

EDICIÓN ESPECIAL

# METODOLOGÍA DE LA INVESTIGACIÓN



ARTÍCULOS PUBLICADOS EN LA REVISTA INTERNACIONAL  
“JOURNAL OF ORTHOPAEDIC & SPORTS PHYSICAL THERAPY”

GENERANDO PREGUNTAS DE  
INVESTIGACIÓN: EL VÍNCULO  
ENTRE EVIDENCIA Y PRÁCTICA

SESGO

FUNDAMENTOS  
DE MEDICIÓN

INTERPRETACIÓN  
DE LOS RESULTADOS



Tecnología aplicada a la  
Rehabilitación y Estética

# Estamos en cada movimiento



Medita  
oficial

**NUEVO**



### Electromagnetoterapia

**Magnetherp 330**  
Hasta 200 Gauss

- Pantalla táctil
- Interacción sencilla, intuitiva y agradable.

### Nueva Generación en Ondas Radiales

#### Accustec NG

Es un método no invasivo que asegura un tratamiento efectivo



### Ultrasonido

**Sonotherp 990**  
Versión 1 Mhz o 3 Mhz  
Emisión **Continuo o Pulsante**,  
tiempo de sesión, y regulación  
de potencia entre **0.5 y 3.5  
Watt/Cm2**.



### Laser terapéutico infrarrojo

**Laser 320 MW**  
Posee un diodo laser  
de 808 nanometros



### Termoterapia y Electroestimulación

#### Termocel Stim

Permite el tratamiento en simultáneo de  
Termoterapia y Electroestimulación

Ventas 0810-666-9097

[meditea.com/shop](http://meditea.com/shop)



Tel. 54-11 4953-6114 / 4951-9079 - Viamonte 2255/2265  
Buenos Aires - Argentina - CP (C1056ABI)



**MEDITEA**  
[www.medita.com](http://www.medita.com)

# “La importancia de una práctica clínica actualizada, crítica y responsable”

LIC. DIEGO RUFFINO | LIC. GONZALO ECHEGARAY | LIC. GABRIEL NOVOA | LIC. PABLO POLICASTRO

**E**s claro que en la práctica clínica todos perseguimos la premisa de tratar de ofrecerle al paciente el mejor tratamiento posible. Muchas veces sabemos cuál es pero otras tantas solo lo intuimos. Sin embargo, el problema de la “intuición” es que nos puede llevar a considerar teorías y conceptos que han demostrado no ser los correctos.

Citando un claro ejemplo, los estudios complementarios para el tratamiento del dolor lumbar como la radiografía o la resonancia magnética han sido muy importantes para la toma de decisiones clínicas durante muchos años. No obstante, entre 2000 y 2010, comenzaron a multiplicarse la cantidad de estudios que observaban que muchos hallazgos que considerábamos como patológicos y anormales como por ejemplo una hernia de disco, también se encontraban en sujetos sanos y sin historia de dolor lumbar. Estas observaciones han demostrado la dificultad de encontrar una causa a este dolor y comenzamos, a partir de ello, a clasificar a este como inespecífico en más del 80% de las ocasiones. Sumado a ello, en los últimos 10 años se demostró que dichos hallazgos

“anormales” transmitidos a los pacientes en informes extensos, los cuales muchas veces son ratificados por nosotros, tienen un impacto negativo en los pacientes.

Este camino ha permitido obtener información muy valiosa para ayudar a nuestros pacientes lo cual, sin duda, debe ser el mayor logro de la medicina basada en la evidencia.

Hace un tiempo atrás, David Sackett, un médico canadiense, observó que muchas de las decisiones tomadas en la práctica clínica eran basadas exclusivamente en la experiencia, situación que conllevaba a que muchos profesionales de la salud no sepan si sus tratamientos eran los mejores disponibles hasta el momento.

Con el tiempo, el Dr. Sackett, considerado el padre de la práctica basada en la evidencia, definió a esta como la combinación de la mejor evidencia posible, las preferencias del paciente y la experiencia profesional.

En un mundo donde la información crece de manera exponencial y donde trabajar con personas ofrece un desafío versátil y complejo, la actualización constante permite acortar la brecha entre lo que creemos y lo que realmente es.

Se estima que, aproximadamente, 3 millones de investigaciones científicas son publicadas por año a nivel mundial. Algunas realizadas con un correcto método científico y otras no tanto, por lo que resulta sumamente importante estar entrenado en la discriminación de la información.

Persiguiendo la premisa de alcanzar aptitudes y destrezas en la selección de la mejor información, desde la AKD les presentamos este número especial de la Revista dedicado exclusivamente a la práctica basada en la evidencia. En las siguientes páginas podrán disfrutar de conceptos que no solo incumben a los investigadores, sino que también son muy importantes para todos aquellos que busquen la mejor información disponible para poder ayudar a sus pacientes.

Cómo Editores, consideramos que una práctica clínica actualizada, crítica y responsable es el camino para el crecimiento kinésico internacional generando no solo un beneficio profesional individual, sino también como organización y familia que es la AKD.



# BODY CARE®

HEALTH & TECHNOLOGY



## TOBILLERA DEPORTIVA RUNNER PRO CON TIRAS

USO DEPORTIVO  
PREVENTIVO Y CORRECTIVO

Compresión y estabilidad  
firme y pareja en tibia y peroné.

Sustituye las vendas.

Protege tobillos débiles.

Actividades de alto rendimiento.



## MEJOR EN MOVIMIENTO



INMOVILIZADOR DE  
MUÑECA CORTO | **BC1651**



COLLAR ESTABILIZADOR  
TIPO SHANZ | **BC1085**



FAJA ELÁSTICA  
BALLENADA | **BC1352**



RODILLERA CON SOSTEN  
ROTULIANO | **BC1102**

# Evidencia en la Práctica: Nueva serie para fisioterapeutas

STEVEN J. KAMPER

School of Public Health, University of Sydney,  
Camperdown, Australia;

Centre for Pain, Health and Lifestyle, Australia.

Orthop Sports Phys Ther 2018;48(6):512-513. doi:10.2519/jospt.2018.0701

Se espera que los fisioterapeutas presten sus servicios clínicos dentro del paradigma de la práctica basada en evidencia (PBE). Mientras continúa el debate sobre la forma óptima de PBE en los círculos académicos y profesionales, está bien aceptado el principio subyacente: la práctica clínica debe estar informada por la mejor evidencia de investigación disponible (integrada con los valores clínicos y del paciente).<sup>2</sup> A pesar de la simplicidad de la idea, la PBE impone considerables demandas a los fisioterapeutas, y las barreras para la implementación son numerosas.<sup>1</sup>

Una condición previa absoluta para la PBE es la capacidad de cada fisioterapeuta para comprender la investigación científica. Comprender un artículo de investigación requiere cierta familiaridad con el lenguaje y los métodos utilizados en el mundo de la investigación. A partir de esta base, los hallazgos de los estudios se pueden integrar con el conocimiento específico del paciente y la patología que un fisioterapeuta ya posee. No es necesario ser estadístico o epidemiólogo; una comprensión básica de la forma en que funciona la investigación puede permitir a los fisioterapeutas utilizar de forma efectiva la evidencia en su práctica. Queremos que los fisioterapeutas puedan leer un documento (o ver una presentación), comprender lo que se dice, extraer lo que es útil y aplicarlo en la práctica. El objetivo de la serie “Evidencia en la práctica” (Evidence in Practice) es ayudar a los fisioterapeutas practicantes a desarrollar experiencia en la comprensión de la investigación. Cada artículo se enfocará en una característica del proceso de investigación y explicará los

conceptos básicos de manera concisa (menos de 2 páginas) en un lenguaje no técnico. Los artículos de Evidence in Practice no profundizarán en la complejidad de la metodología científica; más bien, el objetivo es proporcionar explicaciones “del tamaño de un bocado” que cubran los temas más importantes para interpretar la evidencia de la investigación. La atención se centrará en la investigación clínica, en oposición a la investigación básica de descubrimiento, para reflejar el propósito central y a los lectores de la revista. Los artículos de la serie cubrirán algunos temas generales: (1) los conceptos básicos para identificar literatura de investigación relevante, como hacer una pregunta, buscar y los tipos de preguntas de investigación; (2) principios y componentes importantes de la investigación, como sesgo, aleatorización, cegamiento, cocientes de probabilidades y riesgos, y medidas de resultado; y (3) los diferentes tipos de estudios que los lectores encontrarán, incluidos ensayos clínicos, metanálisis, subgrupos y estudios de diagnóstico.



FIGURA. La primera entrega de este tema en Evidence in Practice aparece en la página 512.

Esperamos que Evidence in Practice se convierta en un recurso valioso para los fisioterapeutas que intentan enfrentar el desafío de integrar la evidencia de la investigación en la prestación de salud. Al igual que cualquier iniciativa nueva, no hay certeza de que obtengamos el tono y el contenido de inmediato; así que, comenzando con la primera entrega, titulada “Comprometerse con la investigación: vincular la evidencia con la práctica”, en este número de la revista, los invitamos a enviarnos sus comentarios. ●

### REFERENCIAS

1. da Silva TM, Costa LC, Garcia AN, Costa LO. What do physical therapists think about evidence-based practice? A systematic review. *Man Ther.* 2015;20:388-401. <https://doi.org/10.1016/j.math.2014.10.009>
2. Straus SE. *Evidence-Based Medicine: How to Practice and Teach It.* 4th ed. Edinburgh, UK: Elsevier/Churchill Livingstone; 2011.



# Comprometerse con la investigación: vincular la evidencia con la práctica

**E**ste es el primero de una serie de resúmenes breves que cubren distintos aspectos de la investigación clínica. El objetivo es ayudar a los fisioterapeutas a convertirse en consumidores más idóneos de la investigación y fomentar la incorporación adecuada de la evidencia en la práctica.

Los debates sobre el por qué y cómo de la práctica basada en evidencia (PBE) ocupan miles de páginas en libros de texto, revistas e Internet. Estas discusiones van más allá de la medicina en campos tan diversos como la educación, las finanzas y la política social, por nombrar algunas. Independientemente de lo que se piense sobre la PBE, existen al menos 3 razones por las cuales los fisioterapeutas deben participar en la evidencia de la investigación:

1. Los pacientes esperan que la atención médica se base en la ciencia.
2. Si la fisioterapia quiere llamarse a sí misma una profesión científica, entonces se debe generar evidencia relevante y utilizarla en la práctica clínica.
3. Las instituciones que pagan por los servicios, como las aseguradoras y los organismos gubernamentales, están haciendo cada vez más reembolsos contingentes para proporcionar atención basada en evidencia.

A veces, los resultados de la investigación difieren sustancialmente de su propia experiencia sobre la efectividad de un tratamiento en particular. Esto presenta un desafío y plantea la pregunta lógica de por qué se debe esperar que un fisioterapeuta priorice la evidencia de un estudio sobre su propia experiencia clínica.

La información obtenida de la experiencia clínica debe evaluarse y valorarse para determinar el sesgo, al igual que la información de la investigación. Por lo tanto, es importante comprender las limitaciones potenciales y los sesgos relevantes.

## Sesgo de confirmación

Las personas tienden a evaluar e interpretar la información de manera que refuerce sus propias creencias. Sobrevaloramos la información que respalda nuestras creencias, ignoramos u olvidamos la información que las contradice e interpretamos la información ambigua de una manera que favorece nuestros puntos de vista. Este es el sesgo de confirmación; no es un defecto de carácter de un individuo, ni es un intento de justificar las acciones propias. Es una especie de atajo cognitivo que nuestra especie ha desarrollado para una función diaria más eficiente. El problema es que nos lleva por mal camino en ciertas situaciones. Por ejemplo, considere a los fisioterapeutas que creen que (1) son buenos en su trabajo y (2) que tienen las mejores intenciones, de corazón, con respecto a sus pacientes. ¿Qué efecto podría tener el sesgo de confirmación en la recolección de la efectividad de sus tratamientos para estos fisioterapeutas?

## Sesgo de recuerdo

Las personas tienen una tendencia a recordar mejor los eventos sustanciales o impresionantes, en lugar de los eventos promedio o más comunes. En el contexto de la práctica clínica, esto podría llevar a recordar claramente a los pacientes que se recuperaron espectacularmente bien

(o espectacularmente mal) y recordar menos claramente a aquellos con resultado promedio. Entonces, cuando llega el momento de aplicar la experiencia previa al siguiente paciente, el resultado más probable puede ser el que menos se recuerde.

Un problema relacionado es que algunos pacientes se involucrarán con la terapia, mientras que otros dejarán de asistir. Los resultados de estos últimos pacientes son generalmente desconocidos. Dada la probabilidad de sesgo de confirmación, es probable que se suponga que los abandonos han funcionado bien (no necesitaban más tratamiento), aunque quizás lo contrario sea lo correcto.

## Observación clínica

Supongamos que trata a un paciente con un tratamiento particular. El paciente asiste varias veces y, después de 3 semanas, mejora mucho. La interpretación más simple es que el tratamiento fue efectivo. El problema es que la mejora puede, o no, deberse al tratamiento. Existe una diferencia fundamental entre el cambio en el resultado (cambio clínico a lo largo del tiempo) y el efecto del tratamiento (cambio clínico a lo largo del tiempo debido al tratamiento). En la clínica se observa el cambio en el resultado durante 3 semanas, pero solo una parte de ese cambio se debe al tratamiento. Hay una serie de factores que contribuyen al cambio en el resultado, incluidos los siguientes (**FIGURA**).

**Historia natural** Muchas afecciones se resuelven por completo o, en cierta medida, por sí solas, particularmente

<sup>1</sup> School of Public Health, University of Sydney, Camperdown, Australia; Centre for Pain, Health and Lifestyle, Australia.  
Copyright ©2018 Journal of Orthopaedic & Sports Physical Therapy®

las afecciones agudas. Por ejemplo, un paciente que presenta dolor, hinchazón y función limitada el día después de un esguince de tobillo mejorará esos resultados en las próximas 2 semanas. Esto sucederá si el paciente recibe tratamiento intensivo, tratamiento mínimo, tratamiento ineficaz o ningún tratamiento. Obviamente, la historia natural varía según la condición, por lo que la proporción de cambio en el resultado debido a la historia natural variará de un caso a otro.

**Regresión a la media** Muchas condiciones siguen un curso episódico o fluctuante. Los pacientes no tienen síntomas o síntomas de bajo nivel la mayoría de las veces, intercalados por exacerbaciones o brotes. El dolor de espalda y el dolor de cuello son ejemplos comunes. Las personas con estas afecciones generalmente buscan atención cuando los síntomas son peores, es decir, durante una exacerbación. Debido a que la gravedad de los síntomas es generalmente peor o casi máximo cuando consultan al profesional de la salud, es probable que la próxima fluctuación en la gravedad sea una mejora. En la práctica, esto significa que es más probable que mejoren después de una consulta, independientemente del tratamiento.

**Efectos placebo** Los efectos de placebo probablemente se deban a la manipulación de las expectativas del paciente y / o al acondicionamiento clásico. Independientemente del mecanismo, los efectos del placebo están unidos a las acciones de cualquier tratamiento particular, pero son distintas de las mismas; son “portátiles” en diferentes

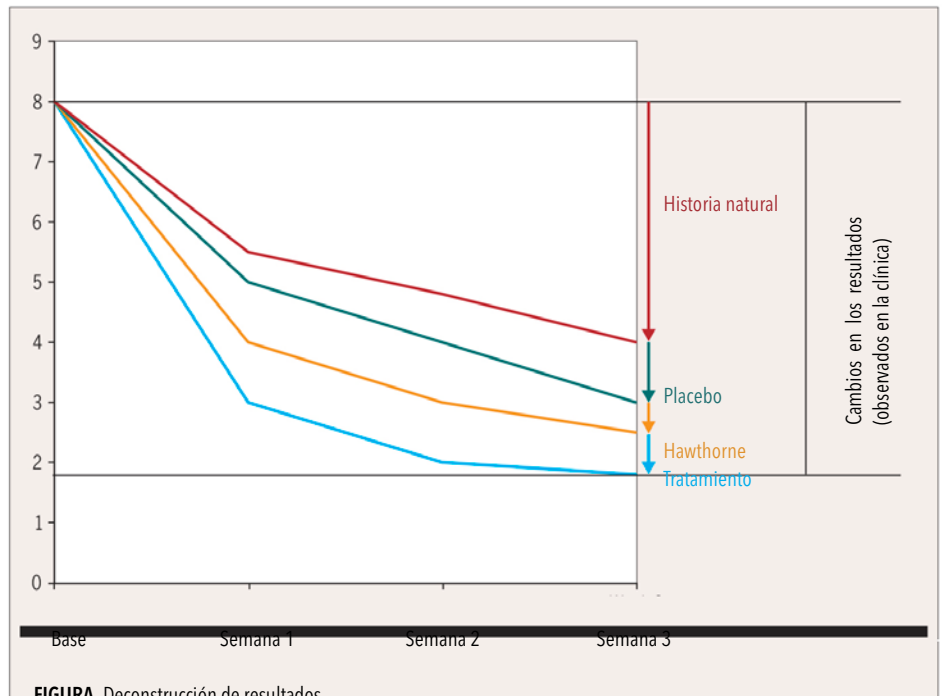


FIGURA. Deconstrucción de resultados.

tratamientos. Pueden contribuir al cambio en el resultado, pero son diferentes del efecto del tratamiento.

**Pacientes respetuosos** La mayoría de los fisioterapeutas intentan establecer una relación con sus pacientes. Como resultado, muchos pacientes no quieren “decepcionar” a su terapeuta al no responder a los esfuerzos del terapeuta para ayudarlos. Este deseo de los pacientes puede sesgar los informes de resultados. Los pacientes pueden parecer (y afirmar que) han mejorado más de lo que realmente han mejorado.

## RESUMEN

La PBE bien ejecutada no descarta la experiencia clínica. Sin embargo, se de-

ben tener en cuenta los posibles sesgos al evaluar y aplicar la información obtenida en la experiencia clínica. Del mismo modo, la PBE no se trata de encontrar errores en estudios publicados. Implica considerar cuidadosamente la naturaleza y la magnitud de los sesgos que se aplican a toda la información (de la experiencia clínica y de la investigación) que un fisioterapeuta debe tener en cuenta en su proceso de razonamiento.

Existe la idea errónea de que la PBE busca transferir la toma de decisiones clínicas a un libro de cocina. En realidad, la PBE le pide más a los fisioterapeutas, no menos. Los desafía a aceptar fallas en su razonamiento y reconocer sus propios sesgos, y exige que desarrollen las habilidades para encontrar, evaluar e integrar evidencia de investigación en su práctica. ●

## DOWNLOAD PowerPoint Slides of JOSPT Figures

JOSPT offers PowerPoint slides of figures to accompany all full-text articles with figures on JOSPT's website ([www.jospt.org](http://www.jospt.org)). These slides are generated automatically by the site, and can be downloaded and saved. They include the article title, authors, and full citation. JOSPT offers full-text format for all articles published from January 2010 to date.

# Generando preguntas de investigación: el vínculo entre evidencia y práctica

**T**anto para los fisioterapeutas como para los investigadores, una pregunta clara y fácil de responder es el punto de partida del proceso de práctica basado en la evidencia. Para los fisioterapeutas, la pregunta tiene como objetivo comprender al paciente y obtener la información necesaria para tratarlo.

Para los investigadores, la pregunta orienta el diseño del estudio e impulsa todo, desde el reclutamiento hasta la presentación de informes. La pregunta también es el punto de anclaje que vincula la práctica clínica con la investigación (FIGURA). Para los fisioterapeutas experimentados, formular una pregunta clínica es probablemente un proceso inconsciente que ocurre durante la evaluación del paciente. Involucrarse en la práctica basada en la evidencia hace que este proceso se vuelva consciente y explícito. La definición clara y concisa de la pregunta conduce directamente a la búsqueda de investigaciones relevantes.

## Definiendo la pregunta: PICO

Existen varios tipos de preguntas clínicas, pero la mayoría de los fisioterapeutas lo que quieren es información sobre la efectividad del tratamiento. Para encontrar este tipo de información, es útil estructurar la pregunta en el formato PICO (Patient/Problem (Paciente o Problema), Intervention (Tratamiento), Comparator (Comparación), Outcome (Resultado)). Para preguntas sobre diagnóstico y pronóstico, se pueden definir partes similares (pero no idénticas) de la pregunta.

La sensibilidad y la especificidad son relevantes al definir la pregunta. Se puede hacer que una parte de la pregunta sea

más específica agregando palabras que definan con mayor precisión al paciente, las intervenciones y el resultado. Esto ayudará a identificar la evidencia que sea más relevante con respecto a la situación en particular. Sin embargo, si su pregunta es demasiado específica, no encontrará ninguna evidencia que cumpla con todos sus criterios. Por otro lado, puede hacer que las partes de su pregunta sean muy amplias. Las preguntas amplias conducen a búsquedas sensibles que identifican evidencia relevante, pero también obtienen mucha información irrelevante. Esto implica rastrear a través de una larga lista

de estudios para encontrar lo que le interesa. Existe cierta habilidad para definir y clarificar preguntas de modo que una búsqueda sea específica y sensible.

## Paciente

Esta parte de la pregunta define el grupo de pacientes en el que está interesado. Lo más probable es que incluya el diagnóstico o la descripción de la afección. Esto puede ser muy amplio (por ej., dolor de hombro) o más específico (por ej., manguito rotador, capsulitis adhesiva, luxación, etc.). El primero conducirá a una búsqueda sensible y el segundo dará resultados más específicos. También se puede hacer que la pregunta sea más específica agregando elementos que definan la población que le interesa. Por ejemplo, puede que solo le interesen las personas mayores de 65 años

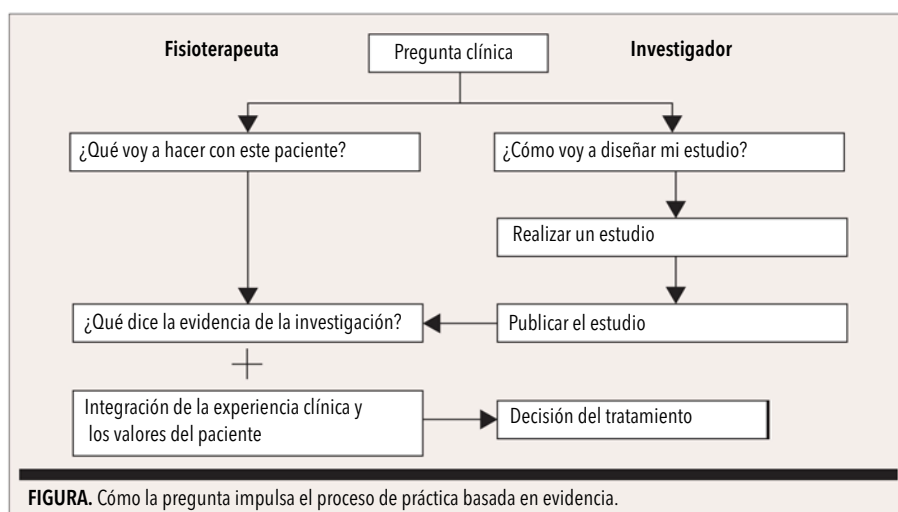


FIGURA. Cómo la pregunta impulsa el proceso de práctica basada en evidencia.

<sup>1</sup> University of Sydney, School of Public Health, Camperdown, Australia; Centre for Pain, Health and Lifestyle, Australia. Copyright ©2018 Journal of Orthopaedic & Sports Physical Therapy®



o aquellas con lesiones relacionadas al fútbol. Todo esto se aplica de igual modo a cuestiones de efectividad del tratamiento, diagnóstico y pronóstico.

### Comparador

Las preguntas sobre la efectividad del tratamiento son siempre preguntas comparativas. Por lo tanto, si está interesado en encontrar información sobre la efectividad del tratamiento, es fundamental definir con qué se debe comparar el tratamiento de interés. Por ejemplo, la pregunta sobre la efectividad de la movilización de la columna cervical versus ningún tratamiento es diferente de la pregunta sobre la efectividad de la movilización de la columna cervical versus los ejercicios de fortalecimiento muscular cervical. Esta sección no es relevante para las preguntas de diagnóstico y pronóstico.

### Resultados

Es importante comprender qué resultado es más importante para su paciente. Los resultados de mayor interés deberían estar especificados en la pregunta. Por ejemplo, los resultados del regreso al trabajo pueden ser críticos para algunos pacientes, pero no son relevantes

para los pacientes jubilados. Las mismas consideraciones se aplican a las preguntas de pronóstico, pero la sección no es relevante para las preguntas de diagnóstico.

### Construyendo una búsqueda

La explosión de la evidencia de investigación en la medicina está bien documentada. El número de estudios publicados ha aumentado a un ritmo increíble: MEDLINE incluye más de 23.000.000 de artículos, y la Physiotherapy Evidence Database indexa más de 40.000 revisiones sistemáticas y ensayos controlados aleatorios relevantes para la fisioterapia. Estos números demuestran la necesidad de poseer habilidades para buscar y evaluar evidencia.

Diseñar una búsqueda implica convertir una pregunta clínica en términos de búsqueda. Esto significa crear una lista de sinónimos para cada parte de la pregunta. Por ejemplo, consulte algunos de los términos de búsqueda que Nascimento y col.<sup>1</sup> usaron para su revisión sistemática (TABLA). Es común realizar la búsqueda en etapas. Comience con una búsqueda sensible (amplia), aumentando gradualmen-

te la especificidad hasta que quede una cantidad manejable de artículos. Esto podría implicar realizar una primera búsqueda que incluya solo los sinónimos de la parte "Paciente" de la pregunta. Si esa búsqueda inicial devuelve demasiados artículos, entonces se pueden agregar los sinónimos para la parte "Intervención/tratamiento" de la pregunta. Si la búsqueda inicial identifica muy pocos artículos, entonces se deben agregar más sinónimos a la parte "Paciente".

### Operadores booleanos

"Y" y "O" son operadores booleanos; conectan los términos de búsqueda y hacen que su búsqueda sea más sensible (usando O) o más específica (usando Y). Los sinónimos dentro de la misma parte de la pregunta se conectan usando O (por ejemplo, lumbar o espalda o columna vertebral o espinal o vértebra). Diferentes partes de la pregunta están conectadas por Y; esto significa que la búsqueda encontrará artículos que incluyan cualquiera de las palabras de los pacientes (lumbar, espalda, etc.) y cualquiera de las palabras de intervención (por ejemplo, ejercicio o fortalecimiento o resistencia). Incluyendo más sinónimos conectados por O se encontrarán más artículos; si se agregan partes de la pregunta usando Y entre ellas, se encontrarán menos.

Incluso si todas las partes de la pregunta original no están incluidas en la búsqueda final, todas deben ser definidas. Esto se debe a que todas las partes de la pregunta son relevantes cuando se decide cómo aplicar los resultados, de los artículos identificados en su búsqueda, a sus pacientes. ●

TABLA		Términos de búsqueda *
Elemento PICO	Descripción	Términos de búsqueda
Paciente (o problema)	Personas con dolor Patelofemoral	Rótula o articulación de la rodilla o rodilla o dolor anterior de la rodilla o (patela * o patelofemoral* o femoropatelar * o retropatelar*) "adj2" (dolor o síndrome o disfunción).
Intervención	Ejercicios de fortalecimiento	Ejercicios de fortalecimiento terapia de ejercicio o ejercicio o rehabilitación o (fisioterapia o terapia física) o entrenamiento de resistencia o levantamiento de pesas o fuerza o (excéntrico o concéntrico o isométrico).

*Abreviatura: PICO (Patient/Problem, Intervention, Comparator, Outcome)*  
 \* El símbolo de asterisco en las cadenas de búsqueda es un comodín que puede representar cualquier letra o grupo de letras. El término "adj2" significa "adyacente a". La pregunta era: "¿El fortalecimiento de la cadera y la rodilla aumenta la fuerza, reduce el dolor y mejora la actividad en individuos con dolor patelofemoral?"

### REFERENCIAS

1. Nascimento LR, Teixeira-Salmela LF, Souza RB, Resende RA. Hip and knee strengthening is more effective than knee strengthening alone for reducing pain and improving activity in individuals with patellofemoral pain: a systematic review with meta-analysis. J Orthop Sports Phys Ther. 2018;48:19-31. <https://doi.org/10.2519/jospt.2018.7365>

## Sesgo: vinculación de la evidencia con la práctica

Si le pide a un epidemiólogo una definición de sesgo, obtendrá algo como: “Una desviación sistemática de la estimación del parámetro de muestra del valor de la población”. A menos que usted también sea epidemiólogo, es poco probable que esto le resulte útil. Desglosado en un lenguaje más accesible, sería algo así como, “Una diferencia [desviación] en una dirección particular [sistemática] entre los resultados de un estudio [estimación de parámetros] y lo que sucede en la vida real [valor de la población]”.

Dicho de otra forma, el sesgo significa que la información que tenemos frente a nosotros da una impresión incorrecta de lo que realmente está sucediendo.

Tenga en cuenta que el sesgo es diferente del error. El error aleatorio ocurre porque la forma en que medimos las cosas, ya sea por observación, cuestionarios, entrevistas, herramientas de medición o imágenes, nunca es perfectamente precisa (FIGURA). Posteriormente, los artículos sobre medición de Evidence in Practice, discutirán este tema.

Cualquier información puede estar sesgada, ya sea que provenga de investigaciones, experiencia clínica, cursos profesionales, colegas, medios de comunicación, etc. El primer artículo de Evidence in Practice de la serie explica algunos de los sesgos relacionados con la información proveniente de la experiencia clínica. El sesgo tiende a sobrestimar los efectos. Por ejemplo, los tratamientos parecen más efectivos de lo que serán en la práctica, las pruebas diagnósticas parecen más precisas de lo que realmente son y los factores pronósticos predicen más fuertemente los resultados. En la clínica, es fácil ver cómo la sobreestimación de la efectividad de un tratamiento podría influir en su manejo.

La sobreestimación de la precisión de una prueba diagnóstica podría llevar a un profesional de la salud a sopesar los resultados de esa prueba más de lo que se justifica. La información de un estudio con una estimación sesgada del pronóstico podría dar lugar a consejos y expectativas imprecisos.

### Algunos tipos de sesgo

Existen muchos, muchos tipos diferentes de sesgo. A continuación se detallan algunos que se aplican a la investigación (y a veces a la experiencia clínica).

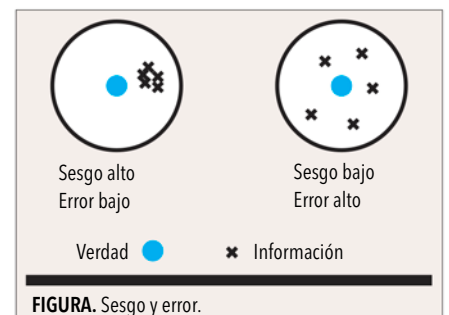
### Sesgo de selección

Considere 2 estudios, (1) un ensayo controlado aleatorio en el que los pacientes con dolor de espalda se asignan aleatoriamente para recibir un tratamiento de ejercicios o un folleto informativo, y (2) un estudio comparativo en el que los pacientes con lumbalgia reciben ejercicio o el folleto, según la elección del profesional a cargo. Las estimaciones de efectividad comparativa se pueden derivar de cualquier estudio al observar la diferencia en el resultado en-

tre el grupo de ejercicio y el grupo de folletos. No es difícil imaginar que el fisioterapeuta tratante elegiría el tratamiento con ejercicios en el estudio 2 para los pacientes que él piensa que se beneficiarían más del ejercicio; después de todo, eso es lo que sucedería en la práctica. Pero es probable que haya diferencias importantes entre las personas que reciben el tratamiento con ejercicios y las que reciben el folleto en el estudio 2, lo que significa que la estimación de la efectividad del tratamiento puede estar sesgada. La aleatorización en el estudio 1 superaría dicho riesgo de sesgo de selección.

### Sesgo de rendimiento

A menudo, las personas involucradas en un estudio, particularmente los terapeutas y los participantes, tienen un “caballito de batalla”, es decir, la creencia de que uno u otro tratamiento es mejor y, por lo tanto, un deseo de que los resultados del estudio se den de una manera determinada. Por ejemplo, un fisioterapeuta puede creer firmemente que el ejercicio es necesario para estos



<sup>1</sup> School of Public Health, University of Sydney, Camperdown, Australia; Centre for Pain, Health and Lifestyle, Australia. Copyright ©2018. Journal of Orthopaedic & Sports Physical Therapy®

pacientes y no debe ser reemplazado por un simple folleto informativo. Del mismo modo, un paciente podría sentirse decepcionado si recibe solo un folleto en lugar de un programa de ejercicios. Si este es el caso, el fisioterapeuta podría no realizar las 2 intervenciones con igual entusiasmo y confianza, y los participantes no leerían ni seguirían la información del folleto. El cegamiento del fisioterapeuta y del paciente superan el sesgo de rendimiento, pero el cegamiento a menudo es difícil o imposible en los ensayos de fisioterapia. El riesgo de sesgo de rendimiento puede reducirse garantizando que la intervención de control parezca tan creíble y valiosa como la intervención de estudio.

### Sesgo de detección

Las creencias que los fisioterapeutas y los pacientes tienen sobre los tratamientos también pueden dar lugar a sesgos a la hora de medir los resultados. Si el profesional está recolectando medidas de resultado y sabe qué tratamiento recibió un participante, entonces puede interpretar o calificar inconscientemente los resultados de una manera que favorezca al grupo

de intervención. Si un participante se siente desanimado por estar en el grupo de control, puede realizar una evaluación más negativa de sus síntomas, o expresar decepción con los puntajes en los resultados autoinformados. El sesgo de detección se supera cegando a los evaluadores de resultados. Sin embargo, esto es difícil para los resultados informados por el paciente, ya que, esencialmente, el paciente también es evaluador de resultados. Como en el caso del sesgo de rendimiento, los investigadores intentan reducir el riesgo de sesgo de detección asegurando que los tratamientos de intervención y control sean igualmente creíbles.

### Sesgo de deserción

No importa qué tan bien se realice un estudio, habrá participantes que abandonen y otros que no completen todas las medidas de seguimiento. La cuestión fundamental es que no sabemos qué pasó con las personas que no completaron todas las medidas de resultado. No podemos simplemente asumir que, en promedio, son iguales que las personas que se quedaron en el estudio. La probabilidad de que

los abandonos y los datos faltantes introduzcan sesgos en un estudio depende de una serie de factores. Estos incluyen la proporción de participantes que completaron el estudio, el equilibrio de los abandonos entre los grupos, la comparabilidad de los que abandonan y los que completan y la forma en que se realizaron los análisis estadísticos. Asegurar una alta tasa de seguimiento (superior al 85%) y la realización de análisis por intención de tratamiento son formas de reducir el riesgo de sesgo de deserción.

## RESUMEN

Todos los estudios tienen algún riesgo de sesgo. La idea de evaluar el riesgo de sesgo es proporcionar una base para dar más o menos peso a esa información: si un estudio tiene un mayor riesgo de sesgo, entonces deberíamos tener menos confianza en los hallazgos. Lo mismo ocurre con la información de la experiencia clínica. Cuando el riesgo de sesgo es bajo, creemos que la información está más cerca de “la verdad”; cuando el sesgo es alto, pensamos lo contrario. La práctica clínica implica recopilar información relevante de varias fuentes y reducirla para tomar una decisión. Al hacerlo, debe prestar más atención a la información que es menos probable que sea parcial. ●

## GO GREEN By Opting Out of the Print Journal

JOSPT subscribers and APTA members of the Orthopaedic and Sports Physical Therapy Sections can help the environment by “opting out” of receiving JOSPT in print each month as follows. If you are:

- A JOSPT subscriber: Email your request to [jospt@jospt.org](mailto:jospt@jospt.org) or call the JOSPT office toll-free at 1-877-766-3450 and provide your name and subscriber number.
- APTA Orthopaedic or Sports Section member: Go to <http://www.apta.org/>, log in, and select My Profile. Next click on Email Management/GoGreen. Toward the bottom of the list, you will find the Publications options and may opt out of receiving the print JOSPT. Please save this preference.



## Cegamiento: vinculación de la evidencia con la práctica

**C**iego simple, doble ciego, triple ciego: ¿qué hace que el cegamiento sea tan importante que se convierte en un sello de superioridad en un ensayo clínico? El cegamiento no es ni remotamente una parte de la práctica clínica, ya que no juega ningún papel en el tratamiento diario de los pacientes. Entonces, ¿por qué debería preocuparnos tanto cuando se trata de investigación?

El cegamiento (a veces llamado “enmascaramiento”) generalmente se menciona en el contexto de un ensayo clínico y significa que 1 o más de los grupos de personas involucradas en un ensayo no saben si los participantes del mismo recibieron el tratamiento A o el de control. Hablando en términos prácticos, esto significa que los pacientes en el ensayo no saben si están en el grupo de tratamiento o en el de control (cegamiento de los participantes), no saben los terapeutas que administran las intervenciones (cegamiento del terapeuta) y / o los investigadores que miden los resultados

(cegamiento del evaluador). Hay varios grupos de personas involucradas en el estudio (TABLA) que pueden cegarse, y no hay una terminología estándar para simple, doble o triple ciego; se debe buscar en la sección Métodos para saber quién está cegado. Los estudios que no ciegan a los participantes, terapeutas y evaluadores tienen un mayor riesgo de sesgo; son más propensos a informar un efecto que difiere de lo que se puede esperar en la práctica. Es por eso que todas las escalas de calidad de estudio, como la herramienta Cochrane de riesgo de

sesgo y la escala Physiotherapy Evidence Database (PEDro), incluyen elementos sobre el cegamiento.

### ¿Por qué es importante el cegamiento?

La falta de cegamiento puede introducir sesgos en los resultados del estudio de diferentes maneras, dependiendo de qué personas involucradas en el estudio no estén cegadas. Si los participantes saben que están en el grupo de tratamiento, en lugar del grupo de control, entonces pueden tener expectativas más positivas sobre cómo les irá, o informar mejores resultados para complacer a su terapeuta tratante. Por otro lado, si un paciente sabe que está en el grupo de control, podría tener expectativas más bajas o estar decepcionado por lo que no es el tratamiento real / mejor / más reciente e informar consecuentemente resultados más pobres.

Los fisioterapeutas aportan sus propias creencias sobre los tratamientos a los ensayos clínicos y, por lo tanto, pueden estar más entusiasmados y hacer un mejor trabajo aplicando un tratamiento en comparación con otro. También pueden transferir sus expectativas sobre los diferentes tratamientos a los pacientes en el estudio. Por ejemplo, considere un ensayo en el que un grupo recibe terapia manual y ejercicio, y el otro recibe ultrasonido descalibrado e información inespecífica. Imagine la dificultad de un fisioterapeuta para implementar estas 2 intervenciones con igual entusiasmo, confianza y convicción.

Debido a que prácticamente todas las medidas (incluidas las más objetivas)

TABLA	Quién puede ser cegado en una prueba y por qué
QUIEN	POR QUÉ
Participante	De esta manera las personas que reciben tratamiento participan en la misma medida, independientemente de si reciben tratamiento o control.
Terapeuta	De esta manera las personas que ofrecen el tratamiento lo hacen con el mismo esfuerzo y entusiasmo, independientemente de si están implementando tratamiento o control.
Evaluador	De esta manera los resultados se miden exactamente de la misma manera en los grupos de tratamiento y control
Estadístico	De esta manera las decisiones sobre cómo tratar con datos poco claros o faltantes y los métodos de análisis se toman solo en función del uso más eficiente de los datos disponibles
Escritor	De esta manera la interpretación de los resultados no está influenciada por sesgos personales y creencias del equipo de estudio

<sup>1</sup> School of Public Health, University of Sydney, Camperdown, Australia; Centre for Pain, Health and Lifestyle, Australia. Copyright ©2018. Journal of Orthopaedic & Sports Physical Therapy®

implican cierto grado de influencia de la persona que informa y / o la persona que realiza la evaluación, la evaluación de resultados no cegada puede introducir sesgo. Esto ocurre cuando la interpretación o puntuación del resultado se sesga hacia mejores resultados en el grupo de tratamiento que en el grupo de control, y ocurre porque los evaluadores tienen una creencia o deseo preexistente de demostrar que el tratamiento es más efectivo que el control. Otra cosa a tener en cuenta aquí es que cuando el resultado es reportado por el paciente (por ejemplo, cuestionario sobre dolor o calidad de vida), el evaluador a menudo solo se considera cegado cuando el participante está cegado

Es importante tener en cuenta que los sesgos mencionados anteriormente no son (necesariamente) debido a la deshonestidad de la gente involucrada en la prueba; ellos suceden a nivel subconsciente.

### Definición de cegamiento

El cegamiento se puede definir de maneras ligeramente diferentes, especialmente el cegamiento de los participantes. Según la definición más estricta, los pacientes sólo pueden ser cegados cuando no saben en qué grupo están (p. ej., tratamiento o control) y no son capaces de notar la diferencia entre las 2 intervenciones. Esta significa que los pacientes en el hipotético ensayo de

terapia manual versus ultrasonido descalibrado no pueden ser cegados.

El cegamiento se puede definir de manera menos estricta diciendo que los pacientes están cegados si, cuando se les pregunta, no saben si recibieron tratamiento o control. Entonces, en el caso del hipotético ensayo de arriba, se puede considerar a los participantes cegados. Esta segunda definición de cegamiento nos lleva a considerar el tema de la credibilidad del tratamiento. El argumento es que si los pacientes en ambos grupos piensan que están recibiendo una intervención real y ambas intervenciones son igualmente creíbles, entonces se puede considerar a los participantes cegados. Este punto de vista tiene mérito cuando pensamos en la fuente de sesgo que viene con los participantes no cegados. Si credibilidad y expectativas asociadas con las 2 intervenciones son más o menos iguales, entonces el riesgo de sesgo asociado con los pacientes no cegados es probablemente bastante pequeño. La credibilidad del tratamiento también puede influir sobre el riesgo de sesgo asociado con el terapeuta. Una prueba de tratamiento versus un placebo (por ejemplo, terapia manual versus ultrasonido descalibrado) tiene un mayor riesgo de sesgo debido a la falta de cegamiento del terapeuta que una prueba que compare 2 tratamientos actuales (por ejemplo, terapia manual versus ejercicio).

### Cegamiento en ensayos de tratamientos físicos

Puede resultar obvio que, en base a una definición estricta, el cegamiento de los participantes, terapeutas y evaluadores no sea posible para la mayoría de los ensayos en fisioterapia. Si bien es relativamente sencillo crear una píldora o inyección de placebo para cegar un paciente, e incluso un terapeuta, en ensayos de drogas, no lo es cuando hablamos de tratamientos manuales, educación, programas de ejercicio, etc. Esto ha llevado a algunos a sugerir que los elementos cegadores no deberían ser considerados cuando se evalúa la calidad de una prueba de rehabilitación. Pero ese es el punto. El objetivo de la evaluación de la calidad del estudio es determinar el riesgo de sesgo tal como se aplica a los resultados de un ensayo en particular y los riesgos asociados con la falta de cegamiento son reales, sin importar cuán difícil (o imposible) sea el cegamiento.

El cegamiento es un problema más complejo que lo que parece a simple vista y aún más en el contexto de los tipos de ensayos relevantes para los fisioterapeutas. Ensayos en el que los participantes, terapeutas y / o los evaluadores no están cegados están en riesgo de sesgo, pero ese no es el final de la historia. La magnitud del riesgo también depende de los tipos de intervenciones y la naturaleza de los resultados. ●

## SEND Letters to the Editor-in-Chief

JOSPT welcomes letters related to professional issues or articles published in the Journal. The Editor-in-Chief reviews and selects letters for publication based on the topic's relevance, importance, appropriateness, and timeliness. Letters should include a summary statement of any conflict of interest, including financial support related to the issue addressed. In addition, letters are copy edited, and the correspondent is not typically sent a version to approve. Letters to the Editor-in-Chief should be sent electronically to [jospt@jospt.org](mailto:jospt@jospt.org). Authors of the relevant manuscript are given the opportunity to respond to the content of the letter.

# Aleatorización: vincular evidencia con la práctica

Los kinesiólogos/as podrían pensar que los investigadores están obsesionados con los ensayos clínicos controlados aleatorizados (ECA). ¿Pero por qué es eso? ¿Hay algún tipo de magia en la aleatorización? La respuesta es sí, más o menos. La aleatorización nos permite sacar algunas conclusiones específicas que no son posibles en otros diseños de estudio.

## ¿Tratamiento A o Tratamiento B?

Supongamos que estamos tratando de decidir qué hacer con el paciente que se encuentra frente a nosotros y necesitamos saber si los resultados pueden ser mejores después del tratamiento A o el tratamiento B. Existen varios diseños de estudio diferentes que podrían brindarnos información acerca de esto, pero no son todos iguales. Es esencial, para evaluar estos estudios, saber cómo llegaron los pacientes a recibir el tratamiento A o el B. Y de las opciones disponibles, la aleatorización nos da la estimación más precisa sobre la efectividad del tratamiento.

Pero primero, un punto importante es prestar atención sobre la diferencia entre la aleatorización (o asignación aleatoria) y el muestreo aleatorio. La aleatorización es una forma de decidir qué individuos entran en el grupo A y cuáles en el grupo B. El muestreo aleatorio es algo completamente diferente y se refiere a un método para reclutar una muestra de una determinada población. Este artículo trata sólo sobre la aleatorización en grupos de tratamiento; el muestreo se abordará en otro artículo.

## Asignación al tratamiento

La aleatorización asegura que el juicio subjetivo del investigador, el kinesiólo-

go y el participante no desempeñe un papel en la decisión de qué intervención recibe este último. En la práctica, esto generalmente significa que el número de identificación de cada participante (el número asignado a los participantes cuando ingresan al estudio) se asigna a un grupo u otro a través de una lista aleatoria generada por computadora (llamada calendario de aleatorización) antes de que comience el estudio. El programa de aleatorización se utiliza para crear una caja llena de sobres opacos sellados, con números de identificación de estudio en el exterior y una hoja de papel con el grupo de tratamiento en el interior. Otros métodos válidos de aleatorización incluyen lanzar una moneda, usar dados o sacar cartas de una baraja.

**Asignación oculta** De forma ideal, un estudio tendrá una “asignación oculta”, lo que significa que la persona que decide si un individuo está incluido en el estudio no ha visto el calendario de aleatorización. El siguiente sobre no se abre hasta que una nueva persona se incluya realmente en el estudio. Si el investigador sabe a qué grupo se asignará la siguiente persona, esto podría influir en la decisión

de incluir o excluir a ciertas personas e introducir sesgos en el estudio. Se ha demostrado que influye en el tamaño del efecto del tratamiento en ese estudio saber si la asignación está oculta o no; los estudios sin asignación oculta en general informan un efecto mayor para el grupo de intervención.

Algunos métodos tienen la apariencia de aleatorización pero son menos destacados y los estudios que usan estos métodos se denominan ensayos cuasialeatorios. Dichos métodos incluyen asignación alternativa o asignación en función de la fecha de nacimiento, día de presentación o número de registro médico. Por lo general, se considera que los ensayos cuasialeatorios tienen un mayor riesgo de sesgo que los ECA, porque a menudo es difícil ocultar la asignación.

## Factores de confusión

Existen varias formas en que los pacientes en estudios no aleatorios pueden terminar recibiendo un tratamiento u otro. El kinesiólogo puede decidir, el paciente puede decidir, el tratamiento puede ser recetado por una institución, o el paciente puede recibir una intervención de acuerdo con un protocolo de investigación. En todos los casos, los resultados pueden medirse y compararse entre grupos. Pero para concluir que las diferencias en los resultados se deben al tratamiento real, debemos realmente ser capaces de afirmar que los pacientes que recibieron el tratamiento A tenían las mismas características que los

<sup>1</sup> School of Public Health, University of Sydney, Camperdown, Australia; Centre for Pain, Health and Lifestyle, Australia. Copyright ©2018. Journal of Orthopaedic & Sports Physical Therapy®



que recibieron el tratamiento B. Los investigadores usan el término factores de confusión para referirse a características que pueden ser diferentes entre los grupos e influir en el resultado, independientemente del tratamiento que reciba el paciente. Por ejemplo, sabemos que, en promedio, las personas que han tenido dolor durante períodos largos en general no lo hacen tan bien como los que tuvieron dolor durante un período más corto. Entonces, si las personas en el grupo A tienen una duración del dolor más larga en promedio que las personas en el grupo B al inicio del tratamiento, no sabremos si la diferencia en el resultado del estudio se debe a la du-

ración del dolor o a la mayor efectividad del tratamiento B por sobre el tratamiento A.

Conocemos algunos factores de confusión, y podemos medirlos; como la duración del dolor, en estos casos es bastante fácil verificar que son iguales entre los grupos. Pero también hay factores de confusión que conocemos pero que no se miden en el estudio y, lo que es peor, hay factores de confusión que ni siquiera conocemos. Este es un problema importante para los estudios comparativos que no utilizan la aleatorización. Aquí también es donde ocurre la magia (de la aleatorización). La aleatorización hace que sea probable que los grupos estén

equilibrados en factores de confusión que conocemos y hemos medido, en factores de confusión que conocemos y no hemos medido, y también en factores de confusión que ni siquiera sabemos que están. En el discurso de los investigadores, se dice que un ECA nos da una estimación imparcial del efecto de la asignación a un tratamiento u otro. Los estadísticos también han demostrado que esto es así a través de una prueba matemática, ¡pero es mejor dejar el tema para los nerds!

Esto no quiere decir que cualquier estudio que se aleatorice automáticamente nos dé la respuesta correcta a la pregunta; hay otros aspectos de la calidad del estudio que deben considerarse. Uno de ellos es el tamaño de la muestra: las ventajas de la aleatorización solo son verdaderas si hay suficientes personas en el estudio. Sin embargo, probablemente no sea demasiado afirmar que la aleatorización es la característica de diseño de estudio más importante para los estudios de comparación de tratamientos. ●

*“La aleatorización hace que sea probable que los grupos estén equilibrados en factores de confusión que conocemos y hemos medido, en factores de confusión que conocemos y no hemos medido, y también en factores de confusión que ni siquiera conocemos”.*

## PUBLISH Your Manuscript in a Journal With International Reach

JOSPT offers authors of accepted papers an international audience. The Journal is currently distributed to the members of APTA's Orthopaedic and Sports Physical Therapy Sections and 32 orthopaedics, manual therapy, and sports groups in 24 countries who provide online access either as a member benefit or at a discount. As a result, the Journal is now distributed monthly to more than 37,000 individuals around the world who specialize in musculoskeletal and sports-related rehabilitation, health, and wellness. In addition, JOSPT reaches students and faculty, physical therapists and physicians at more than 1,250 institutions in 60 countries.

Please review our Information for and Instructions to Authors at [www.jospt.org](http://www.jospt.org) in the Info Center for Authors and submit your manuscript for peer review at <http://mc.manuscriptcentral.com/jospt>. distributed monthly to more than 37,000 individuals around the world who specialize in musculoskeletal and sports-related rehabilitation, health, and wellness. In addition, JOSPT reaches students and faculty, physical therapists and physicians at more than 1,250 institutions in 60 countries.

Please review our Information for and Instructions to Authors at [www.jospt.org](http://www.jospt.org) in the Info Center for Authors and submit your manuscript for peer review at <http://mc.manuscriptcentral.com/jospt>. Relevant manuscripts are given the opportunity to respond to the content of the letter.

# Confiabilidad y validez: vincular la evidencia con la práctica

**E**l artículo anterior de Evidence in Practice introdujo la idea del “constructo”, o lo que interesa medir, por ejemplo, dolor, incapacidad o fuerza. Como a menudo hay numerosas medidas disponibles para cualquier constructo dado, ¿cómo elegimos cuál usar? Del número de consideraciones que se incluyen en esto, ninguna es más importante que la confiabilidad y la validez. No es exagerado decir que si una medida no es lo suficientemente confiable y válida, entonces no es adecuada para su propósito.

## Confiabilidad

Formalmente, la confiabilidad es el grado en que una medición está libre de errores. En la práctica, una medida confiable es aquella que le da la misma respuesta cuando se mide el mismo constructo varias veces. Considere el ejemplo de medir la altura (el constructo) con una cinta métrica (la medida). Puede medir la altura de una persona en milímetros 3 veces con la cinta métrica; el grado en que el número de milímetros es el mismo en cada ocasión es la confiabilidad de la medida.

Las implicaciones de las medidas poco confiables son graves. Si se aplicó una prueba diagnóstica poco confiable (medida) a un paciente varias veces, entonces el mismo paciente podría ser diagnosticado con o sin la condición en diferentes ocasiones o por diferentes personas. Si se obtuviera una medida poco confiable de la gravedad de los síntomas de un paciente antes y después de un tratamiento, sería imposible saber si ese síntoma mejoró, permaneció igual o empeoró. Esencialmente, los datos recopilados de medidas no confiables no proporcionan información útil; una medida que no es

confiable no puede ser válida.

Existen varios tipos diferentes de confiabilidad, cada uno de los cuales es especialmente relevante para las situaciones en las que se pueden usar las medidas. La confiabilidad intraevaluador se refiere a la situación en la que el mismo evaluador toma la medida en un paciente en varias ocasiones, y la confiabilidad es el grado en que los puntajes de las medidas sucesivas son los mismos. La confiabilidad interevaluador es relevante cuando varios evaluadores usan la misma medida en una sola persona, y la confiabilidad es el grado en que los puntajes de los diferentes evaluadores son iguales.

## Validez

La validez es el grado en que el puntaje de una medida realmente refleja el constructo que se supone que debe medir. Esto es relativamente sencillo cuando se trata de cosas como altura o fuerza, pero las aguas se vuelven rápidamente turbias cuando consideramos constructos no observables o “latentes” como el dolor, la calidad de vida o la discapacidad. Para este tipo de constructos, recopilamos medidas indirectas, como las experiencias y comportamientos auto reportados, o

el recuerdo de creencias y emociones, y suponemos que reflejan el constructo. Por ejemplo, podríamos pedirle a un paciente que responda las 24 preguntas del Roland-Morris Disability Questionnaire (la medida) y califique su nivel de discapacidad relacionado con el dolor de espalda (constructo) sumando el número de “sí” en sus respuestas. La puntuación del paciente de las 24 respuestas es válida en la medida en que las preguntas realmente reflejen el constructo de discapacidad y en la medida en que tener dificultades con más ítems refleja una mayor discapacidad.

Existen varios tipos diferentes de validez relevantes para las medidas clínicas, la más comúnmente evaluada es la validez de constructo. Cuando los investigadores evalúan la validez de constructo de una medida, son idealmente capaces de comparar su medida con un “estándar de oro”. Por ejemplo, la visualización artroscópica del ligamento cruzado anterior se considera un estándar de oro de la ruptura del ligamento cruzado anterior, por lo que un estudio podría comparar los resultados de la prueba de Lachman con los resultados de la artroscopía para evaluar la validez de la prueba de Lachman.

Desafortunadamente, no hay estándares de oro para muchos constructos en los que estamos interesados (por ejemplo, constructos latentes como discapacidad y dolor). En estos casos, la validez de constructo se prueba con un “estándar de referencia”, que es una

<sup>1</sup> School of Public Health, University of Sydney, Camperdown, Australia; Centre for Pain, Health and Lifestyle, Australia.  
Copyright ©2018. Journal of Orthopaedic & Sports Physical Therapy®

especie de estándar de oro imperfecto. Cuando no hay un estándar de oro, la mejor manera de probar la validez es a través de la prueba de hipótesis. Esto implica establecer una serie de hipótesis antes de recopilar los datos. Estas hipótesis son relaciones teóricas entre una puntuación en la medida y otras características, por ejemplo, que el puntaje estará fuertemente correlacionado con los puntajes en otra medida del mismo constructo y menos fuertemente correlacionado con puntajes en un constructo diferente pero relacionado. El grado en que los datos son consistentes con las hipótesis predeterminadas será la evidencia que respalde la validez de la medida.

### Estadísticas

La confiabilidad y validez de las pruebas generalmente implican evaluar el acuerdo entre 2 puntajes, ya sean puntajes en la misma medida recopilados dos veces (confiabilidad) o puntajes

en diferentes medidas (validez). Las estadísticas utilizadas para describir el acuerdo dependen de si las medidas son dicotómicas (por ej. kappa, sensibilidad / especificidad) o continuas (por ej., coeficientes de correlación intraclase, correlaciones, límites de acuerdo, R2).

### Conclusión

Hay un par de puntos generales importantes a tener en cuenta sobre la confiabilidad y la validez. Primero, ambos están en un espectro, por lo que las medidas no son “poco confiables” o “confiables”, sino más o menos confiables y más o menos válidas. Por supuesto, esto dificulta la elección de medidas; necesitamos hacer un juicio subjetivo sobre si una medida es “suficientemente confiable” y “suficientemente válida” en una situación particular. Existen pautas generales para ayudar a interpretar las estadísticas de confiabilidad y validez, pero estas

pautas no reemplazan ni deben reemplazar el juicio clínico. El segundo punto es que ninguna medida se encuentra en el final del espectro; no existe una medida perfectamente confiable y perfectamente válida. Incluso al medir la altura con una cinta métrica es probable que las medidas sucesivas difieran en unos pocos milímetros para arriba o abajo. Finalmente, existen preocupaciones prácticas a la hora de elegir una medida, incluido el tiempo que lleva administrar, si el paciente puede comprender el texto o instrucciones y cómo se almacenarán y utilizarán los datos.

La medición es un campo completo de investigación en sí mismo. Aunque los conceptos generales son bastante sencillos, no hace falta ahondar mucho por debajo de la superficie antes de que las cosas se compliquen. Al leer una investigación, se debe buscar información que le asegure que las medidas utilizadas son lo suficientemente confiables y válidas. El mensaje para llevar a casa: tenga mucho cuidado al usar, o al intentar interpretar la información de una medida si no tiene información sobre su confiabilidad y validez. ●

*“Esencialmente, los datos recopilados de medidas poco confiables no proporcionan información útil”.*

## BROWSE Collections of Articles on JOSPT's Website

JOSPT's website ([www.jospt.org](http://www.jospt.org)) offers readers the opportunity to browse published articles by Previous Issues with accompanying volume and issue numbers, date of publication, and page range; the table of contents of the Upcoming Issue; a list of available accepted Ahead of Print articles; and a listing of Categories and their associated article collections by type of article (Research Report, Case Report, etc). Features further curates 3 primary JOSPT article collections: Musculoskeletal Imaging, Clinical Practice Guidelines, and Perspectives for Patients, and provides a directory of Special Reports published by JOSPT.



# Fundamentos de medición: vinculación de la evidencia con la práctica

La medición es fundamental para la ciencia, lo que significa que los fisioterapeutas deben comprometerse con la medición si desean que su profesión se considere científica. Sin embargo, aunque es sencillo estar de acuerdo en que la medición es una buena idea, hay más en la medición de lo que parece.

Desde el punto de vista del investigador, la razón para medir factores demográficos, personales y clínicos es evidente: los puntajes de las mediciones responden la pregunta de la investigación. En la clínica, los fisioterapeutas recopilan mediciones (a menudo de manera informal) desde el comienzo del encuentro clínico, ya sea haciendo una pregunta como, “¿Qué te molesta hoy?” u observando la forma en que el paciente se levanta de una silla y camina. Esta información guía nuevas evaluaciones y gestión. Los profesionales de la salud y los investigadores enfrentan problemas similares a la hora de interpretar las medidas que recopilan. Para hacer esto con precisión, es necesario comprender algunos conceptos básicos sobre medición.

## Constructos y medidas

Un “constructo” es lo que interesa medir. Una medida (a veces llamada herramienta o instrumento) es cómo se mide el constructo. Por ejemplo, se puede estar interesado en el constructo de “incapacidad” en un paciente con dolor en el hombro, y optar por utilizar el cuestionario Disabilities of the Arm, Shoulder and Hand como medida. Puede haber varias medidas diferentes para el mismo constructo; por ejemplo, el Shoulder Pain and Disability Index y el Shoulder Disability Questionnaire también son

medidas de incapacidad para personas con dolor de hombro (FIGURA).

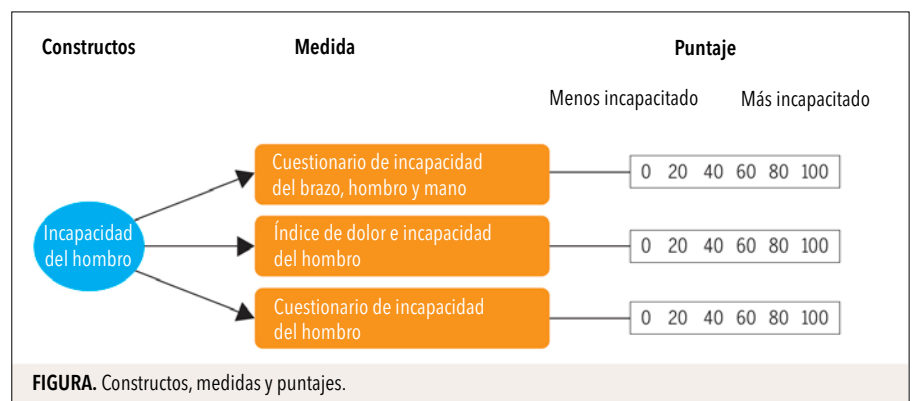
## Propósito

Existen muchas razones para medir algo y gran cantidad de medidas disponibles. Las medidas de *detección* están diseñadas para estimar la probabilidad de que una persona sana tenga una cierta condición en el futuro y si se deben realizar más investigaciones (por ejemplo, detección de riesgo cardiovascular). Las medidas de *diagnóstico* (pruebas) están diseñadas para determinar si alguien tiene o no una determinada afección, un ejemplo es la prueba de Lachman para la ruptura del ligamento cruzado anterior. Las herramientas de *pronóstico* están diseñadas para ayudar a predecir si

un paciente se recuperará o cuándo, como el Örebro Musculoskeletal Pain Questionnaire. Las herramientas de *clasificación basadas en el tratamiento* están diseñadas para dirigir a un paciente hacia cierto tipo de tratamiento, como la herramienta STarT Back. Las medidas de *resultado* están diseñadas para rastrear el nivel o la presencia de un síntoma, función o marcador de enfermedad, por ejemplo, la Patient-Specific Functional Scale. Algunas medidas pueden tener varios propósitos; por ejemplo, la intensidad del dolor medida en una escala de calificación numérica puede formar parte de una prueba de diagnóstico, informar el pronóstico probable y ser rastreada con el tiempo como una medida de resultado.

## Medidas subjetivas y objetivas

Un error común es llamar a las medidas calificadas por el paciente como “subjetivas” y aquellas calificadas por un observador “objetivas”. La *subje-*



<sup>1</sup> School of Public Health, University of Sydney, Camperdown, Australia; Centre for Pain, Health and Lifestyle, Australia. Copyright ©2018. Journal of Orthopaedic & Sports Physical Therapy®

*tividad* se refiere al alcance del juicio personal involucrado en la toma de una medida, y el juicio personal podría ser en nombre del paciente o del observador; por ejemplo, la calificación de un fisioterapeuta sobre la cantidad de hinchazón también es subjetiva. Además, es mejor considerar las medidas como más o menos subjetivas, a lo largo de un continuo hipotético, en lugar de totalmente subjetivas u objetivas. Otro error es considerar automáticamente que las medidas más objetivas son superiores a las más subjetivas, basándose en el supuesto de que las medidas objetivas son más confiables y válidas. La investigación ha demostrado que la fiabilidad y la validez no dependen de cuán objetiva o subjetiva sea una medida. Finalmente, es importante reconocer que algunos constructos son inherentemente subjetivos, como la experiencia del dolor de una persona, y en tales casos una medida más subjetiva será la más válida.

### Métodos de medición

Hay muchas formas de recopilar medidas y algunas de las más comunes se enumeran a continuación:

- *Medidas informadas por el paciente.* Como su nombre lo indica, estas son calificaciones proporcionadas por el paciente. Esto puede implicar completar un cuestionario escrito o responder preguntas verbalmente. Las medidas informadas por los pacientes se usan comúnmente para

calificar la gravedad de los síntomas, los impactos de una afección o para medir constructos psicológicos como calidad de vida, depresión y autoeficacia.

- *Medidas calificadas por el observador.* Estas medidas implican observaciones realizadas por el fisioterapeuta. Pueden incluir medidas de capacidad física como la fuerza o el rango de movimiento, la calidad del movimiento o la capacidad de realizar tareas particulares. También pueden incluir observaciones como la existencia de escoliosis, niveles de activación muscular, características de la marcha y resultados de pruebas clínicas. Tenga en cuenta que muchas de estas medidas evalúan no sólo el rendimiento físico, sino también la motivación del paciente. Estas medidas también implican un juicio subjetivo del fisioterapeuta.
- *Escaneos, imágenes, pruebas y dispositivos de monitoreo.* Estos pueden usarse para la detección, el diagnóstico o la medición de constructos como la actividad física habitual. A pesar de que este tipo de medidas generalmente se dirigen hacia el extremo más objetivo del espectro, su adecuación para el propósito, la confiabilidad y la validez no se deben suponer de manera automática. Por ejemplo, la interpretación de escaneos e imágenes es altamente subjetiva y los hallazgos pueden no

tener consecuencias funcionales importantes.

- *Datos administrativos.* Se usan con gran frecuencia en la investigación e incluyen métricas como asistencia hospitalaria, ausencia laboral, datos de reclamos de seguro y fallecimiento.

### Relevancia para el paciente

Otro aspecto de las medidas de interpretación es determinar la relevancia para el paciente. Esto implica juzgar qué tan importante es el constructo del resultado para su paciente. Por ejemplo, un estudio podría concluir que una intervención es efectiva porque mejora la fuerza, pero la fuerza en sí misma puede no ser tan importante para un paciente a menos que se traduzca en la capacidad de realizar tareas importantes.

### Conclusión

Las opciones de medición están limitadas sólo por la imaginación, pero el uso de medidas escasas o inadecuadas tiene consecuencias críticas tanto para la práctica clínica como para la investigación. Al leer una investigación, es importante verificar que el constructo coincida con la pregunta del estudio. Desde una perspectiva clínica, es clave poder articular claramente por qué se desea medir un determinado constructo. Una vez que haya decidido el constructo adecuado, debe seleccionar la mejor medida. Esto implica la consideración de fiabilidad y validez, que será el tema del próximo artículo de Evidencia en la Práctica. ●

## BROWSE Collections of Articles on JOSPT's Website

JOSPT's website ([www.jospt.org](http://www.jospt.org)) offers readers the opportunity to browse published articles by Previous Issues with accompanying volume and issue numbers, date of publication, and page range; the table of contents of the Upcoming Issue; a list of available accepted Ahead of Print articles; and a listing of Categories and their associated article collections by type of article (Research Report, Case Report, etc). Features further curates 3 primary JOSPT article collections: Musculoskeletal Imaging, Clinical Practice Guidelines, and Perspectives for Patients, and provides a directory of Special Reports published by JOSPT.

## Grupos de control: vinculación de la evidencia con la práctica

**A** pesar de que un ensayo clínico controlado aleatorizado (ECA), por definición, no puede existir sin un grupo de control, éste es el hijastro olvidado del ECA. La mayoría de las veces, lo que contiene ni siquiera aparece en el título del estudio. Sin embargo, no es posible interpretar los resultados de un ECA a menos que realmente

comprendamos lo que le está sucediendo a las personas en el grupo de control.

El lugar del ECA, cerca de la parte superior de la jerarquía de evidencia de efectividad del tratamiento, en cierta medida se basa en la presunción de que sólo se puede conocer el beneficio de una intervención al compararla con otra cosa. El grupo de control en un ECA es el “algo más”. Entonces, en el contexto de un ECA, la efectividad de una intervención es la diferencia entre lo que les sucede a las personas que reciben la intervención de interés (tratamiento A) y lo que les sucede a las personas que reciben el tratamiento de control. Se deduce que la interpretación de la efectividad del tratamiento A, como se informa en cualquier ECA particular, también depende de la efectividad del tratamiento de control.

Obviamente, no todos los grupos de control son igualmente efectivos. El punto clave a entender es que diferentes intervenciones de control responden a diferentes preguntas de investigación. Por ejemplo, un ECA que compara un programa de ejercicios para el dolor de hombro con una ecografía simulada responde una pregunta diferente de la de un ECA que compara el mismo programa de ejercicios con un tratamiento de movilización articular.

Aunque hay un número casi ilimitado de intervenciones de control diferentes, po-

demos agruparlas en tipos similares y considerar los diferentes tipos de preguntas de investigación que van con cada tipo.

### Sin tratamiento o lista de espera

A veces, las personas que reciben el tratamiento A se comparan con un grupo que no recibe nada o está en una lista de espera. Se recopilan las mismas medidas de resultado, en los mismos puntos de tiempo, pero los controles no reciben tratamiento. El objetivo de este estudio es comparar el efecto del tratamiento A con el curso natural de la afección. Responder esta pregunta puede ser especialmente relevante cuando la condición de salud es aguda o se caracteriza por episodios recurrentes que van y vienen con el tiempo. Esto puede informar las decisiones sobre si los beneficios del tratamiento A tienen valor en relación con los costos, riesgos e inconvenientes más allá de la recuperación natural.

### Cuidado usual

Aquí, el tratamiento A se compara con un grupo de control que recibe la atención que normalmente recibirían si no estuvieran involucrados en un estudio, a veces llamado el “curso clínico” de la condición. Es difícil interpretar estos estudios porque puede

ser que, al igual que en la vida real, no todos en el grupo de control reciban el mismo tratamiento. Esto significa que el ensayo compara los resultados del tratamiento A con los resultados de una variedad de intervenciones. Otro desafío implica la generalización, porque la atención habitual en un lugar puede ser muy diferente de la atención habitual en otro. Pero hay una recompensa por estos desafíos. Estos estudios responden a una pregunta muy pragmática y clínicamente relevante: ¿deberíamos dejar la situación actual de la atención médica (para esa afección) como está, o implementar un nuevo estándar de atención en un lugar en particular? Cuando se hace bien, estos estudios pueden tener poderosas implicaciones clínicas y políticas.

### Placebo o Control de Atención

Un placebo, o intervención simulada, puede diseñarse para un estudio en particular. La idea general es que las personas en el grupo de control piensan que están recibiendo un tratamiento real, pero en realidad no lo están. El placebo clásico en estudios médicos es la píldora de azúcar; en los ensayos de fisioterapia, podría tratarse de una electroterapia desactivada, una acupuntura falsa donde las agujas no penetran en la piel o una cinta que no sirve para ningún propósito estructural. Un tipo especial de placebo se llama control de atención, que a veces se usa en ensayos que evalúan intervenciones educativas o de asesoramiento. Aquí es donde el grupo de control recibe la misma cantidad de interacción con los profesionales de la salud que el grupo de tratamiento, pero la interacción no tiene un propósito terapéutico específico. Esto podría im-

<sup>1</sup> School of Public Health, University of Sydney, Camperdown, Australia; Centre for Pain, Health and Lifestyle, Australia. Copyright ©2018. Journal of Orthopaedic & Sports Physical Therapy®

plicar una discusión general sobre la condición sin asesoramiento específico, educación o información. Tenga en cuenta que los placebos constituyen un tema complejo y profundo, y no hay espacio para analizar todos los temas aquí. Los ensayos controlados con placebo están diseñados para aislar y medir el efecto de un mecanismo específico propuesto relacionado con el tratamiento A. Por ejemplo, considere un ECA que compare el ultrasonido terapéutico real con el ultrasonido terapéutico desactivado para mejorar la inflamación después de un esguince de tobillo. La diferencia entre grupos estima el efecto de la

transmisión real de las ondas sonoras al tejido lesionado. La intervención con placebo “controla” la información proporcionada por el fisioterapeuta, el roce de la cabeza del ultrasonido sobre la piel y los efectos inespecíficos del encuentro terapéutico.

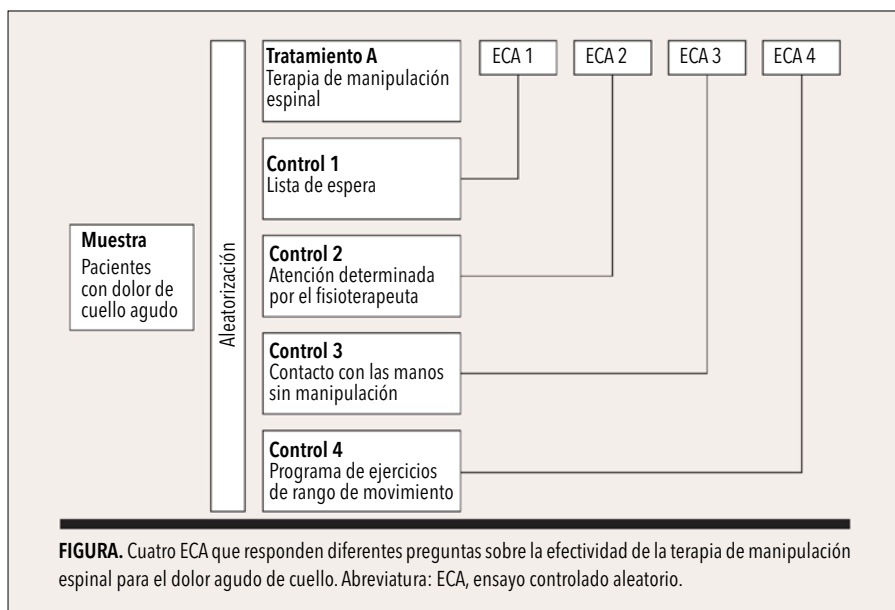
### Otras intervenciones

El tratamiento A también se puede comparar con otra intervención que se administre comúnmente para esa afección. En este caso, las personas en ambos grupos reciben un tratamiento real, por lo que el efecto es la diferencia en el resultado entre el grupo que recibió el tratamiento A y el

grupo que recibió el tratamiento B. Por ejemplo, las personas con dolor de espalda podrían ser aleatorizadas para recibir ya sea un ejercicio de estabilización espinal específico o un programa general de acondicionamiento aeróbico. Estos ensayos buscan responder la pregunta sobre si es mejor ofrecer a los pacientes un tratamiento u otro.

### Resumen

Casi por definición, el tratamiento A es el acto principal en cualquier ECA, pero el papel desempeñado por el grupo de control es de importancia crítica, por lo que el diseño de la intervención de control requiere una cuidadosa elaboración y planificación. La naturaleza del grupo de control determina en parte la pregunta de investigación que responde el ECA (FIGURA), y sólo entendiendo la pregunta precisa se pueden interpretar los resultados con sensatez. Dado que los ECA siempre miden la efectividad relativa del tratamiento y las intervenciones de control, es importante que la interpretación refleje esto. Eso significa que las conclusiones de que un tratamiento sea efectivo siempre deben indicar “en comparación con la intervención de control”. ●



**FIGURA.** Cuatro ECA que responden diferentes preguntas sobre la efectividad de la terapia de manipulación espinal para el dolor agudo de cuello. Abreviatura: ECA, ensayo controlado aleatorio.

## FIND Author Instructions & Tools on the Journal's Website

JOSPT's instructions to authors are available at [www.jospt.org](http://www.jospt.org) by clicking Complete Author Instructions in the right-hand Author Center widget on the home page, or by visiting the Info Center for Authors, located in the site's top navigation bar. The Journal's editors have assembled a list of useful tools and links for authors as well as reviewers.



# Intervalos de confianza: vinculación de la evidencia con la práctica

**D**os artículos anteriores de Evidence in Practice describieron el cambio en la investigación clínica hacia el uso de las diferencias entre grupos como medida de la efectividad del tratamiento.<sup>2,3</sup> Una ventaja clave para informar la diferencia entre grupos (la estimación del efecto), en lugar de solo proporcionar el valor *P* de

una prueba de hipótesis, es que informa al lector sobre el tamaño del efecto.

El efecto en un estudio se denomina “estimación” porque los datos se recopilan de una muestra de personas, no de todas las personas de la población. Pero lo que los lectores necesitan es una medida del efecto del tratamiento en toda la población: el concepto de generalización se tratará en el próximo artículo de Evidence in Practice. La estimación del efecto del tratamiento proporcionada por un estudio tiene algún error asociado, por lo que es importante saber algo sobre la precisión de la estimación. Un efecto estimado y su intervalo de confianza le dan al lector información importante sobre el tamaño, la extensión y la dirección del efecto del tratamiento de la población.

Los intervalos de confianza se pueden aplicar a las estimaciones de muchos tipos diferentes de preguntas de investigación, por ejemplo, sobre la efectividad del tratamiento, la prevalencia, los factores de riesgo o pronóstico, o la precisión de las pruebas de diagnóstico. Los mismos principios se aplican a los intervalos de confianza en cualquier estudio, pero para simplificar, este artículo se centra en los estudios de efectividad del tratamiento.

## ¿Qué son los intervalos de confianza?

Los intervalos de confianza abarcan un rango de valores por encima y por de-

bajo de la estimación del efecto. Las matemáticas detrás del cálculo de un intervalo de confianza están fuertemente influenciadas por el tamaño de la muestra y la variabilidad del efecto del tratamiento. Un estudio pequeño generalmente tendrá un intervalo de confianza más amplio en comparación con un estudio grande. Si el tratamiento es muy efectivo para algunas personas y el control es más efectivo para otras, entonces el intervalo de confianza será amplio.

Un intervalo de confianza del 95% le dice al lector lo siguiente: si los mismos tratamientos se compararon en 100 ensayos controlados aleatorios en la misma población, y los investigadores generaron 100 intervalos de confianza, la verdadera diferencia entre grupos caería dentro de 95 de esos intervalos de confianza. Cinco de los 100 intervalos de confianza no contendrían la verdadera diferencia entre grupos.

Debido a que los investigadores generalmente solo realizan un estudio sobre la misma población una vez, el lector no puede estar completamente seguro de que el intervalo de confianza contenga el verdadero efecto, por lo tanto, 95% de confianza. Los investigadores generalmente informan el intervalo de confianza del 95%, pero se pueden calcular otros intervalos.

Un intervalo de confianza del 90% sería más estrecho (más preciso), pero el lector tendría menos confianza de que contuviera el efecto verdadero. Un intervalo de confianza del 99% sería más amplio (menos preciso), pero el lector estaría más seguro de que el intervalo contiene la estimación real de la población.

Los intervalos de confianza son comúnmente mal interpretados. Un intervalo de confianza no es el rango de efectos en el que caerá el 95% de los pacientes de la población. Tampoco es estrictamente cierto decir que hay un 95% de posibilidades de que el intervalo de confianza incluya el verdadero efecto de la población.

El intervalo de confianza es el rango de efectos que probablemente contendrá el verdadero efecto medio del tratamiento, en comparación con el control. El lector puede confiar en el tamaño del efecto medio del tratamiento cuando el intervalo de confianza es estrecho, pero no está seguro cuando el intervalo de confianza es amplio. El verdadero efecto del tratamiento medio de la población es más probable que esté cerca de la estimación del efecto (en el medio del intervalo de confianza) en lugar de cerca de cualquier extremo del rango.

## Uso de intervalos de confianza en la práctica

El intervalo de confianza proporciona el rango de efectos plausibles que los fisioterapeutas y los pacientes pueden esperar de un tratamiento. Esta información debe formar una parte importante de la conversación cuando se llega a una decisión de tratamiento compartida.

<sup>1</sup> School of Public Health, University of Sydney, Camperdown, Australia; Centre for Pain, Health and Lifestyle, Australia. Copyright ©2018. Journal of Orthopaedic & Sports Physical Therapy®

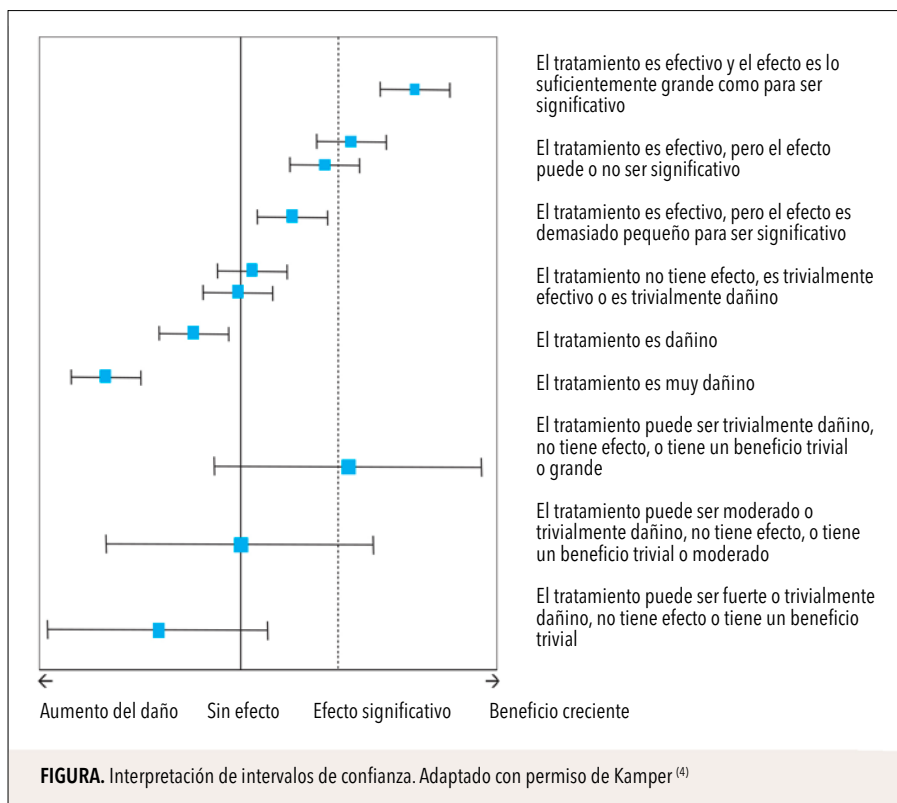
El concepto de efectos clínicamente significativos, descrito en 2 artículos anteriores de Evidence in Practice <sup>2,3</sup> también es relevante para interpretar los intervalos de confianza. Si el extremo inferior del intervalo de confianza es inferior al efecto clínicamente significativo, es posible que el tratamiento no tenga un beneficio valioso, independientemente del tamaño de la estimación del efecto. Cuando un intervalo de confianza cruza la línea sin efecto (es decir, contiene el valor de cero al evaluar las diferencias entre grupos), es posible que el tratamiento no sea más efectivo que el control (FIGURA).

Un estudio con un amplio intervalo de confianza no proporciona información útil sobre la efectividad de un tratamiento (FIGURA). En parte, esto explica por qué los estudios pequeños no son muy útiles cuando se toman decisiones sobre el tratamiento, y por qué los metanálisis bien realizados son valiosos. La combinación adecuada de los resultados de varios estudios en un solo análisis (es decir, metanálisis) reduce el intervalo de confianza y proporciona una estimación más precisa del efecto del tratamiento. Los investigadores en el campo de la fisioterapia están informando estimaciones de efectos con intervalos

de confianza más regularmente <sup>1</sup> ayudando a los lectores a interpretar con precisión la evidencia.

## Conclusión

Informar los intervalos de confianza es parte de un cambio de juzgar la efectividad del tratamiento únicamente por los valores de *P* a estimar el tamaño de un efecto. Los intervalos de confianza le dan al lector información crítica sobre la precisión de una estimación del efecto informada en un ensayo. La integración de la información sobre el efecto probable y su precisión, junto con la comprensión del concepto de significado clínico, ayudan al fisioterapeuta a involucrar a los pacientes en un proceso de toma de decisiones compartido e informado. ●



## REFERENCIAS

1. Freire APCF, Elkins MR, Ramos EMC, Moseley AM. Use of 95% confidence intervals in the reporting of between-group differences in randomized controlled trials: analysis of a representative sample of 200 physical therapy trials. *Braz J Phys Ther.* 2019;23:302-310. <https://doi.org/10.1016/j.bjpt.2018.10.004>
2. Kamper SJ. Interpreting outcomes 2—statistical significance and clinical meaningfulness: linking evidence to practice. *J Orthop Sports Phys Ther.* 2019;49:559-560. <https://doi.org/10.2519/jospt.2019.0704>
3. Kamper SJ. Interpreting outcomes 3—clinical meaningfulness: linking evidence to practice. *J Orthop Sports Phys Ther.* 2019;49:677-678. <https://doi.org/10.2519/jospt.2019.0705>
4. Kamper SJ. Showing confidence (intervals). *Braz J Phys Ther.* 2019;23:277-278. <https://doi.org/10.1016/j.bjpt.2019.01.003>

## NOTIFY JOSPT of Changes in Address

Please remember to let JOSPT know about changes in your mailing address. The US Postal Service typically will not forward second-class periodical mail. Journals are destroyed, and the USPS charges JOSPT for sending them to the wrong address.

You may change your address online at [www.jospt.org](http://www.jospt.org). Visit Info Center for Readers, click Change of Address, and complete the online form. We appreciate your assistance in keeping JOSPT's mailing list up to date.

# Interpretación de los resultados 1— Cambio y diferencia: vinculación de la evidencia con la práctica

La escritura científica es por naturaleza técnica, por lo que es importante contar con la terminología y redacción correctas. Cuando hay inconsistencia en las definiciones o mal uso de las palabras, surgen problemas. La inconsistencia puede ocurrir cuando no hay una definición universalmente aceptada para un concepto o simplemente debido a un descuido por parte del autor.

Pero esto es más que un simple problema de semántica; cuando se usan las mismas palabras para describir diferentes conceptos, los resultados del estudio pueden malinterpretarse fácilmente y sin saberlo.

La inconsistencia es común cuando se describen resultados. Los resultados, ya sean evaluados al inicio del estudio, durante o después del tratamiento, o en el seguimiento, se pueden medir de varias maneras<sup>(2)</sup>. Pero existe una división crítica que siempre es relevante, y la interpretación de los hallazgos de un estudio no es posible a menos que el lector entienda de qué lado de una cerca metafórica están los números. Los 2 lados de esta cerca pueden llamarse cambio y diferencia.

La distinción entre cambio y diferencia es crítica con respecto a la interpretación de los resultados de los estudios de efectividad del tratamiento. Los 2 tipos de hallazgos le dan al lector diferentes tipos de información.

## Cambio

En pocas palabras, el cambio es la puntuación en una medida de resultado (por ejemplo, en el seguimiento) menos la puntuación en la misma medida en un punto de tiempo anterior (por ejemplo,

al inicio del estudio). Esto es un cambio en una persona, o un cambio medio dentro de un grupo de personas a lo largo del tiempo. El problema surge cuando el cambio medio dentro del grupo se denomina “efecto del tratamiento” o “respuesta al tratamiento”. Como se mencionó en un artículo anterior de Evidencia en la práctica<sup>1</sup>, los cambios desde el inicio del estudio (baseline) hasta el “seguimiento” incluyen cambios que se deben a la historia natural de la condición, regresión a la media, efectos inespecíficos y efectos del tratamiento. Este no es el lugar para ahondar sobre las complejidades de todo esto, pero es suficiente decir que todos son relevantes, independientemente de la condición y el tratamiento; es decir, esto se aplica a todos los estudios! El cambio dentro del grupo a lo largo del tiempo no es lo mismo que el efecto del tratamiento. A menudo, una persona con un gran cambio en el resultado se denomina “respondedor” al tratamiento A, pero este lenguaje también es engañoso, porque es muy probable que la misma persona también responda al tratamiento B. Esto es algo básico, pero es extremadamente tentador atribuir todos los cambios observados al tratamiento y términos como *efecto del tratamiento*, *respuesta*

*al tratamiento* y *respondedor*, alimentan la tentación. Esto no quiere decir que los puntajes de cambio no brinden información útil. Son una estimación de lo que es probable que suceda cuando un paciente recibe el tratamiento del estudio, pero son diferentes del efecto del tratamiento y no son la respuesta al tratamiento.

## Diferencia

La diferencia requiere datos de 2 grupos de personas. La diferencia entre grupos es la puntuación media en una medida de resultado en el grupo de tratamiento A menos la puntuación media en el grupo B. Por lo general, es la diferencia entre las puntuaciones en el seguimiento o la diferencia en el cambio entre los 2 grupos. La diferencia puede llamarse *efecto del tratamiento* o *respuesta al tratamiento*, porque (suponiendo que el estudio esté bien diseñado) el tamaño de la diferencia no incluye la historia natural, la regresión a la media y los efectos inespecíficos. De manera crítica, el “efecto del tratamiento” es un efecto comparativo. Es lo que se puede esperar si un paciente recibió el tratamiento A en comparación con lo que se puede esperar si ese paciente recibió el tratamiento B. Por lo tanto, la “efectividad” del (o la respuesta al) tratamiento A, como se informó en un estudio en particular, se interpreta a la luz de lo que el tratamiento B involucraba. La diferencia en las puntuaciones de resultado entre los grupos cuantifica el efecto del tratamiento (FIGURA).

<sup>1</sup> School of Public Health, University of Sydney, Camperdown, Australia; Centre for Pain, Health and Lifestyle, Australia.  
Copyright ©2018. Journal of Orthopaedic & Sports Physical Therapy®

Las cosas se vuelven difíciles aquí para quienes leen una investigación porque los autores a veces informan conclusiones basadas en cambios dentro del grupo en ensayos controlados aleatorios (ECA). Esto generalmente ocurre cuando no hay diferencia entre los grupos. Por ejemplo, podrían concluir que el tratamiento A es efectivo, o que tanto el tratamiento A como el tratamiento B son efectivos, en función de las mejoras dentro del grupo. Afirmar que ambos tratamientos son efectivos casi nunca es una conclusión válida de un ECA, excepto en algunas circunstancias muy especiales. El

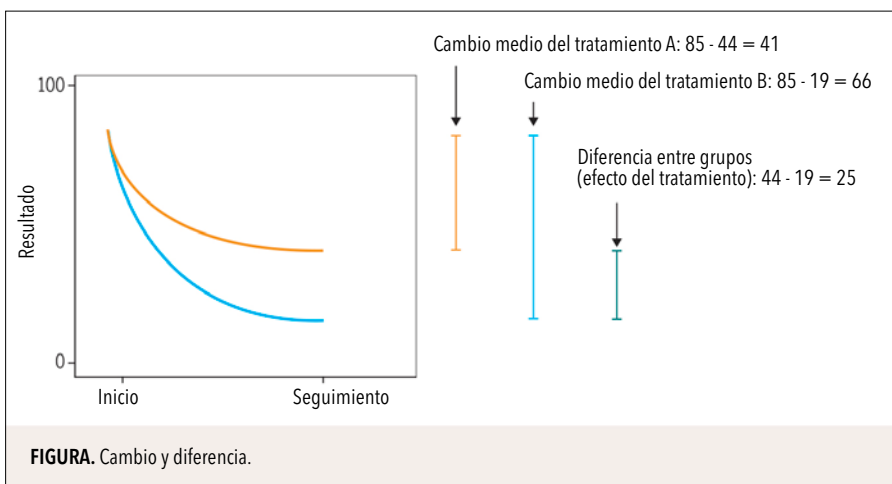
cambio dentro del grupo en un ECA no es más que el “efecto del tratamiento” de los resultados del estudio de un grupo único (no controlado): todavía incluyen historia natural, regresión a la media y efectos inespecíficos. Interpretar los hallazgos de un ECA de esta manera anula el propósito de la aleatorización y de tener un grupo de comparación.

### En suma

Al leer los resultados de un estudio, debe poder responder a la pregunta de si los autores están hablando sobre un cambio dentro del grupo o una dife-

rencia entre grupos. El primero incluye la recuperación natural, la regresión a la media, los efectos inespecíficos y los efectos del tratamiento. El último es el efecto del tratamiento.

La mayoría de las veces, la información necesaria para responder a la pregunta se encuentra en los métodos de estudio y requiere un conocimiento práctico de cómo se realizó el estudio y qué métodos y análisis diferentes nos pueden decir. Lo importante para alguien que lee un artículo no es tanto identificar si un autor está usando un lenguaje “correcto” sino determinar el verdadero significado detrás de las palabras utilizadas. ●



## REFERENCIAS

1. Kamper SJ. Engaging with research: linking evidence with practice. *J Orthop Sports Phys Ther.* 2018;48:512-513. <https://doi.org/10.2519/jospt.2018.0701>
2. Kamper SJ. Fundamentals of measurement: linking evidence to practice. *J Orthop Sports Phys Ther.* 2019;49:114-115. <https://doi.org/10.2519/jospt.2019.0701>

## EARN CEUs With JOSPT's Read for Credit Program

JOSPT's Read for Credit (RFC) program invites readers to study and analyze selected JOSPT articles and successfully complete online exams about them for continuing education credit. To participate in the program:

1. Go to [www.jospt.org](http://www.jospt.org) and click on Read for Credit in the top blue navigation bar that runs throughout the site.
2. Log in to read and study an article and to pay for the exam by credit card.

3. When ready, click Take Exam to answer the exam questions for that article. 4. Evaluate the RFC experience and receive a personalized certificate of continuing education credits. The RFC program offers you 2 opportunities to pass the exam. You may review all of your answers—including your answers to the questions you missed. You receive 0.2 CEUs, or 2 contact hours, for each exam passed. JOSPT's website maintains a history of the exams you have taken and the credits and certificates you have been awarded in My CEUs and Your Exam Activity, located in the right rail of the Read for Credit page listing available exams.



# Interpretación de los resultados 2— Importancia estadística y significado clínico: vinculación de la evidencia con la práctica

Los puntajes en las medidas de resultado son importantes en el mundo clínico porque influyen en las decisiones de tratamiento y porque quienes financian la salud piden a los kinesiólogos/as, cada vez más, que justifiquen sus decisiones de tratamiento. Para los investigadores, los puntajes en las medidas de resultado son

la moneda con la que comercian; los puntajes de resultados proporcionan las respuestas a las preguntas de investigación. Muchos constructos de resultados pueden ser de interés para los kinesiólogos, investigadores, pacientes y financiadores de la salud: dolor, función, discapacidad, actividad física, trabajo, depresión, ansiedad, calidad de vida, productividad laboral, etc. Cuando se habla de un resultado, generalmente consideramos que esto significa el nivel del constructo (p. ej., intensidad de dolor baja / media / alta o función física buena / mala). Utilizamos una puntuación en la medida de resultado para proporcionarnos esta información.<sup>1</sup> Los kinesiólogos necesitan interpretar las puntuaciones en las medidas de resultado, ya sea que las obtengan de sus pacientes en la práctica o de los participantes en los estudios para guiar la atención clínica.

La serie Evidencia en la Práctica ha abordado la importante distinción entre cambio y diferencia.<sup>2</sup> En la práctica clínica, la medición generalmente implica calcular el cambio en las puntuaciones de resultado desde el inicio (pretratamiento) hasta el final (postratamiento), e interpretar si el cambio es importante o significativo. En la investigación, un estudio podría informar un cambio medio en 1 grupo de personas a lo largo del tiempo, o la diferencia media entre 2 grupos.

Interpretar una puntuación de resultado implica hacer un juicio informado sobre lo que realmente significa un cambio o diferencia de ese tamaño. Esto implica preguntar si el cambio o la diferencia es clínicamente significativo, estadísticamente significativo o ambos (TABLA). Para simplificar, este artículo se referirá a las diferencias entre grupos a fin de ilustrar los conceptos.

## Importancia estadística

¿Qué es un valor  $P$ ? Un recuerdo duradero de las clases de estadísticas es que un valor  $P$  menor que 0.05 significa que algo importante ha sucedido. El valor  $P$  representa la probabilidad de que la verdadera diferencia entre las medias de 2 grupos sea tan grande

o mayor que la diferencia reportada en el estudio, si los grupos provienen de la misma población y suponiendo que los datos cumplen con varios supuestos estadísticos. El valor  $P$  no es la probabilidad de que los 2 grupos sean diferentes, ni la probabilidad de que la diferencia se deba al azar.<sup>3</sup> Si esto le suena sospechosamente complicado, no está solo; el concepto no es fácil de entender. Tradicionalmente, los investigadores han interpretado un valor de  $P$  inferior a 0,05 en un ensayo controlado aleatorio que significa que estamos bastante seguros de que la diferencia informada no se debe a una variación aleatoria sino al tratamiento.

Una interpretación común e incorrecta de un valor  $P$  menor que 0.05 es que la diferencia entre los grupos es 95% probable que sea “real”, en oposición a un hallazgo casual. Esta interpretación no es precisa debido a la forma en que se calculan los valores  $P$ , lo que significa que un valor  $P$  en

TABLA	Importancia estadística y significado clínico
	<b>Descripción</b>
Diferencia estadísticamente significativa	La probabilidad, bajo un modelo estadístico específico, de que un resumen estadístico de los datos (por ejemplo, diferencia de medias entre 2 grupos) sea igual o más extremo que su valor observado <sup>(3)</sup>
Diferencia clínicamente significativa	Una diferencia media entre grupos que es lo suficientemente grande como para que los pacientes consideren que la diferencia es importante

<sup>1</sup> School of Public Health, University of Sydney, Camperdown, Australia; Centre for Pain, Health and Lifestyle, Australia.  
Copyright ©2018. Journal of Orthopaedic & Sports Physical Therapy®

sí mismo no es una buena medida de evidencia con respecto a las hipótesis sobre la efectividad del tratamiento. Este punto de vista está ganando aceptación gradualmente en el mundo de la investigación clínica. Los autores, revisores y editores de revistas están cambiando la forma en que se interpretan los valores  $P$  en los artículos publicados; esto implica menos énfasis en la importancia estadística para determinar la efectividad del tratamiento.

### Tamaño del efecto

El proceso matemático utilizado para calcular los valores  $P$  es sensible al tamaño de la muestra y la precisión de la medida. Supongamos que realizamos 2 ensayos controlados aleatorios en la misma población y realizamos las mismas intervenciones de tratamiento y control. Un ensayo tiene 50 participantes en cada grupo y el otro 500 en cada grupo. En la medida de discapacidad hubo una diferencia de 10 puntos entre grupos en el ensayo más pequeño y una diferencia entre grupos de 5 puntos en el ensayo más grande. Es posible que la diferencia (de 5 puntos) sea estadísticamente significativa en el ensayo más grande y que la diferencia (de 10 puntos) en el ensayo más pequeño no sea significativa, simplemente porque el ensa-

yo más pequeño tenía menos poder estadístico (es decir, capacidad para detectar una diferencia estadística). Por lo tanto, podríamos concluir que el mismo tratamiento fue efectivo cuando se probó en un ensayo clínico grande, pero no fue efectivo cuando se probó en un ensayo más pequeño. Esto significa que, dados suficientes participantes, uno puede encontrar un valor  $P$  estadísticamente significativo pero que no tenga significado o valor clínico.

### Significado clínico

Interpretar el cambio o la diferencia en una puntuación de resultado requiere considerar qué tan grande es realmente el cambio o la diferencia y si tiene implicaciones clínicas. Los investigadores se han interesado cada vez más en el concepto de significado clínico. Es posible que haya leído términos como cambio mínimo importante, diferencia mínima clínicamente importante y el efecto más pequeño que valga la pena. El objetivo es determinar si es probable que el tamaño de la diferencia valga la pena desde el punto de vista del paciente. Definir una diferencia clínicamente significativa significa determinar un umbral para una medida de resultado (por ej. 2 puntos en una escala de intensidad del dolor de 0 a 10, o una

mejora del 30% de la puntuación inicial). Si la diferencia entre los grupos es mayor que esto, entonces el tratamiento se considera “efectivo” en comparación con el control. La precisión del efecto del tratamiento, generalmente informada con intervalos de confianza, también es relevante al interpretar cuán significativo es un cambio. La precisión será tema de un próximo artículo de Evidencia en la Práctica.

Al igual que cuando se asigna un umbral  $P < .05$  para la importancia estadística, hay desafíos reales para determinar los umbrales de “significación clínica” para las medidas de resultado. La definición de la importancia clínica para evaluar la efectividad del tratamiento es compleja, y el próximo artículo de Evidencia en la práctica profundiza en el tema.

### Conclusión

La medición de los resultados es crítica. Pero medir es sólo la mitad del trabajo: los datos deben analizarse e interpretarse adecuadamente. Este es el caso en la investigación y en la práctica clínica. Determinar si un tratamiento es efectivo implica hacer juicios sobre (1) el tamaño del efecto informado, (2) la precisión del efecto y (3) si el efecto es clínicamente significativo. ●

## FIND Author Instructions & Tools on the Journal's Website

JOSPT's instructions to authors are available at [www.jospt.org](http://www.jospt.org) by clicking Complete Author Instructions in the right-hand Author Center widget on the home page, or by visiting the Info Center for Authors, located in the site's top navigation bar. The Journal's editors have assembled a list of useful tools and links for authors as well as reviewers.

## REFERENCIAS

1. Kamper SJ. Fundamentals of measurement: linking evidence to practice. *J Orthop Sports Phys Ther.* 2019;49:114-115. <https://doi.org/10.2519/jospt.2019.0701>
2. Kamper SJ. Interpreting outcomes 1—change and difference: linking evidence to practice. *J Orthop Sports Phys Ther.* 2019;49:357-358. <https://doi.org/10.2519/jospt.2019.0703>
3. Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. *Am Stat.* 2016;70:129-133. <https://doi.org/10.1080/00031305.2016.1154108>

# Interpretación de los resultados 3— Significado clínico: vinculación de la evidencia con la práctica

**P**ara juzgar si un tratamiento es más efectivo que otro, no alcanza con sólo saber si existe una diferencia. Necesitamos saber qué tan grande es la diferencia.<sup>4</sup> Una forma de juzgar el tamaño de la diferencia informada en un ensayo es preguntar si es “clínicamente significativa” (o clínicamente valiosa): “El cambio más

pequeño que es importante para los pacientes”.<sup>6</sup> A menudo se dice que un tratamiento es “efectivo” y recomendado para la práctica cuando la diferencia entre grupos en un estudio es mayor que el efecto clínicamente significativo. Pero esta idea simple esconde cierta complejidad.

## Cambio y diferencia (nuevamente)

Un artículo anterior de Evidencia en la práctica distinguía entre el cambio en el resultado dentro de un grupo y la diferencia entre grupos.<sup>3</sup> Se debe reservar el término *efecto* para la diferencia entre grupos y evitarse cuando se habla de cambio dentro de un grupo o persona. Cuando se trata de efectos clínicamente significativos, se usan varios términos, no siempre de manera consistente, y a menudo en formas que dificultan la interpretación.

El *cambio* mínimamente importante (CMI) y el *cambio* mínimo clínicamente importante (CMCI) son fundamentalmente diferentes de la *diferencia* mínima clínicamente importante (DMCI). Los investigadores también usan el término *efecto de menor valor*, que describe el mismo concepto que DMCI y se refiere a la diferencia entre grupos. El problema es que estos términos a menudo se usan indistintamente, lo que genera confusión para los lectores.

## Métodos para definir CMCI y DMCI

Los investigadores generalmente expresan CMCI o DMCI en unidades de una medida particular (por ej., 2 puntos en una escala de dolor de 0 a 10) o como una proporción del cambio con respecto al valor inicial (por ej. mejora del 30% en la puntuación del Oswestry Disability Index). Aunque el concepto suena simple, es difícil calcular una diferencia clínicamente significativa. Hay 4 métodos comunes.

## Propuesta

Los investigadores han propuesto definiciones preestablecidas. Un ejemplo común es el tamaño de los efectos de Cohen: 0.2, 0.5 y 0.8 son umbrales para efectos pequeños, medianos y grandes, respectivamente. Los tamaños de efecto de Cohen son múltiplos de la desviación estándar. Por ejemplo, la desviación estándar en una escala numérica de calificación del dolor de 0 a 10 suele ser de aproximadamente 2 puntos. Entonces, un tamaño de efecto de 0.5 equivaldría a aproximadamente 1 punto en una escala de 0 a 10. Un problema con los tamaños del efecto de Cohen es que los umbrales son arbitrarios: no sabemos si los pacientes piensan que las diferencias de 0.2, 0.5 y 0.8 corresponden a efectos peque-

ños, medianos y grandes. Además, los tamaños de “efecto” definidos de esta manera se aplican con bastante frecuencia, pero erróneamente, a los cambios dentro del grupo, destacando el problema de la terminología confusa.

## Métodos basados en la distribución

Los métodos basados en la distribución utilizados para definir un CMI se basan en la difusión de los datos en un estudio y dependen de la confiabilidad del instrumento de medición.<sup>5</sup> Los métodos basados en la distribución proporcionan un umbral que se interpreta mejor como “cambio mínimo detectable” en lugar de un cambio clínicamente importante. Un cambio informado en un estudio que es más pequeño que el cambio mínimo detectable podría no ser un cambio real, sino más bien “ruido” debido a un error (poca confiabilidad) en la medida. Este método no se usa habitualmente para definir diferencias significativas.

## Métodos basados en el anclaje

Los métodos basados en el anclaje comparan los cambios en una medida a lo largo del tiempo con los puntajes del cambio general de los pacientes durante ese período. Se les puede pedir a los pacientes que califiquen su dolor antes del tratamiento y nuevamente 2 semanas después. A las 2 semanas, a los pacientes también se les pregunta cuánto ha mejorado su condición en ese período. La información de los pacientes se dividirá en 2 gru-

<sup>1</sup> School of Public Health, University of Sydney, Camperdown, Australia; Centre for Pain, Health and Lifestyle, Australia.  
Copyright ©2018. Journal of Orthopaedic & Sports Physical Therapy®

pos: los que consideran que han mejorado en una cantidad significativa y los que no. El CMI es el cambio medio en las puntuaciones de dolor para el grupo que se consideró mejorado. Este método no se usa habitualmente para definir diferencias significativas entre grupos.

### Método de compensación beneficio-daño

Este método implica entrevistas en las que se presentan a los pacientes los costos, la cantidad de tiempo y esfuerzo, y los riesgos de daños asociados con una intervención. Se les pregunta cuanta mejora necesitarían experimentar para considerar que la intervención vale la pena. Por ejemplo, "Si tiene que ir a 6 sesiones de kinesiología durante 3 semanas, a \$800 por sesión, y hacer 20 minutos de ejercicio en casa por día, ¿cuánta mejora en el dolor necesitaría para que esto valga la pena?"

### Interpretando el cambio significativo y la diferencia

Un desafío importante para la interpretación es que la propuesta, los métodos basados en la distribución y en el anclaje calculen un cambio mínimamente importante, no una diferencia mínimamente importante. A pesar de esto, los CMI a menudo se usan para interpretar el tamaño de los efectos (entre grupos). La excepción

es el método de compensación beneficio-daño, que se puede utilizar para comparar costos, tiempo y riesgos de daños de un tratamiento con los de otro.<sup>1</sup> Desafortunadamente, el método de compensación beneficio-daño no se ha utilizado con mucha frecuencia para definir tamaños de efectos clínicamente significativos en los campos ortopédico y deportivo.

Habitualmente, un CMI está conectado a un instrumento de medición (por ejemplo, el CMI en una escala numérica de calificación del dolor de 0 a 10, es 2), y este número se utiliza para interpretar la diferencia entre grupos en ensayos controlados aleatorios. Un problema es que la interpretación no tiene en cuenta los factores contextuales importantes. Un CMI aplicado a un tratamiento con riesgos significativos de daños, o un tratamiento costoso, debe ser mayor que el CMI aplicado a un tratamiento sin tales riesgos ni costos. Los pacientes con síntomas más severos pueden requerir un cambio mayor antes de considerar el cambio significativo. Por lo tanto, simplemente no es recomendable considerar CMI o DMCI como características de un instrumento de medición.

### Determinar qué es clínicamente significativo en la práctica

Los métodos utilizados para calcular los umbrales para el cambio clínicamente

importante y la diferencia clínicamente importante no son ideales. Sin embargo, los kinesiólogos que desean aplicar evidencia de investigación en la práctica necesitan evaluar e interpretar el tamaño de los efectos del tratamiento. Un enfoque sensato podría ser conocer el rango general de estimaciones para un cambio o diferencia clínicamente significativo, y usar el rango como un filtro grueso al leer la investigación. Puede preguntarse: "¿El tamaño del efecto en un estudio es la estimación aproximada para la diferencia clínicamente significativa? Si no (es decir, es mucho más pequeño), entonces el tratamiento probablemente no será útil para su paciente. En caso afirmativo, entonces el tratamiento podría ser una opción. Este probable tamaño del efecto puede ser incorporado en una discusión con el paciente donde se incluyan los costos, el período de tratamiento, las expectativas del paciente y los posibles daños de las opciones de tratamiento, para culminar en una decisión de tratamiento compartida. ●

*"Una diferencia es una diferencia sólo si hace la diferencia."*<sup>2</sup>

### DOWNLOAD PowerPoint Slides of JOSPT Figures

JOSPT offers PowerPoint slides of figures to accompany all full-text articles with figures on JOSPT's website ([www.jospt.org](http://www.jospt.org)). These slides are generated automatically by the site, and can be downloaded and saved. They include the article title, authors, and full citation. JOSPT offers full-text format for all articles published from January 2010 to date.

### REFERENCIAS

1. Barrett B, Brown R, Mundt M, et al. Using benefit harm tradeoffs to estimate sufficiently important difference: the case of the common cold. *Med Decis Making*. 2005;25:47-55. <https://doi.org/10.1177/0272989X04273147>
2. Huff D. *How to Lie With Statistics*. New York, NY: Norton; 1954.
3. Kamper SJ. Interpreting outcomes 1—change and difference: linking evidence to practice. *J Orthop Sports Phys Ther*. 2019;49:357-358. <https://doi.org/10.2519/jospt.2019.0703>
4. Kamper SJ. Interpreting outcomes 2—statistical significance and clinical meaningfulness: linking evidence to practice. *J Orthop Sports Phys Ther*. 2019;49:559-560. <https://doi.org/10.2519/jospt.2019.0704>
5. Kamper SJ. Reliability and validity: linking evidence to practice. *J Orthop Sports Phys Ther*. 2019;49:286-287. <https://doi.org/10.2519/jospt.2019.0702>
6. Stratford PW, Binkley JM, Riddle DL, Guyatt GH. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 1. *Phys Ther*. 1998;78:1186-1196. <https://doi.org/10.1093/ptj/78.11.1186>



# Generalización: vinculación de la evidencia con la práctica

**E**l concepto básico de generalización es simple: los resultados de un estudio son generalizables cuando se pueden aplicar (son útiles para informar una decisión clínica) a los pacientes que concurren a recibir atención. La generalización es la cara de la práctica basada en la evidencia, donde la evidencia de la investigación influye en el manejo del paciente.

## Validez interna y validez externa

Es importante distinguir entre generalización (a veces denominada validez externa) y la calidad del estudio (validez interna o calidad metodológica). La validez interna es qué tan seguro puede estar el lector de que el estudio proporciona un resultado imparcial. Los investigadores emplean características de diseño como la aleatorización y el cegamiento para reducir el riesgo de sesgo y mejorar la validez interna. Estas características de diseño son el tema de artículos anteriores de Evidencia en la práctica<sup>1-3</sup> y el siguiente artículo describirá cómo evaluar la validez interna. Evaluar la validez interna y la generalización tiene dos propósitos críticos pero diferentes; la validez interna determina si es probable que la estimación del efecto en el estudio sea precisa, y la generalización determina si los hallazgos son aplicables a personas fuera de la muestra del estudio.

## Muestras y Poblaciones

Entender la generalización requiere una apreciación de la diferencia entre muestras y poblaciones. Los términos: *muestra* y *población*, a veces se usan indistintamente, pero significan cosas diferentes. Una población es un grupo de personas que comparten una característica importante. En términos demográficos

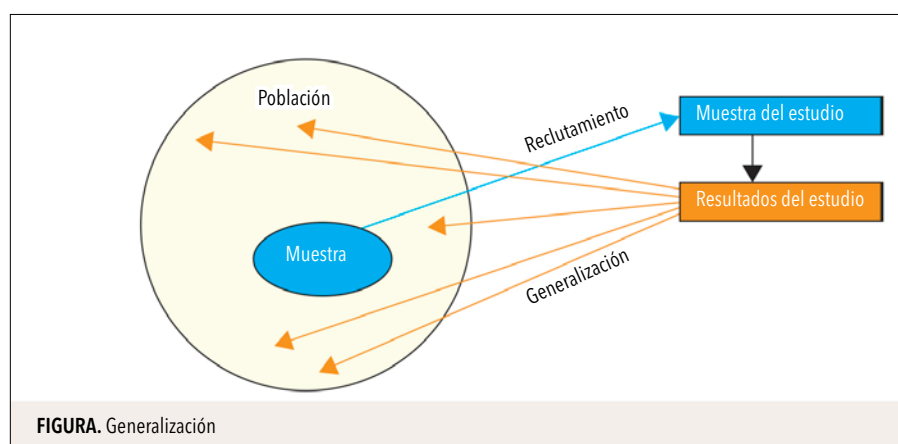
esta característica podría ser el lugar donde viven. La población de un país incluye a todos los que viven en ese país. Con respecto a la investigación clínica, la población suele estar definida por una determinada afección de salud, por ejemplo, todos los que han tenido una luxación de hombro.

Como en general es imposible incluir a una población completa en un estudio, los investigadores utilizan una muestra representativa. Una muestra es el grupo de personas de la población de interés que participan en un estudio. El estudio se realiza sobre la muestra y los resultados sólo son útiles en la medida en que reflejan lo que sucede en toda la población. Esto es generalización (**FIGURA**).

## Representatividad

La muestra ideal es una “muestra representativa”; es lo mismo que la población en todos los sentidos, excepto que hay menos individuos. Por ejemplo, la proporción de mujeres, la media y la variabilidad de la gravedad de la afección y la distribución de creencias y actitudes son las mismas en la muestra que en la población. El concepto de representatividad es la medida en que la muestra es representativa de la población con respecto a todas las características importantes para el estudio.

Los investigadores tienen como objetivo conseguir una muestra que sea lo más representativa posible de la población. La representatividad de la muestra dependerá de los criterios de inclusión y exclusión del estudio y los métodos utilizados para reclutar a los participantes del mismo. Estos serán temas de futuros artículos de Evidencia en la Práctica.



<sup>1</sup> School of Public Health, University of Sydney, Camperdown, Australia; Centre for Pain, Health and Lifestyle, Australia. Copyright ©2018. Journal of Orthopaedic & Sports Physical Therapy®

### Aplicación práctica

La generalización es un continuo, no una dicotomía; los resultados de un estudio son más o menos generalizables, a diferencia de lo generalizable o no. Los kinesiólogos/as siempre pueden argumentar que los hallazgos de un estudio no son generalizables para sus pacientes. Esto ocurre a menudo cuando los hallazgos entran en conflicto con la práctica habitual. En la práctica basada en la evidencia, el objetivo es evaluar con sensatez cómo guiarán los resultados de la investigación las decisiones de tratamiento e informarán el pronóstico de un paciente en particular.

#### ¿Debo aplicar los hallazgos de este estudio al paciente frente a mí?

Al evaluar un estudio, un lector se debe preguntar si la muestra del estudio es representativa de la población de interés. Esto a menudo no es sencillo porque la población de estudio puede no estar definida de forma explícita por los autores. Determinar la representatividad implica comprender los criterios de inclusión y exclusión, leer los métodos de reclutamiento y observar los datos descriptivos que resumen las características clínicas de las personas en la muestra.

La capacidad, habilidades y recursos disponibles para el kinesiólogo también son relevantes para la generalización. Los resultados del estudio serán aplicables al paciente frente a usted siempre que se pueda realizar el tratamiento de la misma manera que en el estudio.

Si cree que la muestra es suficientemente representativa de la población de interés, los hallazgos pueden considerarse representativos de la población. El siguiente paso es evaluar si el paciente frente a usted es parte de la población; si es así, los hallazgos deben ser generalizables para el paciente. En este caso, el estudio proporciona evidencia sobre el tratamiento que se puede incorporar junto con las preferencias del paciente y la experiencia del profesional para guiar las decisiones clínicas.

#### ¿Qué evidencia debo usar para aplicar un tratamiento a un paciente que se encuentra frente a mí?

La idea de la generalización como un continuo es importante. Aquellos que aplican la investigación no deben ser demasiado dogmáticos o limitados en su interpretación de la evidencia de la investigación, especialmente en los casos en que puede no haber

un estudio de alta calidad realizado en una población de pacientes en particular. En tales situaciones, comprender la generalización puede ayudar a identificar la mejor evidencia disponible, según el principio central de la práctica basada en la evidencia. La ausencia de un estudio en la población que defina con precisión a su paciente no significa que se pueda ignorar por completo la evidencia de la investigación para ese encuentro clínico.

Los kinesiólogos/as deben tomar decisiones razonadas sobre la generalización de los resultados de la investigación más allá de una población de estudio. Esto requiere una comprensión matizada de la condición que define a la población, el tratamiento que forma parte del estudio y el paciente. Por ejemplo, podría ser razonable aplicar los resultados de un estudio que investiga un programa de ejercicios para pacientes luego de una fractura de tobillo a un paciente con artritis de tobillo. Dependiendo del tipo de tratamiento, puede ser válido aplicar los hallazgos de un estudio de pacientes con dolor de espalda crónico a personas con dolores crónicos diferentes. Finalmente, los kinesiólogos deben decidir el nivel de generalización de un estudio y determinar cómo evaluar la información para llegar a una decisión de tratamiento. ●

### NOTIFY JOSPT of Changes in Address

Please remember to let JOSPT know about changes in your mailing address. The US Postal Service typically will not forward second-class periodical mail. Journals are destroyed, and the USPS charges JOSPT for sending them to the wrong address. You may change your address online at [www.jospt.org](http://www.jospt.org). Visit Info Center for Readers, click Change of Address, and complete the online form. We appreciate your assistance in keeping JOSPT's mailing list up to date.

### REFERENCIAS

1. Kamper SJ. Bias: linking evidence with practice. *J Orthop Sports Phys Ther.* 2018;48:667-668. <https://doi.org/10.2519/jospt.2018.0703>
2. Kamper SJ. Blinding: linking evidence to practice. *J Orthop Sports Phys Ther.* 2018;48:825-826. <https://doi.org/10.2519/jospt.2018.0705>
3. Kamper SJ. Randomization: linking evidence to practice. *J Orthop Sports Phys Ther.* 2018;48:730-731. <https://doi.org/10.2519/jospt.2018.0704>

# Riesgo de sesgo y evaluación de la calidad del estudio: vinculación de la evidencia con la práctica

La práctica basada en la evidencia requiere el uso de la mejor evidencia disponible, lo que implica que alguna evidencia es mejor que otra. Una buena evidencia proviene de los resultados de una investigación que tengan un menor riesgo de sesgo. Recordando un artículo anterior de Evidencia en la práctica <sup>2</sup>,

sesgo significa que los resultados del estudio no reflejan el verdadero efecto de una intervención. Un estudio sesgado puede sobreestimar (o subestimar) el tamaño de un efecto o asociación.

Separar lo bueno de lo malo cuando se trata de evidencia de investigación requiere una evaluación de los métodos y resultados del estudio. Existen varias herramientas disponibles para ayudar al lector a evaluar la calidad del estudio o el riesgo de sesgo. La calidad del estudio y el riesgo de sesgo son conceptos superpuestos pero sutilmente diferentes. El riesgo de sesgo es la probabilidad de que los resultados del estudio estén sesgados, mientras que la calidad del estudio es un concepto más amplio que incluye el riesgo de sesgo, entre otras características, como la generalización <sup>4</sup> y los informes adecuados. Los ítems incluidos en las herramientas de evaluación de riesgo de sesgo están más focalizados, mientras que los de las herramientas de calidad del estudio cubren más terreno.

Las herramientas de evaluación de la calidad del estudio y el riesgo de sesgo tienen como objetivo ayudar al lector a evaluar la probabilidad de que los resultados de un estudio sean verdaderos y, por lo tanto, determinar hasta qué punto el estudio debe informar las decisiones clínicas.

## Medición del riesgo de sesgo y calidad del estudio

Los investigadores han desarrollado listas de verificación, escalas y herramientas basadas en el dominio que identifican indicadores (criterios) de calidad del estudio y sesgo potencial.

- Lista de verificación: un conjunto de criterios que identifican las características del estudio que indican la calidad del estudio o el riesgo de sesgo. Las listas de verificación no deben sumarse para producir una puntuación total.
- Escala: un conjunto de criterios que se suman para producir una puntuación total.
- Herramienta basada en el dominio: una lista de tipos de sesgo (dominios); los evaluadores juzgan el riesgo de cada tipo de sesgo.

Los principios que subyacen a la medición de la calidad del estudio y el riesgo de sesgo no son diferentes de los que miden los resultados clínicos. <sup>3,5</sup> Al usar cualquier herramienta para evaluar la calidad del estudio o el riesgo de sesgo, nos preocupa más la confiabilidad (si diferentes personas tienen mismo puntaje en la escala / lista de verificación) y validez (si la herramienta realmente mide el riesgo

de sesgo). No todas las herramientas han sido probadas por su confiabilidad y validez. Confíe más en las herramientas establecidas y conocidas para guiar su evaluación.

## Herramientas de evaluación

Se aplican diferentes herramientas de calidad de estudio y riesgo de sesgo a los diferentes diseños de estudio, porque ciertos tipos de sesgo pueden o no ser relevantes dependiendo de la pregunta de investigación. Por ejemplo, la identificación, la medición y el ajuste estadístico de los factores de confusión son críticos para los estudios pronósticos, pero no para los ensayos clínicos aleatorios (ECA). El análisis de intención a tratar es una característica importante de los ECA, pero no es relevante para los estudios de diagnóstico. La herramienta utilizada para medir la calidad del estudio o el riesgo de sesgo debe coincidir con el diseño del estudio.

La medición de la calidad del estudio está mejor desarrollada para los estudios de efectividad del tratamiento. Las herramientas más comunes son la Physiotherapy Evidence Database (PEDro) y la herramienta Cochrane de riesgo de sesgo. El primero se usa comúnmente como una escala y el segundo es una herramienta basada en dominios. La escala PEDro incluye una lista de características del estudio que están marcadas como presentes, ausentes o poco claras, y el número de elementos "presentes" se suma para representar el puntaje de calidad del estudio. La herramienta Cochrane de riesgo de sesgo incluye una lista de tipos de sesgo (dominios), y el evalua-

<sup>1</sup> School of Public Health, University of Sydney, Camperdown, Australia; Centre for Pain, Health and Lifestyle, Australia.  
Copyright ©2018. Journal of Orthopaedic & Sports Physical Therapy®

El autor juzga si el estudio está en riesgo de sesgo (alto, bajo o poco claro). La escala PEDro y la herramienta de riesgo de sesgo Cochrane identifican muchas de las mismas características del estudio: ambas incluyen elementos relacionados con la aleatorización, asignación oculta, cegamiento e informes apropiados de resultados. Si bien ambas herramientas se desarrollaron utilizando métodos consistentes<sup>1, 6, 7</sup>,

no hay pruebas sólidas de que ninguna sea más confiable y válida que la otra. Si bien hay numerosas herramientas de evaluación disponibles para la mayoría de los tipos de estudio, generalmente incluyen elementos similares y, por lo general, no hay consenso sobre cuál es el preferido. Esto se debe a que las características más importantes que denotan la calidad del estudio y conducen al sesgo son bien reco-

nocidas. Algunas herramientas aceptadas para los tipos de estudio incluyen la Quality In Prognostic Studies (QUIPS) para estudios observacionales que investigan factores pronósticos, la Quality Assessment of Diagnostic Accuracy Studies-revised (QUADAS-2) para estudios de diagnóstico, y A Measurement Tool to Assess systematic Reviews-revised (AMSTAR-2) para revisiones sistemáticas (TABLA).

TABLA		Ítems en las herramientas de calidad de estudio			
Tipo de estudio	Herramienta de riesgo de sesgo Cochrane	Escala PEDro	QUIPS	QUADAS-2	AMSTAR-2
Eca	<ul style="list-style-type: none"> <li>Sesgo de selección</li> <li>Sesgo de deserción</li> <li>Sesgo de rendimiento</li> <li>Sesgo de detección</li> <li>Sesgo de informes</li> <li>Otro sesgo</li> </ul>	<ul style="list-style-type: none"> <li>Elegibilidad</li> <li>Aleatorización</li> <li>Asignación oculta</li> <li>Comparabilidad entre grupos</li> <li>Cegamiento de los participantes</li> <li>Cegamiento del terapeuta</li> <li>Cegamiento del evaluador</li> <li>Análisis por intención de tratar</li> <li>Pérdida de seguimiento</li> <li>Comparaciones estadísticas</li> <li>Medidas de punto y variabilidad</li> </ul>			
Pronóstico			<ul style="list-style-type: none"> <li>Participación</li> <li>Desgaste</li> <li>Medición del factor pronóstico</li> <li>Medida de confusión</li> <li>Medición de resultados</li> <li>Análisis e informes</li> </ul>		
Pruebas de diagnóstico				<ul style="list-style-type: none"> <li>Selección de pacientes</li> <li>Prueba de índice</li> <li>Estándar de referencia</li> <li>Flujo y sincronización</li> </ul>	
Revisiones sistemáticas					<ul style="list-style-type: none"> <li>Pregunta de investigación</li> <li>Protocolo a priori</li> <li>Criterios de inclusión</li> <li>Búsqueda integral</li> <li>Justificación de exclusiones</li> <li>Selección de estudios duplicados</li> <li>Extracción de datos duplicados</li> <li>Descripción del estudio incluido</li> <li>Evaluación de riesgo de sesgo</li> <li>Impacto del sesgo en la interpretación</li> <li>Impacto del sesgo en la síntesis</li> <li>Metaanálisis apropiado</li> <li>Fuentes de financiamiento</li> <li>Heterogeneidad</li> <li>Sesgo de publicación</li> <li>Conflictos de interés</li> </ul>

Abreviaturas: AMSTAR-2, A Measurement Tool to Assess systematic Reviews-revised; PEDro, Physiotherapy Evidence Database; QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies-revised; QUIPS, Quality In Prognostic Studies; ECA, ensayo clínico aleatorizado.



## Directrices

Los primos de las herramientas de evaluación de la calidad del estudio son las directrices, aunque no están diseñadas para evaluar el riesgo de sesgo. Las directrices ayudan a los autores a garantizar que sus artículos incluyan suficiente información para que los lectores juzguen la calidad y la aplicabilidad de su investigación. Los elementos a menudo se superponen con los de las listas de verificación de calidad del estudio, pero tienen un propósito diferente. Debido a que las directrices se centran en lo que aparece en el manuscrito (en lugar del diseño del estudio), a menudo los editores de revistas también las utilizan para mejorar los artículos de investigación durante la revisión entre pares. Existen directrices específicas para diferentes diseños de estudio. Las más conocidas son la CONSolidated Standards of Reporting Trials (CONSORT) para ECA, la STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) para estudios observacionales y el Preferred Reporting Items for Systematic Reviews and Meta-Analyses

(PRISMA) para revisiones sistemáticas. La red Enhancing the QUALity and Transparency Of health Research (EQUATOR) (<https://www.equator-network.org/>) posee un listado actualizado de directrices.

## Aplicación

Tamizar la investigación de baja calidad y prestar atención a lo que está bien es un requisito de la práctica basada en la evidencia. La clave para evaluar la calidad es comprender cómo y en qué medida el sesgo puede afectar los resultados del estudio. Reconozca qué tipos de sesgo son más relevantes para el estudio en cuestión, juzgue el grado en el cual el sesgo debería afectar la confianza en los hallazgos del estudio y decida cuándo el riesgo de sesgo es tan grande que los resultados deben ignorarse por completo. Las revisiones sistemáticas pueden ser un recurso útil, porque la mayoría de las revisiones informan el riesgo de sesgo o las evaluaciones de calidad de los estudios incluidos.

La calidad del estudio y las herramientas de evaluación del riesgo de sesgo pueden ayudar al lector a comprender cuánta confianza se debe depositar en los resultados de un estudio. Sin embargo, el lector debe seleccionar la herramienta adecuada para el trabajo, y la aplicación adecuada requiere una comprensión de los principios que sustentan los elementos de la herramienta. ●

## BROWSE Collections of Articles on JOSPT's Website

JOSPT's website ([www.jospt.org](http://www.jospt.org)) offers readers the opportunity to browse published articles by Previous Issues with accompanying volume and issue numbers, date of publication, and page range; the table of contents of the Upcoming Issue; a list of available accepted Ahead of Print articles; and a listing of Categories and their associated article collections by type of article (Research Report, Case Report, etc). Features further curates 3 primary JOSPT article collections: Musculoskeletal Imaging, Clinical Practice Guidelines, and Perspectives for Patients, and provides a directory of Special Reports published by JOSPT.

## REFERENCIAS

1. Higgins JP, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928. <https://doi.org/10.1136/bmj.d5928>
2. Kamper SJ. Bias: linking evidence with practice. *J Orthop Sports Phys Ther*. 2018;48:667-668. <https://doi.org/10.2519/jospt.2018.0703>
3. Kamper SJ. Fundamentals of measurement: linking evidence to practice. *J Orthop Sports Phys Ther*. 2019;49:114-115. <https://doi.org/10.2519/jospt.2019.0701>
4. Kamper SJ. Generalizability: linking evidence to practice. *J Orthop Sports Phys Ther*. 2020;50:45-46. <https://doi.org/10.2519/jospt.2020.0701>
5. Kamper SJ. Reliability and validity: linking evidence to practice. *J Orthop Sports Phys Ther*. 2019;49:286-287. <https://doi.org/10.2519/jospt.2019.0702>
6. Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. *Phys Ther*. 2003;83:713-721. <https://doi.org/10.1093/ptj/83.8.713>
7. Verhagen AP, de Vet HC, de Bie RA, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol*. 1998;51:1235-1241. [https://doi.org/10.1016/s0895-4356\(98\)00131-0](https://doi.org/10.1016/s0895-4356(98)00131-0)