

Herramientas ROBINS para evaluar el riesgo de sesgo de estudios no aleatorizados

ROBINS tool to assess the risk of bias in non-randomized studies

Comentado de:

Sterne JA, et al. *BMJ*. 2016;355:i4919¹ y ROBINS-E Development Group. Launch version, 1 June 2022²

Introducción

Los estudios observacionales diseñados de forma adecuada han brindado información útil sobre la efectividad de intervenciones tales como el uso obligatorio de los cascos, el rastreo de cáncer de cuello uterino, o la difusión de guías de práctica, así como sobre los efectos de eventos adversos infrecuentes. Sin embargo, la calidad de los estudios observacionales continúa siendo difícil de evaluar.

Por este motivo, miembros de los grupos Cochrane de métodos de sesgo y de estudios no aleatorizados desarrollaron la herramienta ROBINS (acrónimo de *Risk Of Bias In Non-Randomized Studies*), con el objetivo de valorar el riesgo de sesgo de estudios no aleatorizados (ENA). En la actualidad, existen dos versiones de esta herramienta: ROBINS-I¹, para ENA de intervenciones (ENAI), y la recientemente lanzada ROBINS-E², que evalúa el efecto de una exposición en estudios de cohortes. A continuación se resumen las principales características informadas por los

autores de cada instrumento.

ROBINS-I

Los ENAI que se pueden evaluar con ROBINS-I incluyen diseños de estudios en los que la asignación ocurre durante el curso de las decisiones terapéuticas habituales o las elecciones personales (estudios observacionales), tales como estudios de cohortes y estudios de casos y controles; estudios antes-después controlados; estudios de series temporales interrumpidas (también llamados estudios cuasi-experimentales); y ensayos controlados en los que los grupos de intervención se asignan mediante un método aleatorización imperfecto (cuasi-aleatorizados).

ROBINS proporciona preguntas de señalización cuyas respuestas indican el potencial de sesgo, brindando así una forma sistemática de organizar y presentar la evidencia disponible relacionada con el riesgo de sesgo en los ENA. La Tabla 1 enumera los dominios de sesgo cubiertos por la herramienta ROBINS-I para la mayoría de los tipos de ENAI. Según las respuestas a las preguntas de señalización, las opciones para cada dominio son: riesgo de sesgo **bajo, moderado, grave o crítico**, con una opción adicional: **sin información**.

Tabla 1. Dominios de sesgo incluidos en la herramienta ROBINS-I. Fuente: Sterne JA, et al. *BMJ*. 2016;355:i4919¹

Tiempo	Sesgo relacionado al dominio	Explicación
Pre-intervención	Confusión: sesgo vinculado a la presencia de algún factor con efecto en la ocurrencia de la intervención y también en el desenlace	La confusión ocurre cuando una o más variables tienen influencia sobre la magnitud o la probabilidad de que suceda la intervención y también en la magnitud y/o probabilidad del desenlace. También ocurre cuando los individuos cambian de rama de intervención y algún factor pronóstico post-basal tiene alguna influencia sobre la intervención recibida a posteriori.
	Selección de participantes	Cuando la exclusión de participantes elegibles, el tiempo de seguimiento inicial de algunos participantes, o los eventos medidos están relacionados tanto con la intervención como con el resultado, habrá una asociación entre la intervención y el resultado incluso si el efecto de las intervenciones es idéntico. Un ejemplo es el sesgo debido a la inclusión de usuarios frecuentes, en lugar de nuevos usuarios, de una intervención.
Intervención	Clasificación errónea de intervenciones: sesgo de información, recuerdo, medición y del observador	Sesgo introducido por la clasificación errónea diferencial o no diferencial del estado de la intervención. La clasificación errónea no diferencial no se relaciona con el desenlace y normalmente sesga el efecto estimado de la intervención hacia la nulidad. La clasificación errónea diferencial ocurre cuando este error está relacionado con el resultado, y es probable que lleve al sesgo.
Post-intervención	Desviación de las intervenciones planeadas: sesgo de desempeño y confusión por variación temporal	Surge cuando hay diferencias sistemáticas en el cuidado previsto entre los grupos intervención y control, representando una desviación de la intervención planeada. El sesgo dependerá de cuál sea efecto de interés: el efecto de la asignación a la intervención (por intención de tratar) o el efecto de adherir a la intervención (por protocolo).
	Datos faltantes: sesgo de desgaste y de selección	Surge por pérdidas de seguimiento de individuos inicialmente incluidos (p. ej., pérdida diferencial cuando el seguimiento está afectado por factores pronósticos), o la exclusión de individuos con información perdida sobre el estado de la intervención o de variables confundidoras.
	Medición de resultados: sesgo de detección de recuerdo, información, clasificación errónea, de observador y de medición	Sesgo introducido por clasificación errónea diferencial o no diferencial de los datos de resultados. Puede ocurrir cuando los evaluadores son conscientes del estado de la intervención, si en los distintos grupos se utilizaron diferentes métodos para evaluar resultados, o si los errores de medición están relacionados al grupo asignado o a los efectos de la intervención.
	Reporte selectivo de resultados	Reporte selectivo de resultados que depende de los hallazgos y evita que la estimación se incluya en un metanálisis



La herramienta ROBINS-I se basa en la herramienta Cochrane RoB para ensayos controlados aleatorizados (ECA). Al igual que ésta, el riesgo de sesgo se evalúa dentro de dominios específicos y se solicita a los revisores que documenten la información en la que se basan los juicios. Tanto ROBINS-I como la herramienta Cochrane RoB se centran en la validez interna de los estudios, y estiman el sesgo de un estudio evaluando en cuánto difieren sus resultados en comparación con los resultados esperados de un ECA realizado de manera impecable en el mismo grupo de participantes (p. ej., un ensayo grande que logró el ocultamiento de la asignación aleatoria; mantuvo el cegamiento de los pacientes, los profesionales de la salud y los evaluadores de resultados a la intervención; no tuvo pérdidas de seguimiento; e informó los efectos de la intervención para todos los resultados medidos). Este sesgo es distinto de los problemas de generalización (aplicabilidad) a individuos excluidos del estudio. P. ej., restringir la muestra del estudio a individuos sin comorbilidades puede limitar la utilidad de sus hallazgos porque no pueden generalizarse a la práctica clínica, donde las comorbilidades son comunes.

Entre los sesgos evaluados, vale la pena destacar los que suceden en el período preintervención. Se debe tener en cuenta que la aleatorización es el único medio de asignación que distribuye los confundidores desconocidos (y por lo tanto, no mensurables) de manera equilibrada entre los grupos intervención y control. Es por esto que los estudios no aleatorizados son muy

susceptibles al sesgo de selección (p. ej., mediante la exclusión de pacientes elegibles) generando ramas con diferente pronóstico y distorsionando el efecto en cualquier dirección (aunque con más frecuencia tienden a sobreestimar los efectos del tratamiento). El sesgo de selección puede darse también sobre etapas del seguimiento. P. ej., se sabe que los anticonceptivos orales se asocian a mayor riesgo de trombosis venosa al inicio del tratamiento. Si se comparan nuevos progestágenos recientemente iniciados con otros iniciados hace más tiempo podría generarse un sesgo en contra de los nuevos progestágenos. También está el sesgo por factores de confusión, que ocurre cuando uno o más factores al mismo tiempo influyen sobre la probabilidad o sobre la magnitud de la intervención que recibe una persona y el desenlace de interés. P. ej., esto puede ocurrir en el caso del factor confusor edad avanzada; estos pacientes tienen menos probabilidades de recibir una opción de tratamiento quirúrgico y más probabilidades de presentar síntomas graves con cualquier tratamiento, lo cual puede producir un sesgo en la estimación del efecto causal atribuido a la intervención.

La Tabla 2 describe la interpretación del riesgo de sesgo para un dominio individual. Se deben considerar todos los dominios de sesgo en los ECA, excepto el sesgo derivado del proceso de asignación al azar (por definición, no se realiza la asignación de forma aleatoria); sin embargo, debe considerarse que también hay fuentes adicionales de sesgo.

Tabla 2. Interpretación de la evaluación del riesgo de sesgo para un dominio individual ROBINS-I. Fuente: Sterne JA, et al. *BMJ*. 2016;355:i4919¹

Juicio de riesgo de sesgo	Interpretación
Bajo	Estudio comparable a un ensayo clínico aleatorizado bien realizado con respecto a este dominio
Moderado	Estudio sólido, bien realizado (pero no comparable a un ensayo clínico aleatorizado) con respecto a este dominio
Serío	Estudio con algunos problemas importantes en este dominio.
Crítico	Estudio con demasiados problemas en este dominio como para proporcionar evidencia útil sobre los efectos de la intervención.
Sin información	No hay información sobre la cual basar un juicio para este dominio.

ROBINS-E

Los estudios epidemiológicos observacionales son clave para la evaluación de los efectos de ciertas exposiciones en la salud humana, tales como las exposiciones ambientales, ocupacionales y conductuales, en las cuales la evaluación a través de ECA generalmente no resulta factible. Incluso cuando ha sido posible realizar un ECA, la evidencia que proporciona suele presentar limitaciones. P. ej., los ECA suelen tener un tiempo de seguimiento menor al necesario para observar los efectos de las exposiciones mencionadas. Por lo tanto, la evidencia en estos casos suele provenir de estudios observacionales que implementan métodos apropiados para minimizar el riesgo de sesgo en sus resultados, y de revisiones sistemáticas de este tipo de estudios. Estas revisiones deben incluir, siempre que sea factible y apropiado, un meta-análisis de los resultados reportados. Además, la interpretación de los resultados de las revisiones sistemáticas puede también ser informada por síntesis adicionales, p. ej., sobre la base de la triangulación de resultados de diferentes tipos de estudios, y otros tipos de evidencia, como estudios en animales y mecanicistas.

Las revisiones sistemáticas de los estudios epidemiológicos observacionales, al igual que las revisiones sistemáticas de los ECA, deben incorporar evaluaciones del riesgo de sesgo en los resultados de los estudios presentados. Tales evaluaciones pue-

den ser útiles para seleccionar los resultados a incluir en un meta-análisis, así como para interpretar el conjunto de evidencias resumidas en la revisión y la solidez de la evidencia proporcionada por un meta-análisis. No obstante, los estudios con mayor riesgo de sesgo también pueden proporcionar pruebas importantes del efecto de la exposición sobre el resultado de interés. P. ej., la magnitud estimada del efecto puede ser lo suficientemente grande como para que haya evidencia clara de un efecto importante, a pesar de una cantidad plausible de sesgo.

Las evaluaciones del riesgo de sesgo requieren de habilidades y conocimientos, incluida la experticia en el contenido (conocimiento detallado de la materia) y la experiencia metodológica (conocimiento de los diferentes tipos de sesgo que pueden distorsionar los resultados de los estudios observacionales).

Las evaluaciones de riesgo de sesgo para los estudios de exposición deben considerar la dirección del sesgo y la probabilidad de que su magnitud tenga un efecto importante en la interpretación de los resultados. P. ej., si la dirección predicha del sesgo tiende hacia la nulidad del efecto, un resultado positivo puede proporcionar pruebas sólidas de un efecto importante. Estas mismas consideraciones aplican para la evaluación de los estudios de síntesis o meta-análisis. P. ej., la comparación de estudios de calidad variable puede aclarar en qué medida la clasificación errónea de

la exposición atenúa los efectos estimados.

Suele ser útil describir la capacidad general de un estudio para proporcionar información sobre el efecto de la exposición sobre el resultado. Este proceso es comúnmente conocido como la *sensibilidad del estudio*. La determinación de la sensibilidad del estudio incluye no sólo la evaluación del riesgo de sesgo sino también de otros aspectos del diseño, tales como valorar si el tiempo de seguimiento es lo suficientemente largo en relación con la duración esperada del efecto de la exposición, o si la amplitud del intervalo de niveles de exposición es adecuada en relación a los niveles que pueden dar lugar a diferencias notables en el resultado de interés durante el período de seguimiento estudiado.

La construcción de la herramienta ROBINS-E para la evaluación de estudios de cohorte siguió los principios ya adoptados para el desarrollo de la herramienta ROBINS-I, en la que la evaluación del riesgo de sesgo se subdivide en dominios y se basa en las respuestas a las *preguntas de señalización* de cada uno de ellos.

Un punto de partida para evaluar el riesgo de sesgo es definir el efecto esperado de la exposición sobre el resultado en ausencia de sesgo. La especificación de este efecto debe estar libre de las limitaciones involucradas en el intento de estimarlo utilizando la epidemiología observacional. Para definirla, será necesario recabar información sobre la población en la que se aplica el efecto, el resultado y la exposición, así como realizar comparaciones entre los diferentes patrones de exposición de interés (comparación con *contrafactuales*). P. ej., una metodología para formalizar la especificación del efecto de la exposición sobre el resultado es definir un *experimento objetivo hipotético o emulado* cuyos resultados estimarían este efecto sin sesgo (en el caso de de la herramienta ROBINS-I, este experimento lo constituirían los ECA).

Por este motivo, antes de realizar una evaluación aplicando ROBINS-E (o una serie de evaluaciones, p. ej., en el contexto de una revisión sistemática), los usuarios de la herramienta deben tener en mente un hipotético *experimento objetivo sin sesgo* y especificar los factores de confusión que probablemente influyan en la asociación entre la exposición y el resultado. Además, resultará importante recopilar los métodos que se podrían utilizar para medir la exposición y el resultado, y los tipos de análisis estadísticos que serían apropiados para estimar el efecto de la exposición sobre el resultado. Asimismo, resulta adecuado establecer juicios tempranos sobre la validez de los métodos de medición, para evitar que estos juicios se vean influenciados por los resultados de los estudios bajo consideración.

La evaluación de un estudio con ROBINS-E debe basarse en la máxima cantidad posible de información disponible. Además de los artículos publicados que describen los métodos y resultados de un estudio, esta información puede derivarse también de reportes no publicados o a través de la correspondencia con los investigadores del estudio.

Evaluación de un estudio con la herramienta ROBINS-E

El punto de partida para la evaluación de un estudio es especificar el desenlace que se está evaluando para determinar el riesgo de sesgo (sección A). A continuación, una sección de *cribado* facilita la identificación de los resultados que tienen un riesgo muy alto de sesgo, lo que permite al usuario evitar una evaluación detallada (sección B). Luego, el usuario de la herramienta recopila información sobre los participantes, las medidas de exposición, los resultados y los métodos de análisis en los que se basa ese resultado (sección C). Asimismo, una característica clave del enfoque ROBINS-E es la especificación, para cada estudio, del efecto causal estimado por el resultado considerado

(sección D). Esto es esencial para la evaluación del riesgo de sesgo, porque el efecto causal define el resultado esperado (aparte del impacto de la variación del muestreo) en ausencia de sesgo. La Tabla 3 enumera los dominios de sesgo cubiertos por la herramienta ROBINS-E.

Tabla 3. Dominios de sesgo incluidos en la herramienta ROBINS-E. Fuente: ROBINS-E Development Group, 2022²

Dominio	Riesgo de sesgo
Dominio 1	Por confusión
Dominio 2	Derivado de la medición de la exposición
Dominio 3	Selección de participantes en el estudio (o el análisis)
Dominio 4	Debido a las intervenciones posteriores a la exposición
Dominio 5	Debido a la falta de datos
Dominio 6	Derivado de la medición del resultado
Dominio 7	Por selección del resultado informado

Cada dominio de sesgo en esta herramienta se aborda utilizando una serie de preguntas de señalización que tienen como objetivo recopilar información importante sobre el estudio y el análisis que se está evaluando. Una vez respondidas las preguntas de señalización pertinentes, se dictan tres juicios para cada dominio:

1. **El riesgo de sesgo**, entendiéndose por tal el riesgo de sesgo importante que tiene el potencial de afectar el efecto estimado de la exposición en el resultado. Esta evaluación se realiza siguiendo un algoritmo basado en las respuestas a las preguntas de señalización. La Tabla 4 describe los potenciales resultados. ROBINS-E tiene por objeto proporcionar un marco para hacer juicios informados y razonables sobre el riesgo de un sesgo concreto de los efectos de la exposición sobre el resultado. En ocasiones, las respuestas a las preguntas de señalización pueden no conducir a una valoración adecuada del riesgo de sesgo. Por lo tanto, dado que la transparencia y la razonabilidad priman sobre la adhesión mecanicista a cada palabra de la herramienta, se puede obviar el riesgo de sesgo sugerido esquemáticamente por esta herramienta de evaluación si se proporciona una justificación para ello.
2. **La dirección prevista del sesgo**. Las opciones de respuesta dependerán del tipo de sesgo que se aborde.
3. **El riesgo de amenazar las conclusiones**. Este juicio hace referencia a la posibilidad de encontrar un riesgo de sesgo en un dominio en particular lo suficientemente alto como para amenazar las conclusiones del estudio (acerca de si la exposición tiene un efecto importante sobre el resultado). Este juicio debe tener en cuenta tanto el resultado del estudio (incluida su magnitud y la fuerza de la evidencia a su alrededor) como una evaluación amplia de sesgo (a través de la probabilidad de que esté presente, su dirección probable y su magnitud probable).

Después de completar los siete dominios de sesgo, se hace un juicio general para cada uno de los tres aspectos descritos: 1) la presencia o ausencia de sesgo, 2) su dirección prevista y 3) el riesgo de que su presencia amenace las conclusiones.

La evaluación para el primero y el tercero se realiza a nivel de cada dominio mediante la utilización de un algoritmo. La categoría que arroja dicho algoritmo puede ser modificada por el autor, pero se debe proporcionar una justificación cuando se anula el juicio general sugerido por el algoritmo.



Tabla 4. Interpretación de la valoración del riesgo de sesgo para un dominio individual ROBINS-E. Nota: *Para el Dominio 1 (sesgo debido a confusión), en el caso de que el riesgo de sesgo se considere bajo, la valoración del riesgo debe denominarse *Bajo riesgo de sesgo (excepto por las preocupaciones sobre la confusión no controlada)*, debido a que no se puede excluir un riesgo de sesgo debido a la confusión no controlada en un estudio observacional. Fuente: ROBINS-E Development Group, 2022²

Juicio riesgo de sesgo	Interpretación
Bajo *	Hay poca o ninguna preocupación por el sesgo con respecto a este dominio
Algunas preocupaciones	Existe cierta preocupación por el sesgo con respecto a este dominio, aunque no está claro que exista un riesgo importante de sesgo.
Alto	El estudio tiene algunos problemas importantes en este dominio que dan lugar a un alto riesgo de sesgo
Muy alto	El estudio es muy problemático en este ámbito. Las características del estudio dan lugar a un riesgo muy alto de sesgo

Comentario

De manera desafortunada, no hay herramientas universalmente aceptadas para evaluar el riesgo de sesgo de los estudios observacionales comparativos. El hecho de que una revisión sistemática haya identificado 86 herramientas específicas refleja este problema^{3,4}. Además, varias de ellas incluyen elementos de calidad de reporte que no son necesariamente relevantes para el riesgo de sesgo. Sin embargo, la mayoría de las herramientas abarcan los siguientes seis ítems fundamentales:

1. Selección de participantes. ¿Es la muestra adecuada y similar a la población de interés; se minimiza la posibilidad de sesgo de selección?
2. Definición, medición y recolección de las variables. ¿Se recolectaron todas las variables de confusión conocidas, y el análisis está ajustado por las mismas?
3. Métodos de control de confundidores. ¿Se han identificado y tenido en cuenta de forma adecuada en el diseño y en el análisis del estudio los principales elementos de confusión posibles? Valora además si se realiza un ajuste por los factores pronósticos importantes ¿Se ha realizado un modelo de análisis multivariable que ajuste por dichos factores de confusión?
4. Comparabilidad entre los grupos. ¿Los grupos son comparables; se minimiza la posibilidad de sesgo de selección diferencial?
5. Métodos estadísticos. Excluyendo el control de confundidores, ¿el análisis es adecuado?

Conclusiones del comentarador

El desarrollo de ROBINS-I para ENAI y la recientemente lanzada ROBINS-E para exposición en estudios de cohortes constituyen sin duda avances metodológicos clave que permitirán una evaluación más precisa de la confianza en la evidencia proveniente de estudios no experimentales. Aunque la versión actual de ROBINS-E está diseñada para estudios de cohorte, será esencial extenderla luego a otros diseños, como los estudios de casos y controles.

Agustín Ciapponi [Servicio de Medicina Familiar y Comunitaria, Hospital Italiano de Buenos Aires; Centro Cochrane Argentina, Instituto de Efectividad Clínica y Sanitaria. agustin.ciapponi@hospitalitaliano.org.ar]

Ciapponi A Herramientas ROBINS para evaluar el riesgo de sesgo de estudios no aleatorizados. Evid Actual Pract Ambul. 2022;25(3):e007024. Available from: <https://dx.doi.org/10.51987/EVIDENCIA.V25I4.7024>. Comentado de: Sterne JA, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919. doi: 10.1136/bmj.i4919. PMID: 27733354 y ROBINS-E Development Group (Higgins J, et al). Risk Of Bias In Non-randomized Studies - of Exposure (ROBINS-E). Launch version, 1 June 2022. Available from: <https://www.riskofbias.info/welcome/robins-e-tool>

Referencias

1. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919. Available from: 10.1136/bmj.i4919.
2. Higgins J, Morgan R, Rooney A, ROBINS-E Development Group, et al.. Risk Of Bias In Non-randomized Studies - of Exposure (ROBINS-E); 2022. Available from: <https://www.riskofbias.info/welcome/robins-e-tool> [Last access: 2022-08-08].
3. Sanderson S, Tatt ID, Higgins JPT. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. Int J Epidemiol. 2007;36(3):666–676. Available from: 10.1093/ije/dym018.
4. Ciapponi A. Guía de lectura crítica de estudios observacionales en epidemiología (segunda parte). Evid Actual Pract Ambul. 2011;14(1):7–9. Available from: 10.51987/evidencia.v14i1.6075;<https://www.evidencia.org.ar/index.php/Evidencia/article/view/6075>.