

De la opinión a la evidencia en ortopedia. Parte III: medicina basada en la evidencia.

MIGUEL MARÍA SÁNCHEZ MARTÍN

HOSPITAL CLÍNICO UNIVERSITARIO. FACULTAD DE MEDICINA. UNIVERSIDAD DE VALLADOLID.

Resumen. En el momento presente, la medicina basada en la evidencia es un imperativo para bien de la ciencia y de la mejor selección de tratamientos para nuestros pacientes. La realización de ensayos controlados aleatorizados precisa del conocimiento de principios de epidemiología y bioestadística. Además, obliga a utilizar información completa e imágenes que proporciona internet. Finalmente, es imprescindible que el cirujano moderno valore con precisión los datos de los artículos de revista basados en la evidencia.

From the opinion to the evidence in orthopedics. Part III: based-evidence medicine.

Summary. The based-evidence medicine is a necessity at present time, not only for science but in every way to choose the best treatment for our patients. To do a controlled aleatorized assay we need to know the principles of epidemiology and biostatistics. Besides, it needs to use complete information and images through the Internet. The surgeon must know with precision the data offered by journals' papers.

Correspondencia:
MM. Sánchez Martín
Regalado, 13, 6º
47002 - Valladolid
Telf. 983 300 121

Evolución histórica de la medicina basada en la evidencia

El origen de la medicina basada en la evidencia (MBE) viene de hace unos 30 años¹. No obstante, ha sido Gordon Guyatt², un internista canadiense de la Universidad McMaster, quien acuñó el término a comienzos de los años noventa. La MBE se describió inicialmente como una actitud de "escepticismo ilustrado" hacia la aplicación de tecnologías diagnósticas, terapéuticas y pronósticos. Indicado de otra manera, la MBE es el "uso concienzudo, explícito y juicioso de la mejor demostración actual para tomar decisiones en el cuidado de pacientes concretos". Recientemente se ha discutido un "desplazamiento paradigmático" más allá de las prácticas de enseñanza basada en la eminencia (opiniones y experiencias de expertos) hacia otra enseñanza basada en la evidencia^{3,4}. No se trata de conceptos excluyentes

entre sí, y claramente las opiniones y experiencias de expertos son parte importante del proceso de toma de decisiones cuando se trata a pacientes. Ciertamente, la práctica de la MBE integra la experiencia clínica individual² y las preferencias del paciente con la mejor evidencia clínica externa disponible. Esto lo hace para adoptar la aptitud de hacer una pregunta clínica pertinente, buscar la literatura, valorarla y aplicar los hallazgos en la atención al paciente de un caso clínico. Estos principios se han promovido recientemente en cirugía ortopédica y han dado lugar a series de "guías del usuario" en el Journal of Bone and Joint Surgery americano⁵.

Los cirujanos ortopédicos se han adherido clásicamente a innovaciones o técnicas nuevas basándose en una evidencia limitada. Con el crecimiento de los costes sanitarios por las nuevas técnicas y por el envejecimiento de la población, se ha hecho necesario moverse hacia el empleo de intervenciones que se basan en una investigación clínica de alta calidad con importantes resultados en el paciente y que demuestren coste-efectividad⁶.

Niveles de evidencia

Los investigadores han intentado minimizar posibles problemas a pacientes basando decisiones clínicas en

tipos de evidencia que sean menos fácilmente equívocos. Se ha definido lo que se piensa que es la evidencia que proporciona el cálculo menos sesgado del efecto de una situación: uno es la revisión sistemática que documenta la homogeneidad en resultados de un gran número de ensayos controlados aleatorizados de alta calidad (aleatorizado con seguimiento oculto, ciego, completo y análisis en la intención de tratar). A esto se denominó evidencia de nivel 1. Adicionalmente se categorizaron estudios de una intervención basada en un grado cada vez mayor de sesgos posibles, o sea, la revisión sistemática con ensayos controlados aleatorizados, que revelan diferencias en la consecuencia del tratamiento (heterogeneidad); ensayos controlados aleatorizados individuales de gran calidad (evidencia de nivel 1B); ensayos controlados aleatorizados menos rigurosos; estudios de cohorte u observacionales (evidencia de nivel 2); estudios de control de casos (evidencia de nivel 3); series de casos (evidencia de nivel 4); y opinión de expertos (evidencia de nivel 5) (Oxford Guidelines: <http://minerva.minervation.com/cebm/docs/levls.html>).

La medicina basada en la evidencia no es un fin en sí misma, sino más bien un conjunto de principios y herramientas que ayudan al clínico a distinguir ignorancia de evidencia de la seguridad científica auténtica, distinguir evidencia de opiniones injustificadas y, finalmente, proporcionar mejor atención al paciente.

Metodología de ensayos controlados aleatorizados (ECA)

La investigación clínica puede ser observacional (estudios de cohorte, control de casos, series de casos, publicación de casos) o experimental (ensayos controlados aleatorizados= ECA)⁷. Los estudios observacionales dominan la literatura quirúrgica⁸. La aleatorización elimina sesgos al elegir tratamiento, es el único medio de controlar factores pronósticos desconocidos y facilita el doble ciego.

Sin embargo, la aleatorización no garantiza que los resultados sean válidos, es decir que no tengan sesgos. Para eliminar estos sesgos, Devereaux y cols.⁹ presentan una guía para evaluar la validez de los ECA, dando especial atención a estas posibilidades confrontadas en ensayos de tipo quirúrgico, tal y como se muestra en la tabla 1, que pueden utilizarse para determinar si los resultados de los ECA representan un cálculo no sesgado del verdadero efecto del tratamiento. Hay dos soluciones de método que afectan a la validez de un ECA antes de que se produzca la designación de la intervención terapéutica, y cinco que se producen después de la misión del tratamiento.

¿Los pacientes se aleatorizan adecuadamente?

El proceso de aleatorización tiene que asegurar la impredecibilidad de las asignaciones de tratamiento. La echada de una moneda al aire, una secuencia de asignación de ordenador generada al azar y la tabla de números al azar, todos ellos, pueden generar secuencias de aleatorización. A veces, un ECA describe un proceso para determinar la asignación del tratamiento, que no es al azar sino más bien sistemático. La misión del paciente basada en la fecha de ingreso, número de historia clínica, fecha de nacimiento o la asignación de alternativas no asegura que se produzca aleatorización. La limitación de estos métodos de asignación de tratamiento tiene que ver con la siguiente guía: la ocultación de la aleatorización.

¿Había ocultación en la aleatorización?

Esto indica que los pacientes incluidos en un ECA son ignorantes de que el siguiente paciente va a estar aleatorizado para tratamiento o control. Si esto no es así, la asignación del paciente no se basará en un proceso al azar sino más bien en una elección del que investiga.

El riesgo de no utilizar métodos no aleatorizados de distribución del tratamiento es que fácilmente se produz-

Tabla 1. Guía para evaluar la validez de un ensayo controlado aleatorizado.

Posibilidades metodológicas hasta el momento de asignación de tratamiento
¿Los pacientes están aleatorizados adecuadamente?
¿Hay encubrimiento de aleatorización?
Posibilidades metodológicas después de asignación de tratamiento
¿Los pacientes se encuentran en varios grupos de tratamiento similares a factores pronósticos conocidos al inicio del estudio?
¿Los cirujanos tienen experiencia en las intervenciones que se evalúan?
¿Los pacientes, cuidadores, recogedores de datos y asesores de resultados estaban cegados a la distribución del tratamiento?
¿Los pacientes están analizados en los grupos en que fueron aleatorizados?
¿Los pacientes tienen un seguimiento completo?

Tabla 2. Grupos con posibilidad de que puedan ser cegados en ensayos controlados aleatorizados.

Participantes	Sujetos que se aseguran al azar para intervenciones que se evalúan
Cuidadores	Médicos, enfermeras, fisioterapeutas o cualquier otra persona dispensadora de cuidados a los participantes durante el período de estudio o que administra las intervenciones
Recolectores de datos	Sujetos que coleccionan datos para los resultados del estudio. La recogida de datos podría incluir la administración de un cuestionario, la toma de una medición física o sonsacar síntomas
Asesores de resultados	Sujetos que deciden al final si un participante ha soportado el resultado o resultado de interés

ca encubrimiento de aleatorización. Otros factores son la utilización de sobres transparentes que contienen el tratamiento. Debido a los sesgos de los investigadores, no sorprende que estudios empíricos hayan mostrado que tales ensayos, cuando se comparan con ECA adecuadamente ocultados, pueden sobrestimar el efecto del tratamiento de manera sustancial. Ciertamente, tales sesgos pueden ser mayores que el efecto para el que fue diseñado el estudio para detectarlo.

Para asegurar una ocultación de aleatorización lo más fiable es un proceso de aleatorización por teléfono central o unos frascos de medicación administrados en farmacia a doble ciego. Otros métodos para mantener el cegado de aleatorización -como sellado opaco de sobres numerados y aumento de confianza de los lectores- son los investigadores publican que se hizo revisión de cuentas y no mostró interferencia alguna.

¿Los pacientes están en varios grupos de tratamiento con factores pronósticos conocidos al comienzo del estudio?

El proceso de aleatorización es crear grupos de pacientes al comienzo del estudio con un pronóstico similar. Cuando así se hace y la metodología del ensayo es buena, los resultados pueden atribuirse a las intervenciones quirúrgicas que se evalúan. A veces, con mala suerte y ensayos pequeños, la aleatorización no produce grupos de similar pronóstico. Por ello, conviene, al margen del factor suerte, que los factores pronóstico conocidos estén razonablemente bien equilibrados. Si se discuten las diferencias entre ellos es esencial comprobar diferencias clínicas importantes y no estadísticamente marcadas. Incluso, cuando hay diferencias clínicas importantes no se ha perdido todo, ya que hay técnicas estadísticas de ajuste de diferencias, y el clínico tiene más confianza cuando tanto los análisis ajustados como no ajustados coinciden en la misma conclusión. De ahí que haya que buscar ensayos amplios, para ayudar a conseguir mejores resultados.

¿Los cirujanos tienen experiencia en las operaciones que evalúan?

El resultado del ensayo puede ser engañoso cuando el cirujano no tiene buena experiencia al intentar comparar el resultado de dos operaciones quirúrgicas. Es preciso que haya tenido una amplia curva de aprendizaje antes de emprender el estudio¹⁰.

¿Los pacientes, los cuidadores, los que recogen datos y los asesores de resultados están cegados en la distribución del tratamiento?

El cegado o enmascaramiento (blinding) en ECA es el proceso de ocultación de información sobre distribución del tratamiento en los que podrían estar influenciados por esta información¹¹. En la tabla 2 se exponen los grupos que pueden enmascarse en los ECA. Por ello, es importante que al realizar un ECA se valore el estado de cegamiento de los grupos en discusión para evitar sesgos. Una nota de precaución a propósito de la terminología del cegamiento es que clínicos y epidemiólogos en los textos sobre definiciones muestran variaciones marcadas en las definiciones sobre lo que es un enmascaramiento de un ensayo único, doble o triple¹². Lo mejor es fijarse en los grupos de la tabla 2. Si se mantiene como ciego o no ciego permitirá evitar aceptaciones incorrectas, ignorando esta inconsistente terminología de único, doble o triple ciego.

¿Los pacientes analizados están en grupos que fueron aleatorizados?

Si los pacientes no se analizan en los grupos que fueron aleatorizados, puede entonces destruirse el equilibrio en el pronóstico, ya que el objetivo de aleatorizar es crear grupos con similar pronóstico, y cualquier análisis que no se base en la intención de tratar pone en peligro la validez del ensayo.

¿El seguimiento del paciente es completo?

Los pacientes que se pierden para el seguimiento pueden experimentar un resultado de interés y, por tanto, éste

Tabla 3. Tipos de estudios observacionales.

Diseño del estudio	Propiedades
Publicación de un caso	Presentación de un caso
Series de casos	Colección de casos de publicación con detalles del procedimiento utilizado para tratamiento de los pacientes y de los resultados de los pacientes
Estudios de cohorte	Resultados que se comparan por asuntos que reciben dos o más tratamientos
Control de casos	Asuntos que tienen un resultado específico (los casos) y se comparan con asuntos que no tienen resultado (los controles), con relación a factores de riesgo de interés.

se altera. Cuando ocurre esto hay varias salidas al problema para comprobar el impacto fácilmente. En el peor de los casos hay que asumir que cualquiera que se pierda para el seguimiento en el grupo de tratamiento, experimento o tiene resultado de interés y ninguno perdido, sigue en el grupo control que experimenta el resultado de interés; en otras palabras, se le pasa al grupo control.

Una única salida relacionada con aleatorización de ensayos clínicos.

Los ECA en cirugía ortopédica han aleatorizado pacientes a los cirujanos que hacen su operación preferente con el fin de que se tenga experiencia con la operación que están haciendo¹⁰, y es importante que en cada centro haya algunos cirujanos que hagan un tipo de operación. No obstante, si por ejemplo ciertos factores institucionales (por ejemplo, cuidados postoperatorios de enfermería) tienen impacto sobre los resultados y hay variación de estos factores entre centros, es posible que los resultados no sean un reflejo directo de la propia intervención quirúrgica, sino más bien de estos factores institucionales. Cuando se busca este equilibrio en cada centro se tiene más confianza en que los resultados estén relacionados con la intervención quirúrgica que se analiza.

La validez frecuentemente es una doble decisión, es decir, saber si un estudio es válido o no lo es. La validez debería entenderse como una solución de continuidad que va desde lo completamente válido, de umbral poco fiable, de umbral fiable y, finalmente, completamente válido.

De una manera ideal, los ensayos controlados aleatorizados deberían ser completamente válidos. Pero esto ocurre rara vez. Lo más real es considerar si se contribuye a la validez del ensayo por encima o debajo del umbral de fiabilidad o en una zona incierta. Tomada esta decisión, el lector del ensayo puede situarse en los resultados del estudio con conocimiento de que se ha juzgado la metodología del ensayo para que sea fiable, no fiable o incierta.

Al final hay que comprender que aunque un ensayo sea válido, es decir, sin sesgos, es posible que aún resulte incorrecto por error de aleatorización, ya que la suerte puede dar lugar a resultados falso-positivos o falso-negativos. Todos los resultados del estudio tienen valores de probabilidad relacionados con la posibilidad de que los hallazgos se atribuyan a la suerte. La posibilidad de error al azar disminuye en ensayos grandes y con gran número de acontecimientos clínicos, especialmente si la significación estadística es extrema (excepcional). Por ello, los ensayos clínicos grandes tienen la ventaja de reducir la aparición de sesgos y errores al azar si sus resultados se diseñan, dirigen e interpretan adecuada y fácilmente y tienen un marcado impacto en la salud del paciente⁹.

En resumen, la capacidad del médico para obtener conclusiones sobre la relativa eficacia de las intervenciones basadas en su experiencia clínica es muchas veces limitada. Por tanto, con frecuencia los cirujanos utilizan la investigación de la evidencia para dirigir su práctica clínica. El ensayo controlado aleatorizado es el estudio más frecuente con diseño más fuerte. Sin embargo, la aleatorización en sí misma no garantiza que los resultados del ensayo sean válidos, o sea que estén exentos de sesgos. El conocimiento del posible impacto de varias características metodológicas en un ensayo controlado aleatorizado permite al clínico determinar la validez de un ensayo.

Metodología en estudios observacionales

Pacientes y cirujanos no permiten asignaciones al azar de los tratamientos quirúrgicos -a diferencia de los ensayos con tratamientos farmacéuticos- y la valoración de los tratamientos ortopédicos a menudo dependen de estudios observacionales en los que el tratamiento lo deciden el cirujano y el paciente conjuntamente, más que bajo las órdenes del investigador. Estos estudios observacionales analizan datos extraídos de historias clínicas, o recogidos de manera prospectiva para ser investigados¹³.

Tipos de estudios observacionales

De una manera escueta se describen en la tabla 3. La literatura quirúrgica presenta de forma más frecuente *series de casos* realizados por cualquier médico que recoge información sobre resultados. Son útiles para determinar resultados de un procedimiento concreto a largo plazo, pero también para resolver cuestiones complejas de factores que influyen en el resultado y mostrar los resultados que son mejores con un procedimiento nuevo frente a otro clásico. Estas series de casos deberán incluir pacientes consecutivos para evitar que sólo los pacientes con mejores resultados se incluyan en los datos obtenidos, ya que si no son consecutivos el resultado puede ser engañoso, pues el riesgo del paciente y de su tratamiento pueden variar a lo largo del estudio de una manera desconocida. En particular, los resultados de una serie de casos pueden ser mejores que los controles clásicos porque muchas técnicas de tratamiento han mejorado y no exactamente la técnica descrita en la serie de casos.

Los buenos *estudios de cohorte* proporcionan mejor información para tratamientos comparados que las series de casos porque la recogida de datos es estandarizada, los pacientes son desde el mismo período y en casos similares estandarizados, coleccionan datos de pautas antes de iniciar el tratamiento y nuevamente durante el seguimiento. Los tratamientos se comparan respecto a los resultados de los pacientes después de utilizar métodos estadísticos para tener en consideración diferencias en el riesgo del paciente antes del tratamiento. Los estudios de cohorte retrospectivos utilizan datos que se extraen de las historias clínicas sobre el tratamiento, riesgo apuntado, obteniendo así el resultado del paciente.

Los estudios de control de casos representan un método eficaz para evaluar cuáles son los factores de riesgo que ocurren menos frecuentemente y que proporcionan resultados adversos; se utilizan raramente para evaluar la efectividad del tratamiento.

Fuente de riesgos

Los errores que hacen que un estudio observacional no sea válido se denominan sesgos. Se definen como errores sistemáticos que producen tendencia hacia resultados erróneos¹⁴.

Vías de selección. Es la inclusión preferente de sujetos con ciertos resultados del tratamiento. En los estudios de cohorte se producen cuando la información del seguimiento no es tan fácil de recoger en sujetos que tienen resultados buenos o malos. Por ejemplo, cuando en historias clínicas y sujetos con menos de seis meses de

seguimiento, algunos sujetos con muy buenos resultados pueden no estar incluidos porque no tuvieron visitas de seguimiento después de seis meses. También es posible en algunos sujetos que hayan tenido malos resultados sin información del seguimiento en las historias por haber cambiado de cirujano.

Sesgos de información. Sirven para medir errores por definición imperfecta de un estudio variable. Procedimiento de recogida de datos defectuosos o imperfectos. Por ejemplo, el cirujano valora resultados del estado funcional de su tratamiento preferido más favorablemente que los resultados de otros tratamientos. Estos sesgos de información no se aplican a resultados objetivos, como, por ejemplo, si un paciente con artroplastia total de cadera precisa de cirugía de revisión. No obstante, una vez que el fallo es complicado (se trata de un implante radiológicamente flojo), entonces el observador que está a favor de un tratamiento particular puede estar más inclinado a adscribir el resultado más favorable al tratamiento. Las medidas de resultados bien especificadas no sólo disminuyen sesgos de información diferencial que están a favor de un tratamiento, sino que también reducen la mala clasificación al azar que hace más difícil detectar una diferencia auténtica en los tratamientos.

Los sesgos de selección e información se aplican teóricamente igual a ensayos controlados aleatorizados que a estudios observacionales. No obstante, estos sesgos pueden ser menos frecuentes en los primeros debido a protocolos desarrollados más rigurosamente y más patrocinio para contratar al personal de apoyo preciso para realizar cuidadosamente el seguimiento sin sesgos de sujetos y de valoración de resultados.

Confusión. Con confusión los efectos del tratamiento se alteran con los de los factores de riesgo, como ocurre cuando pacientes que siguen un tratamiento A se encuentran en riesgo por mal resultado antes del tratamiento, por ser de más edad o estar más enfermos que los pacientes que siguen el tratamiento B. Debido a esto, el tratamiento A puede parecer peor que el B, aunque en realidad sea igual de eficaz o parece similar al tratamiento B si realmente es superior.

La única ventaja inherente de los ECA en comparación con los estudios observacionales es la disminución de confusión.

Sesgos de publicación. Este tipo de sesgos se encuentra más allá del control del investigador y es debido a la tendencia a publicar estudios con resultados positivos (interesantes). Cuando hay sesgo de publicación, el cuerpo de literatura tiende a sobrestimar la eficacia del tratamiento. Como los ensayos controlados aleatorizados

Tabla 4. Elementos de datos requeridos para estudios observacionales.

1.	Factores que influyen en la relevancia
a.	Específicos de ejecución del tratamiento
b.	Características demográficas y salud del paciente
c.	Eliminación de medidas de resultados
2.	Factores que influyen en la validez
a.	Métodos de valoración de resultados
a.1.	Métodos para conectar con el paciente
a.2.	Personas que valoran los resultados
a.3.	Índice de seguimiento
b.	Factores que influyen en la confusión
b.1.	Métodos para seleccionar pacientes para cada tratamiento
b.2.	Diferencias entre grupos de tratamiento respecto a factores de riesgo importantes
	- Factores no diferentes estadísticamente
	- Factores diferentes estadísticamente
b.3.	Utilización de metodologías de ajuste estadístico, como análisis de estratificación o regresión
	- Sujetos con fracturas de cadera estratificados como estables o inestables

(ECA) negativos tienen más probabilidad de ser publicados que un estudio observacional negativo, los sesgos de publicación pueden ser más fáciles de presentarse en estudios observacionales.

Valoración de la calidad en estudios publicados. Los factores que influyen en los resultados se exponen en la tabla 4, referidos a los resultados de estudios observacionales, y se dividen en dos tipos: los que influyen en la relevancia y los que lo hacen sobre validez. La relevancia se refiere a los resultados que se aplican a situaciones clínicas frecuentes, y la validez a la precisión de los resultados para tal situación clínica de donde provenían. Los factores que influyen en la relevancia son interesantes para ensayos controlados aleatorizados porque estos estudios a menudo escogen una estrecha clase de pacientes y examinan el tratamiento en condiciones rigurosamente controladas, impracticables en la mayoría de pacientes. Cuando pacientes o afecciones en los ensayos controlados aleatorizados difieren mucho de la práctica normal, los resultados pueden ser menos relevantes que los que proceden de buenos estudios observacionales. La validez habitualmente es de mayor trascendencia en estudios observacionales que en ensayos controlados aleatorizados.

Papel de los estudios observacionales. Los estudios observacionales tienen varias ventajas sobre los ensayos controlados aleatorizados (ECA), pero la mayor es que no interfieren en la elección del paciente y del médico. También pueden diseñarse y llevarse a cabo con más celeridad y con menos coste y, a menudo, comprueban tratamientos en casos relevantes y con pacientes relevantes. Clásicamente tienen más sesgos que los ECA, pero pueden superarse en su mayoría con planificación y eje-

cución cuidadosas. La única desventaja inherente a estos estudios es su vulnerabilidad para la confusión al no reconocer o anotar los factores de riesgo. Por ello, los ECA con grandes ensayos en pacientes representativos proporcionan una evidencia más convincente que los estudios observacionales. Sin embargo, aquéllos a menudo no son prácticos de hacer cuando se necesitan. Si la evidencia de los ECA no está disponible, entonces la evidencia de los estudios observacionales hace que pueda ser útil.

Ningún estudio observacional ni ensayo controlado aleatorizado pueden considerarse definitivos. Mediante estudios observacionales comparados se puede obtener más información, siendo especialmente interesante los resultados de buenos estudios observacionales con diferente diseño que apoyan de manera consistente la superioridad de un tratamiento frente a otro, incluso aunque estos estudios no parezcan tener sesgos similares.

Se podría pensar que los estudios observacionales sustituyen a los ECA si los resultados de los estudios observacionales muestran de manera convincente que un tratamiento es muy superior y que la confusión no parece fácil. No obstante, algunos creen que pueden resultar engañosos y nunca deben tomarse como definitivos, como señala Sackett y cols.¹⁵. Otro papel de los estudios observacionales es proporcionar información que podría ayudar a diseñar ECA, en modificaciones del tratamiento, medidas de resultados y adecuación del paciente.

En resumen, la valoración de tratamientos en ortopedia depende sobre todo de estudios que analizan los datos obtenidos de historias clínicas, o recogidos prospectivamente con fines de investigación. Como en estos estudios el tratamiento lo deciden el cirujano y el pacien-

Tabla 5. Metodología de una revisión sistemática.

<p>Definir la cuestión</p> <ul style="list-style-type: none"> Especificar la inclusión y exclusión de criterios <ul style="list-style-type: none"> Población Intervención o exposición Resultado Metodología Establecer una hipótesis a priori para explicar la heterogeneidad <p>Dirigir la búsqueda de la literatura</p> <ul style="list-style-type: none"> Decidir sobre fuentes de información: bases de datos, expertos, agencias de fondos económicos, compañías farmacéuticas, búsqueda manual, archivos personales, registros, listas de citas de artículos recuperados Resolver limitaciones: de tiempo, datos no publicados, lenguaje Títulos de identidad y resúmenes <p>Aplicar criterios de inclusión y exclusión</p> <ul style="list-style-type: none"> Aplicar criterios de inclusión y exclusión a títulos y resúmenes Obtener artículos completos para títulos y resúmenes deseables Aplicar criterios de inclusión y exclusión a artículos completos Seleccionar artículos finales deseables Fijar acuerdos sobre selección del estudio <p>Crear abstracción de datos</p> <ul style="list-style-type: none"> Abstracción de datos: participantes, intervenciones, intervenciones para comparar, diseño de estudio Resultados Calidad metodológica Establecer acuerdo sobre comprobación de validez <p>Análisis de conducta</p> <ul style="list-style-type: none"> Determinar método para agrupar resultados Agrupar resultados, si son apropiados Decidir sobre el manejo de datos ausentes Explorar la heterogeneidad, sensibilidad y análisis de subgrupos Explorar posibilidad de sesgos de publicación
<p>Fuente: A manual for evidence-based clinical practice. American Medical Association. 2002.</p>

te, sin el control del investigador, estos estudios se conocen como observacionales. La mayoría de la investigación ortopédica utiliza diseños de estudios de observación porque son menos costosos y lleva menos tiempo realizarlos, incluyen a más pacientes representativos y deja la toma de decisiones médicas bajo control de médico y paciente. Los sesgos que se producen en estos estudios están muy relacionados con el diseño del estudio, los métodos de recogida de datos y el análisis estadístico. Todos los estudios observacionales deberían publicar factores que influyan en la relevancia (características de los pacientes, tratamiento y medidas de resultados) y factores que influyan en la validez (métodos de obtención de información y factores de riesgo del paciente que varíen con los tratamientos). Los mejores modelos de publicación de los estudios observacionales facilitarían

la interpretación de resultados y la comparación de estudios con resultados conflictivos¹³.

Metodología en revisiones sistemáticas y metaanálisis

Sólo los mejores diseños de experimentos pueden conseguir la mejor salvaguardia frente a la aparición de sesgos y proporcionar una estimación razonable de la verdadera asociación entre dos acontecimientos (co-tratamiento y curación), teniendo en cuenta que esto representa la evidencia. Entre los niveles de evidencia las observaciones no sistemáticas se encuentran en la parte baja de la escala y los ensayos clínicos aleatorizados con base metodológica, en la parte alta. Por tanto, para conseguir estimaciones de certeza precisa y sin sesgos, los investigadores necesitan evidencia aleatorizada a gran escala y las revisiones no sistemáticas tradicionales no resultan no ser la orientación apropiada para resumir la evidencia. La solución sería, pues, hacer resúmenes sistemáticos de toda la evidencia disponible con intención de obtener una medida precisa y no sesgada de la verdadera magnitud y dirección de la asociación entre resultados que podrían aplicarse ampliamente. Tales sumarios sistemáticos se denominan revisiones sistemáticas (RS) y el agrupamiento cuantitativo (estadístico) de estimaciones de los estudios individuales se denominan metaanálisis y, aunque los términos podrían intercambiarse, deben diferenciarse pues siempre es adecuado resumir la evidencia disponible en una revisión sistemática, pero no siempre es deseable o apropiado agrupar cuantitativamente resultados para alcanzar una estimación. Por tanto, no todas las revisiones sistemáticas incluyen metaanálisis; lo que les diferencia es su metodología.

Revisión sistemática

La realización de una revisión sistemática queda reflejada en sus diferentes pasos en la tabla 5. El sine qua non de la revisión sistemática es la búsqueda e identificación sistemática de estudios. Montori y cols.¹⁶ recomiendan el uso de un protocolo de búsqueda para hacer una lista de todas las fuentes de datos posibles y múltiples estrategias -que con frecuencia se superponen- para consultarlas. Las bases de datos electrónicas, como MEDLINE, proporcionan el grueso de datos de la mayor parte de las revisiones sistemáticas, ya que en sus índices aportan un gran número de publicaciones, la mayoría de Estados Unidos. Sin embargo, EMBASE, otra base de datos popular que se superpone con MEDLINE, tiene una mayor cobertura europea y multidisciplinaria.

La Cochrane Collaboration ha sido y es un punto de contacto para investigadores interesados en producir y difundir revisiones sistemáticas actualizadas sobre eficacia de tratamientos. Pues bien, Cochrane Collaboration somete de forma rutinaria protocolos de revisión por informes de expertos, y publica protocolos para que se realicen críticas abiertas y amplias así como comentarios, para evitar sesgos, como paso previo, en el proceso de revisión. Los desacuerdos con frecuencia se resuelven mediante discusión o arbitrio por una tercera persona.

Metaanálisis y heterogeneidad

El objeto de las revisiones sistemáticas es contestar a una cuestión concreta con toda la evidencia disponible. Para hacer un metaanálisis, sin embargo, la decisión vuelve a ser revisada después de análisis cuidadoso de los estudios que se han agrupado. Si se encuentran estudios que difieren en la estimación de consecuencias pero que tienen amplios intervalos de confianza que hacen pensar que la suerte es buena explicación para diferencias entre estudios, la revisión debería ser razonable. Pero si la recopilación de hechos no parece satisfactoria, los que la revisan pueden buscar comodidad de pruebas estadísticas formales para las diferencias entre estudios. En estas pruebas estadísticas la hipótesis nula establece que no hay diferencia alguna en los resultados entre estudios. Lamentablemente, estas pruebas tienden a tener poder limitado para detectar marcada heterogeneidad estadística cuando la revisión incluye pocos estudios con pocos pacientes en cada estudio. Por tanto, las pruebas que no rechazan la hipótesis nula no indican necesariamente ausencia de gran heterogeneidad. Si ésta fuera importante, los que hacen la revisión pueden decidir no agruparlas, a lo que seguiría un resumen cualitativo de la evidencia. En otro sentido, pueden decidir que es susceptible de reunir las y proponer a priori las características de una exploración del estudio para explicar que podría producirse.

Las explicaciones de heterogeneidad proceden de dos fuentes de variabilidad. La primera es clínica y se debe a diferencias en pacientes, intervenciones y resultados comprobados en cada estudio; su valoración precisa de estimación crítica y sensibilidad clínica, pero no de prueba estadística. El segundo tipo de heterogeneidad es metodológico, que se refiere a diferencias en los resultados del tratamiento entre estudios que resultan de diferencias en salvaguardias metodológicas frente a sesgos.

En resumen, las revisiones sistemáticas de investigación original se utilizan cada vez más, y difieren de las revisiones denominadas narrativas en que tratan una

cuestión clínica específica, precisan de una amplia investigación de la literatura, utilizan criterios de selección explícitos para identificar estudios relevantes, valoran la calidad metodológica de estudios incluidos, exploran diferencias entre resultados de estudios y sintetizan de manera cualitativa y cuantitativa los resultados de estudios. Se denomina metaanálisis a las revisiones sistemáticas que reúnen resultados cuantitativamente de más de un estudio. Existen varias organizaciones que están funcionando en colaboración para producir revisiones sistemáticas y metaanálisis de alta calidad. Familiarizarnos con ello nos conducirá a tener una mayor habilidad al utilizar este tipo de artículos y, en cualquier caso, para que las fuentes de evidencia sean útiles.

Metaanálisis o revisión sistemática cuantitativa es un sistema que emplea métodos estadísticos para combinar los resultados de dos o más estudios¹⁷. Todas las revisiones sistemáticas son retrospectivas y observacionales y, por tanto, sujetas a error sistemático y aleatorio¹⁸. De ahí que la calidad de una revisión sistemática, de acuerdo con su validez, dependa de los métodos científicos que se hayan utilizado para disminuir error y sesgo.

Un metaanálisis bien dirigido es de gran valor para el cirujano, ya que no es usual que estudios únicos proporcionen respuestas definitivas a cuestiones clínicas. Incluso, una revisión cuantitativa bien dirigida puede resolver discrepancias entre estudios con resultados conflictivos¹⁹.

Los principios que sirven de guía para conducir metaanálisis son varios: preguntas específicas sobre salud, uso de estrategia de una búsqueda completa, valoración de reproductibilidad de la selección del estudio, valoración de la validez del estudio, evaluación de la heterogeneidad, diferencia de resultados en estudios cruzados, inclusión de todas las medidas relevantes y clínicamente útiles del resultado del tratamiento y pruebas de robustez de los resultados relativos a características de los estudios primarios (análisis de sensibilidad)²⁰.

La popularidad de las revisiones sistemáticas ha producido un incremento de 500 veces en el número de metaanálisis publicados en la pasada década. Lamentablemente, su mayor utilización no se ha acompañado de apreciación de la importancia de la metodología clínica sin la cual los metaanálisis pueden producir estimaciones imprecisas, con sesgos y desorientaciones sobre la efectividad de una intervención médica o quirúrgica concreta^{21,22}, que puede tener serias complicaciones en términos de calidad y coste en la atención al paciente. Muchos autores, en un esfuerzo para describir fuentes de sesgos en la dirección de metaanálisis e instrumento, los han

desarrollado para graduar la calidad científica de tales estudios^{23,24}.

Los cirujanos ortopédicos deben darse cuenta de las limitaciones y sesgos del metaanálisis y esforzarse en disminuirlos. Igualmente, los editores de revistas tienen que asegurar que los metaanálisis que publican se adhieren a la metodología científica aceptada.

La metodología del metaanálisis difiere de la revisión de literatura narrativa en varios puntos: 1) el metaanálisis a menudo se ocupa de cuestiones clínicas determinadas; 2) supone una estrategia de búsqueda extensa y explícita; 3) la selección de artículos se basa en un paquete de criterios de elegibilidad; 4) la validez de los estudios es comprobada; y 5) se realiza un resumen cuantitativo de datos¹⁸.

La actual "responsabilidad" de la metodología puede haber facilitado en parte el empleo incrementado de metaanálisis. Aunque conceptualmente parece fácil, la producción de revisiones sistemáticas de alta calidad es muy exigente. A veces, sujetos con conocimiento limitado de un tratamiento, de la biología de la enfermedad o las circunstancias clínicas relacionadas con un tema específico, pueden realizar un metaanálisis que tiene poca relevancia clínica. Por tanto, mientras que la metodología estadística del metaanálisis ha avanzado mucho en la década pasada, ha habido también un incremento en metaanálisis relacionados utilizando métodos que no son los mejores. Esto puede evitarse fácilmente en la mayoría de metaanálisis asegurando que los investigadores atienden las soluciones que se presentan en cada artículo en el sistema de puntuación de Oxman y Guyat (1991) (tabla 6).

Aprender a leer artículos de revista basados en la evidencia

La medicina basada en la evidencia (MBE) es un proceso que utiliza información clínica veraz, además de la experiencia práctica del cirujano para tomar decisiones prácticas²⁵. En teoría, también las pautas prácticas que se generan por estudios clínicos científicos pueden reducir el número de complicaciones y malos resultados quirúrgicos que proporcionan los más altos costes sanitarios²⁶. A la larga, para trabajar en MBE será preciso aprender estadística, probabilidad, investigación clínica, encuestas guiadas, revisiones sistemáticas y niveles de evidencia. Por suerte, por el momento es suficiente con estar preparado para que el cirujano ortopédico se levante para tomar velocidad y leer sobre niveles de evidencia en las revistas revisadas por expertos. Los editores de estas revistas han adoptado sistemas de grados para sim-

plificar el proceso de valoración por contenidos²⁷. Los niveles de evidencia han sido creados para ayudar a los clínicos a comprender que se ha creado la evidencia para investigación científica que dé respuesta a cuestiones o resuelva problemas. La aplicación de artículos de revista a la práctica clínica es el cuarto estadio del algoritmo que resuelve problemas que comienzan por: 1) formular cuestiones contestables; 2) recoger evidencia; 3) valorar la evidencia; 4) poner evidencia en la práctica; 5) valorar los resultados de poner en práctica la evidencia. Cuando los resultados clínicos no son aceptables, deben revisarse las situaciones buscando errores en los cuatro primeros estadios. Las revistas ayudan a utilizar mejor la evidencia ya que están haciendo que los resultados de los ensayos clínicos estén disponibles, y se los asigne un nivel de importancia²⁸.

A continuación se presenta la forma en que el lector debe familiarizarse con ciertos términos, demostrar algunos principios de la MBE y ayudar al cirujano ortopédico a ubicar la evidencia indispensable²⁹.

Evidencia médica y errores: consideraciones estadísticas

El médico práctico suele ser escéptico de las estadísticas que aparecen en la literatura médica y las contempla con frecuencia como un complejo método de apoyo a la información incorrecta, "masajeando" datos y manipulando resultados para apoyar algo que un autor considera significativo. En otras palabras, las estadísticas son una herramienta diabólica que da respuesta a lo que el autor del artículo pretende, y que puede variar de la "verdad". Cuando el clínico se informe para utilizar y conocer medidas estadísticas y resultados, entonces se le permitirá aumentar los conocimientos de lo que se presenta en la literatura médica clínica y, en cierto modo, llegar a conclusiones más correctas de la base de datos que presenta un artículo.

Por tanto, es recomendable conocer algunas fuentes de error y la manera más conveniente de tratarlas para crear evidencia. Conviene advertir que todos los estudios tienen algún error y es necesario hacer un análisis crítico del artículo para, con los medios disponibles, valorar, reducir y poner en evidencia el error. Lo más importante es saber la relevancia de lo que se pregunta en el trabajo a los pacientes. Así, surgen varias preguntas. ¿La cuestión es adecuada? ¿El estudio realizado se basa en la población adecuada? ¿Habrá sesgos al hacer la selección? ¿Habrá sesgos de técnica? ¿Cuándo se valora una radiografía de manera estándar, debería utilizarse similar método para valorar el artículo científico? ¿Existe un

mejor enfoque sistemático en varios artículos utilizados por varios clubs de revistas ortopédicas?³⁰.

Lo primero que hay que tener en cuenta en una valoración estándar es decir qué tipo de estudio se ha reali-

zado: ¿serie de casos, ensayo clínico aleatorizado o una serie de cohorte prospectiva?

La serie de casos es muy útil para algunos problemas concretos: una amplia población de pacientes con un pro-

Tabla 6. Índice de calidad científica en revisiones de investigación.

1.	¿Se han utilizado métodos de investigación para encontrar evidencia (investigación original sobre cuestión o cuestiones primarias)?	no	parcialmente	sí
2.	¿La investigación de evidencia fue razonablemente completa?	no	parcialmente	sí
3.	¿Se utilizaron criterios para decidir qué estudios se incluyeron en la revisión publicada?	no	parcialmente	sí
4.	¿Se evitaron sesgos en la selección de los estudios?	no	no se puede decir	sí
5.	¿Se utilizaron criterios para comprobar la validez de los estudios incluidos publicados?	no	parcialmente	sí
6.	¿Se refirió la validez de todos los estudios en el texto valorado con utilización de criterios apropiados, tanto en seleccionar estudios para inclusión o en analizar los estudios que se citaron?	no	parcialmente	sí
7.	¿Se utilizaron métodos para combinar los hallazgos de los estudios relevantes (para alcanzar una conclusión) publicados?	no	parcialmente	sí
8.	¿Se combinaron los hallazgos de los estudios relevantes apropiadamente con relación a la cuestión primaria a que se refiere la revisión?	no	no se puede decir	sí
9.	¿Se hicieron las conclusiones por el autor o autores apoyándose en los datos y/o análisis publicados en la revisión?	no	parcialmente	sí
10.	¿Cómo podría valorarse la calidad científica de esta revisión?			
	Extensa		Mayor	
	Defectos		Defectos	
	1	2	3	4
				5
				6
				7
				Mínima
				Defectos

Instrucciones para puntuación del Índice

El propósito de este índice es evaluar la calidad científica, es decir, la adhesión a principios científicos de revisiones de investigación (artículos de revisión) publicados en la literatura médica. No se pretende medir calidad, importancia, relevancia, originalidad literaria u otros atributos de las revisiones.

El índice se diseña para valorar revisiones de investigación originales sobre cuestiones prácticas con relación a la etiología, diagnóstico, pronóstico, terapéutica o profilaxis. Una revisión de investigación es una encuesta de investigación. Los mismos principios que se aplican a las encuestas epidemiológicas se aplican a las revisiones, dejando muy claramente especificada una cuestión, una población diana debe identificarse, valorarse, obteniendo información apropiada de esta población, de una manera no sesgada, debiendo derivarse conclusiones, a veces con ayuda de un análisis estadístico formal, como se hace en metaanálisis. La diferencia fundamental entre revisiones y encuestas epidemiológicas es la unidad de análisis, no las salidas científicas de las preguntas que se atienden en este índice.

Como la mayor parte de revisiones publicadas no incluyen una sección de métodos, es difícil contestar a algunas de las preguntas del índice. Las contestaciones (respuestas) deberían basarse, en lo posible, en la información aportada en las revisiones. Si los métodos que se han utilizado se consideran relativamente incompletos a un asunto concreto, la puntuación se considera "parcial". De igual manera, si nos aporta ninguna información con relación a los métodos utilizados relativos a una cuestión particular, se puntúa como "no se puede decir" salvo que haya información en la revisión que sugiera si reúne o no un criterio.

En la cuestión 8, si no se hace intento alguno para combinarse los hallazgos y no se hace ninguna afirmación con relación a lo inapropiado de combinar los hallazgos se marca un "no". Si un resumen de cálculo se hace en cualquier parte del contenido, discusión o sección del resumen del trabajo publicado, y el método utilizado para sacar el cálculo no se publica, se marca un "no", incluso si no hay un juicio relacionado con las limitaciones de combinar los hallazgos de los estudios revisados; en caso de duda se marca un "no se puede decir".

En la cuestión 9 para una revisión que reciba un "sí", los datos (no exactamente las citaciones) deben publicarse diciendo que apoyan la principal conclusión con relación a la cuestión o cuestiones originales a que va dirigida la revisión.

La puntuación para la cuestión 10, la calidad científica global debería basarse en las contestaciones a las nueve primeras cuestiones. Si la opción "no se puede decir" se utiliza una o más veces en las cuestiones precedentes, es fácil que la revisión tenga pequeños defectos a lo mejor y sea difícil descartar defectos mayores (es decir, una puntuación de 4 puntos o menos). Si se utiliza la opción "no" en las cuestiones 3, 4, 6 u 8, la revisión es fácil que tenga defectos mayores, esto es, una puntuación de 4 puntos o menos, dependiendo del número e importancia de los defectos.

blema concreto tratado con una operación concreta, o un algoritmo de intervenciones con un alto porcentaje de seguimiento y con importantes medidas de resultados. Este tipo de serie proporciona importante información al médico sobre sus pacientes, pero tiene mayor utilidad cuando se incluye una población como la vista por el médico y con una operación que puede reproducirla él mismo. Los resultados obtenidos pueden ser transmitidos al paciente antes de realizar la operación para su mejor conocimiento y consentimiento informado.

Finalmente, el médico debe ser objetivo y justo al comparar sus propias habilidades con las de los autores de la serie publicada, ya que hay operaciones muy complejas que sólo algunos cirujanos están capacitados para realizar, de manera que se precisará de entrenamiento especial en habilidades y juicio, y una mayor experiencia ulterior para ello.

Otra situación es cuando hay múltiples opciones terapéuticas y hay que desarrollar la mejor evidencia posible para seleccionarlas. Entonces el mejor método es un ensayo clínico aleatorizado. Hay que saber lo que es un valor p : no es más que una prueba de probabilidad. De tal manera que cuando se comparan dos métodos para tratar un problema, el valor estándar de p es 0.05. Para que un tratamiento sea mejor que otro, hay que observar un 95% de seguridad, si bien es preciso que la muestra aleatorizada pueda irse a la derecha o izquierda de la curva de población. A continuación hay que entender cómo el trabajo de las muestras es el 95% de los límites de confianza. Esto quiere decir que en el 95% del tiempo (límite de confianza del 95%), la muestra se superpone con la media real de la población. Una vez comprendido el concepto de muestra en estudio, se ponen en evidencia por sí mismos los errores de la muestra. Cuando se comparan dos tratamientos hay dos muestras de población que proceden de la población global con el problema concreto y que deberían caer bajo esta curva. Asumiendo que los resultados de estos dos tratamientos son iguales, pueden coincidir o casualmente estar separados. Si ambas muestras son de hecho las mismas y por suerte al azar, uno sale al encuentro de las muestras de la principal población que no se solapan, podría concluirse que estos dos tratamientos eran diferentes en base a los datos apreciados en las muestras aleatorizadas. Tomando $p=0.05$, se acepta que hay un 5% (5 entre 100) de posibilidades de que este ocurra. De ahí que en un caso particular dos métodos de tratamiento parecen ser diferentes cuando de hecho son lo mismo. Es lo que se llama error alfa, que ocurre cuando la población para tratamientos A y B son lo mismo realmente, pero las dos muestras, por suerte, no

se solapan. Es decir, concluimos incorrectamente que hay una diferencia entre tratamientos cuando de hecho no hay diferencia alguna. Una verdadera diferencia de una población en estudio debería reflejarse en una diferencia entre dos muestras de población tratadas con métodos diferentes. Los grupos de muestra no deberían solaparse pero esto sería una auténtica diferencia en resultados, llegando a la conclusión de que hay diferencia en los tratamientos.

El error alfa aumenta cuando en el estudio entran múltiples variables llegando hasta el 37% de los ensayos clínicos aleatorizados analizados en 40 estudios de estas características, llevados a cabo por Bhandari y cols., en 2005, de manera que se llega a la conclusión de que había diferencia cuando en realidad nunca existió³¹. En principio esto se debe a que las evaluaciones múltiples se realizaron sin corregir estas variables; en estos casos de evaluaciones múltiples debe hacerse la corrección. El lector puede para ello utilizar la pauta de dividir el valor de p por el número de evaluaciones hechas, con lo que se demuestra claramente que existe una diferencia entre los dos métodos de tratamiento respecto a la variable específica ($p=0.005$). Además de detectar un error alfa, el cirujano decide si se trata de una diferencia que es cierta y probada pero inútil clínicamente en la actualidad. Esto se refiere al tamaño del efecto, es decir, que la diferencia entre tratamientos tenga una diferencia medible, pero no es de importancia del cirujano para el paciente. De esta manera, comparando dos técnicas de osteosíntesis en fracturas tibiales, por ejemplo, y asumiendo que el tiempo medio de consolidación de la fractura es de 250 días con un tratamiento y de 260 días con otro, y que la desviación estándar es de 60 días, aquí puede haber una diferencia auténtica: por ejemplo, si $p=0.03$; sin embargo, en el gran esquema del tratamiento de las fracturas de tibia, 10 días con una desviación estándar de 60 días en los grupos de tratamiento, probablemente no es muy importante. El tamaño del efecto aquí podría ser $10/60$ o sólo 0.16. En general, con un tamaño de efecto igual o mayor de 0.8 se considera bastante importante para que no pueda cambiar la decisión práctica, dan similares perfiles de riesgo al paciente y otros factores. En estudios con variables dicotómicas, el empleo de reducción de riesgo relativo es un método excelente de evaluar la importancia clínica. La reducción absoluta del riesgo o el número de casos que están afectados por el tratamiento es otra herramienta útil para valorar la importancia clínica. Si una afección prevalente, como la no consolidación de la fractura, disminuye un 5% con un tratamiento dado, entonces un 5 por ciento de los pacientes tienen un ries-

go reducido de no consolidación con este tratamiento.

También es importante fijarse en las diferencias de los parámetros clínicos publicados. Si el estudio muestra un 95% de posibilidades de que el tratamiento puede aportar una mejora de 4 puntos en la puntuación sobre una escala de 100 puntos, esto no es de gran importancia para el cirujano ni para el paciente. En particular, ¿qué riesgos deberían tomarse en consideración para ganar una media de 4 puntos sobre 100? El valor de un resultado está muy influenciado por la forma en que se presenta la información al cirujano al leer el artículo y textos científicos. Esto tiene peligro³². Cuando uno lee solamente artículos de revisión o textos generales, las opiniones quedan ensombrecidas o camufladas por un autor que no tiene el apoyo de datos de estudios relevantes.

Otro tipo es el error beta que se aprecia en ensayos clínicos, y es mucho más frecuente que el error alfa. Se produce cuando se concluye que no hay diferencia alguna entre dos tratamientos, cuando de hecho tal diferencia puede existir. Por ejemplo, dos poblaciones de pacientes pueden tener una auténtica diferencia pero dos muestras al azar de poblaciones diferentes no son diferentes. El error beta se relaciona con el número de pacientes de la muestra. Es lo que se denomina infravaloración de un estudio. El poder de un estudio es en esencia su capacidad de demostrar que dos cosas son lo mismo. De ahí que la comparación de 3 fracturas de tibia tratadas con un método y otras 3 con otro, podrían no ser la mismísima información convincente. Sin embargo, si 600 pacientes tratados con un método se comparan con 600 tratados con otro, esto podría ser una información muy precisa.

Los ensayos clínicos aleatorizados no establecen frecuentemente el poder del estudio o su capacidad de demostrar diferencias. Y esto es más evidente en problemas que tienen variables dicotómicas, como por ejemplo en el caso de la trombosis venosa profunda. Esto hace que uno se pregunte cómo deberían publicarse los resultados negativos, ya que un error beta puede enmascarar una diferencia auténtica. Para ello hay dos maneras de realizarlo. La primera es cuando no hay diferencia entre dos resultados comparativos el valor p y la amplitud de los valores pueden publicarse, de manera que aportan la probabilidad de mostrar cómo son de diferentes los dos resultados. Además, el autor puede publicar el número de pacientes que pueden precisarse para demostrar una diferencia sí, de hecho, el número real de pacientes del estudio y el número requerido para mostrar una diferencia son diferentes. Por ejemplo, puede ocurrir que la conclusión del estudio es que no se encontró ninguna diferencia con los datos disponibles; sin embargo, si se incluyó un

número adicional de pacientes en el estudio, se podría posiblemente encontrar una diferencia, lo cual daría al lector una idea de cómo en estos dos tratamientos pueda existir una gran diferencia. Establecido de manera diferente, con poder del 80%, ¿qué diferencia podría haberse detectado en los resultados en dos grupos? Por ejemplo, un estudio sobre dos tratamientos diferentes en fracturas de tibia pueden concluir con un poder del 80% detectando una diferencia de 6 semanas para el tiempo de consolidación. Luego el lector ve más claramente que si estos dos tratamientos tienen diferentes resultados la posible diferencia en el tiempo de consolidación de la fractura sería de 6 semanas.

A medida que pasa el tiempo se aprecian mejores artículos de ortopedia censados por expertos, ya que los autores comprenden mejor la estadística y los expertos lo emplean para su propio uso. Antes, las publicaciones clínicas tenían errores que proporcionaban menos buenos tratamientos a los pacientes. Ahora, es claro que quien lee artículos debe tener cierto nivel de sofisticación y conocimiento de los métodos estadísticos utilizados en estudios clínicos y, al mismo tiempo, esto ha de repercutir en que nuestros pacientes sean mejor tratados. Sin embargo, incluso la revisión basada en la evidencia y los niveles de evidencia no tienen en consideración todas las cosas, ya que a veces la mejor evidencia es una serie simple de casos que muestra los resultados de un método de tratamiento particular para una población de pacientes particular. Como cirujanos, deberíamos llegar a ser, de alguna manera, expertos en el análisis crítico de la literatura ortopédica.

Dónde encontrar evidencia ortopédica

Puede encontrarse evidencia en revistas revisadas por expertos y medios de investigación médica con expertos censores que interpretan información clínica y reúnen evidencia. Las revistas publican los resultados de ensayos clínicos y ensayos de revisión sistemática y, en cambio, los medios de investigación acumulan éstas y otras publicaciones similares para proporcionar una imagen más completa de la evidencia. Pocas veces se presenta evidencia en una presentación científica o un centro de formación. Los que hacen revisiones de temas y los cirujanos expertos que escriben o hablan en su campo de experiencia son realmente expertos, pero no proporcionan signos de la mejor evidencia.

La evidencia proviene de ensayos clínicos que demuestran eficacia en una prueba diagnóstica o en el tratamiento. La prueba diagnóstica certera es clave para un tratamiento exitoso, después de hacer un proceso

diagnóstico. Luego, la decisión terapéutica proviene de la participación entre cirujano y paciente en inclinarse por una u otra opción. Cuando el cirujano informa a su paciente tiene que conocer las posibilidades de éxito de tal opción para tal paciente. Esto se basa en ensayos clínicos que comparan las alternativas terapéuticas. A diferencia de lo que ocurría en el pasado, actualmente hay necesidad de tener ensayos clínicos que proporcionan al cirujano la evidencia que necesita para practicar en un mundo basado en la medicina de la evidencia.

Decidirse por una opción quirúrgica, aunque a veces es fácil, en ocasiones resulta difícil pues hay que contar, en ausencia de evidencia y coincidencia de opiniones e información, con que no son precisas. Un cirujano muy ocupado decide con más facilidad recoger la opinión de un colega y preguntar a un experto y tener respuesta a sus correosas preguntas, que formularse preguntas sin respuesta y buscar en la literatura. Es más fácil leer un libro o un artículo de revisión y hacer lo que el "experto" recomienda que esforzarse en aclarar un problema clínico y buscar ensayos clínicos relevantes y explicar los resultados a un paciente escéptico. Aquí es donde la medicina basada en la evidencia resulta útil y relevante, gracias ampliamente a los medios de búsqueda electrónica y a la creciente disponibilidad de niveles de evidencia. Además, la tendencia a utilizar medición de resultados en toda la medicina clínica es reafirmación del concepto que señala que el resultado final es un resultado empírico que interesa a la mayoría de los pacientes, en opinión de Codman³³.

Cómo aplicar la evidencia

Aunque cada cirujano quiere lo mejor para su paciente, en cualquier circunstancia, con frecuencia discrepa en cómo tratar mejor los problemas clínicos dentro de esta divergencia que puede atribuirse a muchos factores, como la seguridad en series de casos no controlados y la aceptación sin decisión de la opinión de los colegas de más edad que están en desacuerdo en muchos asuntos clínicos, como le ocurriría a cualquiera. Como nadie puede tener razón, la variación de opiniones impacta casi con seguridad de forma negativa en la atención clínica global. El coste de las malas decisiones clínicas, sean diagnósticas o terapéuticas, las paga el paciente y la sociedad.

La práctica basada en la evidencia proporciona el ejemplo alternativo frente a la práctica basada en la opinión³⁴. La práctica basada en la evidencia supone aplicar evidencia de alta calidad a partir de ensayos aleatorizados para tratar a los pacientes.

Los elementos esenciales de la práctica basada en la evidencia es el planteamiento de una cuestión clínica, identificación y evaluación de la evidencia y aplicación de la misma a la decisión clínica. De ahí que el cirujano necesite reconocer la necesidad de evidencia, buscar y reconocer la evidencia y utilizar la evidencia.

Para practicar evidencia en ortopedia, el primer paso que tiene que adoptar el cirujano ante un dilema clínico es aceptar que no conoce la respuesta (si no lo sabe) y buscar la mejor, o admitir que hay una respuesta mejor cuando se conoce su existencia. Identificar la mejor evidencia no siempre es recto o sincero. Los libros suelen estar pasados de moda, pero son la base de opinión más usual. La literatura proporciona evidencia más definitiva, si bien el número de publicaciones es abrumador. Hay que utilizar búsqueda en Medline que, para mejor aplicación dispone de "clinical queries" en los servicios de PubMed, en su columna izquierda. Las revistas como el Journal of Bone and Joint Surgery americano también proporcionan datos sobre la medicina basada en la evidencia en un apartado al final de cada volumen mensual. Los niveles de evidencia establecen cuatro tipos: terapéuticos, pronósticos, diagnósticos y análisis de decisión y económicos. Los estudios se establecen en cinco niveles de acuerdo con su diseño, en tanto que los ensayos aleatorizados son los de más nivel o nivel I de evidencia y la opinión del experto, es el nivel más bajo, nivel V. No obstante, un problema práctico para cirujanos que utilizan investigación de la literatura y las "clinical queries" o que quieren perfeccionar la investigación por niveles de evidencia, lo mejor es identificar la evidencia más segura para la mayoría de las investigaciones clínicas.

Otras fuentes de revisión sistemática son las siguientes:

- el sitio en la red "Clinical evidence" del British Medical Journal's: www.clinicalevidence.org/ceweb/conditions/index.jsp
- y la base de datos Cochrane: www.cochrane.org/index.htm.

Para muchas cuestiones clínicas, los ensayos aleatorizados faltan, pero esto está cambiando.

Para consultar casos clínicos en revisiones sistemáticas o al lado de la cama del paciente es necesario entrar en internet. Si el cirujano no tiene acceso a internet en la clínica puede responder a preguntas cuando pueda tener el acceso más cercano a la red. En el futuro, los ordenadores personales podrán proporcionar al cirujano acceso en tiempo real a internet en "zonas calientes". El paso final en la práctica basada en la evidencia precisará que el cirujano adopte la mejor evidencia, intercambiando su

opinión y sobre todo, su práctica cuando coteje con buena evidencia, paso este último el más difícil, ya que no todos los tratamientos quirúrgicos cuando se someten a ensayos aleatorizados pueden considerarse beneficiosos³⁵ y, por respeto al paciente, tales tratamientos deberían perfeccionarse o abandonarse²⁹.

Bibliografía:

1. Poolman RN, Petrisor BA, Marti RK et al. Misconceptions about practising evidence-based orthopedic surgery. *Acta Orthop* 2007; 78:2-11.
2. Guyatt GH, Rennie D. The evidence-based medicine working group. User's guide to the medical literature. A manual for evidence-based clinical practice. Chicago: AMA Press.. 2002.
3. Bhandari M, Tornetta P III. Evidence-based orthopedics: a paradigm shift. *Clin Orthop* 2003; 413:9-10.
4. Hurwitz SR, Tornetta P III, Wright JG. An AOA critical issue how to read the literature to change your practice: an evidence-based medicine approach. *J Bone Joint Surg* 2006; 88A:1873-9.
5. Bhandari M, Guyatt GH, Swiontkowski MF. User's guide to the orthopaedic literature: how to use an article about a surgical therapy. *J Bone Joint Surg* 2001; 83A:916-26.
6. Brighton B, Bhandari M, Tornetta MD III, Felson DT. Hierarchy of evidence from case reports to randomized controlled trials. *Clin Orthop* 2003; 413:19-24.
7. Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. *Lancet* 2000; 359:57-61.
8. Schultz KF, Grimes DA. Generation of allocation sequences in randomized trials. Chance, not chance. *Lancet* 2002; 359:515-9.
9. Devereaux PJ, McKee MD, Yusuf S. Methodologic issues in randomized controlled trials of surgical interventions. *Clin Orthop* 2003; 413:25-32.
10. McCulloch P, Tayler I, Sassako M et al. Randomized trials in surgery: problems and possible solutions. *BMJ* 2002; 321:1448-51.
11. Devereaux PJ, Bhandari M, Montori VM et al. Double blind, you are the weakest link: good bye! *ACP. J Club* 2002; 136:A11-A12.
12. Devereaux JP, Manns BJ, Ghali WA et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA* 2001; 285:2000-3.
13. Martz A, Marsh JL. Methodologic issues in observational issues. *Clin Orthop* 2003; 413:33-42.
14. Szklo M, Nieto FJ. *Epidemiology: beyond the basis*. Gaithersburg, MD. Aspen Publishers, Inc. 2000.
15. Sackett DL, Richardson WS, Rosenberg W, Haynes RB. *Evidence based medicine: how to practice and teach EBM*. New York: Churchill Livingstone. 1997.
16. Montori JM, Swiontkowski MF, Cook DJ. Methodologic issues systematic review and meta-analyses. *Clin Orthop* 2003; 413:43-54.
17. Bhandari M, Morrow F, Kulkarni AV, Tornetta P. Meta-analysis in orthopaedic surgery. A systematic review on their methodologies. *J Bone Joint Surg* 2001; 83A:15-24.
18. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for health care decisions. Mulrow C, Cook D (eds). Philadelphia: American College of Physicians. 1998. pp. 5-12.
19. Gerbarg ZB, Horwitz RI. Resolving conflicting clinical trials: guidelines for meta-analysis. *J Clin Epidemiol* 1988; 41:503-9.
20. Cook DJ, Sackett DL, Spitzer WO. Methodological guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on meta-analysis. *J Clin Epidemiol* 1995; 48:167-71.
21. Felson DT. Bias in meta-analytic research. *J Clin Epidemiol* 1992; 45:885-92.
22. Thompson SG, Pocock SJ. Can meta-analysis be trusted? *Lancet* 1991; 338:1127-30.
23. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol* 1991; 44:1271-8.
24. Oxman AD, Guyatt GH, Singer J et al. Agreement among reviewers of review articles. *J Clin Epidemiol* 1991; 44:91-8.
25. Evidence-based medicine working group. Evidence-based medicine. A new approach to teaching the practice of medicine.. *JAMA* 1992; 268:2420-5.
26. Wright JG, Swiontkowski MF, Heckman JD. Introducing levels of evidence to the journal. *J Bone Joint Surg* 2003; 85A:1-3.
27. Wright JG, Einhorn TA, Heckman JD. Grades of recommendation. *J Bone Joint Surg* 2005; 87A:1909-10.
28. Kreder HJ. Evidence-based surgical practice: what is it and do we need it? *World J Surg* 1999; 23:1232-5.
29. Hurwitz SR, Tornetta P III, Wright JG. How to read the literature to change your practice: an evidence-based medicine approach. *J Bone Joint Surg* 2006; 88A:1873-9.
30. Dirschl DR, Tornetta P 3rd, Bhandari M. Designing, conducting and evaluating clubs in orthopaedic surgery. *Clin Orthop* 2003; 413:146-57.
31. Bhandari M, Montori VM, Schemitsch EH. The undue influence of significant p-values of perceived importance of study results. *Acta Orthop* 2005; 76:291-5.
32. Thoma A, Farrokhyar F, Bhandari M, Tandan V. User's guide to the surgical literature. How to assess a randomized controlled trial in surgery. *Can J Surg* 2004; 47:200-8.
33. Mallon WJ, Ernest Amory Coadman. *The end result of a live in medicine*. Philadelphia: W.B. Saunders. 1999.
34. Sackett DL, Richardson WS, Ronsenberg W, Haynes RB. *Evidence-based medicine. How to practice and teach EBM*. New York: Churchill Livingstone. 1998.
35. Moseley JB, O'Malley K, Petersen NJ et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med* 2002; 347:81-8.