basal (sesgo de susceptibilidad) o cómo intervienen las comaniobras (sesgo de ejecución).

Seguimiento de los sujetos

Algunos autores consideran que el estudio de casos y controles es longitudinal cuando existen registros previos al desenlace, tanto para los casos como para los controles. Sin embargo, es difícil que suceda así, de no ser el registro de vacunas que se lleva en la totalidad de la población o cuando el estudio se desarrolla en una cohorte; en esas situaciones, la evidencia será de mayor calidad, puesto que se contará con la medición de la exposición antes de que aparezca el desenlace.

En la mayoría de las ocasiones, la reconstrucción se realiza por medio del interrogatorio, con lo que el registro de lo sucedido en la exposición y el desenlace

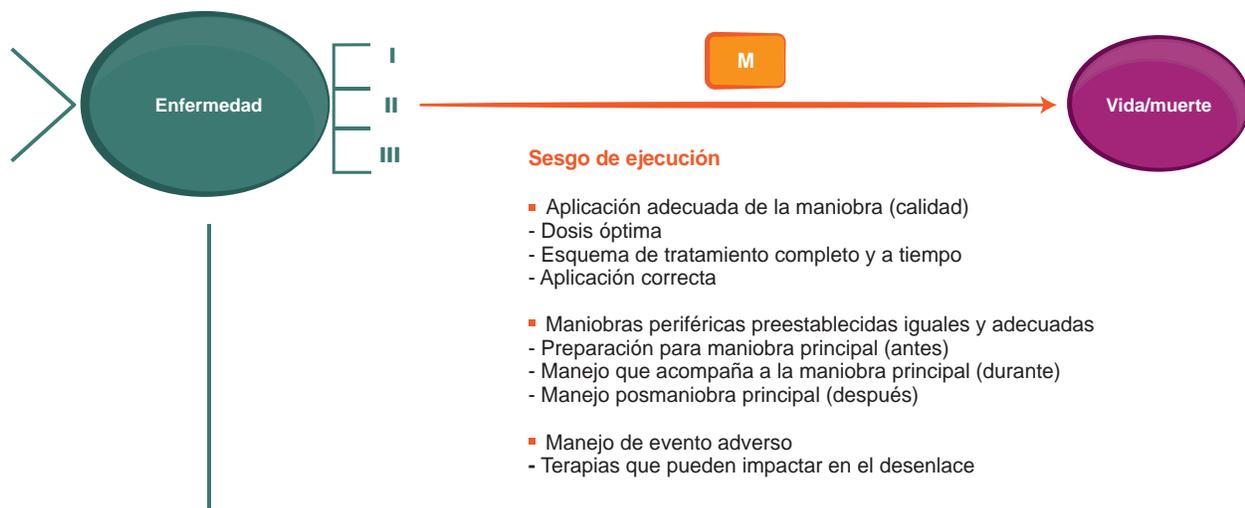


Figura 2 Características que se deben considerar para evitar el sesgo de ejecución

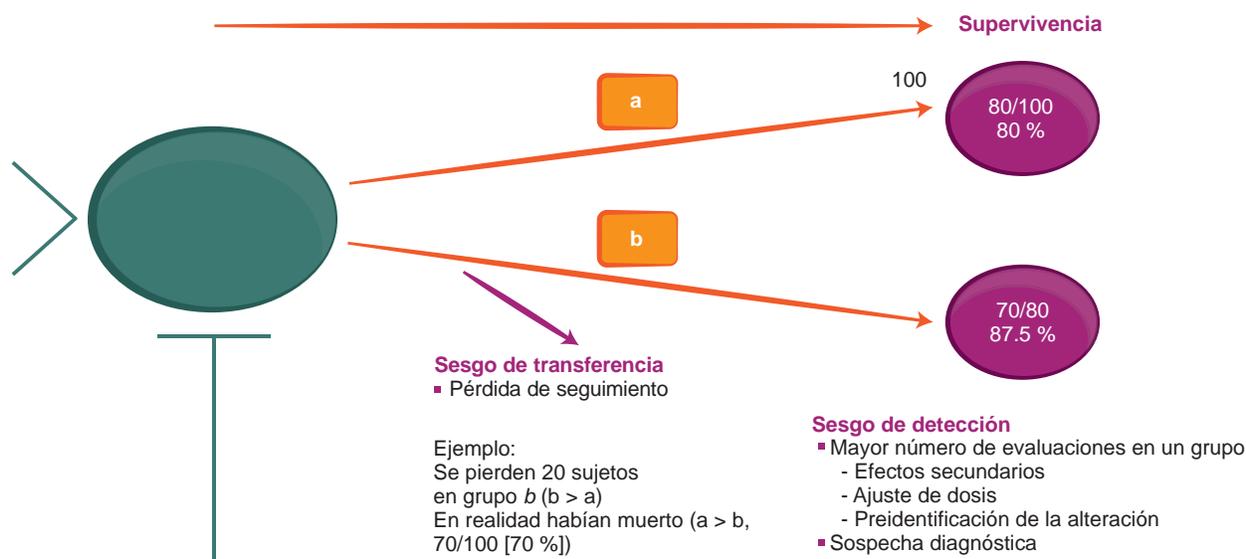


Figura 3 Características que hay que considerar para evitar el sesgo de detección y transferencia

son simultáneos (transversal). Esta forma de obtener la información es común cuando los integrantes del grupo control están relacionados con los casos o cuando aceptan participar en el estudio a través de llamadas telefónicas o por Internet; incluso, puede suceder con controles hospitalarios, aunque en estos últimos ocasionalmente se puede reconstruir la información en forma longitudinal si se dispone de registros previos. Obtener la información en forma transversal puede provocar sesgos por la mala calidad de los datos en todos los componentes del fenómeno de causalidad (estado basal, maniobra, desenlace), debido, comúnmente, al recordatorio diferencial entre los integrantes del grupo de casos y del grupo control.

Direccionalidad en las mediciones

El diseño de casos y controles es retrolectivo (retrospectivo). A diferencia de la cohorte histórica —que también es retrolectiva, pero cuyo ensamble de la población se realiza a partir del estado basal—, el ensamble de la población se realiza a partir del desenlace (sea caso o control). Es decir, en la mejor de las situaciones, la calidad de la información no solo depende de su recolección previa con fines ajenos al objetivo de interés (por ejemplo, la cartilla de vacunación no se diseñó pensando en evaluar posteriormente su asociación con alguna patología y, de igual forma, se omitió una gran cantidad de variables de confu-

Primero se tiene una serie de casos (IAM= infarto agudo al miocardio) y se identifica un grupo control (sin IAM)

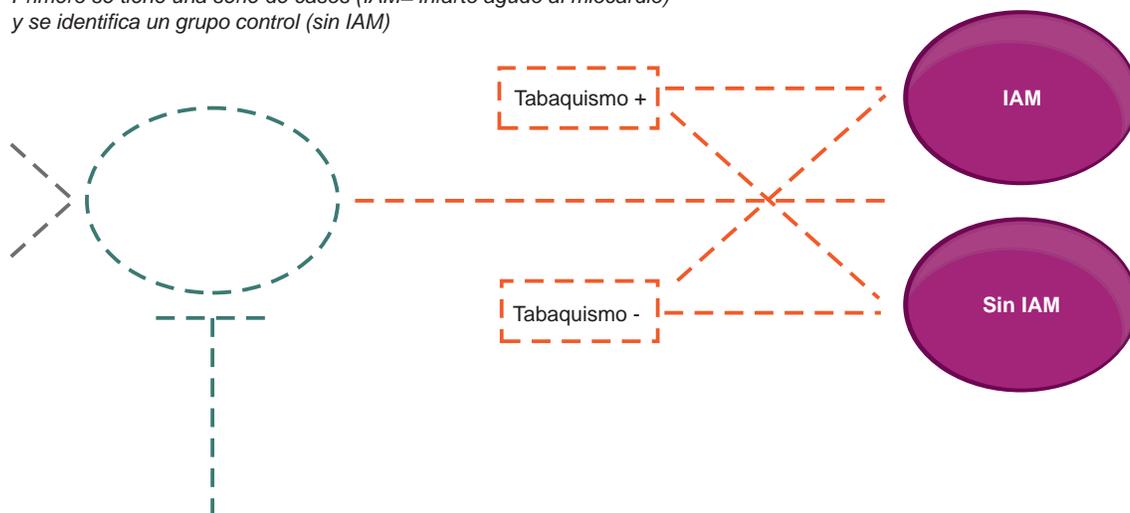
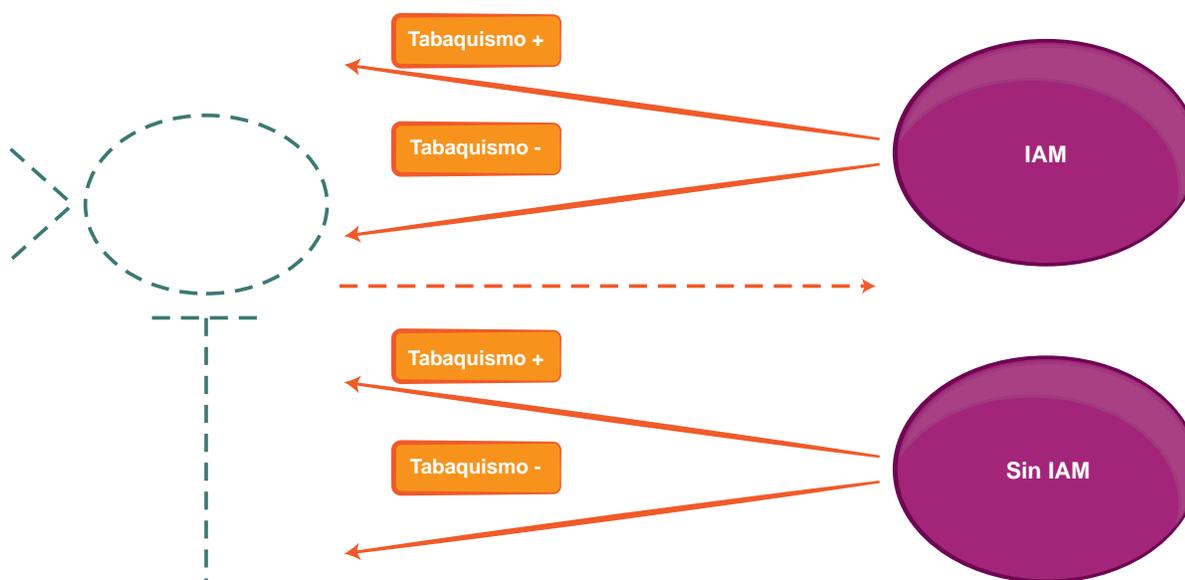


Figura 4 Estudio de casos y controles. Identificación del caso y selección del control



Se documenta la presencia o no de exposición al factor de interés. Se parte del desenlace y se trata de identificar la probable causa

Figura 5 Estudio de casos y controles. Documentación de la exposición

sión), sino también de los sesgos de transferencia al tratarse de una cohorte de supervivientes (en una población definida a partir del estado basal y no del desenlace, cabe la posibilidad de incluir tanto casos vivos y muertos, como controles vivos y muertos).

Búsqueda de asociación

La búsqueda de un grupo control para una serie de casos se lleva a cabo siempre con la intención de establecer asociaciones.

Selección del grupo control

La selección del grupo control constituye el proceso más difícil de este tipo de diseño y puede inducir sesgos en todos los apartados del fenómeno de causalidad, especialmente el sesgo de transferencia.

Habitualmente, los integrantes del grupo de casos son seleccionados entre pacientes que si bien son atendidos en una misma unidad médica, proceden de distintas regiones geográficas. Son pacientes preseleccionados: en principio, buscaron atención médica por

Cuadro I Características principales de los estudios de casos y controles

Diseño	Observacional/ Experimental	Longitudinal/ Transversal	Prolectivo /Retrolectivo/ Retroprolectivo	Comparativo/ Descriptivo	Medición
Cohorte	Observacional	Longitudinal	Prol/Retro/Rp	Comparativo	Incidencia
Casos y controles	Observacional	Long/Trans	Retrolectivo	Comparativo	Proporción caso/ control

El enfoque metodológico considera cuatro características: 1. La imposición o no de la maniobra con fines de investigación: estudio observacional o experimental. 2. Seguimiento (longitudinal) o no (transversal) del paciente a lo largo del tiempo. 3. Direccionalidad en la obtención de la información: prolectiva, retrolectiva y retroprolectiva. 4. La búsqueda o no de asociación de dos o más variables: comparación o descripción.

La medición de la ocurrencia del desenlace se determina mediante la incidencia, la prevalencia o la proporción caso-control

distintas razones; después, tuvieron que ser evaluados por lo menos por un médico antes de llegar al hospital; además, deben aceptar o no participar en el estudio y cubrir una serie de criterios de selección. De tal forma, es difícil determinar de qué población provienen o a quién representan.

Población definida

Si se conoce la población de origen de los casos y esta a su vez resulta claramente definida, la mayor dificultad del diseño del estudio está resuelta. Esto ocurre cuando el estudio de casos y controles es de base poblacional o analiza un grupo anidado en una cohorte. En ambas situaciones se cuenta con el total de la población que dio origen a los casos y, evidentemente, de donde se seleccionarán los controles. Incluso, es posible determinar a qué grupo corresponden las muertes (si existen). Cuando el número de sujetos de la población excede el tamaño que se ha calculado con la muestra, también es posible hacer una selección aleatoria de los casos al igual que de los controles.

Dado que generalmente en los estudios de cohorte se documenta información de la población en análisis —que fue medida antes de que ocurriera el desenlace que se examinará en el estudio de casos y controles—, se evitan errores en la documentación de dicha información. Los *estudios de casos y controles anidados en una cohorte* tienen características adicionales: suelen restringirse al análisis de los elementos de interés que se obtienen durante la evaluación inicial de la cohorte (que correspondería al estado basal del estudio de casos y controles), en lugar de abocarse a los elementos del total de la cohorte. De esta forma se examinan únicamente los sujetos que han desarrollado el desenlace y un grupo control. Lo anterior permite optimizar los recursos y preservar en el resto de los sujetos de la cohorte los elementos que se estudian (muestras sanguíneas, tejidos, etcétera).

Población no definida o de fuente secundaria

Como es común que no se disponga de una población definida, existen diferentes estrategias para obtener sujetos control con probabilidad de pertenecer a la misma población de los casos. Lo más usual es incluir a vecinos o amigos de los casos, a individuos invitados telefónicamente o por Internet (de quienes previamente se ha identificado que provienen de la misma región geográfica que los casos) y, en otras ocasiones, a controles hospitalarios. Cualquiera que sea la situación, suele existir una sub o sobrerrepresentación de la exposición que alterará los resultados.

Reconstrucción fenomenológica de los hechos

Los hechos deben ser reconstruidos de acuerdo con el fenómeno de causalidad, sin importar las limitantes propias de cómo se ensambla la población (del desenlace a la exposición) y cómo se recolectan los datos (retroactiva y transversalmente). Para ello existe una serie de recomendaciones:

- Establecer claramente los criterios para integrar la población que se va a estudiar, aplicables tanto para los casos como para los controles (figura 1). Es necesario trasladar el interrogatorio o la búsqueda de la información del expediente al periodo que para cada caso o control correspondería al estado basal, y habrá que procurar lo siguiente para el total de la población:
 - a) Limitar, en lo posible, el universo de trabajo solo a los sujetos que pertenezcan a una misma región.
 - b) Definir los criterios diagnósticos, es decir, la población que se desea analizar.
 - c) Definir los criterios de selección, es decir, las características que deben reunir los sujetos en quienes no se ha manifestado el desenlace o, si el interés es evaluar su progresión más que su aparición, en quienes todavía es incipiente. Aunque resulta obvio, debe cuidarse que esos criterios no incluyan a sujetos con indicación o contraindicación de la maniobra, pero sí a aquellos en los que sea posible que ocurra el desenlace. Es importante recordar que el estado basal, aun el grupo de casos, debe estar libre del desenlace. De hecho, los criterios son iguales para ambos.
- Documentar todas las variables en el estado basal que pueden modificar el efecto de la maniobra sobre el desenlace, o que independientemente de la maniobra participen en la aparición del desenlace (figura 1).
- Definir claramente la exposición y, de ser posible, graduarla en magnitud y tiempo, así como todas las posibles comaniobras (figura 2).
- Especificar los criterios que definen el caso y el control.
- Procurar seleccionar casos diagnosticados recientemente, para evitar que la exposición a la maniobra haya sido modificada después del diagnóstico.
- Determinar cuáles serán las fuentes de documentación para obtener los datos de los casos. Estas deberán ser las mismas que en los controles.
- Homologar la manera de reconstruir la información en los casos y los controles, ya sea que se base en

datos previamente obtenidos o en un interrogatorio. Resultaría erróneo conseguir la información de los casos en el expediente y la de los controles mediante un interrogatorio.

- Asignar a personas diferentes las tareas de reconstrucción de los hechos. Resulta ideal que quien obtiene la información del estado basal sea ajeno a quien documenta la exposición a la maniobra y ambos, a su vez, diferentes de quien documenta el desenlace.
- Obtener la información en el orden en que se presenta el fenómeno de causalidad (estado basal, maniobra y desenlace).

Comentarios

Sin duda, además de los errores mencionados, la reconstrucción de los hechos a partir del desenlace conlleva sesgos de transferencia, dado que en los casos y en los controles por lo general solo se evalúan supervivientes.

Es recomendable evitar el diseño de casos y controles como estrategia para documentar el fenómeno de causalidad cuando la respuesta se puede obtener por medio de

un ensayo clínico o una cohorte. Lo que tiene en común con los demás diseños de investigación es que solo se trata de una herramienta para documentar el fenómeno de causalidad; por lo tanto, la sugerencia principal es mantener siempre la estructura mental del juicio clínico, con la que se conceptualizan tres consabidos elementos: un estado basal en el que las características propias de un grupo de sujetos llevan a distribuirlos en subgrupos de acuerdo con su probabilidad de sufrir el desenlace aun antes de la exposición a cualquier maniobra (demarkación pronóstica); una maniobra principal con características propias, acompañada de una serie de acciones alrededor de ella (comaniobras); y la medición del cambio en la condición basal o la aparición de nuevas características, que conocemos como desenlace.

Esa estructura fenomenológica habitual para los clínicos —juicio clínico/arquitectura de la investigación— es universal y no se ve modificada por la forma como se obtiene la información, ya sea en un ensayo clínico o un estudio observacional. Al realizar la evaluación estructurada de un artículo o al intentar responder un cuestionamiento mediante un estudio de investigación, siempre debe pensarse en el fenómeno de causalidad desde el punto de vista clínico.

Lecturas recomendadas

- Feinstein AR. *Clinical biostatistics*. Saint Louis: Mosby; 1977.
- Feinstein AR. *Clinical epidemiology. The architecture of clinical research*. Philadelphia: WB Sanders; 1985.
- Feinstein AR. Directionality and scientific inference. *J Clin Epidemiol*. 1989;42:829-33.
- Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. Third edition. Baltimore: Williams & Wilkins; 2008.
- Portney LG, Watkins MP. *Foundations of clinical research: applications to practice*. Third edition. New Jersey: Pearson-Prentice Hall; 2009.
- Talavera JO, Wachter-Rodarte NH, Rivas-Ruiz R. Clinical research III. The causality studies. *Rev Med Inst Mex Seguro Soc*. 2011;49(3):289-94.
- Talavera JO. Clinical research I. The importance of research design. *Rev Med Inst Mex Seguro Soc*. 2011;49 (1):53-8.
- Cruz-Anguiano V, Talavera JO, Vázquez L, Antonio A, Castellanos A, Lezana MA, *et al.* The importance of quality of care in perinatal mortality: a case-control study in Chiapas, Mexico. *Arch Med Res*. 2004;35(6):554-62.

XII. Del juicio clínico a la encuesta transversal

Juan O. Talavera, Rodolfo Rivas-Ruiz

Un estudio longitudinal, trátase de un ensayo clínico o de cohorte, tiene la virtud de seguir la secuencia lógica en la que ocurren los componentes del fenómeno de causalidad. Sin embargo, en un estudio transversal no existe esta secuencia lógica y coincide la medición de los tres componentes (estado basal, maniobra y resultado). El “juicio clínico” nos ayuda a reconstruir artificialmente dichos componentes en la secuencia temporal en que ocurrieron. Sin embargo, la forma como se ensambla la población y como se obtiene la información —transversal y retrospectivamente— conlleva el riesgo de caer en sesgos. Si bien el uso de la encuesta transversal para asociar una maniobra con un probable desenlace resulta difícil y frecuentemente genera errores (sobre todo cuando se estudian fenómenos patológicos), es extraordinaria para mostrar el desarrollo del sujeto sano simulando un estudio longitudinal, como ocurre con la estimación de la talla y el peso de acuerdo con la edad; a este tipo de diseño se le ha denominado “longitudinal cross sectional study”.

Palabras clave

estudios transversales
estudios de cohorte
estudios de casos y controles
ensayo clínico

Este artículo fue publicado originalmente en Rev Med Inst Mex Seguro Soc 2012; 50 (6): 641-644 y ha sido revisado para la presente edición.

Un estudio longitudinal, trátase de un ensayo clínico o de cohorte, tiene la virtud de seguir la secuencia lógica en que ocurre un fenómeno (en un estado basal se observa el efecto de una maniobra para generar un desenlace). En contraste, en un estudio transversal esta secuencia lógica no existe, ya que al momento de la medición coinciden los tres componentes: estado basal, maniobra y resultado.

El diseño arquitectónico (juicio clínico) nos ayuda a reconstruir artificialmente los componentes en la secuencia temporal en que ocurrieron. De esta forma, en los diseños transversales incluso podemos hacer evaluaciones de causalidad, a sabiendas de las limitaciones y los riesgos (figuras 1 a 3). Entre los diseños transversales se incluyen el estudio de casos y controles y la encuesta transversal.

La encuesta transversal es probablemente el diseño más utilizado en la investigación médica. En general, con excepción del análisis de maniobras terapéuticas (en el que se utiliza generalmente el diseño de ensayo clínico), en la mayoría de los estudios de causalidad se emplea la encuesta transversal y en solo algunas ocasiones el diseño de cohorte, que resulta complejo y costoso por la amplia población que debe seguirse durante periodos prolongados.

La encuesta transversal se caracteriza por estudiar a una población específica o a una muestra de dicha población, en la cual los datos son recolectados a un mismo tiempo. Es decir, la información del estado basal, la maniobra y el desenlace se obtiene en forma retrospectiva; al comenzar el análisis ya ocurrió el desenlace y la exposición a la maniobra. De tal forma, no existe la posibilidad de observar las condiciones basales del sujeto de estudio y su cambio a lo largo del tiempo. Sin embargo, conforme la secuencia lógica del fenómeno de causalidad, se asume que el desenlace no existía antes de aplicar la maniobra. Así, también se puede reconstruir la intensidad y el tiempo de exposición a la maniobra para establecer la magnitud de su asociación con el desenlace. No obstante que todos los componentes del fenómeno de causalidad se miden en un tiempo, la reconstrucción de los hechos deberá hacerse siguiendo la secuencia temporal lógica (figura 4).

Exposición a la maniobra

En la encuesta transversal solo se mide la exposición a la maniobra, a diferencia del ensayo clínico, en el que el investigador asigna la maniobra. Y a diferencia de la cohorte histórica, en la que la exposición a la maniobra ha sido medida, aun cuando con fines ajenos a la investigación, en la encuesta transversal, al igual que en el estudio de casos y controles, la calidad en la medición de la maniobra es baja. En la precisión de los datos



Figura 1 Características que hay que considerar en el estado basal, para evitar un ensamble inadecuado y el sesgo de susceptibilidad

influye el estado del paciente al momento de la medición (con lo que se resiente el efecto o el conocimiento del desenlace) y la lejanía de esta con los componentes del fenómeno de causalidad (a mayor tiempo de transcurrida la exposición a la maniobra, menos precisa es la información). Lo mismo sucede con la medición de las variables que pueden confundir el efecto de la maniobra —condiciones previas a la maniobra (estado basal) y condiciones que acompañaron en su momento a la maniobra (maniobra periférica)— (figura 2).

Seguimiento del sujeto

Cuando la observación de los componentes del fenómeno de causalidad concuerda con su secuencia

temporal (estado basal, maniobra y desenlace), permite prever y evitar una serie de errores; sin embargo, esto solo ocurre en el ensayo clínico y en el diseño de cohorte. En la encuesta transversal, la evaluación de todos los componentes es simultánea —lo que la caracteriza como un estudio transversal—, y la secuencia temporal se reconstruye artificialmente, pero con el riesgo de anteponer la maniobra al desenlace o de medir una supuesta maniobra que en realidad es consecuencia del desenlace o una característica que acompaña al desenlace (en un paciente diabético, por ejemplo, atribuir la hipertrigliceridemia al descontrol glucémico, cuando los dos pueden ser consecuencia de otro factor).

Si bien asociar un desenlace a una probable causa resulta difícil y frecuentemente se generan errores, el diseño de encuesta transversal es extraordinario para

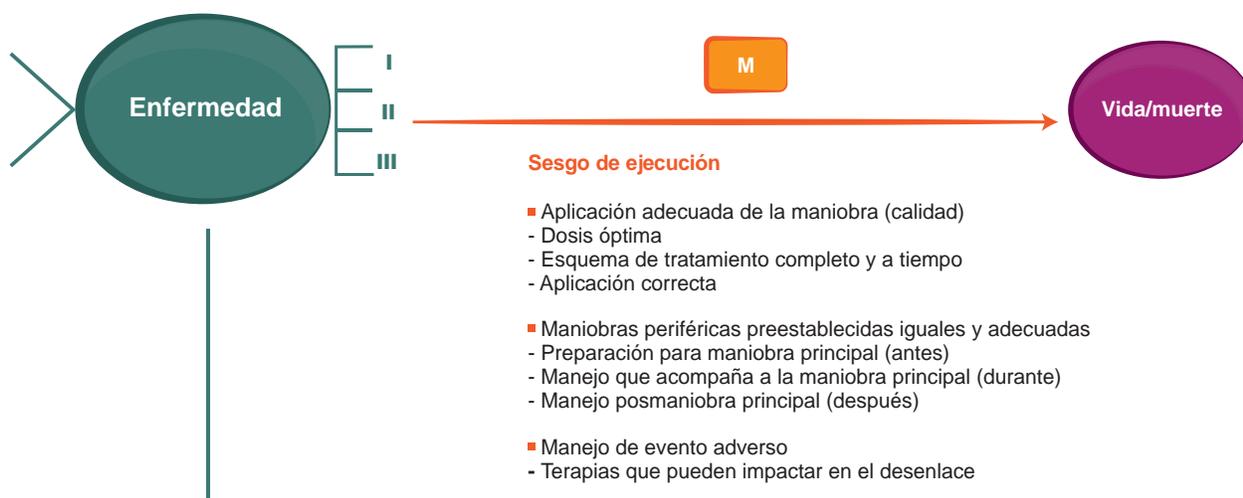


Figura 2 Características que hay que considerar durante la maniobra para evitar el sesgo de ejecución

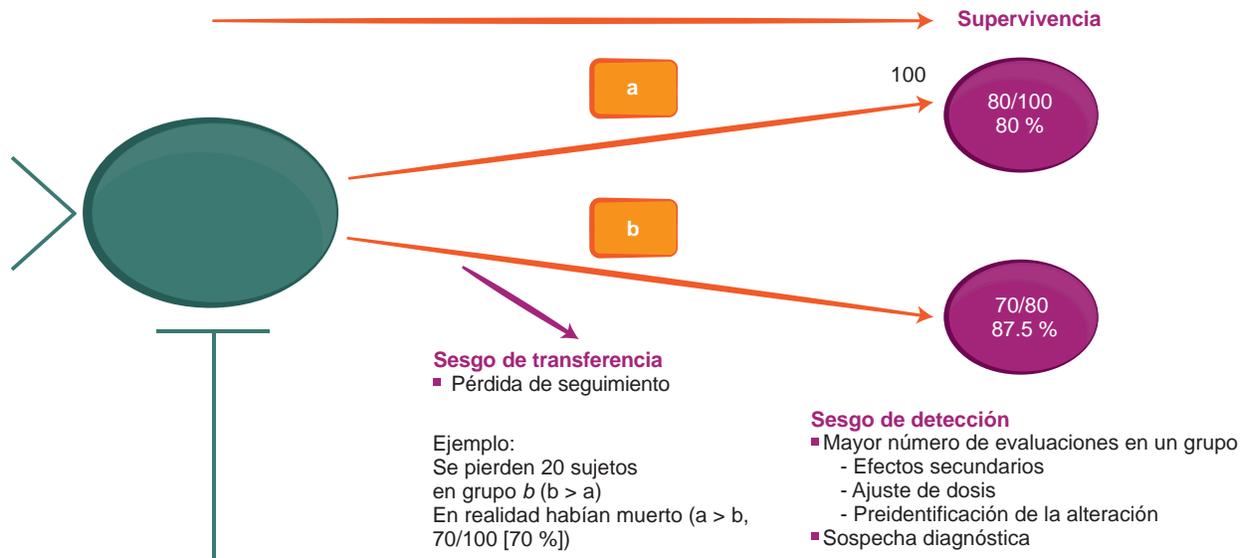


Figura 3 Características que hay que considerar durante la medición del desenlace para evitar el sesgo de detección y transferencia

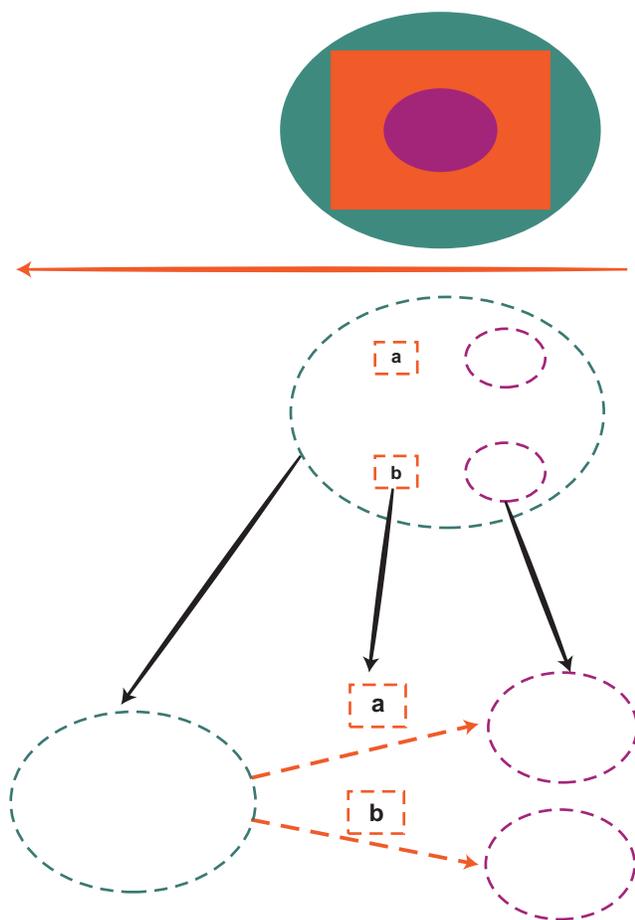


Figura 4 Reconstrucción artificial del fenómeno de causalidad en la encuesta transversal

conocer el desarrollo del sujeto sano. Un ejemplo son las tablas de talla y peso de los niños de acuerdo con la edad y el sexo. Estas tablas fueron realizadas con mediciones transversales de niños de cada sexo con distintas edades; posteriormente, se simuló una cohorte en la que el niño o la niña cambiaba de talla y peso de acuerdo con los años de vida. Este diseño recibe el nombre de *longitudinal cross sectional study* y es adecuado para mostrar el desarrollo del sujeto sano, pero no permite conocer la historia natural o el curso clínico de una enfermedad, ya que en el camino se pierden los sujetos más enfermos y en las evaluaciones posteriores quedan solo los sobrevivientes, por lo que se obtienen resultados falsos de la evolución de la enfermedad. Sin embargo, este diseño puede ser útil en enfermedades con baja mortalidad, siempre y cuando se controle el posible efecto del desenlace sobre la medición de las características que le anteceden.

Direccionalidad en las mediciones

La medición en un solo tiempo de todos los componentes del fenómeno de causalidad se ve influida por el hecho de que la exposición de la maniobra ha ocurrido previamente sobre ciertas condiciones basales, al igual que el desenlace; es decir, la direccionalidad de las mediciones convierten a la encuesta transversal en un estudio retrolectivo (retrospectivo). A diferencia de la cohorte histórica (o cohorte retrolectiva —cuya direccionalidad de las mediciones también la convierte en retrolectiva—), en la que el registro de los hechos fue realizado en forma secuencial conforme fueron ocurriendo longitudinalmente, aunque por razones ajenas

a la investigación, la reconstrucción de los hechos en la encuesta transversal se realiza en un mismo tiempo, de tal forma que la temporalidad y la magnitud de la exposición a la maniobra y comaniobras, así como las condiciones basales —aquellas que antecedían a la maniobra— dependerán, en la mayoría de las ocasiones, de la memoria del sujeto en estudio, lo que afecta la precisión de los datos y las atribuciones de causalidad debido a posibles sesgos en el estado basal, la maniobra y el desenlace (figura 4).

Búsqueda de asociación

Buscar causalidad siempre implicará comparar sin importar el diseño. De igual forma, la encuesta transversal implica la comparación del efecto de la maniobra de interés sobre el estado basal, contra la ausencia de ella o contra el efecto de otras maniobras.

Reconstrucción fenomenológica de hechos

Por tratarse de un estudio transversal y retrolectivo, se prevén recomendaciones con el fin de reconstruir los hechos lo más cercano posible al fenómeno de causalidad.

El proceso de reunir la información se deberá empezar siempre con lo que correspondería al estado basal, específicamente con los criterios de selección, que deberán ser los mismos para toda la población en estudio. De igual forma, del estado basal deberán docu-

mentarse todas las características que pudiesen influir en el desenlace, independientemente de la maniobra o por interacción con esta última.

Deberán definirse dentro de lo posible las características de la maniobra y de las comaniobras, así como las del desenlace.

Es necesario procurar que entre los sujetos en quienes haya acontecido el desenlace de interés solo se incluyan los de reciente diagnóstico, de tal forma que en este se pueda evaluar el efecto de la maniobra principal y disminuir la probabilidad de que el desenlace modifique lo que pudo ser la maniobra.

Es indispensable cuidar que la estructura de obtención de la información siempre sea la misma y no se favorezca una tendencia, para no sesgar las respuestas de los sujetos.

Por último, la obtención de la información deberá ser segmentada, iniciando con las condiciones basales, continuando con la maniobra y finalizando con el desenlace.

Comentarios

Aun cuando los diseños transversales (casos y controles y encuesta transversal) resultan incómodos, gran parte de la investigación utilizada para resolver las molestias de los pacientes proviene de estudios con estos diseños. Si bien la estructura real del fenómeno de causalidad y la reconstrucción de sus componentes en la encuesta transversal son artificiales, resultan lógicas y necesarias al usar el juicio clínico.

Lecturas recomendadas

- Feinstein AR. Clinical biostatistics. Saint Louis: The CU Mosby Co; 1977.
- Feinstein AR. Clinical epidemiology. The architecture of clinical research. Philadelphia: WB Sanders; 1985.
- Feinstein AR. Directionality and scientific inference. *J Clin Epidemiol.* 1989;42:829-33.
- Rothman KJ, Greenland S, Lash TL. Modern epidemiology. Third edition. Baltimore: Williams & Wilkins; 2008.
- Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Third edition. New Jersey: Pearson-Prentice Hall; 2009.
- Talavera JO, Wachter-Rodarte NH, Rivas-Ruiz R. Clinical research III. The causality studies. *Rev Med Inst Mex Seguro Soc.* 2011;49(3):289-94.
- Talavera JO. Clinical research I. The importance of research design. *Rev Med Inst Mex Seguro Soc.* 2011;49(1): 53-8.

XIII. El diseño de investigación en la revisión estructurada de un artículo

Juan O. Talavera, Rodolfo Rivas-Ruiz

La calidad de la información obtenida de acuerdo con el diseño de investigación se integra a la revisión estructurada conforme el modelo de causalidad. Para la ejemplificación, se utiliza el artículo *Reduction in the incidence of poststroke nosocomial pneumonia by using the 'Turn-mob' Program*, cuyo diseño corresponde al ensayo clínico. Entre los aspectos que hay que identificar y analizar están los éticos, con los que se busca resguardar la seguridad y el respeto al paciente; la asignación aleatoria, mediante la que se busca generar grupos con condiciones basales homogéneas, integrados por sujetos con la misma probabilidad de recibir cualesquiera de las maniobras en comparación y con la misma probabilidad premaniobra de apego a ellas y la misma probabilidad de abandono del estudio por causas ajenas a la maniobra. Otros aspectos son la relatividad de la comparación, el cegamiento de la maniobra, la aplicación en paralelo de la maniobra comparativa, la detención anticipada y el análisis de acuerdo con el grado de adherencia. El análisis conforme al diseño de investigación resulta complementario al realizado a partir del modelo arquitectónico de causalidad y de las consideraciones estadísticas y de relevancia clínica.

Palabras clave

diseño de investigación
ensayo clínico
causalidad
sesgo

Este artículo fue publicado originalmente en *Rev Med Inst Mex Seguro Soc* 2012; 51(1): 68-72 y ha sido revisado para la presente edición.

El presente texto integra la revisión estructurada de un artículo (figuras 1 a 3 de la parte VIII de esta serie), las características del diseño de investigación y la consecuente calidad de la información que se obtiene (partes IX y XII también de esta serie).

Nuevamente utilizaremos el artículo "*Reduction in the incidence of poststroke nosocomial pneumonia by using the 'Turn-mob' Program*" (publicado en *J Stroke Cerebrovasc Dis.* 2010;19:23-8), cuyo objetivo fue demostrar la eficacia de un programa de movilización en cama para disminuir la incidencia de neumonía nosocomial en pacientes con accidente vascular cerebral isquémico. El diseño de investigación utilizado fue el ensayo clínico; por lo tanto, analizaremos sus características (figura 4) y las integraremos al ejemplo desde el abordaje arquitectónico de causalidad descrito por el doctor Alvan R. Feinstein.

Características del diseño. Ensayo clínico

Aspecto ético

Si bien el primer aspecto que debe analizarse en cualquier texto es el ético, dada su extensión y su distinta naturaleza será discutido en otro artículo.

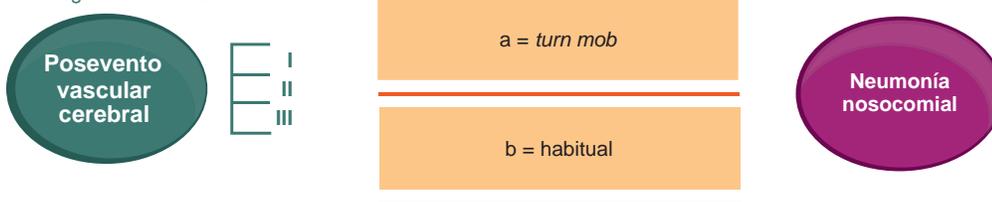
Asignación aleatoria

Un apartado que define al ensayo clínico es la *asignación aleatoria*. Con ella se busca generar grupos con condiciones basales homogéneas para evitar el sesgo de susceptibilidad; integrar en los grupos a sujetos con la misma probabilidad de recibir cualesquiera de las maniobras en comparación, y con la misma probabilidad premaniobra de apego a ellas, para evitar el sesgo de ejecución; facilitar el cegamiento en la evaluación del desenlace, y reducir, en consecuencia, el sesgo de detección diagnóstica. La asignación aleatoria también distribuye a los sujetos entre los grupos con la misma probabilidad de abandonar el estudio por causas ajenas a las maniobras, con lo que se reduce el sesgo de transferencia.

Por su parte, el Programa *Turn-mob* contó con aleatorización y logró grupos balanceados en el estado basal, con excepción de la enfermedad pulmonar obstructiva crónica, que pudo haber favorecido la maniobra experimental. Gracias a la asignación aleatoria, se generaron grupos con la misma posibilidad de apego a la maniobra, aunque en ese estudio nunca fue verificado el apego a la maniobra estándar, con lo cual es posible que esta fuese la ausencia de toda movilidad del paciente. En cuanto a la evaluación del desenlace, no se especifica si la realizó un segundo evaluador que

Forma de selección de la población

Paciente con déficit neurológico agudo, duración: más de 12 horas en Urgencias o Medicina interna



Demarcación diagnóstica

- Más de 48 horas de evolución
- Sin requerimiento de apoyo ventilatorio
- Primer evento vascular
- Sin evidencia clínica de IVR superiores e inferiores
- Sin agitación psicomotora
- Diagnóstico tomográfico de EVC isquémico
- Se excluyeron aquellos que presentaron IVR en las primeras 48 horas

Estratificación pronóstica: grupo a frente a grupo b

<i>Cronométrica</i>	Edad 72 y 74 años
<i>Estatus IMC</i>	Normal 18 frente a 17 %; sobrepeso 69.4 frente a 70.5 %; Obesidad 12.6 frente a 12.5 %
<i>Clínica</i>	Déficit motor, hemiparesia 66.7 frente a 75.9 % Hemiplejía 33.3 frente a 24.1 %; afasia 50.5 frente a 40.2 % Déficit sensitivo: 56.8 frente a 40.2; reflejo nauseoso 82 frente a 79.5 Glasgow puntuación 15, 40.5 frente a 32.1 % NIHSS puntuación 2-7, 30.6 frente a 32.1 % 8-13, 41.4 frente a 43.8 % 14-18, 16.2 frente a 17.9 % 19-23, 11.7 frente a 6.3 %
<i>Morfológica</i>	Subtipo de enfermedad cerebrovascular Infarto parcial de circulación anterior 88.3 frente a 90.2 %
<i>Comorbilidad</i>	DM 50.5 frente a 42 %; HTA 83 frente a 84 %; EPOC 7 frente a 14 %; ECV 39 frente a 40 %
<i>Tratamiento previo</i>	Corticosteroides, antibiótico
<i>Socioeconómicos, culturales y hábitos</i> = tabaquismo 31 frente a 35 % y alcoholismo 24 frente a 24 %	

Figura 1 Características que se deben considerar en el estado basal: demarcación diagnóstica (universo de trabajo, definición de accidente vascular cerebral, criterios de selección), y estratificación pronóstica (variables que impactan en el desenlace independientemente de la maniobra). En el Programa Turn-mob, si bien la aleatorización logró balancear las características de los grupos, con

excepción de la enfermedad pulmonar obstructiva crónica (EPOC) —discretamente mayor en el grupo b (14 frente a 7 %, $p = 0.088$) y que puede impactar en el resultado final—, no puede observarse el efecto de cada una de las maniobras dependiendo de los distintos factores de riesgo, con lo que el resultado observado se debe atribuir principalmente a las características promedio de la población

desconociera el grupo de pertenencia del paciente. Finalmente, se observa que no existieron pérdidas que pudiesen haber provocado sesgo de transferencia.

Relatividad de la comparación

Si bien el Programa de *Turn-mob* se planeó como un estudio de efectividad al comparar la nueva maniobra contra la estándar, pudo haber derivado en un análisis de eficacia al existir la posibilidad de que la maniobra comparativa consistiese precisamente en no aplicar ninguna acción.

Cegamiento

En el Programa de *Turn-mob* resultó imposible el cegamiento de la maniobra y si bien se pudo haber promovido un segundo evaluador del desenlace, esto no se refiere. Por lo tanto, existió la probabilidad de sesgo de detección diagnóstica.

Maniobra comparativa en paralelo

El requisito de realizar una maniobra comparativa en paralelo (durante el mismo tiempo calendario) se cubrió y se logró al evitar que existieran diferencias en la demarcación diagnóstica o de estratificación (para evitar sesgos de ensamble inadecuado y de susceptibilidad pronóstica), diferencias en la accesibilidad a maniobras periféricas (para evitar sesgo de ejecución) y diferencias en criterios de diagnóstico del desenlace (con lo cual se reduce la posibilidad de sesgo de detección).

Detención anticipada

No hubo la presencia de eventos adversos a las maniobras. Tampoco hubo diferencias anticipadas en el desenlace. De haberse presentado eventos o diferencias, hubieran podido detener el Programa *Turn-mob*.

Forma de selección de la población

Maniobras periféricas

- Intubación 7.2 frente a 8 %
- Alimentación enteral 19.8 frente a 21.4 %
- Catéter intravascular 3.6 frente a 6.3 %

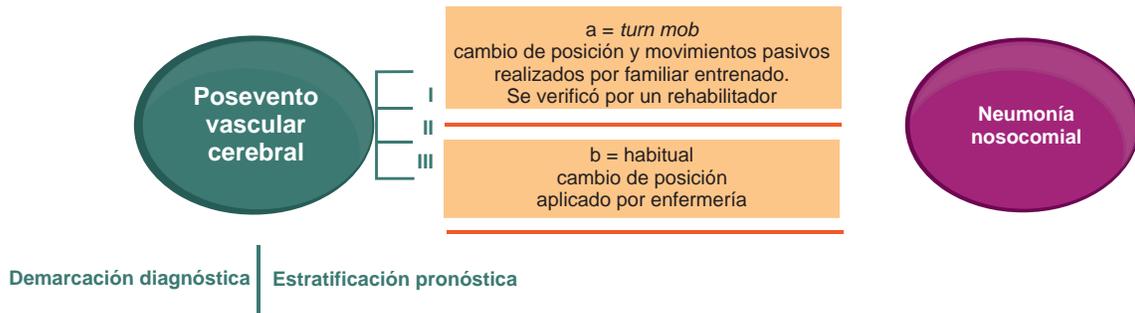


Figura 2 Características que se deben considerar durante la aplicación de la maniobra: la calidad de la aplicación de la maniobra principal (Programa *Turn-mob* comparado con cambios de posición habitual) y verificar que se apliquen las maniobras periféricas en forma similar en ambos grupos. Si bien no existió diferencia en las maniobras periféricas, la aplicación del Programa *Turn-mob* fue estandarizada inicialmente y verificada día

a día. En cambio, el tratamiento habitual nunca fue estandarizado ni se verificó su aplicación día a día, por lo cual no hay garantía de que se haya llevado a cabo; más aún, al egresar el paciente a su domicilio ya no existió el apoyo de enfermería. Esto pudo representar más que una superioridad del Programa *Turn-mob* sobre el tratamiento habitual, el resultado de la aplicación del programa contra ninguna acción

Análisis de acuerdo con la adherencia

El último aspecto consiste en el análisis de acuerdo con la adherencia, que deja en claro que en el Programa de *Turn-mob* se realizó por intención al tratamiento, dado que se evaluaron todos los pacientes en cada uno de los grupos a los que fueron asignados, sin importar que en el grupo de la maniobra estándar la hubiesen o no recibido, como pudo haber sido el caso con el consecuente sesgo de ejecución.

Comentarios finales

Como podemos observar, el análisis de un artículo o trabajo de investigación de acuerdo con el diseño utilizado es complementario al análisis que se hace desde el modelo arquitectónico de causalidad; por otro lado, habrá que tomar en cuenta las consideraciones estadísticas y de relevancia clínica. Sin duda alguna, para efectuar un análisis estructurado se necesita tiempo y conocimiento, y sin duda alguna

Forma de selección de la población

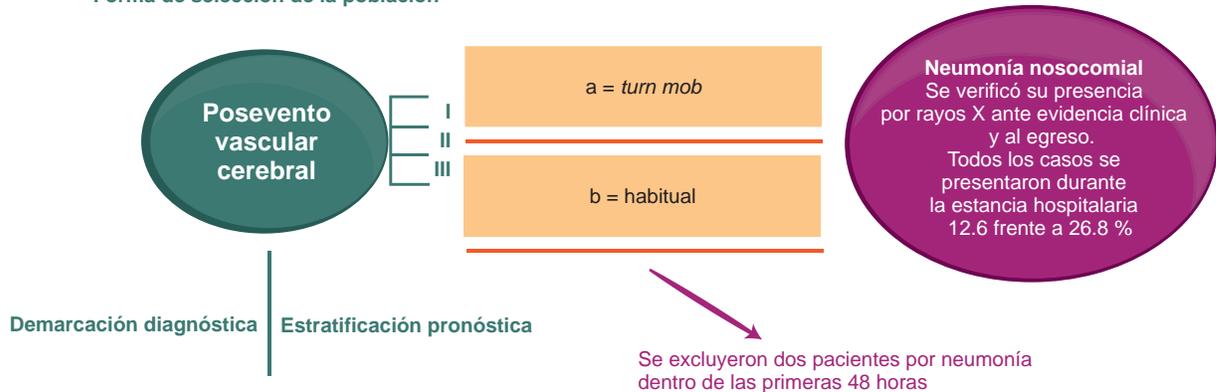


Figura 3 Características que se deben considerar en el resultado: no existe la posibilidad de haber detectado en forma diferencial la neumonía nosocomial, dado que a todos los pacientes se les tomaron rayos X de tórax al egreso o

ante la mínima sospecha clínica. De igual forma, no existe problema por pérdida de pacientes; solo se excluyeron dos casos de un total de 225 y por la presencia de neumonía en las primeras 48 horas del ingreso hospitalario

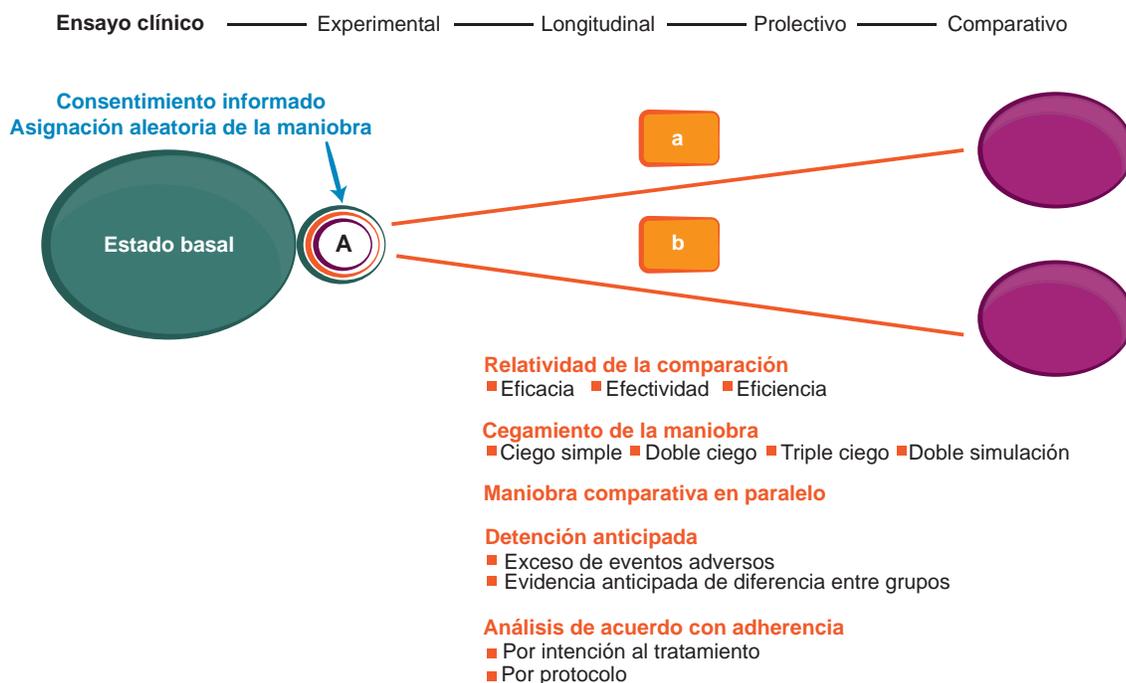


Figura 4 Características del ensayo clínico en paralelo al razonamiento clínico

resulta más enriquecedor que solo aceptar un juicio de calidad ajeno y superficial, como se pretende en la clasificación por nivel de evidencia. Recordemos, por otro lado, que si bien cada artículo intenta contestar

específicamente una pregunta, resulta que contiene una gran cantidad de información útil para el clínico, como los aspectos epidemiológicos y clínicos de la patología en estudio.

Referencias

- Talavera JO. Clinical research I. The importance of the research design. *Rev Med Inst Mex Seguro Soc.* 2011;49(1):53-8.
- Talavera JO, Wachter-Rodarte NH, Rivas-Ruiz R. Clinical research III. The causality studies. *Rev Med Inst Mex Seguro Soc.* 2011;49(3):289-94.
- Feinstein AR. *Clinical epidemiology. The architecture of clinical research.* Philadelphia: WB Saunders; 1985.
- Feinstein AR. *Clinical biostatistics.* Washington: C.V.Mosby; 1977.
- Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology.* Third ed. Baltimore: Williams & Wilkins; 2008.
- Portney LG, Watkins MP. *Foundations of clinical research: applications to practice.* Third edition. Pearson/Prentice Hall; 2009.
- Talavera JO, Rivas-Ruiz R. Clinical research VIII. Structured review of an article. *Rev Med Inst Mex Seguro Soc.* 2012;50(2):163-6.
- Talavera JO, Rivas-Ruiz R. Clinical research IX. From the clinical judgment to the clinical trial. *Rev Med Inst Mex Seguro Soc.* 2012;50(3):267-72.

Autores



Juan O. Talavera

Médico internista dedicado a la docencia y a la investigación en el área clínica. Nació en la Ciudad de México en 1965. Desde 2010, es parte del equipo del Centro de Adiestramiento en Investigación Clínica (CAIC) de la Coordinación de Investigación en Salud del Instituto Mexicano del Seguro Social (IMSS), en el Centro Médico Nacional Siglo XXI (CMN SXXI).

jotalaverap@uaemex.mx



Rodolfo Rivas Ruiz

Médico pediatra neonatólogo del Hospital de Pediatría del CMN SXXI, IMSS. Maestro en Investigación por la Universidad Autónoma del Estado de México (UAEM) y candidato a Doctorado en Epidemiología Clínica por la Universidad Nacional Autónoma de México (UNAM). Miembro del Sistema Nacional de Investigadores (SNI) y profesor de la Facultad de Medicina de la UNAM, y de la Maestría en Investigación Clínica IMSS-Instituto Politécnico Nacional (IPN). Es parte del equipo del CAIC.

rivasrodolfo@gmail.com



Niels H. Wachter Rodarte

Especialista en Medicina interna (UNAM), maestro en Ciencias médicas (UNAM). Jefe de la Unidad de Investigación en Epidemiología Clínica, Hospital de Especialidades, CMN SXXI. Profesor de Epidemiología clínica y Medicina basada en evidencias (UNAM), de los programas de maestría y doctorado en Ciencias Médicas y en Ciencias de la Salud (Epidemiología clínica), y del curso Clinimetría (Maestría en Ciencias de la Salud, UNAM). Es miembro del SNI.

nwacher@hotmail.com



Carla Martínez Castuera Gómez

Licenciada en Relaciones internacionales y maestra en Estudios de América del Norte por la Universidad de las Américas-Puebla. Jefa del Área de Vinculación Internacional de la Coordinación de Investigación en Salud, IMSS. Tiene experiencia en el sector público. Es responsable de los programas de apoyo y la gestión para la cooperación internacional. Es colaboradora del CAIC.

cmartinez.imss@gmail.com



Marcela Pérez Rodríguez

Nutrióloga de la Escuela de Dietética y Nutrición del Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado (ISSSTE). Maestra en Epidemiología Clínica y estudiante del Doctorado en Epidemiología Clínica (ambos de la UNAM). Es profesora del programa de Maestría en Investigación clínica IMSS-IPN de la asignatura de Epidemiología clásica. Perteneció al equipo del CAIC.

marxelapr@gmail.com



Lino Palacios Cruz

Médico psiquiatra general y de la adolescencia del Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz (INPRFM)-UNAM. Maestro en Ciencias Médicas (INPRF-UNAM). Actualmente, cursa el Doctorado en Ciencias de la Salud-Epidemiología Clínica (UNAM). Es profesor de varios cursos en el área de Psiquiatría (INPRF-UNAM), del seminario de investigación de la Maestría en Epidemiología Clínica (UNAM) y de la Maestría en Investigación clínica (IMSS-IPN). Es miembro del SNI y colabora con el CAIC.

palacioslino@gmail.com



Laura Paola Bernal Rosales

Licenciada en Nutrición por la Universidad Iberoamericana. Maestra y doctora en Ciencias de la Salud por la UNAM. Ha trabajado en el Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán (INCMNSZ), en el CMN SXXI, IMSS, así como en diversas instituciones privadas. Actualmente practica la consulta privada y es directora de la Licenciatura en Nutrición de la Universidad Tecnológica de México (UNITEC). Es colaboradora del CAIC.

lau.pao.br@gmail.com