

EVIDENCIAS EN PEDIATRÍA

Toma de decisiones clínicas basadas en las mejores pruebas científicas

www.evidenciasenpediatria.es

Artículos Valorados Críticamente

La valoración no enmascarada de escalas de medida subjetivas sobrestima el efecto

Ochoa Sangrador C¹, Andrés de Llano JM²

¹Servicio de Pediatría. Hospital Virgen de la Concha. Zamora (España).

²Servicio de Pediatría. Hospital General del Río Carrión. Palencia (España).

Correspondencia: Carlos Ochoa Sangrador, cochoas@meditex.es

Palabras clave en inglés: outcome assessment; systematic review; double-blind method; randomized controlled trial; observer bias.

Palabras clave en español: valoración de resultados; revisión sistemática; doble ciego; ensayo clínico aleatorizado; sesgo del observador.

Fecha de recepción: 1 de mayo de 2013 • **Fecha de aceptación:** 8 de mayo de 2013

Fecha de publicación del artículo: 16 de mayo de 2013

Evid Pediatr.2013;9:27.

CÓMO CITAR ESTE ARTÍCULO

Ochoa Sangrador C, Andrés de Llano JM. La valoración no enmascarada de escalas de medida subjetivas sobrestima el efecto. Evid Pediatr. 2013;9:27.

Para recibir Evidencias en Pediatría en su correo electrónico debe darse de alta en nuestro boletín de novedades en <http://www.evidenciasenpediatria.es>

Este artículo está disponible en: <http://www.evidenciasenpediatria.es/EnlaceArticulo?ref=2013;9:27>

©2005-13 • ISSN: 1885-7388

La valoración no enmascarada de escalas de medida subjetivas sobrestima el efecto

Ochoa Sangrador C¹, Andrés de Llano JM²

¹Servicio de Pediatría. Hospital Virgen de la Concha. Zamora (España).

²Servicio de Pediatría. Hospital General del Río Carrión. Palencia (España).

Correspondencia: Carlos Ochoa Sangrador, cochoas@meditex.es

Referencia bibliográfica: Hróbjartsson A, Thomsen AS, Emanuelsson F, Tendam B, Hilden J, Boutron I, et al. Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors. *CMAJ*. 2013 Mar 5;185:E201-11. Epub 2013 Jan 28.

Resumen

Conclusiones de los autores del estudio: existe un sesgo de observador en los ensayos clínicos aleatorizados con resultados de escala de medición subjetivas. La falta de cegamiento en la evaluación de los resultados en estos ensayos incrementa un cuarto de la desviación estándar de la escala de medición utilizada.

Comentario de los revisores: la valoración no enmascarada de medidas de efecto subjetivas parece originar una sobrestimación del efecto en los estudios epidemiológicos. No obstante, la magnitud del error estimada en los estudios de esta revisión no puede extrapolarse directamente a otros escenarios clínicos. Cualquier estudio que evalúe medidas de escala subjetivas debe considerar el empleo de técnicas de enmascaramiento o, al menos, tratar de controlar o cuantificar el error potencial.

Palabras clave: valoración de resultados; revisión sistemática; doble ciego; ensayo clínico aleatorizado; Sesgo del observador.

Unblinded assessment of subjective measurement scales overestimates the effect

Abstract

Authors' conclusions: there is an observer bias in randomized clinical trials with results of subjective measurement scale. Lack of blinding in evaluating the results in these trials increases a quarter of the standard deviation of the measuring scale used.

Reviewers' commentary: unblinded assessment of subjective effect measures seems to cause an overestimation of the effect in epidemiological studies. However, the magnitude of the estimated error in the studies of this review can not be generalized. Any study that uses subjective scale measures should consider employing masking techniques or at least try to control or quantify the potential mistake.

Keywords: outcome assessment; systematic review; double-blind method; randomized controlled trial; Observer bias.

RESUMEN ESTRUCTURADO

Objetivo: estudiar el efecto de la evaluación de medidas de resultado de forma ciega frente a no ciega sobre el tamaño del efecto en los ensayos clínicos aleatorizados con resultados en los que se incluyeron escalas de medición subjetivas.

Diseño: revisión sistemática de ensayos clínicos.

Fuente de datos: se realizaron búsquedas en PubMed, EMBASE, PsycINFO, CINAHL, el Registro Cochrane Central de Ensayos Controlados, Highwire Press y Google Académico, sin restricción en el lenguaje. Se solicitó a los autores de todos los ensayos incluidos si conocían ensayos

adicionales para identificar otros estudios que debieran ser incluidos.

Selección de estudios: se encontraron 4035 estudios. Se seleccionaron los ensayos clínicos con valoración doble, ciega y no ciega de la misma escala de medida. Dos investigadores decidieron sobre la elegibilidad de los artículos. Los desacuerdos se resolvieron mediante discusión.

Extracción de datos: se seleccionó una única escala de medición de cada ensayo (preferiblemente el resultado primario del ensayo y la primera evaluación tras el final del tratamiento). De cada ensayo se extrajeron la media, la desviación estándar y el número de pacientes en los grupos expe-

rimentales y de control. De los estudios cruzados se extrajo la desviación estándar de la diferencia pareada. Cuando fue posible, se extrajeron datos sobre la correlación (coeficiente de correlación de Spearman) y sobre la variación interobservador. Se completaron datos contactando con los autores y consultando la web de la Food and Drug Administration (FDA) de EE. UU.

Para cada ensayo, se evaluaron cinco potenciales factores de confusión: intervalo de tiempo entre las dos evaluaciones, diferentes tipos de evaluadores, diferentes procedimientos de evaluación, riesgo de ineficacia del cegamiento y evaluación de pacientes diferentes. Se enmascararon los artículos manipulando las versiones en PDF. Además, se evaluaron tres características de los resultados que podrían explicar las variaciones en el sesgo del observador: el grado de subjetividad de los resultados, la participación global del evaluador no ciego en el ensayo y la vulnerabilidad de los resultados de pacientes en estudios no ciegos. Se calculó el tamaño del efecto (diferencia de medias estandarizada) utilizando la desviación estándar combinada de las evaluaciones ciegas como unidad común de normalización. Un tamaño del efecto inferior al cero sugiere un efecto beneficioso de la intervención experimental. Posteriormente se resumieron los efectos de la evaluación de resultados no cegados, como la diferencia entre los dos tamaños del efecto. Una diferencia en el tamaño del efecto menor que cero indica que las evaluaciones no ciegas generan estimaciones más optimistas que las evaluaciones ciegas. Finalmente, se utilizó metarregresión de efectos aleatorios para determinar las variaciones en las diferencias del tamaño del efecto y sus asociaciones.

Resultados: se incluyeron finalmente 16 ensayos (2854 pacientes). El coeficiente de correlación de Spearman promedio entre las evaluaciones cegadas y no cegadas fue de 0,67. Entre los 16 ensayos, la diferencia en el tamaño del efecto varió de -1,10 a 0,14. La diferencia en el tamaño global del efecto fue -0,23 (intervalo de confianza del 95% [IC 95%]: -0,40 a -0,06), que equivale a un cuarto de desviación estándar, con heterogeneidad moderada ($I^2 = 46\%$, $p = 0,02$). El tamaño del efecto combinado sobre la base de las evaluaciones de los evaluadores cegados fue -0,34 (IC 95%: -0,55 a -0,14) lo que implica que los evaluadores no ciegos exageran el tamaño del efecto estimado en alrededor del 68% (IC 95%: 14 a 230). Usando metarregresión univariable no se encuentra asociación con variables relacionadas con el sesgo de observador.

Conclusión: los investigadores muestran evidencias empíricas sobre el sesgo de observador en los ensayos clínicos aleatorizados con resultados de escala de medición subjetiva. La falta de cegamiento en la evaluación de los resultados en estos ensayos implica un alto riesgo de sesgo.

Conflicto de intereses: no existe.

Fuente de financiación: el estudio fue financiado parcialmente por el Consejo Danés de Investigación Independiente: Ciencias Médicas.

COMENTARIO CRÍTICO

Justificación: el enmascaramiento o ciego en un ensayo clínico es importante, no solo en la aplicación de las intervenciones evaluadas sino además en la evaluación de los resultados. Si la medida de efecto es subjetiva y el observador que la clasifica conoce la intervención que ha recibido el sujeto que está evaluando, tiende a dejarse influir de forma consciente o inconsciente por las expectativas del estudio, realizando una clasificación sesgada del efecto. De hecho, existe evidencia suficiente de que los evaluadores no ciegos tienden a sobrestimar el efecto. En estudios que han valorado medidas de efecto dicotómicas por evaluadores enmascarados y no enmascarados, estos últimos sobrestiman el riesgo un 36%¹. Este estudio trata de cuantificar la magnitud del sesgo de observador no ciego para medidas de efecto con escalas subjetivas.

Validez o rigor científico: es una revisión sistemática que define correctamente el escenario clínico, el factor de estudio y la medida de efecto a evaluar. Se ha realizado una búsqueda bibliográfica exhaustiva, aplicándose criterios de inclusión y exclusión adecuados. Se ha evaluado de forma explícita la calidad de los estudios, recogiendo posibles covariables relacionadas con la medida de efecto (sesgo de observador). Se ha valorado la heterogeneidad de los estudios y se han empleado técnicas de metarregresión para explorar posibles fuentes sistemáticas de error. La principal limitación de la revisión es la heterogeneidad de los escenarios clínicos y escalas empleadas en los distintos estudios; al seleccionarse trabajos en los que se hubiera hecho una medición doble del efecto, los artículos analizados solo representan a un grupo de intervenciones y escalas de medida peculiares.

Importancia clínica: se ha estimado que la evaluación no ciega de escalas de medida subjetivas en ensayos clínicos sobrestima la magnitud del efecto en aproximadamente un cuarto de la desviación estándar de las escalas, incrementando el efecto un 68% respecto a la valoración ciega. Aunque la interpretación de la importancia clínica de esta sobrestimación resulta compleja por la heterogeneidad de escenarios clínicos y escalas de medidas analizados, supone una magnitud cuantitativamente importante, si tenemos en cuenta que es un error en medidas repetidas de la misma escala². Resulta curioso observar que las escalas, medidas de efecto cuantitativas, parecen más sensibles a interpretaciones subjetivas, que las medidas dicotómicas¹.

Aplicabilidad en la práctica clínica: los resultados de este estudio, junto con evidencia previa permiten establecer que existe un riesgo de cometer errores sistemáticos de clasificación cuando se realiza una valoración no enmascarada

del efecto en los estudios epidemiológicos. No obstante, si tenemos en cuenta las características de los estudios incluidos en esta revisión, no podemos extrapolar la magnitud del error a otros escenarios clínicos. Cualquier estudio que evalúe medidas de escala subjetivas debe considerar el empleo de técnicas de enmascaramiento o al menos tratar de controlar o cuantificar el potencial error.

Conflicto de intereses de los autores del comentario: no existe.

BIBLIOGRAFÍA

1. Hróbjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, *et al*. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ*. 2012;344:e1119.
2. Hojat M, Xu G. A visitor's guide to effect sizes: statistical significance versus practical (clinical) importance of research findings. *Adv Health Sci Educ Theory Pract*. 2004;9:241-9.