



Acta Medica Colombiana

ISSN: 0120-2448

Asociacion Colombiana de Medicina Interna

Muñoz, Óscar Mauricio; Ruiz Morales, Álvaro J.
Revisiones sistemáticas para la evaluación de intervenciones que
incluyen estudios no aleatorizados. Consideraciones metodológicas
Acta Medica Colombiana, vol. 43, núm. 2, 2018, Abril-Junio, pp. 100-106
Asociacion Colombiana de Medicina Interna

Disponible en: <http://www.redalyc.org/articulo.oa?id=163156703007>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

UAEH  redalyc.org

Sistema de Información Científica Redalyc
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso
abierto

Revisiones sistemáticas para la evaluación de intervenciones que incluyen estudios no aleatorizados

Consideraciones metodológicas

Systematic reviews for the evaluation of interventions that include non-randomized studies

Methodological considerations

ÓSCAR MAURICIO MUÑOZ, ÁLVARO J. RUIZ • BOGOTÁ, D.C. (COLOMBIA)

Resumen

Las revisiones sistemáticas para evaluar intervenciones que incluyan la evidencia aportada por los estudios no aleatorizados (ENA), pueden ser muy útiles para soportar decisiones clínicas en temas que no estén bien cubiertos por experimentos clínicos aleatorizados (ECA), como la evaluación de los efectos de un tratamiento a largo plazo, sus efectos adversos o la evaluación de intervenciones en salud pública u organizacionales. El desarrollo de estas RSL requiere una serie de ajustes metodológicos para la búsqueda y selección de los estudios primarios, la evaluación de la calidad de los mismos y la combinación de la evidencia. Especial cuidado debe tenerse para identificar y manejar adecuadamente el mayor riesgo de sesgos, en especial los derivados de un inadecuado control de los factores de confusión. El presente artículo presenta las particularidades a tener en cuenta a la hora de realizar RSL para evaluar intervenciones que incluyan ENA, las herramientas que se han desarrollado para facilitar su desarrollo, y las consideraciones importantes para reportar y discutir los hallazgos. El desarrollo de este tipo de RSL requiere un alto nivel de experiencia por parte del grupo desarrollador y una mayor cantidad de recursos económicos y de tiempo. Sin embargo, cuando han sido bien diseñadas, conducidas y analizadas, pueden ser de gran utilidad en la toma de decisiones clínicas, y pueden ser integradas en el desarrollo de las recomendaciones presentadas en las guías de práctica clínica. (*Acta Med Colomb* 2018; 43: 100-106).

Palabras clave: *revisiones sistemáticas, estudios no aleatorizados, estudios observacionales, sesgo.*

Abstract

Systematic reviews to evaluate interventions including the evidence provided by non-randomized studies (NRS), can be very useful to support clinical decisions in subjects that are not well covered by randomized clinical trials (RCTs), such as the evaluation of the effects of long-term treatment, its adverse effects or the evaluation of public health or organizational interventions. The development of these RSLs requires a series of methodological adjustments for the search and selection of the primary studies, the evaluation of their quality and the combination of the evidence. Special care must be taken to identify and appropriately manage the increased risk of bias, especially those derived from inadequate control of confounding factors. This article presents the particularities to take into account when performing RSL to evaluate interventions that include NRS, the tools that have been developed to facilitate their development, and the important considerations to report and discuss the findings. The development of this type of RSL requires a high level of experience by the developer group and a greater amount of economic resources and time. However, when they have been well designed, conducted and analyzed, they can be very useful in clinical decision-making, and can be integrated into the development of the recommendations presented in the clinical practice guidelines. (*Acta Med Colomb* 2018; 43: 100-106).

Key words: *systematic reviews, non-randomized studies, observational studies, bias.*

Dr. Óscar Mauricio Muñoz: Jefe Unidad Clínica Hospitalaria, Hospital Universitario San Ignacio. Profesor Asistente Departamento de Medicina Interna, Programa de Doctorado en Epidemiología Clínica. Pontificia Universidad Javeriana; Dr. Álvaro J. Ruiz Morales: Profesor Titular Departamento de Epidemiología Clínica y Bioestadística, Departamento de Medicina Interna, Pontificia Universidad Javeriana. Bogotá, D.C. (Colombia).

Correspondencia. Dr. Óscar Mauricio Muñoz. Bogotá, D.C. (Colombia).

E-mail: o.munoz@javeriana.edu.co

Recibido: 12/IX/2017 Aceptado: 10/IV/2018

Introducción

Existe actualmente gran interés en desarrollar revisiones sistemáticas para evaluar intervenciones que incluyan la evidencia aportada por los estudios no aleatorizados (ENA), y que puedan soportar decisiones clínicas en temas que no estén bien cubiertos por experimentos clínicos aleatorizados (ECA). Ejemplos de estos temas incluyen la evaluación de los efectos de un tratamiento a largo plazo y en especial la evaluación de efectos adversos, dado que éstos suelen ser escasos y aparecer en forma tardía; la evaluación de intervenciones que no puedan ser aleatorizadas por razones éticas o la evaluación de efectos de intervenciones realizadas en salud pública o intervenciones organizacionales (1).

El desarrollo de revisiones sistemáticas de la literatura (RSL) que incluyan ENA puede ser de gran utilidad como fuente de información en la toma de decisiones clínicas, tal como lo ha sugerido la colaboración Cochrane (2) y el grupo de trabajo GRADE (*The Grading of Recommendations Assessment, Development and Evaluation working group*) (3). Sin embargo, requiere mayor esfuerzo por parte de los investigadores, dada la gran cantidad de información que se debe considerar y la mayor dificultad que implica realizar la evaluación de calidad de los ENA primarios, que permita evaluar y manejar adecuadamente el mayor riesgo de un amplio rango de sesgos, así como los riesgos de confusión (4).

Este artículo presenta las particularidades de las RSL para evaluar intervenciones que incluyen ENA, así como las implicaciones que se deben tener en cuenta a la hora de diseñar, conducir y reportar este tipo de estudios clínicos.

Particularidades de los estudios no aleatorizados

En la Tabla 1 se presentan los diseños epidemiológicos no aleatorizados más frecuentemente utilizados. La principal diferencia entre los ECA y los ENA es precisamente la falta de aleatorización de los segundos. Cuando se hace aleatorización (idealmente con un adecuado cegamiento), se reduce la posibilidad de que haya diferencias sistemáticas entre los grupos de intervención y de control, dado que el azar tiende a distribuir equitativamente tanto las características conocidas como las desconocidas de los pacientes y, aún en el caso de que no haya distribución simétrica, las diferencias se deben justamente al azar y no son producto de un sesgo.

En los ENA la asignación de los pacientes a los grupos depende de otros factores, con frecuencia desconocidos, lo que aumenta el riesgo de sesgos de selección. Cuando este sesgo de selección produce desbalances en los factores pronósticos entre los grupos de intervención y de control, y cuando estos factores pronósticos se relacionan con el desenlace se dice que se trata de factores de confusión. A continuación, se presentan ejemplos de algunos de los factores de confusión más frecuentemente encontrados en los ENA:

- **Edad:** si en un estudio de cohortes que busca determinar el impacto de la sarcopenia en la mortalidad de los pacientes hospitalizados se detecta que la edad promedio de los pacientes con sarcopenia es mayor que el promedio de edad de los pacientes sin esta condición, los resultados que demuestren el aumento de la incidencia de mortalidad asociada con la sarcopenia pueden estar sesgados dado que la verdadera razón del aumento de este riesgo puede deberse a la edad de los pacientes.

Tabla 1. Diseños epidemiológicos no aleatorizados más frecuentemente utilizados.

Estudio	Características principales
Casos y controles	Estudio que compara personas con un desenlace específico de interés (casos) con personas que provienen de la misma población, pero sin el desenlace (controles), en busca de asociaciones entre el desenlace y algunas exposiciones previas.
Cohortes (Retrospectivas o prospectivas)	Se realizan observaciones a través del tiempo en una serie de individuos (cohorte), a partir de diferentes exposiciones conocidas y buscando la aparición de los desenlaces de interés.
Serie de casos	Se realizan observaciones de una serie de individuos que generalmente reciben una misma intervención, comparando los desenlaces antes y después de esa intervención, pero sin un grupo de control.
Controlados concurrentemente	Compara un grupo de participantes con un grupo de control contemporáneo. Puede usarse una asignación de intervenciones experimental (pero no aleatorizada).
Controlados históricamente	Compara un grupo de pacientes que recibe una intervención con un grupo similar pasado que no la recibió
Corte transversal	Un estudio que recolecta información, en un mismo momento, de intervenciones (pasadas o presentes) y desenlaces (presentes) buscando asociaciones entre ellas.
Series de tiempo	Un estudio que compara las mediciones antes y después de la implementación de una intervención, con un grupo control que no recibe la intervención.

- **Severidad de la enfermedad:** si en un estudio retrospectivo, diseñado para evaluar el impacto de la vacunación contra neumococo en la frecuencia de exacerbaciones de pacientes con enfermedad pulmonar obstructiva crónica (EPOC), los investigadores encuentran que esta intervención aumenta en lugar de reducir la frecuencia de las exacerbaciones, la explicación puede estar dada porque los pacientes que recibieron la vacunación correspondían a casos más severos de EPOC y no a la intervención como tal.
- **Diferentes momentos temporales:** un estudio de antes y después que evalúa el impacto de la disponibilidad de un equipo de intervencionismo coronario en una institución demuestra que la mortalidad asociada con infarto agudo de miocardio disminuyó después de que estuvo disponible la tecnología; el resultado puede estar sesgado si no se consideran las diferencias en el uso de nuevas terapias farmacológicas, no disponibles en la fase inicial, que también podrían haber influido en la reducción de la tasa de mortalidad

Algunos otros aspectos de los ECA pueden ser limitaciones que los ENA pueden ayudar a resolver (5). No son defectos de los ECA, puesto que son su garantía de calidad y de validez interna, pero sí pueden ser limitantes en el momento de la aplicación a la vida real. Los principales aspectos son:

- **Criterios de selección muy restrictivos:** aunque se busque tener criterios de admisión amplios, el ECA necesita una definición precisa de su población, que puede limitar la validez externa del resultado. Generalmente se omiten comorbilidades y cointervenciones que pueden dar mejor validez interna al estudio pero que limitan las poblaciones a las que pueda generalizarse.
- **Tiempos de estudio cortos:** el costo y dificultades logísticas del experimento pueden dificultar el estudio de eventos que requieren tiempos largos de exposición, o que son de muy baja frecuencia.
- **Alto costo:** la planeación y ejecución de un ECA en general necesita considerablemente más recursos que un estudio de casos y controles, o que un estudio de corte transversal o de cohortes.
- **Estandarización de intervenciones, diagnósticos y procesos:** necesarios para garantizar la calidad de los ECA, pueden ser una limitación porque no reflejan la variabilidad y laxitud de los procesos en la vida real.

Más que contraponerlos, deben verse los ECA y los ENA como complementarios, y la información que provee cada diseño debe ser críticamente evaluada. Se habla de la Investigación de Efectividad Comparativa (IEC, en inglés CER por las iniciales de *Comparative Effectiveness Research*) (6), como una herramienta de gran importancia en la generación de evidencia que pueda guiar las decisiones en salud y en la que se propone que los ENA complementen la información de los ECA (7,8). La mayoría de los estudios en IEC son ENA (54 % en un estudio de Benson) (9) por lo

que se hace necesario tener reglas claras para la inclusión de esos estudios, que tienen alto potencial de sesgo por la ausencia de aleatorización.

Implicaciones para el diseño y la conducción

Las particularidades de los ENA hacen necesarios varios cuidados adicionales cuando se hagan RSL que los incluyan, que los hacen diferentes a las mucho más frecuentes RSL de ECA.

Búsqueda de los estudios primarios

Cuando realizamos RSL de ECA frecuentemente utilizamos filtros para seleccionar exclusivamente este tipo de estudios. El desarrollo de filtros para buscar ENA ha sido difícil, dado que varios estudios con diseños metodológicos similares pueden haber sido descritos de forma muy diferente por diversos autores. De igual forma bajo una misma etiqueta (por ejemplo, cohorte) pueden estar descritos inadecuadamente estudios con características de diseño muy disímiles. La escasez de filtros válidos y confiables lleva a que deban considerarse en muchas ocasiones todos los estudios que se refieran a la intervención y a los desenlaces de interés, lo que con frecuencia puede generar volúmenes inmensos de información. Algunos filtros sin embargo han demostrado utilidad y pueden considerarse para la búsqueda de estudios observacionales controlados (10). Una vez identificados, la pobre descripción del diseño metodológico que se incluye en los resúmenes hace prácticamente imposible determinar todas las características necesarias para clasificar adecuadamente un ENA, por lo que es necesario revisar los textos completos de todos los estudios. Esto implica consumos de tiempo y recursos económicos y humanos muy grandes, que deben ser considerados cuidadosamente a la hora de planear un estudio de este tipo (4).

La búsqueda de estudios no publicados también resulta desafiante si se tiene en cuenta que frecuentemente los ENA fueron realizados con datos recolectados para otro propósito, y que sólo en muy contadas ocasiones los protocolos de los ENA han sido registrados prospectivamente en bases de datos de investigación en curso.

Selección de los estudios

Para determinar los criterios de selección de los estudios para una RSL deben tenerse en cuenta los componentes de la estructura PICO de la pregunta, para que se cumplan criterios previamente definidos de población (P), definición clara de la intervención (I) que se va a estudiar, definición a priori de las intervenciones de comparación (C) de la intervención en estudio, así como los desenlaces de interés (O: *Outcomes*) que deben estar reportados en los estudios primarios. Cuando es relevante, debe también considerarse el tiempo (T) de exposición.

Para el desarrollo de una RSL que incluya ENA deben tenerse en cuenta otros aspectos:

¿Deben incluirse tanto ECA como ENA?

Si existen ECA de buena calidad que den respuesta a la pregunta de interés de manera comprensiva, no es recomendable incluir ENA, dado que existe un mayor riesgo de sesgos. Si la respuesta es negativa, es decir no existen ECA de alta calidad, no es esperable que existan a corto plazo, no es posible realizarlos por razones éticas o hay aspectos no resueltos por los ECA, y la respuesta a la pregunta de investigación es prioritaria para la toma de decisiones clínicas, se debe considerar la inclusión de ENA, siempre buscando garantizar que los estudios que se incluyan tengan el menor potencial de sesgo posible.

¿Deben incluirse todos los ENA disponibles?

Dado que resulta prioritario garantizar que los estudios tengan el menor potencial de sesgos, es necesario hacer una evaluación cuidadosa de los estudios primarios en busca de las características de diseño que permitan minimizar los sesgos. El investigador debe seleccionar aquellos estudios que cumplan con las características mínimas de calidad.

Evaluación de la calidad

La evaluación de la calidad debe estar dirigida a detectar las principales fuentes de sesgo y el manejo que los autores de los ENA primarios dieron a éstos o a su riesgo potencial.

Aunque hasta el momento se han desarrollado múltiples herramientas para la evaluación de la calidad de los ENA (11), sólo algunas de ellas han alcanzado amplia aceptación. Se destacan la escala de *Newcastle Ottawa* (NOS) (12), muy usada en la década pasada, la lista de chequeo de Wells (13), y la herramienta desarrollada por la colaboración Cochrane, *ACROBAT-NRSI* (por sus siglas en inglés, *A Cochrane Risk of Bias Tool for Non-Randomized Studies of Interventions*) (14) que fue inicialmente construida como una adaptación de la herramienta Cochrane para la evaluación del riesgo de sesgos en ECA (15), y que fue actualizada en 2016 y se conoce ahora como *ROBINS-I* (*Risk Of Bias in Non-randomized Studies - of Interventions*) (16).

Como se mencionó previamente, el principal riesgo de los ENA es la alta probabilidad de que los factores de riesgo hayan quedado repartidos inequitativamente entre los grupos de intervención y de control (sesgo de selección), y que dichos factores se relacionen directamente con el desenlace (factores de confusión). La herramienta NOS (12) propone un abordaje inicial de este riesgo: se debe identificar el principal factor pronóstico que pueda comportarse como factor de confusión y debe hacerse un listado de otros factores que pudieran tener el mismo comportamiento; finalmente, debe evaluarse si el estudio los tuvo o no en cuenta. Wells (13) propone complementar este abordaje con la evaluación de cómo los investigadores decidieron sobre los factores de confusión para considerar, si los investigadores controlaron por los factores de confusión desde el diseño (por ejemplo si realizaron emparejamiento utilizando índices de propensión) (17),

o si ajustaron por los factores de confusión en la fase de análisis, y en caso de que se haya hecho se debe haber definido el método utilizado (por ejemplo estratificación, regresión multivariable o, nuevamente, utilizado índices de propensión).

Como puede notarse este proceso requiere un alto grado de conocimiento en el área clínica en la que se está desarrollando la RSL, que permita identificar correctamente los posibles factores de confusión relevantes, y además un alto grado de pericia metodológica para evaluar si el manejo que se dio a estos factores de confusión es el adecuado. Todas estas consideraciones deberán ser tenidas en cuenta a la hora de seleccionar el recurso humano necesario para realizar la investigación.

Adicionalmente, la herramienta *ROBINS-I* (16) permite evaluar otras potenciales fuentes de sesgo, similares a aquellas que se encuentran en los ECA, como: la clasificación de las intervenciones o exposiciones, el tratamiento diferencial, la evaluación diferencial de los resultados, el manejo de los datos perdidos y el reporte selectivo. En la Tabla 2 se presentan los dominios para la evaluación de sesgos de la herramienta *ROBINS-I* y una corta explicación de cómo éstos pueden afectar los ENA.

Merece especial atención el riesgo de sesgo por reporte selectivo del análisis y del desenlace, dado que, al tratarse de estudios exploratorios, la probabilidad de que en los ENA sólo se reporten los resultados con desenlaces positivos es mucho más alta. Para evaluar el riesgo de sesgo por reporte selectivo, se propone intentar recopilar la información del protocolo del estudio y de toda la documentación previa al inicio del reclutamiento de los pacientes. Las claves que sugieren la presencia de reporte selectivo incluyen diferencias entre las secciones de métodos y resultados de la publicación, diferencias entre los desenlaces propuestos en el protocolo y los presentados en el reporte final del estudio, o un lapso de tiempo muy largo entre el registro del protocolo, la finalización de la recolección de los datos y la publicación final (18).

De acuerdo con la evaluación de todos estos potenciales sesgos, la herramienta *ROBINS-I* permite clasificar para cada estudio y para cada desenlace el riesgo de sesgo global, definiendo si éste es bajo, moderado, serio o crítico, lo que permite articular esta evaluación con modelos formales de toma de decisiones o generación de recomendaciones como los propuestos por el grupo *GRADE* (16).

Métodos para controlar el efecto de los factores de confusión dentro de los ENA

Existen nuevos diseños y nuevas formas de análisis que buscan controlar el efecto de los factores de confusión dentro de los ENA. Vale la pena destacar los diseños basados en modelos de regresión discontinua, los diseños con base en técnicas de variables instrumentales, y los ajustes que se pueden realizar en la selección de pacientes y en el análisis de la información con las pruebas de propensión (17, 19).

Tabla 2. Dominios incluidos en la herramienta ROBINS-I. Adaptado de (16).

	Dominio	Explicación
Previo a la intervención	1. Sesgo por factores de confusión	Uno o más factores pronósticos se asocian tanto con la intervención recibida al inicio del estudio, como con el desenlace
	2. Sesgo por selección de los pacientes en el estudio	Cuando la exclusión de algunos participantes, o el tiempo de seguimiento inicial se relacionan con exposición y desenlace. Por ejemplo. Cuando en un estudio se incluyen indistintamente casos prevalentes y casos nuevos
Al momento de la intervención	3. Sesgo en la clasificación de las intervenciones	Sesgos introducidos por mala clasificación, diferencial o no diferencial del estado de intervención. Si es no diferencial (no relacionada con el desenlace) el efecto estimado está sesgado hacia el resultado nulo. Si la mala clasificación es diferencial (relacionada con el desenlace) puede condicionar la presencia de sesgos en cualquier dirección.
Posterior a la intervención	4. Sesgo por desviación de las intervenciones planeadas	Cuando hay diferencias sistemáticas entre el cuidado administrado a los grupos de intervención y de control, que representa una diferencia con las intervenciones planeadas
	5. Sesgo por datos perdidos	Cuando el seguimiento, y los datos perdidos al seguimiento se distribuyen de forma diferentes entre pacientes con diferentes factores de riesgo. O cuando se excluye pacientes por tener datos perdidos en cuanto a su exposición o factores de confusión.
	6. Sesgo en la medición de los desenlaces	Cuando existen errores diferenciales o no diferenciales en la medición de los desenlaces. Por ejemplo, cuando el evaluador conoce el estado de intervención, o hay diferentes métodos para evaluar los desenlaces en los diferentes grupos de intervención.
	7. Sesgo en la selección de los desenlaces reportados.	Reporte selectivo en los resultados de una forma que lo hace dependiente de los hallazgos, evitando que los estimativos puedan ser incluidos en un metaanálisis.

En el diseño de regresión discontinua, la asignación de los participantes al grupo de tratamiento depende de la puntuación obtenida en una variable. Específicamente los participantes son asignados al tratamiento si el puntaje en una covariable excede un determinado punto de corte (por ejemplo, los pacientes con una severidad de la enfermedad mayor que un determinado punto de corte son tratados, y los menos severos no). El análisis busca relacionar esta covariable con el desenlace para obtener un estimativo del efecto, con base en las diferencias detectadas alrededor de este punto de corte (20). En las técnicas de variables instrumentales se busca determinar el estimativo del efecto a partir de la información obtenida al aislar el efecto específico de una covariable (generalmente una intervención terapéutica), del posible error derivado de otras variables de confusión. Para ello se selecciona un instrumento que se asocia con la covariable de interés, pero no con los términos de error o con el desenlace (21).

Todos estos diseños generan estimativos del tamaño del efecto a partir de análisis específicos, que buscan controlar el efecto de los potenciales sesgos de selección a partir de una medición cuidadosa y válida de las variables de confusión conocidas, pero son incapaces de controlar el efecto de variables no medidas o no conocidas.

Combinación de la evidencia

Al momento de combinar la evidencia proporcionada por los ENA de manera cuantitativa en un metaanálisis, se

debe tener en cuenta que existen diferentes diseños metodológicos, y que todos ellos pueden proporcionar diferente fortaleza de la evidencia. Como regla general debe aceptarse que los ENA proporcionan menores niveles de evidencia que los ECA, y que en general los sesgos potenciales son mayores, con una magnitud y dirección inciertas (22).

De igual forma los diferentes diseños pueden ofrecer diferentes formas de análisis para controlar el efecto de las variables de confusión. Por ejemplo, la mayor parte de los ENA corresponde a estudios controlados, lo que los hace más parecidos a los ECA y facilita el metaanálisis de la información que proporcionan. Los ENA con grupo control generalmente presentan el control de las variables de confusión utilizando análisis de covarianza (ANCOVA). El análisis ANCOVA representa un intento de estimar el efecto promedio del tratamiento, como si los grupos hubiesen sido equivalentes en cuanto a las covariables al momento de ingreso al estudio. Los resultados ajustados que proporciona ANCOVA pueden ser interpretados como el mejor estimativo del tamaño del efecto, y pueden ser utilizados en conjunto con las desviaciones estándar no ajustadas para ser incluidas en el metaanálisis. Obviamente ANCOVA sólo puede ajustar los estimativos por las covariables que conoce el investigador, dejando potenciales fuentes de sesgo derivadas de covariables no medidas o no conocidas.

Dado que cada estudio puede tener sesgos diferentes, los metaanálisis que incluyen ENA tendrán una mayor heterogeneidad. Así pues, será indispensable explorar cuidadosa-

mente las fuentes de heterogeneidad y en especial evaluar si resultan diferentes los estimativos derivados de los ECA de aquellos obtenidos a partir de los ENA. Como regla general se prefiere presentar de forma separada el resultado del metaanálisis de ECA y del metaanálisis ENA, siendo opcional la presentación del resultado de la combinación entre ellos. Este análisis deberá ir acompañado de las posibles explicaciones a la heterogeneidad encontrada. Así mismo deberá realizarse la aproximación a partir de modelos de efectos aleatorios, que tiene en cuenta la heterogeneidad entre los resultados de los estudios. Cabe resultar que los estimativos de los intervalos de confianza calculados bajo estas condiciones sólo representan la variación estadística identificable, y no reflejan toda la incertidumbre asociada a la falta de información en cuanto a la dirección o la magnitud de los sesgos en los estudios primarios (22-23).

Dado que cada análisis puede controlar por diferentes factores de confusión, puede resultar incorrecto intentar combinar los resultados con los de ENA con diseños diferentes. Resulta pues recomendable considerar la comparación de los resultados ajustados y no ajustados, y presentar esta información de forma diferencial (19).

Algunos diseños de ENA como las series de tiempo temporal, en los que se comparan dos momentos del tiempo, antes y después de una intervención, sin un grupo de control resultan imposibles de combinar matemáticamente con otros ENA y ECA en un metaanálisis.

Discusión y reporte final

Todas las limitaciones de las RSL de ENA previamente comentadas obligan a ser muy cautelosos a la hora de presentar la discusión de los resultados y de escribir el reporte final para publicación.

En general se recomienda hacer una interpretación muy conservadora de los resultados, presentando explícitamente las limitaciones de la evidencia proporcionada por los ENA primarios, especialmente los potenciales sesgos encontrados, y también presentar las limitaciones encontradas a la hora de hacer la combinación de la información.

Si se presenta un estimativo global de resumen, en la discusión deberá hacerse énfasis en la interpretación de los intervalos de confianza, dado que ellos buscan explicar los posibles errores del estimativo, derivados del muestreo, pero no permiten controlar los errores dependientes de los sesgos de los estudios de base.

Si los investigadores deciden presentar en la misma RSL información proveniente de ECA y de ENA, se recomienda analizarla de forma independiente y comentar sobre las similitudes y diferencias encontradas en los resultados, a manera de análisis de sensibilidad.

Los editores de las revistas científicas han presentado llamados a estandarizar y mejorar la calidad del reporte de las RSL de intervenciones que incluyen ENA (24), con recomendaciones que sean de utilidad, para los autores, revisores y editores de revistas. Para ello se ha propuesto

una lista de chequeo análoga a herramienta PRISMA (25) que utilizamos para optimizar el reporte de RSL y metaanálisis de ECA. Para las RSL de intervenciones que incluyan ENA, la herramienta propuesta es MOOSE (*Meta-analysis of Observational Studies in Epidemiology*) (26). La lista de chequeo propuesta contiene especificaciones para el reporte de la estrategia de búsqueda, métodos, resultados discusión y conclusiones.

Utilidad de las RSL que incluyen ENA para la generación de recomendaciones

A pesar de las múltiples limitaciones, las RSL de ENA tienen un papel importante para informar a los tomadores de decisiones en el desarrollo de guías de práctica clínica, tal como lo ha propuesto el grupo de trabajo GRADE (3).

Se ha propuesto que las RSL de ENA pueden ayudar en este proceso de tres formas: como fuente de evidencia indirecta para complementar los juicios, como fuente de evidencia secuencial e incluso como reemplazo de la evidencia proporcionada por los ECA (27).

Con frecuencia los miembros de los grupos desarrolladores de GPC deben generar recomendaciones con evidencia que no se corresponde exactamente con la pregunta que desean responder. Por ejemplo pueden querer crear recomendaciones específicas para población mayor de 65 años, pero la información disponible no tiene en cuenta a pacientes de esa edad, o pueden querer definir el impacto de un tratamiento en desenlaces duros (por ejemplo mortalidad), pero la única información que proporcionan los ECA está dada para desenlaces intermedios (por ejemplo niveles de hemoglobina glucosilada); en estos casos la información contenida en los ENA puede fortalecer las conclusiones y las recomendaciones.

Para el ejemplo presentado, un metaanálisis de estudios observacionales a largo plazo que documente una relación entre los niveles de hemoglobina glucosilada (HbA1c) y mortalidad podría fortalecer la recomendación de dar un tratamiento que baje los niveles de HbA1c.

Se ha propuesto que la información contenida en los ENA pueda ayudar en la toma de decisiones al proporcionar información específica para subgrupos de la población, con diferentes riesgos basales, o puedan ayudar a resolver inconsistencias entre la evidencia proveniente de ECA.

En otras ocasiones los ENA pueden complementar la evidencia insuficiente proporcionada por los ECA. Por ejemplo, en ocasiones un ENA puede complementar la información de los desenlaces a largo plazo (con frecuencia de efectos adversos), de un estudio que inicialmente se planeó solo como un ECA a corto plazo para evaluar un desenlace intermedio.

En algunas ocasiones los ENA pueden proporcionar un nivel de evidencia mayor que la de los ECA. Por ejemplo, si los únicos ECA disponibles no incluyen poblaciones, intervenciones o desenlaces similares a aquellos que se quiere identificar para responder a la pregunta de una GPC,

la evidencia indirecta que proporcionan los ECA puede ser incluso de menor calidad que la evidencia proporcionada por un ENA bien diseñado y por tanto con un potencial de sesgo de selección bajo.

Conclusiones

El desarrollo de RSL de ENA tiene un espacio importante para resolver los vacíos de información que no es posible de responder en ECA, ya sea por razones éticas o logísticas. La RSL de intervenciones que incluye ENA requiere una serie de cuidados importantes para poder identificar y manejar adecuadamente los riesgos mayores de sesgos, en especial los derivados de un inadecuado control de los factores de confusión. El desarrollo de este tipo de RSL requiere un alto nivel de experiencia por parte del grupo desarrollador y una mayor cantidad de recursos económicos y de tiempo. Sin embargo, cuando han sido bien diseñadas, conducidas y analizadas, pueden ser de gran utilidad en la toma de decisiones clínicas.

Referencias

1. Reeves BC, Higgins JPT, Ramsay C, Shea B, Tugwell P, Wells G a. An introduction to methodological issues when including non-randomised studies in systematic reviews on the effects of interventions. *Res Synth Meth*. 2013; 4 (1): 1–11.
2. Reeves B, Deeks J, Higgins J, Wells G. Chapter 13: Including non-randomized studies. In: Higgins JPT, Green S (editors). In: Collaboration TC, editor. *Cochrane Handbook Syst Rev Interv*. Version 510 (updated March 2011). 2011. p. Available from www.handbook.cochrane.org.
3. GRADE working group. Grade handbook. 2016. Available from <https://gradepr.org/handbook/>
4. Higgins JPT, Ramsay C, Reeves B, Deeks J, Shea B, Valentine JC, et al. Issues relating to study design and risk of bias when including non-randomized studies in systematic reviews on the effects of interventions. *Res Synth Meth*. 2013; 4 (1): 12–25.
5. Silverman SL. From randomized controlled trials to observational studies. *Am J Med*. 2009; 122 (2): 114–20.
6. Goulart BHL, Ramsey SD, Parvathaneni U. Observational study designs for comparative effectiveness research: an alternative approach to close evidence gaps in head-and-neck cancer. *Int J Radiat Oncol Biol Phys*. 2014; 88 (1): 106–14.
7. Alemayehu D, Cappelleri JC. Revisiting issues, drawbacks and opportunities with observational studies in comparative effectiveness research. *J Eval Clin Pract*. 2013; 19 (4): 579–83.
8. Garabedian LF, Chu P, Toh S, Zaslavsky AM, Soumerai SB. Potential bias of instrumental variable analyses for observational comparative effectiveness research. *Ann Intern Med*. 2014; 161 (2): 131–8.
9. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med*. 2000; 342 (25): 1878–86.
10. Furlan AD, Irvin E, Bombardier C. Limited search strategies were effective in finding relevant nonrandomized studies. *J Clin Epidemiol*. 2006; 59 (12): 1303–11.
11. 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): 666–76.
12. Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M. The Newcastle–Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Ottawa: *Ottawa Hosp Resea Insti*. 2014. Available from http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
13. Wells G a, Shea B, Higgins JP, Sterne JA, Tugwell P, Reeves B. Checklists of methodological issues for review authors to consider when including non-randomized studies in systematic reviews. *Res Synth Meth*. 2013; 4 (1): 63–77.
14. Sterne J, Higgins J, Reeves B, Editors. A Cochrane risk of bias assessment tool: for non-randomized studies of interventions (ACROBAT-NRSI). Version 1.0.0. 2011.
15. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011; (343): 889–93.
16. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016; (355): i4919.
17. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res*. 2011; 46 (3): 399–424.
18. Norris S, Moher D, Reeves B, Shea B, Loke YK, Garner S, et al. Issues relating to selective reporting when including non-randomized studies in systematic reviews on the effects of healthcare interventions. *Res Synth Meth*. 2013; 4 (1): 36–47.
19. Valentine JC, Thompson SG. Issues relating to confounding and meta-analysis when including non-randomized studies in systematic reviews on the effects of interventions. *Res Synth Meth*. 2013; 4 (1): 26–35.
20. Moscoe E, Bor J, Bärnighausen T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: A review of current and best practice. *J Clin Epidemiol*. 2015; 68 (2): 122–33.
21. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol*. 2000; 29 (4): 722–9.
22. MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AMS. A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies. *Health Technol Assess (Rockv)*. 2000; 4 (34): 1–154.
23. Turner RM, Spiegelhalter DJ, Smith GCS, Thompson SG. Bias modelling in evidence synthesis. *J R Stat Soc Ser A Stat Soc*. 2009; 172 (1): 21–47.
24. Van Zuuren EJ, Fedorowicz Z. Moose on the loose: checklist for meta-analyses of observational studies. *Br J Dermatol*. 2016; 175 (5): 853–4.
25. Moher D, Liberati A, Tetzlaff J, Altman D. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2010; (8): b2535.
26. Stroup D, Berlin J, Morton S, Olkin I. MOOSE Statement: Meta-analysis of Observational Studies in Epidemiology: A Proposal for Reporting. *JAMA*. 2000; 283 (15): 2008.
27. Schünemann HJ, Tugwell P, Reeves BC, Akl E a., Santesso N, Spencer F a., et al. Non-randomized studies as a source of complementary, sequential or replacement evidence for randomized controlled trials in systematic reviews on the effects of interventions. *Res Synth Meth*. 2013; 4 (1): 49–62.