
BÚSQUEDA, EVALUACIÓN Y SÍNTESIS DE LA EVIDENCIA DE EFECTIVIDAD Y SEGURIDAD EN EVALUACIONES DE TECNOLOGÍA: MANUAL METODOLÓGICO



División de Protección Social y Salud
Banco Interamericano de Desarrollo
www.iadb.org/salud - scl-sph@iadb.org

Copyright © 2017 Banco Interamericano de Desarrollo. Esta obra se encuentra sujeta a una licencia Creative Commons IGO 3.0 Reconocimiento-NoComercial-SinObras-Derivadas (CC-IGO 3.0 BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) y puede ser reproducida para cualquier uso no-comercial otorgando el reconocimiento respectivo al BID. No se permiten obras derivadas.

Cualquier disputa relacionada con el uso de las obras del BID que no pueda resolverse amistosamente se someterá a arbitraje de conformidad con las reglas de la CNUDMI (UNCITRAL). El uso del nombre del BID para cualquier fin distinto al reconocimiento respectivo y el uso del logotipo del BID, no están autorizados por esta licencia CC-IGO y requieren de un acuerdo de licencia adicional.

Note que el enlace URL incluye términos y condiciones adicionales de esta licencia.

Las opiniones expresadas en esta publicación son de los autores y no necesariamente reflejan el punto de vista del Banco Interamericano de Desarrollo, de su Directorio Ejecutivo ni de los países que representa.



AUTORES

Hernando G. Gaitán Duarte MD, MSc.

Profesor Titular del Departamento de Obstetricia y Ginecología. Instituto de Investigaciones Clínicas.
Universidad Nacional de Colombia
Coordinador Editorial Grupo Cochrane de Infecciones de Trasmisión Sexual
Director del grupo de evaluación de Tecnologías y políticas en salud
Miembro de la Alianza Centro Nacional de Investigación en Evidencia y Tecnologías en Salud .CINETS

María Ximena Rojas Reyes RN, MSc, PhD.

Profesora Asociada del Departamento de Epidemiología y Bioestadística. Pontificia Universidad Javeriana
Miembro de la Alianza del Centro Nacional de Investigación en Evidencia y Tecnologías en Salud. CINETS

John Feliciano-Alfonso MD, MSc.

Médico Especialista en Farmacología Clínica
Maestría en Epidemiología Clínica
Especialista de la Información Grupo Cochrane. Infecciones de Trasmisión Sexual. STI

EDITORES

Hernando G Gaitán Duarte MD, MSc.

María Ximena Rojas Reyes RN, MSc, PhD.

ÍNDICE

Índice de cuadros	6
Agradecimientos	7
Prólogo	8
1. Introducción	10
1.1 ¿Qué es la evaluación de tecnologías sanitarias?	10
1.2 ¿Qué es transferencia del conocimiento?	10
1.3 ¿Qué es evidencia y qué es información en transferencia del conocimiento?	10
1.4 ¿Qué son las tecnologías en la salud?	12
1.5 ¿Qué aspectos considerar para informar al tomador de decisiones en tecnologías sanitarias?	12
1.6 Conceptos clave	14
2. Sistemas para formular la pregunta a responder con la evaluación de tecnología	16
2.1 ¿Qué son las preguntas básicas?	16
2.2 ¿Qué son las preguntas avanzadas?	16
2.3 Clasificación de los desenlaces	17
2.4 Conceptos clave	18
3. Búsqueda y selección de la evidencia e información que soporta la evaluación de tecnología	19
3.1 Aproximación general a la evidencia	19
3.2 Identificación de evaluaciones de tecnología previas y Guías de Práctica Clínica	21
3.3 Búsqueda de evidencia sobre efectividad y seguridad de la tecnología	23
3.3.1 Bases de datos y motores de búsqueda	23
3.3.2 Estrategia de búsqueda	24
3.4 Selección de los estudios para efectividad y seguridad	28
3.5 Conceptos clave	28
4. Estimación del efecto de una intervención	30
4.1 Medición del efecto	30
4.2 Medidas de efecto relativo	31
4.2.1 Riesgo relativo	32
4.2.2 Razón de oportunidades (OR)	35
4.2.3 Razón de peligros	37
4.3 ¿Cuál estimador de efecto relativo seleccionar?	38
4.4 Medidas de efecto absoluto	38
4.5 Precisión de los estimadores de efecto (relativo o absoluto)	41
4.6 Conceptos clave	41

5. Calidad metodológica de los estudios y validez de las observaciones	41
5.1 ¿Cómo evaluar la calidad de los ensayos clínicos aleatorizados?	43
5.1.1 Generación de la secuencia aleatoria (sesgo de selección)	44
5.1.2 Ocultamiento del código o de asignación (sesgo de selección)	45
5.1.3 Enmascaramiento de la intervención: cegamiento de los participantes y el personal que asigna la intervención o sus resultados (sesgo de desempeño y sesgo de detección)	45
5.1.4 Datos incompletos de desenlace (sesgo de desgaste)	46
5.1.5 Reporte selectivo (sesgos de reporte)	47
5.1.6 Otros sesgos (búsqueda de posibles sesgos no cubiertos en las otras secciones)	48
5.2 ¿Cómo evaluar la calidad de las revisiones sistemáticas?	50
5.2.1 ¿Qué herramienta debo utilizar para evaluar las RS?	51
5.2.2 ¿Cómo proceder si existen varias RS que evalúan la misma pregunta?	51
5.3 Evaluación de estudios no aleatorizados como evidencia del efecto de intervenciones	54
5.4 Conceptos clave	54
6. Evaluación de la seguridad de las tecnologías	55
6.1 ¿De dónde obtener la evidencia sobre los efectos adversos de las tecnologías?	55
6.2 ¿Qué sesgos pueden afectar los estudios no aleatorizados?	55
6.3 Evaluación de los estudios de cohortes	58
6.3.1 Criterios para calificación de ítems o las preguntas a evaluar en los estudios de cohorte	58
6.4 Evaluación de los estudios de casos y controles	60
6.4.1 Criterios para calificación de ítems o las preguntas a evaluar en los estudios de casos y controles	60
6.5 Conceptos clave	63
7. Evaluación del conjunto de la evidencia	64
7.1 ¿Qué es el conjunto de la evidencia y cómo evaluarlo?	64
7.2 ¿Cómo seleccionar y calificar los desenlaces relevantes para la evaluación de tecnologías?	65
7.3 ¿Cómo juzgar la calidad del cuerpo de la evidencia?	66
7.3.1 Limitaciones del diseño (riesgo de sesgos)	66
7.3.2 Inconsistencia	67
7.3.3 Evidencia indirecta	69
7.3.4 Imprecisión	70
7.3.5 Sesgo de publicación	72
7.4 ¿Cómo generar una tabla de resumen de hallazgos a partir de esta evaluación?	76
7.4.1 El perfil de evidencia GRADE	76
7.4.2 Resumen de hallazgos	76
7.5 ¿Cómo juzgar la calidad del cuerpo de la evidencia sobre estudios no aleatorios que evalúan intervenciones?	76

7. 6 Conceptos clave	78
Bibliografía	79
Anexos	92

ÍNDICE DE CUADROS

Cuadro 2.1 Planteamiento de problemas como pregunta de primera línea en formato PICO (t)	17
Cuadro 2.2 Clasificación de los desenlaces de efectividad y seguridad en la evaluación del trastuzumab en cáncer de seno metastásico	18
Cuadro 3.1 Flujograma para búsqueda y evaluación de evidencia para la ET	20
Cuadro 3.2 Bases y páginas de consulta para reposición de las Evaluaciones de Tecnología	21
Cuadro 3.3 Bases y páginas de consulta para reposición de Guías de Práctica Clínica	22
Cuadro 3.4 Bases de datos bibliográficas especializadas en salud	23
Cuadro 3.5 Lista de bases especializadas en publicación de eventos adversos	25
Cuadro 3.6 Búsqueda de la evidencia a partir de los componentes PICO	27
Cuadro 4.1 Efecto relativo en los que se toman en cuenta los “malos” resultados en el numerador y el grupo expuesto es el que recibe la intervención	33
Cuadro 4.2 Ejemplo de efecto relativo donde se toman en cuenta los “malos” resultados en el numerador y el grupo expuesto es el que no recibe la intervención	33
Cuadro 4.3 Ejemplo de efecto relativo donde se toman en cuenta los “buenos” resultados en el numerador y el grupo expuesto es el que recibe la intervención	34
Cuadro 4.4 Ejemplo de razón de oportunidades. Efecto relativo donde se toman en cuenta los “malos” resultados en el numerador y el grupo expuesto es el que recibe la intervención	35
Cuadro 4.5 Ejemplo de razón de oportunidades. Efecto relativo donde se toman en cuenta los “buenos” resultados en el numerador y el grupo expuesto es el que recibe la intervención	36
Cuadro 4.6 Ejemplo de razón de tasas (razón de peligros por el método del riesgo proporcional de Cox)	38
Cuadro 4.7 Ejemplo de la estimación de la reducción absoluta del riesgo (evaluación de eficacia o efectividad)	39
Cuadro 4.8 Ejemplo de la estimación del incremento absoluto del riesgo (evaluación de seguridad)	40
Cuadro 5.1 Ejemplo de calificación del riesgo de sesgos	49
Cuadro 5.2 Ejemplo de la calificación de una Revisión Sistemática por el instrumento AMSTAR	53
Cuadro 6.1 Efecto de un factor de confusión en la asociación entre la intervención y el resultado	56
Cuadro 6.2 Ejemplo de la calificación de estudios de cohorte con la escala de Newcastle–Ottawa	62
Cuadro 7.1 Criterios de calificación de los desenlaces	65
Cuadro 7.2 Calificaciones de la calidad en la evidencia según la aproximación GRADE	73
Cuadro 7.3 Ejemplo de aplicación de evaluación de la calidad del cuerpo de la evidencia	74
Cuadro 7.4 Criterios para aumentar la calificación de calidad de la evidencia en estudios no aleatorios que evalúan intervenciones	77

AGRADECIMIENTOS

Ana Lucía Muñoz: por su apoyo durante el proyecto y su decisión de posicionar las evaluaciones de tecnología como un elemento para la toma de decisiones en política pública en la Región de Latinoamérica y el Caribe.

Francisco Rossi: por su invitación a participar en este proyecto que facilitará la toma de decisiones informadas en el campo de los medicamentos de altos costos financieros para los países de la región.

Edna Sánchez: por su apoyo en la conducción del proyecto y las observaciones al manuscrito final del documento.

Ana María Díaz: por su ayuda en la edición final del documento.

PRÓLOGO

El Banco Interamericano de Desarrollo (BID) ha venido desarrollando una serie de proyectos y actividades regionales en el sector salud encaminadas a fortalecer los procesos de priorización del gasto sanitario en los países de América Latina y el Caribe. En este marco, y con el objeto de brindar herramientas para una asignación eficiente de este gasto, el Banco se encuentra en el proceso de implementación de un proyecto para mejorar la gestión de medicamentos de alto impacto financiero. El proyecto busca que los tomadores de decisiones a nivel político dispongan de información válida, robusta, actual y oportuna para definir qué medicamentos financiar con recursos públicos y a qué precio.

En este contexto se diseñaron y aprobaron los proyectos ATN/OC-13546-RG y ATN/OC-14703/RG*, que buscan generar un paquete de información sobre evidencia, precios, cobertura por los planes de beneficios financiados con recursos públicos, propiedad intelectual y competencia, que fortalezca los procesos de toma de decisión para lograr una gestión eficiente de los medicamentos de alto impacto financiero.

En la actualidad, hay una gran cantidad de medicamentos nuevos que son producto de innovaciones tecnológicas recientes. Estos ingresan al mercado no siempre precedidos de una adecuada evaluación de su efectividad y seguridad comparada con los tratamientos disponibles, como tampoco

del impacto económico, ético y social de su uso. Adicionalmente, son con frecuencia más costosos sin que se les pueda atribuir un mayor valor terapéutico. Financiar estos medicamentos con recursos públicos resulta más preocupante porque podría conllevar a una menor cobertura de los medicamentos esenciales o básicos para la población, incrementando la inequidad y posiblemente deteriorando los indicadores de salud en nuestras poblaciones.

Para abordar esta situación, uno de los componentes del proyecto es la generación de informes de evaluación de tecnologías sanitarias. Se busca que el conocimiento derivado de la investigación clínica y básica, basada en metodologías repetibles, verificables y robustas, brinde información válida y confiable, que sea útil para la toma de decisiones en la definición de políticas públicas.

En el desarrollo de este componente, se construyó un modelo de evaluación de tecnologías sanitarias desde una aproximación regional, basado en la adaptación al contexto latinoamericano del modelo propuesto por la Red Europea de Evaluación de Tecnologías (EUnetHTA). La idea de tener un modelo regional partió del supuesto de que podría ser una opción más eficiente, al permitir compartir recursos entre países y, a la vez, disponer de un número mayor de evaluaciones por país. Adicionalmente, se busca fortalecer los procesos locales de evaluación de tecnologías, fomentar la creación de equipos de gestión de conocimiento entre los grupos académicos de los países, dar transparencia a los procesos y facilitar los consensos en la región.

Con este documento nos proponemos apoyar los procesos de evaluación de tecnología a desarrollar

* Los dos proyectos comparten el objetivo. La diferencia está en que en el primero se inició el trabajo con Colombia, Ecuador y México, y en el segundo se aprobó la inclusión de Costa Rica, Chile, El Salvador, Perú y República Dominicana.

por los países participantes en el proyecto: Chile, Colombia, Costa Rica, Ecuador, El Salvador, México, Perú y República Dominicana, como también los desarrollados localmente por otros países de la región.

De acuerdo a las anteriores consideraciones, este documento tiene como objetivo orientar al equipo de trabajo que desarrolla informes de evaluación de tecnologías (ET). Asimismo, busca dar al lector elementos que le permitan identificar la naturaleza del problema sanitario que enfrenta y cómo plantearlo. Para esto se recomienda elaborar una serie de preguntas que se puedan contestar por medio de la identificación de estudios publicados en la literatura internacional, que sean relevantes y pertinentes al problema en cuestión. Adicionalmente, procura brindar las herramientas para que el desarrollador de la ET pueda hacer una aproximación crítica y elaborar un informe de evaluación que sea transparente y que contenga los elementos clave para tomar una decisión informada.

Como una primera aproximación a la pertinencia de utilizar la metodología de plantear preguntas para resolver el problema en cuestión, este manual está formulado para contestar los siguientes asuntos: ¿qué es evidencia?, ¿cómo buscar y seleccionar los estudios relevantes?, ¿cómo evaluar críticamente los estudios para determinar la validez y confiabilidad de los resultados y la calidad metodológica?, ¿cómo resumir la evidencia de manera eficiente y sencilla para el lector que la necesita para tomar decisiones?

Para responder estas preguntas, el documento está organizado en siete capítulos, a lo largo de los cuales se desarrollan los conceptos clave en la apreciación crítica de la evidencia que soporta la evaluación de

tecnologías sanitarias. Para un mejor entendimiento de estos conceptos y de cómo se aplican de manera práctica, se ilustran principalmente por medio de ejemplos de evaluación de dos tecnologías: (i) una preventiva, la vacuna para el virus del papiloma Humano (VPH) y (ii) una tecnología para el tratamiento del cáncer del seno, el trastuzumab. Si no se encontraron ejemplos en publicaciones sobre las tecnologías mencionadas, se utilizaron ejemplos con otros medicamentos.

1. INTRODUCCIÓN

Hernando G. Gaitán Duarte, MD, MSc

María Ximena Rojas Reyes, RN, MSc, PhD

Este capítulo tiene por objeto definir el alcance de los conceptos que se utilizarán en el desarrollo del manual con el fin de unificar su lenguaje y facilitar su entendimiento.

1.1 ¿Qué es la evaluación de tecnologías sanitarias?

La evaluación de tecnología (ET) se ha definido como *“la evaluación basada en la investigación y orientada a la utilización del conocimiento relevante disponible acerca de los problemas asociados al uso de una tecnología en relación con la salud y la enfermedad. Por virtud de estos métodos, los cuales son basados en la evidencia, pero también en términos de su intención de apoyar la toma de decisiones, la ET está relacionada con la planeación, administración y gestión de los servicios de salud”*¹. Se considera entonces una forma de transferencia del conocimiento entre los investigadores y los tomadores de decisiones.

1.2 ¿Qué es transferencia del conocimiento?

Se ha definido la transferencia del conocimiento (comúnmente conocido por la frase inglesa Knowledge translation) o apropiación social del conocimiento como *“el intercambio, la síntesis y la aplicación ética del conocimiento, en el marco de un sistema complejo de interacciones entre los investigadores y los usuarios de la investigación, para que los beneficios de esta*

*lleguen de manera más rápida a la población, a través de una mejor salud, servicios y productos más efectivos y un fortalecimiento del sistema de salud”*². Se podría resumir como las actividades que soportan la toma de decisiones basadas en la evidencia, identificando las barreras y los facilitadores del proceso³.

1.3 ¿Qué es evidencia y qué es información en transferencia del conocimiento?

El término “evidencia” (del latín “evidentia”) es aplicado en este documento con el sentido de *“prueba”*⁴. La Real Academia de la Lengua Española lo define como *“certeza clara y manifiesta de la que no se puede dudar”*⁵.

La aplicación del término evidencia en las ciencias de la salud proviene del término *“evidencia en medicina”*, que define la mejor evidencia clínica externa como *“aquella investigación clínicamente relevante, muchas veces originada en la ciencias básicas, pero especialmente de la investigación clínica centrada en el paciente, en aspectos tales como la exactitud y precisión de las prueba diagnósticas, el poder de los marcadores pronósticos y la efectividad y seguridad de los regímenes preventivos, terapéuticos o utilizados en rehabilitación. Incluye el rastreo de la mejor evidencia clínica externa para contestar nuestras preguntas clínicas”*⁶.

Sin embargo, la medicina basada en la evidencia considera, además de la investigación clínica y básica, las preferencias y los valores del paciente y de la sociedad⁷ y se debe acompañar de la experiencia del clínico para que tenga los mejores resultados

en beneficio del paciente. La integración de estos tres niveles de conocimiento —evidencia, valores del paciente y experiencia— permiten una mejor toma de decisiones, ya sea en el campo clínico, en el de salud pública o en el de las evaluaciones de tecnologías.

Respecto a las *preferencias de los pacientes* es más difícil poder encontrar investigación basada en metodologías cuantitativas robustas que permitan tener confianza en que la estimación de las mismas es válida⁸. Se ha sugerido que esta información debería tomarse directamente de los pacientes objeto de la tecnología en evaluación⁹. Esto se debe a que todavía escasean las metodologías sólidas para conocer las preferencias de los pacientes e incluirlas de manera válida en la toma de decisiones en políticas públicas, en la cobertura relacionada con productos farmacéuticos y en guías de práctica clínica¹⁰. Por lo tanto, se podría afirmar que es más limitada la evidencia sobre preferencias del paciente, que sirva de insumo para la toma de decisiones. En consecuencia, las revisiones sistemáticas (RS) publicadas para definir preferencias de pacientes podrían no ser apropiadas por problemas en la heterogeneidad de los diseños incluidos, los resultados y la terminología utilizada¹¹. En este sentido, podríamos considerar que en preferencias de los pacientes más que de “evidencia” se hablará de “información”, entendida como *“la comunicación o adquisición de conocimientos que permiten ampliar o precisar los que se poseen sobre una materia determinada”* (RAE)¹².

La experiencia definida como *“conocimiento de la vida adquirido por las circunstancias o situaciones vividas”*, aunque representa una forma de comprensión de los fenómenos y está más expuesta a sesgos, es útil para la interpretación y la complementación de la evidencia¹³.

La “evidencia” en salud pública tiene un concepto más amplio; se define como *“(…) alguna forma de datos epidemiológicos o de resultados de evaluación de políticas y de información cualitativa y datos de mercadeo que son usados para hacer juicios o tomar decisiones”*¹⁴. Esta aproximación podría ser menos precisa, ya que incluye como pruebas la información y la experiencia que están respaldadas por metodologías menos sólidas.

Brownson (2009) identifica tres tipos de evidencia en salud pública: a) aquella relacionada con lo que se debería hacer, básicamente dirigida a la evaluación de factores de riesgo de enfermedades, para proponer intervenciones preventivas, b) la que muestra el impacto logrado con las intervenciones comunitarias realizadas, es decir, lo que específicamente muestra que ha servido y debería ser hecho, y c) la que muestra cómo debería ser hecho, que se refiere a la descripción de las condiciones de contexto en las que se realizó la intervención y cómo fue recibida¹⁴.

Adicionalmente la evidencia, especialmente la de alta calidad, es transferible y tiene bajo riesgo de sesgos. Por otra parte, la información es más dependiente del contexto y tiene mayor riesgo de sesgos por basarse, por lo menos en la actualidad, en metodologías menos duras. La transferibilidad es un aspecto relevante en las evaluaciones de tecnología, sobre todo cuando se hacen desde una mirada regional¹⁵.

En este documento tendremos en cuenta, en primer lugar, a la “evidencia” que soporta la evaluación de la efectividad y seguridad de las tecnologías utilizadas en el sector salud y, en segundo lugar, la “información” complementaria.

1.4 ¿Qué son las tecnologías en la salud?

La Organización Mundial de la Salud (OMS) define las tecnologías en salud como *“la aplicación de los conocimientos teóricos y prácticos en forma de medicamentos, dispositivos, vacunas, procedimientos y sistemas elaborados para resolver problemas sanitarios y mejorar la calidad de vida”*¹⁶. Por su parte, la International Network of Agencies for Health Technology Assessment (INAHTA) las define como: *“cualquier intervención utilizada para promover la salud, prevenir, diagnosticar o tratar la enfermedad o para la rehabilitación o el cuidado a largo plazo. Incluye productos farmacéuticos, dispositivos médicos, procedimientos y sistemas organizacionales utilizados en el cuidado de la salud”*¹⁷. Estas dos definiciones sobre las tecnologías en salud han sido criticadas por estar más enfocadas a las tecnologías entendidas como una herramienta física para mantener o recuperar la salud, como son los medicamentos, los dispositivos médicos o equipos diagnósticos, sin considerar que las tecnologías en salud también incluyen los métodos para almacenar y procesar la información, métodos para crear conocimiento o que articulan el trabajo entre las máquinas y la experticia¹⁸. Es por esto que, en la actualidad, el término incluye también los dispositivos electrónicos y sistemas de información relacionados con la provisión de servicios de salud¹⁹. En el contexto de este manual, utilizaremos esta última definición más amplia.

1.5 ¿Qué aspectos considerar para informar al tomador de decisiones en tecnologías sanitarias?

La toma de decisiones, tanto a nivel individual como a nivel poblacional, requiere evidencia e información que le permitan al clínico o al formulador de políticas: a) definir la prioridad con que se requiere la tecnología; b) determinar los riesgos y beneficios que esta trae a la población objeto; c) identificar el impacto económico del uso de la tecnología; d) establecer las consecuencias éticas; e) identificar las posibles consecuencias sociales, organizacionales y legales del uso de la tecnología. Dependiendo de los aspectos abordados y el tiempo disponible para elaborar el informe de evaluación de tecnología, los informes podrán ser rápidos o completos. Los informes rápidos de la ET toman en cuenta la condición objeto de tecnología, las características de la tecnología misma, y su efectividad y seguridad comparada. Los informes completos toman en cuenta las implicaciones éticas, sociales, organizacionales y legales que conlleva el uso de la tecnología.

Para definir si la evaluación de una tecnología es prioritaria, es bastante útil la información sobre la importancia de la enfermedad, ya sea en cuanto a la magnitud del problema en términos de su prevalencia o incidencia²⁰ o en términos de mortalidad asociada, años de vida saludable perdidos o costos asociados al manejo de la entidad²¹.

En la ET se necesita el conocimiento acerca del pronóstico de la enfermedad, obtenido a partir de la historia natural o el curso clínico de la misma.

Dicho conocimiento aporta la información sobre los resultados en salud que se derivan de reconocer de manera tardía una entidad o los resultados a esperar al hacer una intervención terapéutica, ya sea en efectividad, seguridad o calidad de vida. También es de gran valor al estimar las probabilidades de los resultados de interés en los análisis de costo-efectividad²².

Otros criterios a tener en cuenta para determinar la prioridad de la ET están relacionados con la tecnología en sí misma, como, por ejemplo, en el caso de medicamentos, su mecanismo de acción, las propiedades físicas o farmacodinamias, el ciclo de vida, o su uso fuera de etiqueta (conocido por la frase inglesa *off label*)¹⁵. También se ha descrito la importancia de disponer de una alternativa tecnológica y existencia de evidencia, como también criterios de tipo social, como por ejemplo, la presión para su inclusión en los sistemas de cobertura o planes de beneficios en salud por parte de grupos de interés o necesidades de tipo político^{23, 24}. Adicionalmente, identificar el efecto potencial final de la aplicación de una tecnología en las personas o las comunidades es otro aspecto necesario para definir la prioridad de la evaluación.

Una vez establecida la prioridad de la ET, el concepto que más interesa evaluar son los beneficios y la seguridad que resultan de la aplicación de la tecnología. En la evaluación de medicamentos se utiliza el término “efectividad relativa” —entendida como la medida según la cual una intervención hace más beneficio que daño— comparada con la mejor alternativa terapéutica disponible, para lograr los resultados deseados, cuando son provistas en las condiciones usuales de atención en salud¹⁵.

Cuando la intención de la aplicación de una tecnología en los individuos es beneficiar a quien la recibe, se podrá llamar intervención. La intervención también ha sido definida como “*todos los actos o métodos terapéuticos destinados a interferir con la etiología, los síntomas o el progreso de una condición de salud*”²⁵. De esta manera, hay intervenciones preventivas, diagnósticas, terapéuticas médicas o quirúrgicas, rehabilitadoras y educativas. Sin embargo, su uso también puede terminar en un daño no intencionado, por esto se habla de los beneficios y riesgos asociados al uso de las tecnologías.

El beneficio es definido por la Real Academia de la Lengua Española como “*el bien que se hace o se recibe*”²⁶ y el riesgo es definido como “*contingencia o proximidad al daño*”²⁷. El Institute for Quality and Efficiency in Health Care define el beneficio como “*los efectos positivos causales*” y el daño como “*los efectos negativos causales*” de una intervención en los resultados en salud²⁸. Desde el punto de vista epidemiológico, el beneficio se ha abordado desde los conceptos de eficacia o efectividad; eficacia es la aproximación al beneficio de una intervención en condiciones ideales y efectividad es el beneficio en las condiciones de la vida diaria²⁹. En cuanto al riesgo, ha sido definido en epidemiología como la probabilidad de que un individuo desarrolle un efecto adverso³⁰.

Por otra parte, con frecuencia, el objeto de las ET son los métodos diagnósticos, ya sean utilizados para a) la detección de estados previos a la enfermedad con la intención de evitar los casos nuevos (ej., la utilización de la prueba de detección de ADN del virus del papiloma humano como complemento de la citología cervical para reducir la incidencia cáncer de cérvix)³¹, b) la detección temprana de la enfermedad para

mejorar el pronóstico (ej., el uso de la mamografía para reducción de la mortalidad por cáncer de seno)³², o c) la aclaración de un diagnóstico (ej., el uso de la laparoscopia para aclarar la causa de dolor pélvico agudo en mujeres en edad reproductiva)³³. En la evaluación de las pruebas diagnósticas interesa su exactitud y reproducibilidad. Sin embargo, también interesa su efectividad en términos del impacto de los posibles resultados (positivos y negativos) de las pruebas diagnósticas sobre los desenlaces en salud³⁴. La necesidad de evaluar la efectividad es más patente en el caso de las intervenciones diagnósticas invasivas (ej., laparoscopia, cateterismos diagnósticos, pruebas que utilizan medios de contraste), porque también interesa el aspecto del beneficio de su aplicación y la seguridad con su uso³³.

Otro aspecto a considerar en la ET son las preferencias de los pacientes respecto al uso de la tecnología, sus beneficios y sus posibles consecuencias. En la actualidad se da cada vez más importancia a la opinión de los pacientes respecto a lo que esperan de la aplicación de tecnologías. Tomar en cuenta sus valores y preferencias puede ayudar a incrementar la adherencia a los tratamientos y a mejorar el empoderamiento de los pacientes con su condición, además de ser considerado como un aspecto ético relevante en la ET³⁵. Los resultados a tener en cuenta son principalmente aquellos de mayor importancia para el paciente y no solo para el trabajador de la salud³⁶. Las preferencias están orientadas a conocer qué tan deseables son los resultados en salud que se evalúan³⁷. Por otra parte, desde el punto de vista económico, la evaluación de las preferencias podrá estar orientada a determinar el valor que percibe el paciente, la sociedad o los tomadores de decisiones con respecto a utilizar una tecnología en relación con

la inversión realizada (conocido por la frase inglesa *value for money*)^{38, 39}.

En el informe completo de la ET es importante abordar el aspecto del impacto económico de la tecnología, que no solo incluye los costos de su adquisición y administración, sino también la relación de estos costos con las consecuencias inmediatas y a largo plazo, tanto para el sistema de salud, como para el paciente. Este problema se aborda generalmente a través de los análisis de costo-utilidad o de costo-efectividad, desde una perspectiva social, del pagador o del sistema de salud. Como se mencionó anteriormente, también se deberán evaluar las consecuencias éticas, sociales y organizacionales de la aplicación de la tecnología. Todas estas consideraciones pueden aportar información importante para la toma de decisiones sobre la cobertura de ciertas tecnologías en salud²⁵.

1.6 Conceptos clave

La ET es una modalidad de transferencia del conocimiento entre la investigación y la toma de decisiones, utilizada a nivel político o institucional.

El conocimiento se obtiene de la evidencia, la información y la experiencia. La evidencia, si es de buena calidad, tiene bajo riesgo de sesgos y es transferible, mientras que la información es, en general, más dependiente del contexto y tiene riesgo de sesgos difíciles de detectar por no estar soportada por metodologías robustas. La experiencia, si bien enriquece la definición de los objetivos de la ET y el abordaje de la misma en relación con la condición, no es transferible y tiene alto riesgo de sesgos, por lo que es necesario utilizar metodologías de consenso al momento de su inclusión en las ET.

Las ET toman en cuenta la información para definir la prioridad del problema respecto a: a) la importancia de la condición, ya sea por su frecuencia o impacto en la salud pública, individual o en el consumo de recursos; b) los aspectos relacionados con la tecnología en sí misma o sus comparadores. Por otra parte, las ET evalúan la evidencia, para determinar los beneficios (efectividad) y riesgos (daño) del uso de la tecnología y, en el caso de los informes completos, establecen las implicaciones económicas, sociales y organizacionales del uso de la tecnología.

2. SISTEMAS PARA FORMULAR LA PREGUNTA A RESPONDER CON LA EVALUACIÓN DE TECNOLOGÍA

Hernando G. Gaitán Duarte, MD, MSc

En el ámbito de la salud, se requiere de una serie de preguntas que nos permitan delimitar el problema que enfrentamos y orientar la indagación para su solución. Para poder definir el problema central se requiere de la desintegración en sus partes, y esto se logra mediante la identificación de las preguntas apropiadas⁴⁰. Estas preguntas —que se pueden plantear como preguntas básicas y preguntas avanzadas— deben ser formuladas de manera que permitan obtener respuestas claras acerca de los diferentes aspectos del problema⁴¹.

2.1 ¿Qué son las preguntas básicas?

Las preguntas básicas nos brindan la información sobre aspectos generales de una situación, una condición o una tecnología. Se caracterizan por tener un pronombre interrogativo junto a un verbo o preposición y un aspecto de la condición de interés. Algunos ejemplos de preguntas utilizadas para describir la condición de interés o subgrupos de población y la tecnología que se quiere evaluar o su impacto son:

¿Qué es el cáncer de seno?, ¿cómo se diagnostica el cáncer de seno metastásico?, ¿cuál es la importancia del cáncer de seno como problema de salud pública en la región?, ¿cuál es el pronóstico del cáncer de seno metastásico?, ¿qué es el trastuzumab?, ¿en qué tipo de pacientes con cáncer de seno está aprobado el uso de trastuzumab en la región?, ¿cuáles son los costos médicos directos de la atención del cáncer de seno?

2.2 ¿Qué son las preguntas avanzadas?

Por otra parte, las preguntas avanzadas están orientadas a identificar la mejor respuesta en una gama de posibilidades, por lo que generalmente van encaminadas a comparar alternativas. Estas preguntas, cuando se formulan en el ámbito de la salud, se caracterizan por tener cuatro componentes: a) la población (P) descrita en términos de la condición o un subgrupo especial; b) la intervención (I) o tecnología en evaluación, que puede ser una tecnología diagnóstica, una vacuna (prevención), un medicamento o una técnica quirúrgica; c) la comparación (C) o las posibilidades alternas al curso de acción a evaluar, que podrán ser: no recibir alguna intervención o recibir un placebo, o recibir otras intervenciones (médicas o quirúrgicas), en el caso de estar disponibles, u otras pruebas diagnósticas; d) los desenlaces en salud esperados ((O) del término outcome en inglés) con la aplicación de la intervención, que buscan describir los resultados primarios y secundarios a tener en cuenta con respecto a la efectividad y seguridad de la tecnología. El conjunto de estos cuatro componentes en una pregunta es conocido en investigación en salud como “PICO”. Algunos autores incluyen en la pregunta PICO el

horizonte de tiempo en el cual se miden los resultados. Este quinto componente se denota con la letra "t" por lo que es posible encontrar también la sigla "PICOt"⁴². El tiempo es un aspecto relevante en enfermedades crónicas como, por ejemplo, el cáncer, en el cual la mortalidad se mide en un periodo de tiempo (ej., sobrevida a 1 año o 5 años). En el Cuadro 2.1 se presentan algunas preguntas PICOt sobre el mismo ejemplo anterior del uso del trastuzumab en cáncer de seno.

Los resultados primarios son los resultados finales más importantes para el paciente desde el punto de

vista clínico e involucran tanto los beneficios como los daños. En enfermedades que amenazan la vida, el resultado primario es la mortalidad o la sobrevida a un tiempo dado, más los eventos adversos serios y la morbilidad, mientras que los resultados relacionados con la calidad de vida serán los resultados secundarios. En enfermedades que no amenazan la vida, la morbilidad y los resultados relacionados con la calidad de vida serán los resultados primarios. También son muy importantes los eventos adversos serios. Idealmente se debe incluir la definición operativa de cada desenlace y la justificación de por qué se incluye⁴³.

Cuadro 2.1 Planteamiento de problemas como pregunta de primera línea en formato PICO (t)

Problema	Población	Intervención	Comparación	Resultado	Tiempo al evento
Beneficio	Mujer con cáncer de seno metastásico	Trastuzumab + otra quimioterapia	Otra quimioterapia	Sobrevida - Tiempo libre de enfermedad	2 y 5 años
Daño	Mujer con cáncer de seno metastásico	Trastuzumab + otra quimioterapia	Otra quimioterapia	Sobrevida - Tiempo libre de enfermedad	2 y 5 años

2.3 Clasificación de los desenlaces

El grupo de trabajo Grading of Recommendations Assessment, Development and Evaluation (GRADE) sugiere hacer una calificación de la relevancia de los desenlaces de seguridad y efectividad a medir en las evaluaciones en salud, de tal manera que se identifiquen como: a) críticos, b) importantes y

c) no importantes para el paciente y la toma de decisiones sobre su salud⁴⁴. La importancia de los desenlaces varía de acuerdo con el problema de salud o condición del paciente; sin embargo, algunos ejemplos de resultados "críticos" serían la muerte, la discapacidad o secuelas permanentes o la presencia de eventos adversos serios como consecuencia de la aplicación o de la exposición a la tecnología; algunos "importantes" serían, por ejemplo, el tiempo

libre de enfermedad o los eventos adversos no serios; y los resultados “no importantes” serían, por ejemplo, los resultados intermedios o subrogados (resultados de laboratorio o aquellos que evalúan la condición del paciente de manera temporal, sin modificar su condición clínica final). Las evaluaciones de tecnología (ET) se deben centrar en los desenlaces considerados como críticos, luego los importantes.

Para calificar la importancia relativa para cada desenlace según GRADE, se evalúa su relevancia en una escala numérica de 1 a 9 puntos, donde 1 representa no importante y 9 representa crítico (ver sección 7.2). En el Cuadro 2.2 se presentan algunos ejemplos de calificación de los desenlaces para el caso de una ET sobre el trastuzumab en cáncer de seno.

Cuadro 2.2 Clasificación de los desenlaces de efectividad y seguridad en la evaluación del trastuzumab en cáncer de seno metastásico

Resultado	Importancia Relativa	Calificación
Sobrevida libre de progresión	6	Importante
Sobrevida global	9	Crítico
Tiempo de progresión	6	Importante
Toxicidad (eventos adversos grado III-IV de la escala WHO o su equivalente)	9	Crítico
Calidad de vida	8	Crítico
Respuesta al tumor	3	No importante

Como resumen, una ET puede responder a una o varias preguntas PICO/PICOt. Una vez identificadas las preguntas PICO a responder en la ET, es deseable que se califique la importancia de los desenlaces a considerar en la evaluación. Esto guiará la búsqueda de la literatura, la extracción de información y la síntesis de la evidencia.

2.4 Conceptos clave

En el abordaje inicial del problema que lleva a una ET, los desarrolladores deben identificar los componentes del problema y sobre ellos formular una serie de preguntas básicas.

Las preguntas iniciales que guían una ET están principalmente relacionadas con las características de la tecnología y la condición del paciente. Luego se formulan preguntas avanzadas, generalmente comparativas, en las que se evaluará la efectividad, la seguridad, el impacto económico y el valor de la tecnología para el paciente en comparación con otras alternativas.

La definición de las preguntas en el formato PICO, facilita la identificación de los componentes que guían la búsqueda de la literatura tanto para definir la efectividad como la seguridad de la tecnología.

3. BÚSQUEDA Y SELECCIÓN DE LA EVIDENCIA E INFORMACIÓN QUE SOPORTA LA EVALUACIÓN DE TECNOLOGÍA

John Feliciano Alfonso, MD, MSc

María Ximena Rojas Reyes, RN, MSc, PhD

El objetivo principal de la evaluación de tecnologías (ET) es proveer información a los tomadores de decisiones acerca de las consecuencias del uso de las tecnologías, ya sea a nivel macro para los sistemas de salud nacionales o regionales, o a nivel de las instituciones prestadoras de servicios de salud. Su función está orientada a soportar cambios en las políticas y prácticas de cuidado de la salud más que la producción de conocimiento per se⁴⁵ (ver sección: 1.2). Estas dos características de las ET, muestran su carácter inmediato (en la mayoría de los casos) y hacen necesario que los procedimientos de búsqueda, evaluación y síntesis de la evidencia o la información, que soporta cada uno de los aspectos que se consideran en una ET sean lo más eficientes posibles sin que por ello se sacrifique la amplitud de la búsqueda y validez en el proceso (ver sección 1.5). Debe prevalecer la garantía de que la ET se basa en la información más actualizada, válida y relevante sobre la pregunta en cuestión.

En este capítulo se desarrolla la metodología a seguir para la búsqueda y selección de la evidencia e identificación de la información que se requiere para la ET.

3.1 Aproximación general a la evidencia

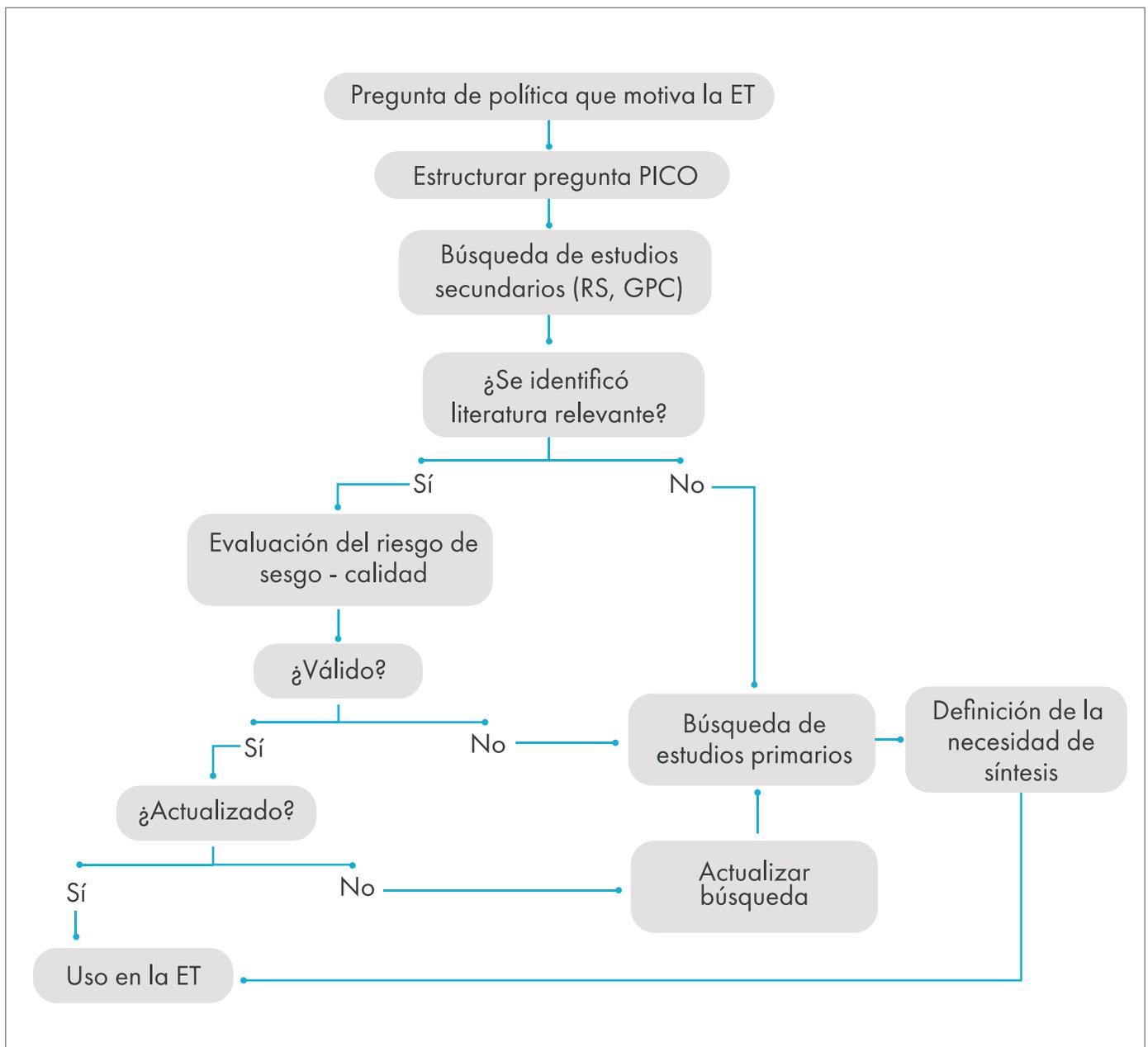
El proceso de búsqueda de información que soporta el desarrollo de una ET se inicia con la definición de la pregunta de política que motiva la ET, para posteriormente obtener la información sobre la condición de interés y uso de la tecnología. Para esto, es preciso identificar publicaciones sobre ET realizadas previamente sobre el tema y en las guías de práctica clínica (GPC) sobre la condición de interés para la cual se propone la tecnología en evaluación⁴⁶. La identificación de este tipo de publicaciones ayudará a definir la condición bajo estudio o el subgrupo de población objeto de la tecnología, y a determinar las prácticas actuales, así como también a definir los componentes de la pregunta en términos de población, intervención, comparador y desenlace (PICO) de las preguntas relacionadas con la tecnología en evaluación (ver: Cuadro2.1).

No hay duda de que para identificar la evidencia sobre efectividad y seguridad de las tecnologías se debe desarrollar una revisión sistemática (RS) de la literatura a partir de la pregunta PICO que debe responder la ET. A partir de esta pregunta se pueden identificar: 1) los términos que guiarán la búsqueda de la información (estrategia de búsqueda) y 2) los criterios que deben cumplir los estudios para ser incluidos como evidencia en la ET.

Una vez definida la estrategia de búsqueda y los criterios de selección, se propone seguir una metodología escalonada, de acuerdo con los resultados que se vayan obteniendo en cada paso y de la evaluación de la validez y pertinencia de los estudios identificados. Esta aproximación parte de la búsqueda inicial de los estudios secundarios o agregativos (como revisiones sistemáticas de ensayos clínicos) que hayan respondido las preguntas de efectividad o seguridad en

cuestión. En caso de no encontrar estudios secundarios o que estos no sean de buena calidad, se seguirá con la búsqueda de estudios primarios, como ensayos clínicos aleatorios para efectividad y seguridad y estudios clínicos de fase IV, cuasi-experimentos o cohortes para seguridad en caso de eventos adversos poco frecuentes (ver Cuadro 3.1).

Cuadro 3.1 Flujograma para búsqueda y evaluación de evidencia para la ET



3.2 Identificación de evaluaciones de tecnología previas y Guías de Práctica Clínica

Con el objetivo de identificar la ET se debe recurrir a los repositorios electrónicos⁴⁷ y, asimismo, para la identificación de guías de práctica clínica (GPC) se debe recurrir a las bases de datos de organismos

elaboradores y compiladores de GPC, así como también a los repositorios de las GPC de la región. Los Cuadros 3.2 y 3.3 muestran algunas de las bases de datos que se recomienda consultar en este paso para garantizar la identificación de la ET y GPC relevantes.

Cuadro 3.2 Bases y páginas de consulta para reposición de las Evaluaciones de Tecnología

La búsqueda de la ET debe incluir de forma obligatoria:

La base de datos de ET es suministrada por el Centro para las Revisiones y Diseminación (Centre for Reviews and Dissemination (CRD)) de la Universidad de York, Reino Unido. En: <https://www.crd.york.ac.uk/CRDWeb/>

La fuente de recursos adicionales (Health Technology Assessment Database) de la Biblioteca Cochrane, disponible en: http://onlinelibrary.wiley.com/o/cochrane/cochrane_clhta_articles_fs.html

Para proyectos a nivel latinoamericano se debe incluir la consulta a agencias de ET de la región como:

La Red Brasileña de Evaluación de Tecnologías en Salud (REBRATS) en: <http://189.28.128.101/rebrats/visao/sociedade/estudo.cfm>

El Instituto de Evaluación Tecnológica en Salud (IETS) en: <http://www.iets.org.co/reportes-iets/Paginas/Resumen.aspx?Cat=Guías+de+Práctica+Clínica>

El Centro Nacional de Excelencia Tecnológica en Salud (CENETEC) en: <http://www.cenetec.salud.gob.mx/>

El Instituto de Efectividad Clínica y Sanitaria (IECS) en: <http://www.iecs.org.ar/home-ets/>

Otras páginas de agencias de ET a nivel mundial:

The National Institute for Health and Care Excellence (NICE) en: <https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/technology-appraisal-static-list>

Swedish Council on Health Technology Assessment (SBU) en: <http://www.sbu.se/en/publications/>

Danish Centre for Health Technology Assessment en: <https://www.sst.dk/en/publications>

Cuadro 3.3 Bases y páginas de consulta para reposición de Guías de Práctica Clínica

La búsqueda de GPC debe incluir de forma obligatoria:

Guidelines International Network (GIN)

www.g-i-n.net/library/international-guidelines-library

National Institute for Health and Clinical Excellence (NICE)

www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-guidelines/nice-clinical-guidelines

National Guideline Clearinghouse (NGC)

<https://www.guideline.gov/syntheses/index>

New Zealand Guidelines Group (NZGG)

[https://www.health.govt.nz/publications?f\[0\]=im_field_publication_type%3A26#find-by-region](https://www.health.govt.nz/publications?f[0]=im_field_publication_type%3A26#find-by-region)

Scottish Intercollegiate Guidelines Network (SIGN) [0]

www.sign.ac.uk

American College of Physicians (ACP)

http://www.acponline.org/clinical_information/guidelines/

Australian Government. National Health and Medical Research Council (NHMRC)

<https://www.nhmrc.gov.au/guidelines-publications>

Institute for Clinical Systems Improvement (ICSI)

https://www.icsi.org/guidelines__more/

Canadian Medical Association (CMA)

<https://www.cma.ca/En/Pages/clinical-practice-guidelines.aspx>

GuíaSalud

<http://portal.guiasalud.es/web/guest/home>

Para proyectos a nivel latinoamericano se debe incluir la consulta a entidades de la región como:

Ministerio de Salud Pública, República de Ecuador

<http://www.salud.gob.ec/guias-de-practica-clinica/>

Ministerio de Salud y Protección Social, República de Colombia

http://gpc.minsalud.gov.co/gpc/SitePages/buscador_gpc.aspx

3.3 Búsqueda de evidencia sobre efectividad y seguridad de la tecnología

Una vez se han definido y refinado las preguntas PICO que se desean responder en la ET y se han priorizado los desenlaces de interés, el paso siguiente será identificar la evidencia sobre la efectividad de la tecnología la cual, en muchos casos, permite también identificar la evidencia sobre seguridad de la tecnología. A continuación se describen los pasos para garantizar la identificación de los estudios relevantes. Como se dijo anteriormente, es preciso seguir una estrategia escalonada para lograr mayor eficiencia en la búsqueda de la evidencia relevante.

3.3.1 Bases de datos y motores de búsqueda

Como se ha dicho, el paso inicial es buscar estudios “agregativos” tipo RS de la literatura, con el fin de hacer más eficiente el proceso (ver Cuadro 3 .3). La identificación de RS se logra a través de búsqueda en las principales bases de datos de la literatura

biomédica, entre las cuales están la versión en línea del Medical Literature Analysis and Retrieval System Online - MEDLARS® llamada MEDLINE, la base Cochrane de RS (Cochrane Database Of Systematic Reviews - CDSR) y la Excerpta Medica Database (EMBASE).

De acuerdo con la región en que se desarrolle el proyecto, es importante considerar la inclusión de otras bases regionales. Por ejemplo, para países de Latinoamérica, es necesario, además, incluir la búsqueda en la base de Literatura Latinoamericana y del Caribe en Ciencias de la Salud (LILACS). De igual forma, es importante considerar el ámbito en el que se aplica la tecnología, por ejemplo, si es de importancia en el ejercicio de la enfermería, es pertinente incluir como base de búsqueda la Cumulative Index to Nursing and Allied Health (CINALH), o si es el ámbito de la psicología, incluir la base de resúmenes creada por la PsycINFO, etc. Las direcciones de Internet vigentes a diciembre de 2016 para acceder a estas bases de datos bibliográficas se presentan en el Cuadro 3.4:

Cuadro 3.4 Bases de datos bibliográficas especializadas en salud

<ul style="list-style-type: none">• MEDLINE (Vía Pubmed) https://www.ncbi.nlm.nih.gov/pubmed/
<ul style="list-style-type: none">• Cochrane Database of Systematic Reviews (CDSR)/ Cochrane Central Register of Controlled Trials (CENTRAL) http://onlinelibrary.wiley.com/cochranelibrary/search/quick
<ul style="list-style-type: none">• EMBASE https://www.elsevier.com/solutions/embase-biomedical-research http://embase.com

- LILACS

<http://lilacs.bvsalud.org/es/>

- CINAHL

<https://health.ebsco.com/products/the-cinahl-database/allied-health-nursing>

<http://www.ebscohost.com/biomedical-libraries/cinahl-plus-with-full-text>

- Psycinfo

<http://www.apa.org/pubs/databases/psycinfo/index.aspx>

Con el fin de aumentar la especificidad de la búsqueda en identificar RS se pueden usar los filtros sobre el tipo de estudio que tienen disponibles estas bases, como “Systematic review”, “metanalysis” y “review”. Ahora bien, es posible que, a pesar de una búsqueda detallada, reproducible y transparente de estudios secundarios, no se obtengan resultados satisfactorios y no se encuentren RS para responder la pregunta de interés o estos no tengan una calidad mínima. Por tanto, el siguiente paso es la identificación de estudios primarios que den cuenta de la eficacia y de la seguridad de la tecnología. En consecuencia, los ensayos clínicos aleatorios (ECA) son el diseño que por excelencia responderán aspectos relacionados con la eficacia y la efectividad de las intervenciones en salud. Para su identificación, se debe seguir recurriendo a las mismas bases de datos bibliográficas ya enunciadas (MEDLINE, EMBASE, LILACS), e incluir el registro central de ensayos clínicos de la Colaboración Cochrane (Cochrane Central Register of Systematic Reviews -CENTRAL) ⁴⁸.

Es común que los ECA y las RS respondan algunos interrogantes relacionados con la seguridad de una tecnología; sin embargo, los desarrolladores de la ET deben asegurar una búsqueda exhaustiva

de la evidencia sobre efectos adversos y daño asociado a la tecnología. En consecuencia, deben recurrir a la búsqueda de otro tipo de diseños, como estudios clínicos de fase IV, cuasi experimentos, estudios longitudinales de seguimiento a largo plazo y reportes de caso. Para ello, se pueden diseñar estrategias encaminadas a la búsqueda de desenlaces indeseables, identificados como críticos o importantes, así como reportes de eventos adversos en las mismas bases de datos anteriores. Además, hay que ampliar la consulta a bases de datos especializadas y de organismos reguladores internacionales (por ejemplo, Food and Drug Administration - FDA, o la European Medicine Agency - EMA). Las bases especializadas en la recopilación de publicaciones o reportes de ensayos clínicos controlados o de eventos adversos que los desarrolladores de la ET deben consultar se encuentran en el Cuadro 3.5.

3.3.2 Estrategia de búsqueda

Para lograr una búsqueda de evidencia eficiente respecto a la efectividad y seguridad de los medicamentos y que cuente con un balance de especificidad y sensibilidad, se debe iniciar con una estrategia amplia pero enfocada a la condición de interés y al medicamento de interés. En este sentido,

los expertos temáticos del grupo desarrollador de la ET deben proporcionar la información necesaria para identificar los términos clave de búsqueda de acuerdo con la pregunta en formato PICO, especialmente para la condición clínica de interés (población blanca) y la tecnología a evaluar. Con este insumo se debe diseñar una estrategia de búsqueda genérica, compuesta por vocabulario controlado (términos Medical SubHeadings - MeSH en MEDLINE, Emtree en EMBASE y Descriptores de Ciencias de la Salud - DeCS en LILACS) y lenguaje libre (variaciones

ortográficas, plurales, sinónimos, acrónimos y abreviaturas) ^{46,48}.

Cuadro 3.5 Lista de bases especializadas en publicación de eventos adversos

<ul style="list-style-type: none">• CENTRAL: Cochrane Central Register of Systematic Reviews http://www.cochranelibrary.com/about/central-landing-page.html
<ul style="list-style-type: none">• MedWatch: el programa de información y reporte de eventos adversos de la Food and Drug Administration, de Estados Unidos. http://www.fda.gov/Safety/MedWatch/
<ul style="list-style-type: none">• Base de datos europea de informes de presuntas reacciones adversas de la European Medicines Agency (EMA). http://www.adrreports.eu/es/search.html
<ul style="list-style-type: none">• Advertencias de seguridad, alertas y retiros de fármacos (o dispositivos médicos) de la Medicines and Healthcare Products Regulatory Agency (MHRA) del Reino Unido. https://www.gov.uk/drug-device-alerts http://www.mhra.gov.uk/Safetyinformation/Safetywarningsalertsandrecalls/DrugAlerts/index.htm
<ul style="list-style-type: none">• Base de datos de notificaciones de eventos adversos del Departamento de Salud del Gobierno de Australia. http://www.tga.gov.au/DAEN/daen-report.aspx

A partir de la estrategia genérica, se deben adaptar estrategias de búsqueda para los diferentes recursos. En el caso de las bases de datos de MEDLINE, EMBASE, LILACS, estas cuentan con filtros propios para búsqueda de RS y metaanálisis, así como de ECA. Sin embargo, es pertinente recordar que la indexación de los ECA y las RS ha cambiado con el tiempo y así como actualmente se reconocen como “randomized controlled trials” (ensayos controlados aleatorios) y “systematic reviews” (revisiones sistemáticas), también pueden estar indexados como “controlled trials” (ensayos controlados), “trials” (ensayos) y “reviews” (revisiones) o “metanalysis” (metaanálisis) respectivamente.

Una forma sencilla de conducir la búsqueda es identificar los términos MeSH a través del motor de búsqueda “Pubmed” para MEDLINE y, a partir de estos, diseñar la estrategia empleando conectores booleanos (AND, OR, NOT) que permitan la inclusión de expansión de términos, identificadores de campo (título y resumen) y truncadores, tanto para la condición de interés como para la intervención en cuestión (lo que aumenta la sensibilidad de la búsqueda). Una vez diseñada esta estrategia, hay que correrla a través de un metabuscador, tipo OVID, que permite la búsqueda simultánea en múltiples bases de datos, tantas como se quieran incluir (ej., MEDLINE, EMBASE LILACS, CENTRAL, Cochrane Library, etc.).

Como se dijo anteriormente, es posible que la información sobre seguridad no pueda ser resuelta en su totalidad tras la búsqueda de efectividad, por lo que se recomienda realizar búsquedas independientes pero complementarias, para asegurar la identificación de la evidencia necesaria para la evaluación de efectividad y seguridad.

La estrategia que parece ser más útil para identificar publicaciones importantes respecto a la seguridad de las intervenciones es adicionar, a la estrategia de búsqueda diseñada para efectividad, los subencabezamientos (subheadings)* relacionados con efectos adversos de la tecnología en evaluación⁴⁹.

En este orden de ideas, una alternativa es utilizar los encabezamientos asociados a términos MeSH ‘therapeutic use’ (tu) y ‘adverse effects’ (ae) para MEDLINE, y los encabezamientos asociados a términos Emtree ‘adverse drug reaction’ (ae) y ‘drug toxicity’ (to) para EMBASE⁴⁶. Entonces, por ejemplo, si se necesita evaluar la seguridad de una intervención como la vacuna del virus del papiloma humano (VPH) y se recurre a MEDLINE bajo la plataforma PUBMED, se debe ubicar su término MESH (disponibles a Diciembre de 2016 en <http://www.ncbi.nlm.nih.gov/mesh/2009893>), y adicionar los subencabezamientos correspondientes a seguridad, de la siguiente manera: (“*Papillomavirus Vaccines/adverse effects*”[Mesh] OR “*Papillomavirus Vaccines/therapeutic use*”[Mesh]).

Adicionalmente, existen varios filtros que aumentan la probabilidad de detección de eventos adversos en MEDLINE y EMBASE que los desarrolladores de la ET pueden consultar para mejorar la sensibilidad de sus búsquedas en este aspecto^{49, 50}.

En el Cuadro 3.6 se presenta como ejemplo la estrategia de búsqueda diseñada para responder la

* Los encabezamientos son herramientas que permiten enfocar un aspecto de interés dentro de un tema en particular (por ejemplo, si se está interesado en la historia de la insulina, el encabezamiento más apropiado que aparece en la página del término MESH “insulin” es “history”; puede verificarlo en <https://www.ncbi.nlm.nih.gov/mesh/68007328>).

pregunta: ¿Cuál es el efecto de la vacuna del virus del papiloma humano (VPH) en comparación con el cuidado usual o no vacunarse?

Cuadro 3.6 Búsqueda de la evidencia a partir de los componentes PICO

Pregunta de evaluación de la vacuna contra el virus del papiloma humano	
P	Población: Mujeres
I	Intervención: Vacuna contra el virus del papiloma humano
C	Comparador: Cuidado usual
O	Desenlaces: Incidencia de lesiones precursoras de cáncer cervical
Términos de búsqueda	
Para la población (condición)	Women, female
Para la intervención	Papillomavirus Vaccines, wart virus vaccine, gardasil, cervarix, silgard
Estrategia (uso de conectores, expansión de términos, términos truncados e identificadores de campo)	
1 exp women/ (31894) 2 wom?n.tw. (927309) 3 female\$.tw. (733167) 4 1 or 2 or 3 (1577568) 5 exp Papillomavirus Vaccines/ (5374) 6 (papilloma\$ adj5 vaccine\$.tw. (2546) 7 (wart\$ adj5 vaccine\$.tw. (145) 8 cervarix.tw. (209) 9 gard?sil.tw. (354) 10 silgard.tw. (14) 11 5 or 6 or 7 or 8 or 9 or 10 (6249) 12 4 and 11 (2476)	
Uso de límites para especificar la búsqueda	
Por tipo de estudios: 13 limit 12 to "reviews (best balance of sensitivity and specificity)" (519) Por fecha: <2010 al presente> 14 limit 13 to yr="2010 -Current" (261)	

3.4 Selección de los estudios para efectividad y seguridad

Si se ha realizado una búsqueda sensible y exhaustiva para cada uno de los dominios de efectividad y seguridad de la tecnología en evaluación, es probable que se tenga una gran lista de referencias para seleccionar. Se sugiere que los desarrolladores de la ET tengan a priori unos criterios de selección claros y predefinidos de acuerdo con los componentes de las preguntas PICO que la ET pretende resolver. Los criterios de selección de los estudios son: tipo de estudios, tipo de participantes y tipo de intervenciones⁵¹.

Para el ejemplo de la vacuna contra el virus del papiloma, los criterios de selección de los estudios serían:

Tipo de estudios: para efectividad se buscarán RS del efecto de la vacuna para el VPH; en caso de no encontrar RS publicadas en los últimos 3 años, se buscarán ECA. Para seguridad se buscarán RS o ECA; en caso de no encontrar, se buscarán estudios clínicos de fase IV, estudios observacionales y registros de EA.

Tipo de participantes: mujeres con vida sexual activa sin evidencia de infección de VPH al ingreso al estudio.

Tipo de intervención: vacuna para el VPH, ya sea polivalente o bivalente, con tratamiento completo (eficacia) o al menos una dosis (efectividad).

El proceso de selección de la evidencia implica seguir varios pasos de forma iterativa con el objeto de reducir de manera gradual el número de estudios. Se parte de la revisión de los títulos o los resúmenes, que permitirá separar los estudios en tres grandes categorías: (i) definitivamente excluido, (ii) definitivamente incluido y (iii) se requiere más información para decidir. En un segundo paso se aplicarán de nuevo los criterios de elegibilidad, empleando el texto completo de los estudios de las dos últimas categorías, con el objeto de decidir de forma definitiva cuáles estudios se incluyen para dar respuesta a cada una de las preguntas de la ET. Como se ha dicho previamente, en el desarrollo de la ET se busca identificar la mejor y más completa evidencia sobre la tecnología en evaluación. Por lo tanto, el proceso de selección de los estudios se basa en la identificación de RS que cumplan los criterios de elegibilidad establecidos previamente.

Esta primera fase del proceso debe ser realizada al menos por dos revisores, de manera que cada uno de ellos decida la inclusión o exclusión de los estudios, a la luz de los criterios de selección, pero de forma independiente (es decir, sin conocer las decisiones del otro).

3.5 Conceptos clave

Una búsqueda amplia y exhaustiva de informes previos de evaluación de tecnología y Guías de Práctica Clínica ayudan a definir de la mejor manera la condición en que será aplicada la tecnología y el manejo más actualizado de la misma.

Para la evaluación de efectividad y seguridad es necesario revisar los repositorios de revisiones

sistemáticas y las bases de datos de estudios primarios y secundarios, así como la consulta a bases de datos especializadas y de organismos reguladores internacionales.

El uso de términos y filtros apropiados para cada base de datos asegura una mayor eficiencia en la búsqueda. Es importante definir previamente unos criterios de selección de los estudios por tipo de diseño, tipo de población y tipos de intervención para una más eficiente elección de los estudios relevantes al problema a resolver.

4. ESTIMACIÓN DEL EFECTO DE UNA INTERVENCIÓN

Hernando G. Gaitán Duarte, MD, MSc

Uno de los problemas más importantes que se enfrenta al hacer una evaluación de tecnologías (ET), es determinar si la aplicación de una tecnología con intenciones preventivas o terapéuticas produce más beneficio que daño en los resultados esperados, en comparación a no hacer nada o en comparación a las otras alternativas disponibles. En otras palabras, se trata de estimar el efecto de la intervención.

Al aplicar una intervención preventiva, se espera reducir el número de sujetos que presentan malos resultados. Por ejemplo, con la aplicación de la vacuna del virus del papiloma humano (VPH), se busca reducir la frecuencia de sujetos que desarrollan condilomas genitales en un periodo de tiempo⁵². En el caso de una intervención terapéutica, igualmente esperamos reducir el número de sujetos con malos resultados derivados de la enfermedad o condición a intervenir. Por ejemplo, con la aplicación del trastuzumab en terapia sistémica para pacientes con cáncer de seno avanzado HER2 (conocido por el término inglés human epidermal growth factor receptor 2) positivo, se espera reducir la mortalidad por cualquier causa a 1 o 2 años. Otra manera de evaluar los resultados sería determinar el número de sujetos con resultados deseados; por ejemplo, mayor supervivencia global o libre de enfermedad⁵³.

Por otra parte, vale la pena recordar que, como resultado de la aplicación de la intervención o como consecuencia de ella, también podremos tener un mayor número de efectos adversos o no deseados o daño, como son los eventos adversos serios. Los resultados de beneficio para el paciente y los efectos positivos causales —ya sea evitar malos desenlaces o incrementar los buenos resultados— son la medida de la eficacia o efectividad de la intervención. Los efectos negativos causales —resultados no deseados o efectos adversos de la intervención— son la medida de la seguridad de la intervención (ver sección 1.5).

4.1 Medición del efecto

Para medir el efecto de la intervención se requiere que el número de sujetos con el resultado de interés (ej., infección por VPH) entre aquellos expuestos a la intervención (ej., quienes reciben la vacuna) se compare con el número de sujetos con el resultado de interés entre aquellos no expuestos a la intervención (ej., aquellos que no reciben la vacuna).

Para medir el efecto en los sujetos expuestos, se mide la incidencia acumulada (casos nuevos) o riesgo absoluto de presentar el evento, dado que se está expuesto a la intervención en un periodo de tiempo⁵⁴. Este riesgo se conoce como **riesgo con la intervención** o **riesgo correspondiente** en la aproximación GRADE⁵⁵.

$$(Rx | +) = (\text{sujetos con el resultado } (+) / \text{sujetos expuestos})$$

Para estimar el efecto en los sujetos que no reciben la intervención (no expuestos) o grupo control, se medirá también el número de sujetos con nuevos eventos no deseados (o deseados) sobre el total de sujetos del grupo no expuesto. Es decir, se medirá la incidencia acumulada que representa el riesgo absoluto de presentar el evento, dado que no se está expuesto a la intervención. Este riesgo se conoce como **riesgo basal o riesgo asumido** en la aproximación GRADE⁵⁵.

$(R_x | (-)) = (\text{sujetos con el resultado (+)} / \text{sujetos no expuestos})$.

Como alternativa, para medir la incidencia, se podrá tener en cuenta el número de sujetos con nuevos eventos en relación con el tiempo que cada persona estuvo o no expuesta (*persona – tiempo*). Eso es el estimador conocido como la tasa de incidencia. Esta medida nos indica la velocidad en la que se presentan los eventos y se mide de igual forma entre los sujetos expuestos y no expuestos a la intervención. Lo importante aquí es determinar el número de sujetos con nuevos eventos en relación con el tiempo que participó o aportó al seguimiento en el estudio⁵⁴. Es utilizado en estudios de cohorte de seguimiento abierta, donde los participantes aportan diferentes tiempos a la observación. Este indicador también se denomina peligro (conocido por el término en inglés hazard). El “peligro” se ha definido como la probabilidad instantánea de que surja un evento en un momento en el tiempo o la probabilidad de que un individuo en observación presente un evento en un periodo de tiempo⁵⁶. Este es el concepto que subyace a la sobrevida total, que es aplicado en las tecnologías que se utilizan para el tratamiento del cáncer y que evalúa el número de sujetos que no mueren a

un tiempo determinado $(1 - \text{mortalidad} / t_n)$ ⁵⁷. Esta aproximación puede ser aplicada a otros eventos de interés, en los que el tiempo de exposición es una variable importante a considerar como, por ejemplo, el desarrollo de una infección asociada al uso de un dispositivo médico⁵⁸.

La medición del efecto se podrá hacer mediante la comparación de los dos grupos por dos alternativas: a) por medio de una razón (el numerador no está incluido en el denominador) y b) por medio de una diferencia.

Vale la pena notar que, para determinar el efecto de una intervención, se requiere evaluar a los sujetos expuestos y no expuestos a la intervención, al menos en dos momentos en el tiempo; al momento inicial cuando administramos la intervención y nos aseguramos de que los sujetos no tienen el evento de interés (t_0), y al menos en un segundo momento, un tiempo después o la última vez que evaluamos a los sujetos en el estudio, cuando verificamos si presentaron o no el evento durante el seguimiento (t_n). Este seguimiento en el tiempo se realiza por medio del ensayo clínico aleatorizado (ECA), los cuasi-experimentos y los estudios de cohortes⁵⁹.

4.2 Medidas de efecto relativo

Nos indican qué tantas veces más se asocia un resultado a estar expuesto, en relación con no estar expuesto. Nos indica la magnitud de la asociación; sin embargo, no detecta cuál es el efecto real de la intervención en la frecuencia real de los resultados deseados o no deseados⁶⁰.

Los estimadores de efecto relativo están dados por tres razones matemáticas, a) la razón de riesgos, también denominado riesgo relativo (RR), b) la razón de oportunidades (OR, odds ratio en inglés) y c) la razón de peligros o HR (hazard ratio en inglés)²⁹. Para hacer una adecuada interpretación del estimador relativo del efecto es importante que el lector tenga claro cuál es el grupo que se considera expuesto y cuál es el grupo que se considera no expuesto, como también cuál es la variable resultante que se desea evaluar. También es importante conocer cuál es la intervención y el comparador o control.

4.2.1 Riesgo relativo

Respecto a la exposición, una primera aproximación, como se mencionó antes, es considerar como expuesto al que recibe la intervención o tecnología en evaluación (ej., el medicamento) y no expuesto a aquel que no la recibe. Otra aproximación es considerar como grupo expuesto, al grupo que no recibe la tecnología. Este es, según la metodología GRADE, el grupo con riesgo basal o riesgo asumido. Esa misma aproximación requiere considerar como grupo no expuesto al grupo que recibe la tecnología, el cual pertenecería, según la metodología GRADE, al grupo con el riesgo correspondiente. Se sugiere esta aproximación cuando se evalúan intervenciones preventivas o aquellas que buscan reducir los malos desenlaces en salud (ejemplo, vacunas, antibióticos).

En el primer escenario:

*El Riesgo relativo (RR) = riesgo del grupo con la intervención / riesgo del grupo control
(Rx. I (+) / Rx. I (-)).*

En el segundo escenario:

*RR = riesgo del grupo control / riesgo del grupo con la intervención
(Rx. I (-) / Rx. I (+)).*

Vale la pena notar que, cuando un estudio busca evaluar una intervención destinada a producir beneficios, es más consistente medir el RR por la estimación de las incidencias de malos desenlaces (los que se quiere evitar), que por la estimación de incidencias de tener los buenos resultados⁶¹. De nuevo, lo importante es tener claro qué define a los grupos expuestos o no expuestos y la interpretación del resultado obtenido.

A continuación, se ilustrará con un ejemplo los conceptos anteriores.

Ejemplo:

En el estudio de Muñoz y colaboradores⁵² se evaluó el efecto de la vacuna contra el VPH (intervención) para prevenir las lesiones del VPH. Se comparó un grupo de sujetos que fueron vacunados con un grupo de sujetos que no fueron vacunados. En el grupo de sujetos que no recibieron la vacuna (8702) un total de 305 presentaron lesiones genitales por VPH (este es el que se denomina grupo "con el riesgo basal o asumido"; se considerará con un mayor riesgo teórico de la infección). Por otra parte, en el grupo que recibió la intervención (8689), un total de 63 sujetos presentaron lesiones genitales (este se considera el grupo "con la intervención o de riesgo correspondiente"; se considerará con un menor riesgo teórico de infección).

Para empezar el análisis de esta información con el fin de estimar el efecto de la intervención, se construye

la tabla de contingencia para comparar los grupos, como se presenta en los Cuadros 4.1 y 4.2.

Cuadro 4.1 Efecto relativo en los que se toman en cuenta los “malos” resultados en el numerador y el grupo expuesto es el que recibe la intervención

Condilomas genitales			
	Sí	No	Total
Grupo riesgo intervención	63	8626	8689
Grupo riesgo control	305	8397	8702

- Riesgo correspondiente (sujetos con el evento entre los vacunados): = $63/8689 = 0,007$
- Riesgo asumido = (sujetos con el evento entre los no vacunados): = $305/8702 = 0,035$
- Riesgo Relativo (RR) = $0,007/0,035 = RR = 0,21$ (IC 95%: 0,16- 0,27)
- Es decir, por lo tanto, los vacunados tienen cerca de un quinto del riesgo de tener lesiones por VPH en relación a los no vacunados

Sin embargo, la tabla de contingencia podría haberse construido al revés. En las casillas superiores se incluiría el grupo de no vacunados (riesgo basal) y en

la inferior, los vacunados (riesgo correspondiente). De esta manera, la estimación del resultado sería como se presenta en el Cuadro 4.2.

Cuadro 4.2 Ejemplo de efecto relativo donde se toman en cuenta los “malos” resultados en el numerador y el grupo expuesto es el que no recibe la intervención

Condilomas genitales			
	Sí	No	Total
Grupo riesgo control	305	8397	8702
Grupo riesgo intervención	63	8626	8689

- Riesgo asumido (sujetos con el evento entre los no vacunados) = $305/8702 = 0,035$
- Riesgo correspondiente (sujetos con el evento entre los vacunados) = $63/8689 = 0,007$
- Razón de riesgos o riesgo relativo (RR) = $0,035/0,007 = 4,83$ (IC 95%: 3,69-6,32)
- Por lo tanto, los no vacunados tienen 4,8 veces más riesgo de tener lesiones genitales que los vacunados

Ahora comparemos el Cuadro 4.1 con el Cuadro 4.3, que presenta el riesgo de no tener malos eventos o número sujetos sin infección.

Cuadro 4.3 Ejemplo de efecto relativo donde se toman en cuenta los “buenos” resultados en el numerador y el grupo expuesto es el que recibe la intervención

	Lesiones genitales		
	No	Sí	Total
Grupo riesgo intervención	8626	63	8689
Grupo riesgo control	8397	305	8702

Aquí consideraremos que quienes tienen el mayor riesgo de tener buenos desenlaces son el grupo de vacuna y los que tienen el menor riesgo de presentar buenos resultados son los del grupo de control:

- Riesgo de grupo de intervención (sujetos con el evento entre los vacunados) = $8626/8689 = 0,99$
- Riesgo de grupo de control (sujetos sin el evento entre los no vacunados) = $8397/8702 = 0,96$
- Razón de riesgos o riesgo relativo (RR) = $0,99/0,96 = 1,029$ (IC 95%: 1,024 - 1,033)

Por lo tanto, bajo esta aproximación, las diferencias entre los dos grupos serían mínimas. Aunque sería mayor el riesgo de tener buenos resultados (sin la infección) en el grupo de la vacuna (el riesgo de no presentar condilomas estaría aumentado en 0,03), la estimación del efecto sería mucho menor al efecto de prevenir la aparición de malos resultados (condilomas) en el grupo que recibe la vacuna. (El riesgo de tener condilomas sería de 0,2 (:1)

Observamos que se reduce de manera importante el efecto protector de la vacuna obtenido del Cuadro 4.1 RR = 0,21 (IC 95%: 0,16 – 0,27) en comparación con el efecto protector obtenido en el Cuadro 4.3 RR = 1.029(IC 95%: 1,024 - 1,033). Entonces, bajo esta aproximación, aunque habría diferencias entre los dos grupos, el efecto de la vacuna sería muy pequeño al incrementar los buenos resultados (1,029:1), en comparación a prevenir los malos resultados (0,21) Por lo tanto, es importante tener en cuenta cómo se estructura la evaluación del RR de una intervención destinada a reducir los malos resultados.

4.2.2 Razón de oportunidades (OR)

El segundo estimador de efecto relativo es la razón de suertes conocida, por su nombre en inglés como (odds ratio - OR), que nosotros hemos denominado “razón de oportunidades”. Esta deriva de una razón de razones: (razón de tener el evento/ no tener el evento) en los expuestos/ (razón de tener el evento / no tener el evento) en los no expuestos. El Cuadro 4.4 muestra cómo se calcularía el OR a partir de los datos del estudio de Muñoz y colaboradores.

Cuadro 4.4 Ejemplo de razón de oportunidades. Efecto relativo donde se toman en cuenta los “malos” resultados en el numerador y el grupo expuesto es el que recibe la intervención

Lesiones genitales			
	Sí	No	Total
Grupo riesgo intervención	63	8626	8689
Grupo riesgo control	305	8397	8702

- Razón del grupo de intervención= (sujetos con el evento entre los vacunados/ sujetos sin en el evento) = 63/8626
- Razón del grupo control = (sujetos con el evento / sujetos sin el evento) entre los no vacunados: = 305/8397
- Y entonces calcularíamos la razón de razones u OR = (axd)/bxc) (63/8626)/ (305/8397) = (63x8397)/ (305X8626) = 0,20 (IC 95%: 0,15 - 0,26)
Por lo tanto, los vacunados tienen 0,2 veces el riesgo de tener lesiones genitales en relación a los no vacunados
- El RR había estimado un efecto de 0,21 (:1) (figura 4.1 a) ligeramente menor a la estimación por medio del OR

El RR había estimado un efecto de 0,21 (:1) (Cuadro 4.1) ligeramente menor a la estimación por medio del OR. Este mayor valor está dado por una posible sobrestimación del efecto a través del OR, que es importante cuando la frecuencia del evento es mayor al 1%⁶². Esta sobrestimación se denomina “sesgo de

construcción” (conocida en inglés como built bias) y el efecto es mayor mientras más frecuente sea el evento⁶³.

Pero miremos qué pasa con el OR cuando se evalúa el resultado de no tener eventos (Cuadro 4.5).

Cuadro 4.5 Ejemplo de razón de oportunidades. Efecto relativo donde se toman en cuenta los “buenos” resultados en el numerador y el grupo expuesto es el que recibe la intervención

Lesiones genitales			
	No	Sí	Total
Grupo riesgo intervención	8626	63	8689
Grupo riesgo control	8397	305	8702

Recordemos que, en esta situación, consideraríamos que quienes tienen el mayor riesgo de no tener eventos son los sujetos del grupo de vacuna y los que tienen menor riesgo de no presentar eventos son los del grupo de control

- Razón del grupo de intervención= (sujetos con el evento entre los vacunados/ sujetos sin el evento) = 8626/63
- Razón del grupo control = (sujetos con el evento / sujetos sin el evento) entre los no vacunados:
= 8397/305

Y entonces calcularíamos la razón de razones u OR $(axd)/bxc)$
 $(8626/63)/(8397/305) = OR = 4,97$ (IC 95%: 3,77 -- 6,64).

- Por lo tanto, quienes reciben la vacuna tienen 4,97 (: 1) de oportunidades de no presentar lesiones en relación a los que no la reciben

Como se puede observar, el OR mantiene un valor muy similar al obtenido por el RR presentado en el Cuadro 4.1. Es decir, el OR no se afecta por la manera como se estimen “malos o buenos” resultados. Este mayor valor está dado por una posible sobrestimación del efecto a través del OR, que es importante cuando la frecuencia del evento es mayor al 1%⁶². Esta sobrestimación es resultado del sesgo de construcción y su efecto es mayor mientras más frecuente sea el evento⁶³.

4.2.3 Razón de peligros

El tercer estimador del efecto relativo, es la “razón de peligros”, conocida como HR (hazard ratio, en inglés) y definida como la comparación de la medida de la sobrevida en los dos grupos o el exceso de “peligro” de presentar el evento en el grupo expuesto en comparación al grupo control⁵⁸. Se utiliza cuando el tiempo hasta el evento es muy importante, como podría ser el tiempo libre de enfermedad o tiempo hasta la muerte (sobrevida)⁶⁰. A continuación, se ilustrará con un ejemplo el concepto de “peligro” y su estimación (ver Cuadro 4.6).

riesgo de tener citología con atípa o un resultado que sugería neoplasia intraepitelial cervical y el desarrollo de NIC II o NIC II estratificando por la cohorte de fecha de nacimiento por medio del modelo de riesgos proporcionales de Cox.

Se construyeron 5 cohortes: Aquellas nacidas entre 1989–1990; 1991–1992; 1993–1994; 1995–1996, y 1997–1999.

Se identificaron 399,244 mujeres danesas nacidas entre 1989 y 1999 que aportaron más de dos millones de años de seguimiento en el análisis del resultado. Un total de 247.313 mujeres fueron vacunadas (62%) principalmente entre aquellas nacidas después de 1993 (85%) cuando se inició el programa de vacunación siendo mucho más bajo entre las nacidas entre 1989- 1990 (14%) y 1991 a 1992 (27%). Al comparar entre las mujeres vacunadas y no vacunadas nacidas entre 1989 y 1999, se encontró menor riesgo de citología anormal entre las nacidas entre 1991 y 1992 y vacunadas, como se presenta en el Cuadro 4.6

Ejemplo:

Baldur-Felskov y cols estudiaron una cohorte de mujeres danesas nacidas entre 1989 y 1999⁶⁴. Determinaron quiénes de ellas fueron vacunadas y no vacunadas entre 2006 y 2012 para infección VPH; a partir del registro nacional de vacunación determinaron quiénes habían sido vacunadas y determinaron quiénes desarrollaron Neoplasia intraepitelial cervical (NIC), a partir del banco nacional de datos de patología. Luego compararon el

Cuadro 4.6 Ejemplo de razón de tasas (razón de peligros por el método del riesgo proporcional de Cox)

Riesgo en las nacidas entre 1991 y 1992 y vacunadas para presentar atipia cervical u otro resultado anormal (en relación a las no vacunadas):

- Razón de peligros (HR) 0,46 (IC 95%: 0,39 a 0,56)
- Riesgo en las nacidas entre 1991 y 1992 y vacunadas para presentar NIC II – III entre las vacunadas (en relación a las no vacunadas)
HR: 0,56 (0,37 a 0,84)
- La interpretación de estos resultados para efectos prácticos es similar a la del RR

4.3 ¿Cuál estimador de efecto relativo seleccionar?

En general el RR es más fácil de interpretar para los clínicos; sin embargo, el OR tiene propiedades matemáticas más estables como se observó en los Cuadros 4.4 y 4.5. El OR hace una estimación similar de la asociación al RR cuando la frecuencia del evento es menor al 1%. Por lo tanto, el OR podría ser considerado como estimador de efecto relativo cuando vamos a agrupar varios estudios, en un metaanálisis, para evaluar el efecto de una medicación, donde los resultados a evaluar tienen baja frecuencia.

Por último, el RR se puede estimar a partir del OR mediante la siguiente fórmula:

$$RR = OR / (1 - RAC) + (RAC \times OR)^{65}$$

Donde RAC es el riesgo asumido del grupo control que proviene de los datos de la literatura médica.

4.4 Medidas de efecto absoluto

Se estiman a partir de la diferencia de riesgos (RAR)
 $RAR = Rx. basal (Rx. I (-)) - Rx. con la intervención (Rx. I (+))^{60}$.

Esta diferencia se conoce como riesgo atribuible⁶⁶. Este se denomina también como reducción absoluta del riesgo⁶⁷. Se utiliza cuando se evalúa la disminución del riesgo de presentar un mal resultado, como sería en el caso del uso de la vacuna de VPH. Si se mide el incremento en el riesgo de presentar un evento indeseado al aplicar la intervención, se denomina incremento absoluto del riesgo⁶⁸.

Se conocen como estimadores de impacto y nos indican cuál es el beneficio absoluto o la proporción de "malos" resultados evitados al aplicar la intervención nueva. Por otra parte, permite determinar la proporción de resultados no deseados (eventos adversos) que se incrementan con la aplicación de la misma. Asimismo, permite detectar los efectos

reales de un tratamiento, ya que distingue entre los efectos grandes y los pequeños⁶⁹. En este análisis del efecto de la intervención es importante tener en cuenta el horizonte de tiempo en el que se miden los resultados⁶⁸.

Otra manera de presentar la información sobre el efecto absoluto de la intervención es por medio del número necesario para tratar (NNT). Este estimador nos indica cuántos sujetos debo tratar para evitar un mal resultado (u obtener un buen resultado). Este se obtiene a partir del inverso de la reducción absoluta del riesgo⁶⁷.

$$NNT = 1/RAR.$$

En cuanto a los efectos no deseados de la intervención, se estima el número necesario para dañar (NND) a partir del inverso del incremento absoluto del Riesgo: $1/IAR$.

En el Cuadro 4.7 se presentan los resultados de realizar estos cálculos a partir del estudio de Muñoz y colaboradores⁵².

Cuadro 4.7 Ejemplo de la estimación de la reducción absoluta del riesgo (evaluación de eficacia o efectividad)

Condilomas genitales			
	Sí	No	Total
Grupo riesgo intervención	63	8626	8689
Grupo riesgo control	305	8397	8702

- Riesgo de grupo de control (asumido) (sujetos con el evento entre los no vacunados) = $305/8702 = 0,035$
- Riesgo de grupo de intervención (correspondiente) (sujetos con el evento entre los vacunados) = $63/8689 = 0,007$
- Reducción absoluta del riesgo: Rx grupo de control – Rx grupo de intervención = $0,035 - 0,007 = 0,028$ o 2,8%
- Por lo tanto, se reduce cerca de un 3% el riesgo de tener lesiones genitales por VPH. Otra manera de presentar los resultados sería el NNT. Este se obtendría del inverso de la RAR o $1/RAR = 1/0,028 = 36$, es decir que prevengo la aparición de lesiones genitales en 1 sujeto por cada 36 sujetos que son vacunados. Este concepto será aplicado al hacer la evaluación la calidad de la evidencia en las tablas GRADE

A continuación, se toma el ejemplo de efectos no deseados de la vacuna para calcular el IAR de presentar eventos nocivos.

este estudio se asignaron 1035 sujetos al grupo de vacuna VPH -16/18 y 1032 al grupo control y fueron seguidos hasta el mes 7 un total de 1035 y 1032 respectivamente. De ellos presentaron (EAS) 13 del grupo de la vacuna y 11 del grupo control.

Ejemplo:

Medina y colaboradores evaluaron la seguridad de la vacuna contra el VPH en 2067 adolescentes en varios países, en cuanto a eventos adversos serios (EAS) en un seguimiento a 12 meses⁷⁰. En

En el Cuadro 4.8 se presenta cómo se construiría la tabla de contingencia para calcular el incremento del riesgo de presentar EAS al comparar estos grupos.

Cuadro 4.8 Ejemplo de la estimación del incremento absoluto del riesgo (evaluación de seguridad)

Eventos adversos serios a 12 meses	
Sí	No
13	1022
11	1021

- Riesgo de grupo de intervención (sujetos con el evento/ entre los vacunados) = $13/1009=0,012$
- Riesgo de grupo de control (sujetos con el evento / entre los no vacunados) = $10/1014 = 0,009$
- Riesgo relativo (RR): 1,3 Incremento absoluto del riesgo: 0,003, es decir, el riesgo de tener EAS se incrementa con la vacuna en 0,3% o 3×1000 . Por ende, el NND sería igual= $1/0,003 = 330$, es decir, por cada 330 sujetos que son tratados con la vacuna, un sujeto presenta al menos un EAS

Por último, el NNT se puede obtener a partir del RR o del OR cuando no se dispone de los datos para estimar directamente el riesgo atribuible. La fórmula para el cálculo varía de acuerdo con el sentido de la asociación. Además, se debe conocer la tasa de eventos esperada del paciente (TEEP) (patient expected events rate – PEER en inglés) o frecuencia de eventos del grupo control. Este se puede obtener a partir de la revisión de la literatura que permita conocer el pronóstico de los sujetos que no reciben tratamiento y son similares a los pacientes a los que queremos aplicar los resultados.

Si el estimador de asociación muestra que la intervención tiene un efecto protector de presentar un mal resultado (RR u $OR < 1$) se utilizan las siguientes fórmulas:

A partir del RR: $NNT = 1 / (1 - RR) \times TEEP$

A partir del OR: $NNT = 1 - (TEEP \times (1 - OR) / (1 - TEEP) \times (TEEP) \times (1 - OR)$

Si el estimador de asociación muestra que la intervención tiene un efecto de incremento del riesgo de resultados indeseados (RR o $OR > 1$) o de presentar un resultado deseado (recuperación de la función) se utilizan las siguientes fórmulas

A partir del RR: $NNT = 1 / (RR - 1) \times PEER$

A partir del OR $NNT = (PEER) \times (OR - 1) / (1 - PEER) \times (PEER) \times (OR - 1)$ ⁶⁸.

Si se decide presentar un estimador agrupado de varios estudios del efecto de una intervención, se prefiere hacerlo mediante los estimadores de efecto

relativo por ser más estables que los de efecto absoluto; es decir, se recomienda presentar como RR, OR o HR tomando en cuenta los criterios antes descritos⁶⁰.

4.5 Precisión de los estimadores de efecto (relativo o absoluto)

Independientemente del estimador que se presente (RR, OR, HR o IAR, RAR) este se debe presentar acompañado del intervalo de confianza, generalmente del 95% (IC 95%), que nos indica el rango en el que varían: la relación o la diferencia de los resultados entre dos grupos. Nos indica con qué probabilidad (en este caso 95%) el verdadero valor del parámetro estimado estará dentro los valores contenidos en el intervalo. De otra manera, si se tomaran muchas muestras de la población y el intervalo de confianza fuera estimado para cada uno de ellas, en un 95% de las muestras, el intervalo contendría el valor del parámetro y solo en el 5% de ellas no estaría contenido; esto, por efecto del azar^{71, 72}. El IC 95% nos da una idea de la incertidumbre que rodea al estimador, del error aleatorio y, además, nos indica si el estimador tiene asociación estadísticamente significativa. Hay que recordar que además de la significancia estadística, es importante considerar si los resultados tienen relevancia clínica⁶⁷.

4.6 Conceptos clave

La estimación del efecto benéfico de una tecnología (reducción de los “malos resultados”) o del riesgo de utilizar una tecnología (incremento de los resultados no deseados asociados al uso de la tecnología) se puede hacer mediante estimadores relativos de efecto (RR,

OR, HR) o estimadores absolutos de efecto (RAR, IAR, NNT o NND).

Los conceptos presentados en este capítulo — tales como riesgo del evento en el grupo control (riesgo asumido), riesgo del evento en el grupo de intervención (riesgo correspondiente), el efecto relativo (RR, OR, HR), el número de participantes que aportan información para el resultado, el número necesario para tratar o el número necesario para dañar (x 1000 sujetos)— son indispensables para generar y entender las tablas de resumen de hallazgos y la evaluación del conjunto de la evidencia siguiendo la aproximación GRADE (ver: sección 7.1).

5. CALIDAD METODOLÓGICA DE LOS ESTUDIOS Y VALIDEZ DE LAS OBSERVACIONES

Hernando G. Gaitán Duarte, MD, MSc
María Ximena Rojas Reyes, MSc, PhD

Desde los años ochenta del siglo pasado, la evaluación de la calidad metodológica fue considerada como un componente crítico, por lo que se sugirió incluir, en los estudios de revisión de la literatura, la evaluación de la calidad metodológica de los estudios primarios seleccionados, en cuanto al diseño, la conducción y el análisis⁷³. Inicialmente la evaluación se centró en determinar si los métodos utilizados eran los considerados válidos para el momento⁷⁴. Se daba una mayor ponderación al diseño epidemiológico de base, dando la mejor calificación inicialmente a los ensayos clínicos aleatorizados (ECA) como diseño para evaluar el efecto de una intervención⁷⁵ y, en años posteriores, la mejor calificación a las revisiones sistemáticas (RS) de ECA con alta homogeneidad seguida por los ECA^{76,77}.

Por otra parte, desde los años noventa ha existido otra tendencia más orientada a evaluar la validez de los estudios, en términos de si los resultados obtenidos eran realmente una adecuada estimación de la realidad, o si más bien podrían estar afectados por sesgos que llevan a distorsionar la dirección o la magnitud real del efecto (en el caso que nos ocupa, de una intervención o tratamiento)⁷⁸.

Estas dos aproximaciones a la calidad de los estudios se mantienen hoy en día. Algunos autores consideran que se debe priorizar la estimación del riesgo de sesgos^{73,79} y otros hacen énfasis en la adherencia a los aspectos metodológicos⁸⁰. Sin embargo, podrían ser más bien dos miradas complementarias a evaluar en los estudios primarios y los estudios integrativos o secundarios, tales como las guías de práctica clínica, las RS y las evaluaciones de tecnología (ET). Adicionalmente, han surgido recomendaciones sobre cómo asegurar la calidad de los reportes de la investigación como un criterio de calidad⁸¹. Esto ha llevado a que algunos autores utilicen las herramientas para verificar la calidad del reporte como instrumentos para evaluar la calidad metodológica, particularmente en las revisiones sistemáticas⁸².

Nuestra intención en este capítulo es describir las herramientas que más se utilizan en la actualidad en la evaluación de los estudios primarios y de las RS de intervenciones, como estudio que integra el conjunto de la evidencia acerca de los efectos benéficos de una tecnología aplicada con fines preventivos terapéuticos⁸³, de acuerdo con el estado del arte en esta materia.

5.1 ¿Cómo evaluar la calidad de los ensayos clínicos aleatorizados?

Iniciamos esta aproximación con la evaluación de los ECA, por ser el estudio en el que se basan las RS de intervenciones. Desde hace tiempo se reconoce que los ECA también son susceptibles de sesgos⁸⁴ y que, por lo tanto, es muy importante evaluar aquellos aspectos del diseño del estudio y su conducción,

elementos que son críticos para garantizar la validez de sus resultados⁷⁸. Con base en esta premisa se han construido múltiples escalas para evaluar la calidad metodológica, por ejemplo, la escala de Jadad⁸⁵, aún utilizada por algunos autores⁸⁶. Sin embargo, la evaluación del comportamiento de estas escalas mostró gran variabilidad en la calificación de la calidad de los estudios⁸⁷. Esto, junto con el hecho de que muchas de ellas no habían seguido un proceso de validación claro, ha llevado a que en la actualidad se desaconseje su uso⁷³.

Como alternativa, y con el objetivo de estar seguros de qué tanto podemos creer en los resultados de los estudios (validez de los resultados), se construyó en el interior de la Colaboración Cochrane el instrumento para la evaluación del riesgo de sesgos de los ECA. Se parte del hecho de que, aunque un estudio puede haber sido hecho con gran calidad metodológica, puede tener riesgo de sesgos. Por lo tanto, se plantea la pregunta sobre las amenazas a la validez de manera directa y separa este aspecto de la calidad metodológica⁷².

El instrumento del riesgo de sesgos está basado en la evaluación de las estrategias utilizadas por los autores, en el diseño del estudio y la conducción del mismo, a fin de minimizar el riesgo de sesgos. Entre las estrategias a evaluar se encuentran: el uso de una asignación aleatoria y el ocultamiento de la asignación (que minimizan el riesgo de sesgos de selección), el enmascaramiento al paciente y los investigadores de la intervención administrada (que minimizan el riesgo de sesgo de desempeño y sesgo de detección), el control de las pérdidas en el seguimiento de los sujetos a estudio (reducen el riesgo de sesgo por datos faltantes) y el reporte de datos de acuerdo a un

protocolo, desarrollado previamente a la conducción del estudio (que reduce el sesgo de reporte). También se ha mencionado que los resultados de los ECA pueden estar en riesgo de otros sesgos derivados por: a) la suspensión temprana del estudio, b) la utilización de resultados complejos de difícil interpretación, y c) el uso de ECA de diseños especiales como los diseños cruzados y en estudios de asignación de conglomerados (cluster), a la intervención, que requieren consideraciones específicas en el análisis⁸⁸. Una copia del instrumento de riesgo de sesgos propuesto por la Colaboración Cochrane se presenta en el Anexo 1 de este Manual.

Para iniciar la evaluación de los estudios seleccionados como evidencia para la ET, se sugiere construir una tabla por cada ECA incluido, en la que se evalúa cada uno de los dominios antes mencionados y se califica el estudio de acuerdo con la información provista en el manuscrito de la publicación y en el protocolo, en caso de estar registrado o publicado. La calificación se hace en tres grados: alto riesgo de sesgos, bajo riesgo de sesgos o no claro cuando no hay información suficiente para hacer el juicio. A continuación, se describen los criterios a tener en cuenta para la calificación.

5.1.1 Generación de la secuencia aleatoria (sesgo de selección)

Método por el cual se hace la asignación de los sujetos incluidos en el estudio al grupo de intervención (expuesto a la tecnología) o al grupo control (no expuesto a la tecnología). La asignación aleatoria tiene la gran ventaja de que distribuye de manera simétrica las variables pronósticas conocidas y no conocidas entre los dos grupos (control de los sesgos de confusión). El método requiere cumplir con algunos

criterios como que todos los sujetos tengan la misma probabilidad de ser asignados a alguno de los dos grupos, que esta probabilidad sea conocida y no se pueda predecir a qué grupo será asignado el siguiente sujeto⁸⁹. Sobre este criterio, la calificación se hace de la siguiente manera: **a) Alto riesgo:** utiliza un método que permite predecir a qué grupo va a ser asignado el siguiente sujeto (métodos que asignan de forma alternativa, por ejemplo, basados en la fecha de cumpleaños, el número del caso, el documento de identificación, disponibilidad de la tecnología).

b) Bajo riesgo: utiliza un método que tiene propiedades matemáticas conocidas y no permite predecir a qué grupo va a ser asignado el sujeto (estudios que tengan métodos como lanzar una moneda, lista de números aleatorios, o códigos de asignación generados por computador). c) Riesgo no claro: no se describe en detalle el método que se utilizó para generar el código de asignación (el uso de afirmaciones tales como “se hizo asignación aleatoria o diseño aleatorizado” es insuficiente para saber la técnica utilizada y no permite su clasificación en “Alto riesgo” o “Bajo riesgo” de sesgos).

5.1.2 Ocultamiento del código o de asignación (sesgo de selección)

Por este método se describen los procedimientos que permiten que la asignación permanezca oculta (en inglés, *concealment*) hasta el momento más cercano al que el sujeto recibe la intervención⁹⁰. La calificación sigue los siguientes criterios: **a) Alto riesgo de sesgo:** los métodos utilizados permiten conocer el grupo al cual serán asignados los sujetos (grupo de intervención o grupo de control) de manera previa a su implementación (ej., listados públicos de códigos de asignación, alternación o rotación o números consecutivos). **b) Bajo riesgo de sesgo:** hay

ocultamiento del código de asignación de los sujetos incluidos en el estudio tanto al personal como a los participantes involucrados en el estudio y es poco probable que haya sido descubierto (ej., uso de sobres sellados, códigos de barras, llamadas a centrales telefónicas para conocer el código de asignación).

c) Riesgo no claro: la información acerca de cómo se mantuvo oculto el código de la asignación aleatoria es insuficiente para permitir el juzgamiento de “bajo riesgo” o “alto riesgo”; por ejemplo, no se describe el método de ocultamiento o se menciona, pero no se describe de forma detallada.

5.1.3 Enmascaramiento de la intervención: cegamiento de los participantes y el personal que asigna la intervención o sus resultados (sesgo de desempeño y sesgo de detección)

Por enmascaramiento de la intervención se conoce a la estrategia que busca que no sea posible conocer qué tratamiento recibe el sujeto en el primer momento del estudio (t_0), ya sea que reciba el medicamento en evaluación o el tratamiento de control. También se denomina cegamiento. Un estudio es conocido como “simple ciego” cuando quien administra la intervención o quien hace la evaluación del resultado desconoce si el paciente recibió el medicamento en evaluación o el de control. Cuando también el paciente lo desconoce se denomina “doble ciego” y cuando, además de los anteriores, el estadístico o quien analiza los datos desconoce la intervención se dice que el estudio es “triple ciego”. Sin embargo, la interpretación de este término varía de manera importante⁹¹. El enmascaramiento se logra por medio de la utilización de un placebo que no contiene tratamiento activo, aunque si tiene apariencia idéntica al tratamiento activo. En el caso de utilizar un tratamiento activo

control, el enmascaramiento se consigue a través de una idéntica apariencia al tratamiento en evaluación. Es más importante en desenlaces subjetivos y poco repetibles (blandos) que en resultados objetivos y repetibles (duros). Por esto se sugiere que este dominio sea dividido según el tipo de resultados (blandos o duros)⁷³. La ausencia de enmascaramiento o estudio abierto afecta el desempeño de los pacientes (pueden informar, por ejemplo, más o menos síntomas si saben que están en el grupo de control) y de los investigadores (estos pueden hacer cointervenciones diferenciales para favorecer a un grupo). Estos comportamientos pueden llevar a sesgos de desempeño. Por otra parte, la ausencia de enmascaramiento también puede afectar de manera diferencial la búsqueda de los resultados a evaluar por parte de los investigadores, esta búsqueda diferencial de resultados lleva al sesgo de detección⁹².

Esta estrategia se califica de la siguiente manera:

a) Alto riesgo de sesgo: cuando el enmascaramiento está ausente o es incompleto y el desenlace tiene una alta probabilidad de estar influenciado por la falta de cegamiento del personal y los participantes (desenlaces subjetivos y poco repetibles). Otras opciones son: hubo enmascaramiento de solo algunas personas involucradas en el ensayo o se intentó el cegamiento pero no fue exitoso (diferente sabor o efectos colaterales) y hay alto riesgo de que la medición de los desenlaces sea influenciada por el conocimiento de la intervención; **b) Bajo riesgo de sesgo:** enmascaramiento adecuado del personal y participantes involucrados en el estudio (medicamentos de igual apariencia y similares efectos en el paciente) y es poco probable que haya sido descubierto. Cegamiento incompleto o ausente: los autores de la revisión consideran que es poco probable que el

desenlace esté influenciado por la falta del mismo (resultados “duros” como pruebas de laboratorio);

c) Riesgo no claro: información insuficiente acerca del enmascaramiento de los participantes y el personal para permitir el juzgamiento de “bajo riesgo” o “alto riesgo”, como cuando no se describe el método o cuando se menciona ciego o doble ciego pero no está descrito de manera detallada.

No se debe confundir el ocultamiento con el enmascaramiento. El primero tiene que ver con la asignación y el segundo con la forma de administración del tratamiento. Por ejemplo, un estudio de “apendicectomía por laparoscopia” versus “apendicectomía abierta” puede tener un adecuado ocultamiento de la asignación por medio de sobres sellados y ser abierto, ya que no se puede enmascarar la intervención (la técnica quirúrgica de incisión abdominal es diferente y tanto el paciente como el investigador conocen qué tratamiento recibió el paciente).

El sesgo de desempeño y el sesgo de detección deben ser evaluados de manera separada, como un dominio diferente, aunque ambos estén relacionados con el enmascaramiento.

5.1.4 Datos incompletos de desenlace (sesgo de desgaste)

Se presenta cuando hay datos ausentes respecto a los resultados del estudio, ya sean los resultados primarios o los secundarios. Es un problema que puede surgir por pérdidas en el seguimiento de los pacientes (que no debe ser mayor al 20%), mala conducción del estudio, problemas de control de calidad de los datos (faltan datos de algunos resultados)⁷⁸ o “desgaste” de los pacientes, los cuales se retiran de manera

voluntaria a través del tiempo. Es conocido que los resultados a un mayor tiempo de seguimiento tienen menos observaciones. Por esto, se sugiere dividir entre resultados a corto y mediano plazo (ejemplo, recurrencia tumoral) y resultados a largo plazo (ejemplo, mortalidad a 5 años). Esta pérdida en los datos puede llevar a sesgos en la estimación del efecto real de la intervención^{73, 93}.

El resultado de la evaluación en este aspecto, se dará también de acuerdo con los siguientes criterios: **a) Alto riesgo de sesgo:** cuando se detecta que la causa de los datos ausentes está probablemente relacionada con el desenlace real. Para datos dicótomos, a menor frecuencia del evento, mayor es el efecto del sesgo en la estimación de la asociación. Para los datos continuos, a mayor número de sujetos con datos ausentes mayor es el efecto del sesgo en la estimación de la asociación. También se debe considerar cuando el análisis por protocolo ha sido realizado con una desviación grande de lo que inicialmente se había elegido en la aleatorización de los tratamientos o si la imputación de datos es potencialmente inapropiada. **b) Bajo riesgo de sesgo:** no se presentaron datos faltantes o es poco probable que las razones de los datos faltantes en los desenlaces se encuentren relacionadas con el verdadero desenlace (datos de supervivencia, etc.); o los datos perdidos se encuentran balanceados en números entre los grupos de intervención, con razones similares para su pérdida. Para los datos de desenlace dicótomos, no es suficiente la proporción de desenlaces faltantes en comparación con las tasas de eventos observadas para tener un impacto clínico relevante en el estimador del efecto de la intervención. Para los datos continuos, los datos perdidos no son suficientes para tener un impacto relevante en el tamaño del efecto observado. Por

último, se puede considerar que los datos perdidos han sido imputados utilizando métodos apropiados.

c) No es claro: información insuficiente acerca de las pérdidas/exclusiones para permitir el juzgamiento de “bajo riesgo” o “alto riesgo”, como cuando no se menciona el número de aleatorizados, o no hay razones que expliquen los datos perdidos.

5.1.5 Reporte selectivo (sesgos de reporte)

Se ha definido como las diferencias sistemáticas entre los resultados reportados y los no reportados⁹³ o el reporte parcial de los resultados en relación con los inicialmente definidos a ser informados, ya sea en cuanto a los resultados primarios o los secundarios (ver sección 2.3). La información sobre el tipo de resultados que se había planeado informar se debe obtener idealmente del protocolo registrado o de la sección de métodos del estudio⁷³.

La evaluación de este aspecto se hará de la siguiente forma: **a) Alto riesgo de sesgo:** en el caso de que no sean reportados todos los desenlaces primarios pre-especificados, o que uno o más de los desenlaces de interés sean reportados de forma incompleta; o que el estudio no reporte un desenlace clave para la condición analizada. **b) Bajo riesgo de sesgo:** está claro que el reporte publicado incluye todos los desenlaces esperados e inicialmente contemplados en la metodología. **c) No es claro:** la información es insuficiente acerca del reporte selectivo como para clasificar en “alto o bajo riesgo de sesgos”.

5.1.6 Otros sesgos (búsqueda de posibles sesgos no cubiertos en las otras secciones)

Otros sesgos pueden surgir por diferencias en la distribución de los factores pronósticos, por análisis inapropiado para un diseño complejo (ensayos aleatorizados de conglomerados, *cluster* en inglés), o diseños cruzados (*cross over trial* en inglés), o por uso de desenlaces complejos difíciles de interpretar o por terminación temprana del estudio⁷².

Para el primer escenario, se recomienda evaluar las diferencias en las características basales del grupo que recibió la intervención en evaluación y el grupo de control; esto permite determinar si hay diferencias clínicamente relevantes en factores pronósticos que pueden afectar la relación entre el tratamiento y (los) desenlace(s) de interés. En cuanto a los diseños especiales, tal como en el caso del ensayo aleatorizado en conglomerado, se debe revisar si se tuvo en cuenta la disminución artificial de la variabilidad entre individuos que son reclutados en un mismo conglomerado⁹⁴. En el caso del ensayo cruzado, es importante determinar si hubo un tiempo de lavado entre el primer y segundo periodo, y si hubo interacción entre el tratamiento y el periodo, ya que ambas situaciones pueden afectar la estimación real del efecto del medicamento y deben ser tomadas en cuenta⁹⁵. Respecto a los resultados complejos es importante evaluar su interpretación⁹⁶. En cuanto a la suspensión temprana de los ECA hay que evaluar las razones aducidas para suspender un estudio de manera temprana con base en los análisis interinos y el papel de los comités monitores en estas decisiones^{97,98}.

La evaluación de la presencia de cualquiera de estos sesgos en el estudio nos llevará a una de tres

opciones de respuesta: **a) Alto riesgo de sesgo:** hay diferencias en las características basales entre los grupos de comparación; o suspensión temprana del estudio o análisis que no toman en cuenta aspectos especiales del diseño cruzado o la asignación en conglomerado; o el estudio fue finalizado en forma temprana sin seguir reglas claras y sin plena información del comité monitor de datos. **b) Bajo riesgo de sesgo:** no hay diferencias en las características basales entre los grupos de comparación. Los análisis toman en cuenta aspectos especiales del diseño cruzado o la asignación en conglomerado. El estudio fue finalizado en forma temprana siguiendo reglas claras y con plena información del comité monitor de datos. **c) No es claro:** no hay suficiente información para evaluar si existen otras fuentes de sesgos o hay insuficiente evidencia de que un problema identificado pueda llegar a introducir sesgo.

A continuación, se ilustran con un ejemplo los conceptos mencionados que subyacen a la evaluación de un ECA con el instrumento de riesgo de sesgos de la Colaboración Cochrane.

Ejemplo:

Realizaremos la evaluación del riesgo de sesgos del estudio de Muñoz y colaboradores⁵². Además de tomar en cuenta los dominios principales, hemos dividido los sesgos de desempeño y de detección, derivados de conocer si el paciente recibió la intervención nueva o la intervención control, de acuerdo con si son resultados o desenlaces blandos o duros y el sesgo de desgaste derivado de las pérdidas en el seguimiento o la ausencia de datos en algunas

variables en resultados a corto y largo plazo. Esto se debe a que esos desenlaces podrán ser afectados de manera diferente en su medición de acuerdo al tiempo de seguimiento como se mencionó antes. Los resultados de la evaluación se presentan en el Cuadro 5.1.

Cuadro 5.1 Ejemplo de calificación del riesgo de sesgos

Referencia: Muñoz N, Kjaer SK, Sigurdsson K, Iversen OE, Hernandez-Avila M, Wheeler CM, et al. Impact of human papillomavirus (HPV)-6/11/16/18 vaccine on all HPV-associated genital diseases in young women. J Natl Cancer Inst. 2010;102(5):325-39 ⁵².

Dominio	Clasificación del estudio			Observaciones
	alto	bajo	no es claro	
1. Generación de la secuencia aleatoria (búsqueda de posibles sesgos de selección)		X		Los autores reportan: "A computer-generated randomized allocation schedule provided by the statistician and an interactive voice response system were used to randomize subjects within each study center in a 1:1. Permuted block sizes of 4 and 13 were employed for Protocols 011 and 012, respectively. "
2. Asignación y ocultamiento (búsqueda de posibles sesgos de selección)			X	No se menciona el método de ocultamiento; sin embargo, es probable que se haya utilizado algún método de ocultamiento de la asignación dado que el estudio fue doble ciego
3.1.a Enmascaramiento de los participantes y el personal (búsqueda de posibles sesgos de desempeño desenlaces duros. (Testing of HPV with NAAT)		X		Los autores reportan: Double-blind, placebo-controlled trials. The quadrivalent HPV-6/11/16/18 L1 virus-like-particle vaccine with amorphous aluminum hydroxyphosphate sulfate (Gardasil, Merck) as an adjuvant and the aluminum-containing placebo were visually indistinguishable for researchers and patients
3.1. b Enmascaramiento de los participantes y el personal (búsqueda de posibles sesgos de desempeño desenlaces blandos)		X		Los autores reportan: Double-blind, placebo-controlled trials. The quadrivalent HPV-6/11/16/18 L1 virus-like-particle vaccine with amorphous aluminum hydroxyphosphate sulfate (Gardasil, Merck) as an adjuvant and the aluminum-containing placebo were visually indistinguishable
3.2. a Cegamiento de la evaluación del desenlace (búsqueda de posibles sesgos de detección, resultados duros)		X		Los autores reportan: Double-blind, placebo-controlled trials. The quadrivalent HPV-6/11/16/18 L1 virus-like-particle vaccine with amorphous aluminum hydroxyphosphate sulfate (Gardasil, Merck) as an adjuvant and the aluminum-containing placebo were visually indistinguishable

3.2.b. Cegamiento de la evaluación del desenlace (búsqueda de sesgos de detección- resultados blandos)		X		Los autores reportan: <i>double-blind, placebo-controlled trials. The quadrivalent HPV-6/11/16/18 L1 virus-like-particle vaccine with amorphous aluminum hydroxyphosphate sulfate (Gardasil, Merck) as an adjuvant and the aluminum-containing placebo were visually indistinguishable</i>
4.1 Datos de desenlace incompletos (búsqueda de los posibles sesgos de desempeño debido a la cantidad, naturaleza y manejo de los datos de desenlace incompletos, con resultados a corto plazo de 3 años)		X		<i>El seguimiento a los 3 años fue superior al 90%</i>
4.2 Datos de desenlace incompletos (búsqueda de los posibles sesgos de desempeño debido a la cantidad, naturaleza y manejo de los datos de desenlace incompletos, con resultados a largo plazo de 4 años)	X			No hay clara información del número de pacientes seguidos a 4 años
5. Reporte selectivo (búsqueda del sesgo de reporte)		X		Se dispone del protocolo que plantean los mismos desenlaces que fueron informados
6. Otros sesgos (búsqueda de posibles sesgos no cubiertos en las otras secciones)			X	No se encuentra la tabla de características basales

5.2 ¿Cómo evaluar la calidad de las revisiones sistemáticas?

Las revisiones sistemáticas (RS) de intervenciones han sido definidas como *“el intento por reunir toda la evidencia empírica que cumpla con unos criterios de elegibilidad previamente especificados con la intención de contestar una pregunta científica específica. Utiliza métodos explícitos y sistemáticos que son escogidos con la intención de reducir los sesgos y, por lo tanto, proveer la información más confiable, de la cual se pueden derivar conclusiones y tomar decisiones”*⁹⁹. Dada esta definición, las RS son la mejor

alternativa para iniciar la búsqueda de evidencia sobre la efectividad y seguridad de intervenciones, dado que las RS producen una estimación más precisa del efecto de la intervención que los ECA individuales y pueden mostrar tendencias de beneficio de manera temprana, lo que potencialmente facilita que el nuevo conocimiento se incorpore a la práctica clínica diaria¹⁰⁰.

Se caracterizan por tener una pregunta de investigación previamente definida a contestar, un protocolo que guía la elegibilidad de los estudios a incluir y una metodología repetible respecto a la búsqueda de estudios disponibles publicados o no

publicados, la selección de los estudios y extracción de los datos, la calificación de la validez o la calidad metodológica y la presentación de la información de los estudios incluidos, los métodos para analizar, y el efecto resumido de la intervención a partir de los resultados obtenidos en los estudios incluidos. Esta metodología está destinada a incrementar la validez de los resultados obtenidos, ya sea que la estimación final del efecto sea presentada agrupando de manera cuantitativa los resultados (metaanálisis) o que esto no sea posible y los resultados de cada estudio se presenten de manera individual¹⁰¹. Como se puede deducir de las anteriores afirmaciones, las RS son también susceptibles de sesgos: en la selección de los estudios (solo escoger artículos de una subcategoría de la población que interesa), en la medición (los estudios incluidos tienen alto riesgo de sesgos), en el desempeño (los autores solo buscan en los estudios los resultados que les interesan) y los sesgos de publicación (solo escoger estudios con resultados positivos o un solo idioma)⁷⁹.

5.2.1 ¿Qué herramienta debo utilizar para evaluar las RS?

Hay dos herramientas que han sido consideradas adecuadas para evaluar la calidad metodológica de las RS⁷⁹. La primera y más antigua es parte de la serie de publicaciones realizadas por JAMA para usuarios de la literatura médica y desarrollada por Oxman y colaboradores en 1994⁹⁹; la segunda y más reciente, es la Assessment of Multiple Systematic Reviews tool (AMSTAR)¹⁰². El instrumento AMSTAR ha demostrado que puede evaluar la calidad metodológica y el riesgo de sesgos de las RS, a partir de su reporte¹⁰³. El AMSTAR evalúa 11 criterios y califica el número de criterios que cumple con relación al total. Una copia de este instrumento se encuentra en el Anexo 2.

Los criterios que evalúa el AMSTAR son los siguientes:

1. ¿Hubo un diseño “a priori”?
2. ¿Hubo duplicación de la selección de los estudios y de la extracción de datos?
3. ¿Se realizó una búsqueda amplia de la literatura?
4. ¿Se utilizó el estado de la publicación como criterio de inclusión (ejemplo, literatura gris)?
5. ¿Se provee una lista de los estudios incluidos y excluidos?
6. ¿Se entregan las características de los estudios?
7. ¿Se evaluó y documentó la calidad científica de los estudios?
8. ¿Se utilizó adecuadamente la calidad de los estudios en la formulación de conclusiones?
9. ¿Fueron apropiados los métodos para combinar los hallazgos de los estudios?
10. ¿Se evaluó la probabilidad de sesgo de publicación?
11. ¿Fueron declarados los conflictos de interés?

5.2.2 ¿Cómo proceder si existen varias RS que evalúan la misma pregunta?

Para definir la mejor RS (aquella en que se va a basar la efectividad y seguridad de la tecnología en evaluación) se sugieren los siguientes criterios:

Evaluar la fecha de la última búsqueda de la literatura (aunque no hay un acuerdo sobre la fecha en la cual se debe actualizar una revisión sistemática)¹⁰⁴. Por una parte, la Colaboración Cochrane considera que debe ser cada 2 años¹⁰¹, pero también se ha informado que la sobrevivencia media de una revisión es de 5 años; sin embargo, se ha descrito que un tercio de las RS requerirán actualización a los 3 años y cerca del 25% a los dos años¹⁰⁵. Por ser la ET un documento integrativo destinado a soportar la toma de decisiones

y elaboración de políticas, para este manual sugerimos considerar una RS actualizada a aquella con un máximo de tres años de publicada y en todos los casos hacer una búsqueda de la publicación de ensayos clínicos posteriormente hasta la fecha de elaboración del informe.

Revisar nuevamente los componentes de la pregunta PICO de la RS para verificar que sean pertinentes a la tecnología en evaluación aplicada en una condición o un subgrupo de población específica.

Seleccionar la RS con la más alta calificación con la escala AMSTAR. Se sugiere revisar cuidadosamente que cumpla favorablemente los ítems relacionados con: la selección de los estudios (ítem 2), la investigación no publicada (ítems 3, 4 y 10), método utilizado para calificar la validez de los estudios primarios o la calidad metodológica de los mismos (ítems 7 y 8) y el método utilizado para el análisis (ítem 9) ⁷⁹.

En caso de que la RS identificada como elegible tenga más de tres años de publicación, se debe actualizar a partir de una búsqueda complementaria. Los lineamientos para realizar este procedimiento se presentan el Anexo 3 de este manual.

A continuación, se ilustra con un ejemplo la evaluación de una RS siguiendo el instrumento AMSTAR.

Ejemplo:

Para evaluar la efectividad y seguridad de la vacuna contra el VPH, se hizo una búsqueda de las revisiones sistemáticas disponibles en la base de datos Medline vía Pubmed y se encontró la siguiente RS: Rambout, L., Hopkins, L., Hutton, B., Fergusson, D. Prophylactic vaccination against human papilloma virus infection and disease in women: a systematic review of randomized controlled trials. CMAJ. 2007 ; 177(5):469-79¹⁰⁶.

Se aplicó el instrumento AMSTAR y se obtienen los resultados que se presentan en el Cuadro 5.2

Cuadro 5.2 Ejemplo de la calificación de una Revisión Sistemática por el instrumento AMSTAR

Referencia: <i>Rambout L, Hopkins L, Hutton B, Fergusson D. Prophylactic vaccination against human papillomavirus infection and disease in women: a systematic review of randomized controlled trials. CMAJ. 2007;177(5):469-79¹⁰⁶.</i>				
EVALUADOR (INICIALES). HGD				
Aspecto/Ítem	Sí	No	No es claro	Observaciones
1. ¿Fue un diseño “a priori”? Criterio: la pregunta de investigación y los criterios de inclusión deben ser establecidos antes de iniciar la revisión. Criterio: es necesario que haga referencia a que un protocolo, la aprobación ética o los objetivos de investigación fueron publicados previamente para marcar “sí”.	X			
2. ¿Hubo duplicación de la selección de los estudios y de la extracción de datos? Criterio: deben existir al menos dos extractores independientes de los datos y un procedimiento para llegar a consenso en caso de que existan desacuerdos.	X			
3. ¿Se realizó una búsqueda amplia de la literatura? Criterio: la búsqueda debe realizarse en al menos dos fuentes electrónicas. El informe debe señalar los años que abarcó la búsqueda y las bases de datos utilizadas (ej. EMBASE, MEDLINE, Lilacs). Las palabras clave y términos MeSH deben estar explicitados y cuando sea posible, debe estar disponible la estrategia de búsqueda. Las búsquedas deben ser complementadas con consulta a registros especializados o expertos en el campo de estudio, y por la revisión de las listas de referencias en los estudios encontrados.	X			
4. ¿Se utilizó el estado de la publicación como criterio de inclusión (ejemplo, literatura gris)? Criterio: los autores debieran declarar que ellos buscaron trabajos sin fijarse en el estado de publicación, el idioma del artículo, etc.	X			
5. ¿Se provee una lista de los estudios incluidos y excluidos? Criterio: debe entregarse una lista de los estudios incluidos y excluidos.	X			No de los estudios excluidos
6. ¿Se entregan las características de los estudios? Criterio: debe incluirse en forma completa, en tablas, la información esencial de los estudios originales, tal como el tipo de participantes, las intervenciones y los desenlaces evaluados en cada estudio.	X			
7. ¿Se evaluó y documentó la calidad científica de los estudios? Criterio: se deben describir los métodos de evaluación a priori.	X			La calidad metodológica se evaluó mediante la escala de Jadad
8. ¿Se utilizó adecuadamente la calidad de los estudios en la formulación de conclusiones? Criterio: los resultados del rigor metodológico y calidad científica deberían considerarse en el análisis y las conclusiones de la revisión, y declararse explícitamente en la formulación de recomendaciones.			X	
9. ¿Fueron apropiados los métodos para combinar los hallazgos de los estudios? Criterio: para los resultados agrupados, se debe realizar un test para evaluar su homogeneidad (ejemplo, prueba de Chi cuadrado para homogeneidad, I ²), para asegurar que los estudios son combinables. La heterogeneidad en los resultados de los estudios primarios puede explicarse por distintas razones: - Por efecto del azar (para ello se utilizan los “test de homogeneidad”, que evalúan si las diferencias en los efectos observados entre los distintos estudios difieren más allá de lo que uno pudiera esperar simplemente por azar) - Por diferencias en los diseños de investigación, o - Por variaciones en los componentes básicos del estudio: población (ejemplo, poblaciones de edades distintas), intervención (ejemplo, el fármaco se usó por un tiempo o a una dosis distinta entre un estudio y otro), o en la forma de medir los resultados. La revisión debería analizar cada uno de estos factores. Si hay heterogeneidad, debe utilizarse un modelo de efecto aleatorio y/o debe considerarse la pertinencia clínica de combinarlos (ejemplo, ¿Es prudente (razonable) combinar? O sea, ¿no son peras y manzanas?).	X			
10. ¿Se evaluó la probabilidad de sesgo de publicación? Criterio: una evaluación del sesgo de publicación debe incluir una combinación de apoyos gráficos (tales como “funnel plot”) y/o pruebas estadísticas (ejemplo, test de regresión de Egger). Si bien esto es lo óptimo, sobre todo si uno sospecha que podría haber un cúmulo importante de estudios no publicados en el tema, a los que no se pudo acceder con la búsqueda; en la práctica son pocas las revisiones que lo evalúan formalmente.		X		
11. ¿Fueron declarados los conflictos de interés? Criterio: las fuentes de ayuda económica deben estar claramente reconocidas, tanto en la revisión sistemática como en los estudios incluidos en ella. Es importante verificar también que las conclusiones de los autores deriven directamente de los resultados de la revisión, y que no se planteen como ciertas asociaciones que no hayan sido debidamente demostradas, ni que, existiendo suficiente evidencia como para sostenerlas, la magnitud de las mismas sea distorsionada (exagerada) por un uso abusivo o poco preciso del lenguaje.	X			
PUNTAJE	9			

Vemos que según el AMSTAR, esta RS tuvo una calificación de 9/11. Los ítems 2, 3, 4, 7 y 9 fueron bien calificados, el ítem 8 no es claro y el ítem 10 no fue evaluado. Por lo tanto, se considera que esta RS puede ser utilizada para evaluar el conjunto de la evidencia publicada hasta el 2007. Sin embargo, habrá que hacer una búsqueda de nuevos estudios publicados desde el 2007 hasta la fecha.

5.3 Evaluación de estudios no aleatorizados como evidencia del efecto de intervenciones

En algunas situaciones en la evaluación de la efectividad de los medicamentos no se dispone de ECA, por lo que se podría recurrir a los estudios no aleatorizados (ENA) tales como los experimentos de fase II sin grupo de control¹⁰⁷, los cuasi-experimentos, que incluyen los estudios con grupo de control con una asignación que no cumple con los criterios de ser aleatoria, y los estudios donde el investigador no intenta ninguna asignación aleatoria⁵⁹, entre los que están las cohortes, los estudios de antes y después, los casos y controles, y los mal llamados “series de casos” que corresponden a series de expuestos a una intervención¹⁰⁸. Esta situación se presenta cuando es poco factible realizar un ECA, ya sea porque no es ético, no es necesario o es difícil conseguir sujetos para ser asignados en el grupo de control¹⁰⁹. Estas situaciones son la excepción que confirma la norma seguida en este manual, según la cual los efectos benéficos se evalúan por medio de ECAS explicativos o de eficacia y, aún de una mejor manera, con ECA pragmáticos o ECA de efectividad²⁴.

Los estudios no aleatorios son generalmente los que más aportan a la evaluación de seguridad de las intervenciones o tecnologías, por lo que la evaluación de la calidad metodológica y validez de este tipo de estudios se evalúa en el capítulo siguiente.

5.4 Conceptos clave

La verificación de la calidad de la evidencia es un paso importante en la evaluación de la efectividad y seguridad que resulta del uso de una tecnología por lo que debe ser realizada tanto en los informes rápidos como los completos de la ET.

La evaluación de la calidad de la evidencia busca determinar la validez de los estudios primarios tipo ensayo controlado aleatorizado. Se recomienda utilizar la herramienta de riesgo de sesgos de la Colaboración Cochrane.

La evaluación de la calidad de la evidencia de las revisiones sistemáticas de intervenciones debe orientarse también a la evaluación del riesgo de sesgos. Para esto se recomienda la herramienta AMSTAR. Se sugiere actualizar la revisión sistemática de la literatura si tiene más de tres años de publicada.

6. EVALUACIÓN DE LA SEGURIDAD DE LAS TECNOLOGÍAS

Hernando G. Gaitán Duarte, MD, MSc

María Ximena Rojas Reyes, MSc, PhD

6.1 ¿De dónde obtener la evidencia sobre los efectos adversos de las tecnologías?

Una vez comercializada una nueva tecnología, generalmente posterior a estudios clínicos de fase I a III, es necesario que se siga vigilando su seguridad durante todo su ciclo de vida. La farmacovigilancia es la responsable de monitorizar la seguridad de los medicamentos en la práctica clínica usual. Sin embargo, el poseedor de la licencia de comercialización es el responsable de la monitorización continua de la seguridad de sus productos, para informar a las autoridades responsables de la vigilancia de las tecnologías, sobre cualquier sospecha de reacción adversa relacionada con el medicamento, así como para asegurar que la información del producto se mantenga actualizada¹¹⁰. Cuando no hay un motivo específico de preocupación, el medicamento entra en el sistema general de farmacovigilancia que suele realizarse mediante el programa de notificación espontánea de sospechas de reacciones adversas. Cuando el perfil tóxico del nuevo medicamento o de los medicamentos de su clase lo aconsejan, suelen iniciarse estudios de poscomercialización específicos para vigilarlos.

Estos estudios pueden ser ensayos clínicos de fase IV o estudios epidemiológicos de tipo cohortes cuya finalidad es cuantificar el riesgo o aportar evidencia para establecer la relación beneficio/riesgo del medicamento o tecnología en las condiciones reales de utilización¹¹¹.

Los estudios no aleatorizados son especialmente útiles cuando se trata de efectos adversos de medicamentos, en especial cuando estos son raros, cuando se presentan a un tiempo largo de seguimiento o cuando son difíciles de reconocer por ser imprevistos¹¹².

En estas situaciones, los estudios observacionales del tipo de los casos y los controles¹¹³ o las series de casos¹¹⁴ pueden ser la única alternativa posible, en ausencia de una mejor evidencia¹¹⁵, aunque se debe reconocer que tienen mayor riesgo de sesgos.

Por otra parte, los desarrolladores de una ET deben hacer un esfuerzo especial en identificar la información de eventos adversos relacionados con la tecnología en evaluación a partir de las bases de datos de literatura médica, de los registros de farmacovigilancia entre otras fuentes, y de los folletos de información del producto que presentan los laboratorios farmacéuticos, en caso de que la tecnología sea un medicamento (ver Cuadro 3.5).

6.2 ¿Qué sesgos pueden afectar los estudios no aleatorizados?

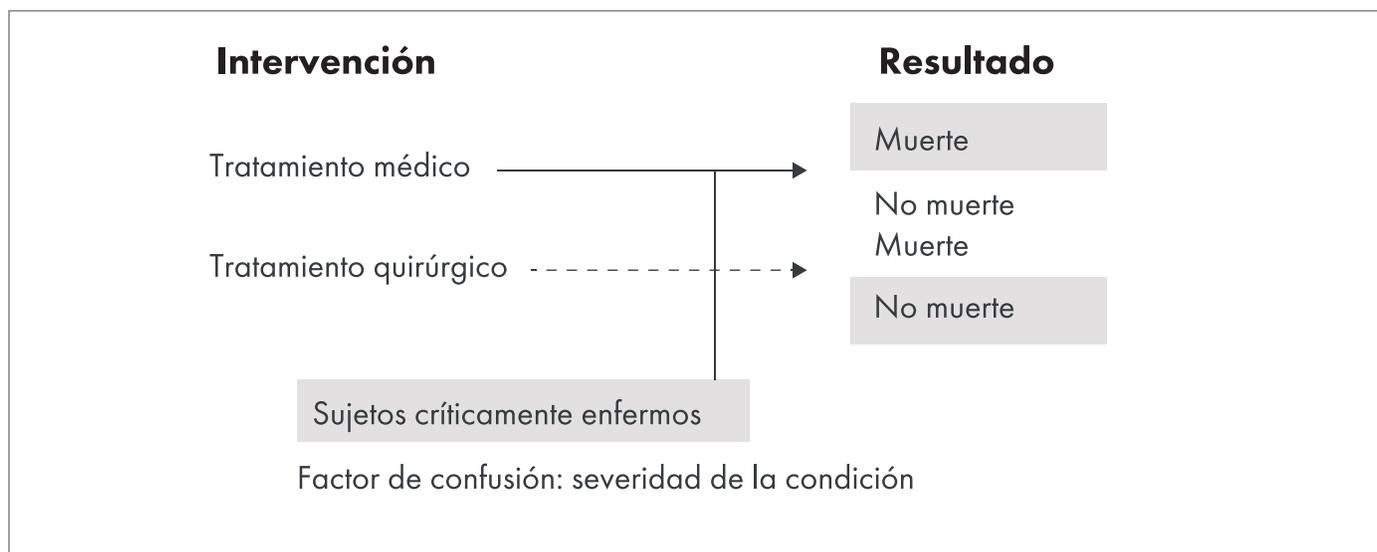
La consideración de “menor evidencia” dada por los estudios no aleatorizados viene del hecho de que estos estudios, además de ser susceptibles de

los mismos sesgos que afectan los estudios clínicos aleatorizados, pueden presentar los sesgos derivados del no control de los factores de confusión que pueden afectar la estimación del efecto de la intervención¹¹⁶. Vale la pena recordar que la asignación aleatoria distribuye todos los factores de confusión conocidos y no conocidos de manera simétrica entre los grupos de asignación (ver sección 5.1.1)²⁹.

En la asignación por métodos no aleatorios, se corre el riesgo de que algunas variables relacionadas con la condición sean las que determinan que el sujeto reciba el tratamiento y que también estén asociadas a los resultados, estas son conocidas como variables de confusión.

Para ilustrar el concepto asumamos que estamos comparando un tratamiento quirúrgico con un tratamiento médico para una condición dada y se utiliza un método de asignación no aleatorio, por ejemplo, por decisión del médico tratante. Supongamos que un paciente, severamente enfermo, es asignado al grupo quirúrgico. El servicio de anestesia ve al paciente y sugiere posponer la cirugía hasta que las condiciones del sujeto mejoren. Como consecuencia, el sujeto sería asignado al grupo de tratamiento médico. De esta manera, los sujetos más severamente enfermos serían más frecuentemente asignados al grupo de tratamiento médico. Recordemos también que los pacientes más severamente enfermos podrían tener más riesgo de morir (ver Cuadro 6.1).

Cuadro 6.1 Efecto de un factor de confusión en la asociación entre la intervención y el resultado



Como resultado de la anterior situación, habría una mayor mortalidad en el grupo de tratamiento médico: esta mayor mortalidad estaría dada por el mayor número de pacientes críticamente enfermos que fueron

asignados al grupo de tratamiento médico, en relación al grupo de tratamiento quirúrgico y no porque el tratamiento quirúrgico fuera más efectivo que el tratamiento médico.

La situación descrita considera un factor de confusión conocido (severidad de la condición); sin embargo, hay muchos otros escenarios en los cuales otros factores de confusión, no conocidos en el momento de realizar el estudio, podrían haber afectado la asociación entre la intervención y el resultado. Por lo tanto, en los diseños de asignación no aleatoria la estimación tiene mayor riesgo de estar sesgada acerca de los verdaderos beneficios o riesgos de aplicar la intervención por la mayor dificultad para el control de los factores de confusión.

Además del sesgo de confusión, los estudios no aleatorizados tienen mayor riesgo de otros sesgos. Por ser la asignación a los tratamientos predecible o ser conocida de antemano y no al azar, se podría manipular el ingreso de los pacientes a uno de los dos brazos, creando un sesgo de selección que, como resultado, lleva a que los factores pronósticos queden diferencialmente distribuidos entre los dos grupos. Por otra parte, por no haber enmascaramiento, se tendría mayor riesgo de cointervenciones diferenciales con mayor riesgo de sesgos de desempeño para parte de los investigadores o los pacientes, y mayor riesgo de sesgo de detección diferencial de los resultados. Además, es posible que los mecanismos de seguimiento sean menos estrictos en un brazo de tratamiento, por lo que podría haber mayor frecuencia de datos ausentes en uno que en otro brazo de tratamiento. Además, es menos probable que en los estudios de cohortes se solicite el registro del protocolo, lo que podría llevar a mayores sesgos de reporte. Todo lo anterior conlleva a considerar inicialmente los estudios no aleatorizados, como evidencia de baja calidad.

Se han sugerido algunas estrategias a ser utilizadas para mejorar la validez interna en los estudios de intervención no aleatorizados. Algunas provienen de los estudios clínicos aleatorizados (ECA), tales como aquellas que buscan no alterar el desempeño de los pacientes o los investigadores, los métodos que buscan no afectar la detección de los resultados y los métodos que están relacionados con el control de las pérdidas de datos en el seguimiento. Además, se utilizan otros métodos que aplican más a los estudios no aleatorizados, orientados a ensamblar grupos que sean más similares en sus características basales, lograr una mejor evaluación de la exposición, el control de los factores de confusión y ajuste por los factores pronósticos¹⁰⁹.

Se han creado múltiples escalas para evaluar la calidad de los estudios no aleatorizados con base en el análisis del cumplimiento de los criterios antes mencionados y la consideración de otros aspectos tales como la de haber sido creados para hacer revisiones sistemáticas (RS). Deeks y colaboradores¹⁰⁸ sugieren que las más adecuadas serían la de Newcastle–Ottawa¹¹⁷, por ser fácil y rápida de usar; y la de Reisch y colaboradores¹¹⁸ que es la más específica para la evaluación de medicamentos. En la actualidad se propone el uso de la herramienta ROBINS I de la Colaboración Cochrane¹¹⁹. Esta herramienta es considerada compleja y demorada para su aplicación, hasta por un 30% de los usuarios¹²⁰ y la confiabilidad interobservador ha sido informada como baja¹²¹. Por lo tanto, en este manual sugerimos el uso de la escala de Newcastle–Ottawa para la evaluación de evidencia de estudios no aleatorizados que evalúen la seguridad de las intervenciones, por ser más sencilla y rápida de aplicar. La herramienta ofrece una escala para evaluar

estudios de cohortes y otra escala para evaluar estudio de casos y controles, que describiremos a continuación. Las escalas tienen una calificación máxima de 8 a 9 puntos. (Anexo 4 ; Anexo 5).

6.3 Evaluación de los estudios de cohortes

La escala de Newcastle–Ottawa¹¹⁷ para la evaluación de estudios de cohortes toma en cuenta los siguientes dominios: la selección de las cohortes (expuestos y no expuestos), la comparabilidad de las cohortes y la medición de los desenlaces.

A. La selección de las cohortes a su vez se subdivide en cuatro preguntas que evalúan:

- La representatividad de los sujetos expuestos seleccionados en relación a los sujetos expuestos en la comunidad.
- La representatividad de los sujetos no expuestos seleccionados en relación a los sujetos no expuestos en la comunidad.
- Cómo se evaluó la exposición.
- Demostración de que el desenlace no estaba presente al inicio de estudio.

B. En cuanto a la comparabilidad de las cohortes, verificar si se aseguró en el diseño o en el análisis.

C. Respecto a la medición de los desenlaces, se subdivide en las siguientes preguntas:

- ¿Se midió adecuadamente el desenlace?
- ¿Fue el seguimiento lo suficientemente largo para que los desenlaces pudieran ocurrir?
- ¿El seguimiento fue adecuado (datos ausentes)?

6.3.1 Criterios para calificación de ítems o las preguntas a evaluar en los estudios de cohorte

A. Dominio de la selección de los sujetos

La representatividad de los sujetos expuestos seleccionados en relación con los sujetos expuestos en la comunidad.

Se pueden tener las siguientes alternativas:

Los sujetos seleccionados son representativos del promedio de sujetos expuestos a _____ (describa la exposición) en la comunidad*

Pertenecen a un grupo especial de usuarios, por ejemplo, voluntarios médicos

No se describe de dónde proviene la cohorte

La representatividad de los sujetos no expuestos seleccionados en relación con los sujetos no expuestos en la comunidad

Se pueden tener las siguientes alternativas:

- a) Proviene de la misma comunidad que la cohorte de expuestos*

- b) Proviene de un origen diferente
- c) No se describe de dónde proviene la cohorte de no expuestos

La evaluación de la exposición
Se pueden tener las siguientes alternativas:

- a) El registro de la exposición es válido y seguro (ej., registro de cirugía)*
- b) Se basa en una entrevista estructurada
- c) Se basa en un autorreporte
- d) No se describe

La demostración de que el desenlace no estaba presente al inicio del estudio
Se pueden tener las siguientes alternativas:

Es un desenlace duro*

Es un desenlace blando

Cada uno de los ítems marcados con el asterisco recibe un punto por ser considerado de alta calidad.

B. Dominio de la comparabilidad que se sugiere tener en cuenta:

- a) El estudio controló en el diseño o en el análisis por el factor de confusión más importante _____ (seleccione el factor que usted considera el primero más importante)*
- b) El estudio controló en el diseño o en el análisis por un segundo factor de confusión más importante _____ (seleccione el factor que usted considera el segundo más importante)*

- c) No se menciona si se estableció comparabilidad o solo se menciona que no hubo diferencias entre los grupos o que estas no fueron estadísticamente significativas

Este dominio es el único que puede recibir dos puntos si dos ítems son considerados de alta calidad.

C. Dominio de la evaluación de los desenlaces

¿Se midió adecuadamente el desenlace?

Se pueden tener las siguientes alternativas:

- a) Hay una evaluación enmascarada del desenlace; se utilizan desenlaces duros y se identifican a través de registros seguros o los desenlaces se identifican a través de códigos (Código internacional de enfermedades CIE X u otros registros)*

b) Autorreporte

c) Otros registros

¿Fue el seguimiento lo suficientemente largo para que los desenlaces pudieran ocurrir?

- a) Una cantidad de tiempo aceptable debe ser definida antes de que el estudio comience*
- b) No hay información de cuándo fue definida y qué tiempo se tuvo como horizonte para la aparición de los desenlaces

El seguimiento fue adecuado (datos ausentes)

a) Las pérdidas en el seguimiento son pocas (describir %) y son seguidas de igual manera entre los expuestos y los no expuestos*

b) Las pérdidas en el seguimiento son importantes (> 20%) y no son seguidas de igual manera entre los expuestos y los no expuestos

c) No se informan las pérdidas en el seguimiento

Cada uno de los ítems marcados con un asterisco recibe un punto por ser considerado de alta calidad.

6.4 Evaluación de los estudios de casos y controles

La evaluación de casos y controles según la escala de Newcastle–Ottawa toma en cuenta tres dominios: la selección de casos y los controles, la comparabilidad de los casos y los controles, y la evaluación de la exposición.

Dominio de la selección de los casos y los controles:
Se subdivide en cuatro preguntas que evalúan:

a) ¿Es adecuada la definición de los casos?

b) ¿Son representativos los casos?

c) ¿Es adecuada la selección de los controles?

d) ¿Es adecuada la definición de los controles?

Dominio de la comparabilidad de los casos y los controles:

Se aseguró en el diseño o en el análisis que:

a) Los controles del estudio fueron emparejados o ajustados por el factor de confusión más importante

Dominio de la medición de los desenlaces:

Se subdivide en las siguientes preguntas

a) ¿Se midió adecuadamente la exposición?

b) ¿Cuál fue la tasa de no respuesta?

6.4.1 Criterios para calificación de ítems o las preguntas a evaluar en los estudios de casos y controles

A. Dominio de la selección de los sujetos

1. ¿Es adecuada la definición de los casos?

a) Requiere una validación independiente, por ejemplo, basada en un registro hospitalario: historia clínica, Rx, imágenes diagnósticas, patología)*

b) Los desenlaces se identifican a través de códigos (Código internacional de enfermedades CIE X u otros registros o por autorreporte)

c) No se menciona cómo se identificaron los casos

2. ¿Son representativos los casos?

a) ¿Se tomaron todos los casos elegibles con el resultado de interés, de manera consecutiva, en el área seleccionada (región, institución, grupo de

hospitales) y en un periodo de tiempo definido o se tomó una muestra apropiada (ejemplo, muestra aleatoria) de los casos?

b) No se satisface este requerimiento o no se menciona

3. ¿Es adecuada la selección de los controles?

Este ítem evalúa si los controles provienen de la misma población que los casos y si los controles eran susceptibles de presentar el desenlace (si hubieran podido ser casos)

a) Los controles provienen de la misma comunidad que los casos y habrían podido tener el resultado (ser casos)*

b) Controles hospitalarios: provienen de la misma comunidad que los casos pero son tomados de población hospitalizada

c) No se describe

4. Definición de los controles

a) Si los casos son casos incidentes, debe ser explícito que los controles no han tenido historia del desenlace

b) No se describe

Nota: Si los casos son nuevos (pueden ser recurrencias, no solo casos incidentes), los controles con previas ocurrencias del desenlace no deberían ser excluidos.

Cada uno de los ítems marcados con un asterisco recibe un punto por ser considerado de alta calidad.

B. Dominio de la comparabilidad que se sugiere tener en cuenta:

a) El estudio controló en el diseño o en el análisis por el factor de confusión más importante. Señale cuál _____*

b) El estudio controló en el diseño o en el análisis por un segundo factor de confusión importante. Señale cuál _____*

c) No se menciona si se estableció comparabilidad o solo se menciona que no hubo diferencias entre los grupos o que estas no fueron estadísticamente significativas

Nota: Si los OR para la exposición a interés son ajustados por los factores de confusión listados se considera que los grupos son comparables por cada variable utilizada en el ajuste.

Este dominio es el único que puede recibir dos puntos si dos ítems son considerados de alta calidad.

C. Dominio de la evaluación de la exposición

Se midió adecuadamente la exposición

a) A partir de registros seguros o entrevistas estructuradas en las que el entrevistado estaba enmascarado sobre el estatus de caso o control
Entrevistas estructuradas en las que el entrevistado no estaba enmascarado sobre el estatus de caso o control*

b) Autorreporte o registro en la historia clínica únicamente

c) No se describe

Tasa de no respuesta

- a) La tasa de no respuesta es igual para ambos grupos
- b) Se describen los sujetos que no aceptaron ingresar al estudio (no respondedores)
- c) No se describe

Cada uno los ítems marcados con un asterisco reciben un punto por ser considerado de alta calidad.

Para este manual solo se otorgará un punto si están presentes los ítems que tienen una estrella. Calificación máxima 8 puntos.

En el Cuadro 6.2, se presenta un ejemplo de aplicación de esta escala al estudio publicado por Schwarz T. y colaboradores¹²² para evaluar la seguridad e inmunogenicidad de esta vacuna contra el VPH.

Cuadro 6.2 Ejemplo de la calificación de estudios de cohorte con la escala de Newcastle–Ottawa

Schwarz T y colaboradores¹²² publicaron en 2012 los resultados de un estudio de cohorte de 4 años de seguimiento en niñas de 10 a 14 años que recibieron la vacuna para VPH cepas 16 y 18, para evaluar la seguridad e inmunogenicidad de esta vacuna.

I. Dominio de la selección de los sujetos

La representatividad de los sujetos expuestos seleccionados en relación con los sujetos expuestos en la comunidad.

Los sujetos seleccionados son parte de los sujetos que fueron incluidos en el ECA que comparó la vacuna del VPH (grupo expuesto) con un grupo de control que recibió la vacuna de hepatitis (no expuestos) y se puede considerar representativo del grupo que entró al estudio. Por lo tanto, consideramos que pertenecen a un grupo especial de usuarios (voluntarios para entrar al estudio).*

La representatividad de los sujetos no expuestos seleccionados en relación con los sujetos no expuestos en la comunidad.

Los sujetos no expuestos provienen de la misma comunidad que la cohorte de expuestos ya que forman parte del ECA.*

Cómo se evaluó la exposición.

El registro de la exposición es válido y seguro.*

Demostración de que el desenlace no estaba presente al inicio de estudio
Se buscaron eventos adversos serios, enfermedades crónicas de inicio reciente y condiciones médicas significativas que no estaban presentes al momento del ingreso al estudio.*

Cada uno de los ítems marcados con una estrella recibe un punto por ser considerado de alta calidad.

II. Dominio de la comparabilidad que se sugiere tener en cuenta:

Si, el estudio controló en el diseño por los factores de confusión, ya que se ensambló sobre un ensayo clínico aleatorizado (ECA).* *

III. Dominio de la evaluación de los desenlaces

Se midió adecuadamente el desenlace.

La evaluación de los resultados a los 4 años no fue enmascarada se basó en los registros de la historia clínica o en los registros de urgencias por autorreporte.*

Fue el seguimiento lo suficientemente largo para que los desenlaces pudieran ocurrir Cuatro años es una cantidad de tiempo aceptable que fue definida antes de que el estudio comenzara.*

El seguimiento fue adecuado (datos ausentes).

Las pérdidas en el seguimiento para seguridad a los 48 meses son del 50%.

Este estudio recibe una calificación de 8/9 como cohorte

6.5 Conceptos clave

Los resultados de seguridad deben ser evaluados en todo informe de evaluación de tecnologías.

Las revisiones sistemáticas de intervenciones y los ensayos controlados aleatorizados son la fuente más válida de evidencia para los desenlaces de seguridad. Sin embargo, los eventos adversos poco frecuentes son difícilmente detectados en este tipo de estudios, por lo que se requieren otros diseños como fuente de evidencia. Estos son: estudios clínicos de fase IV, cuasi-experimentos, cohortes, estudios de casos y controles, y series o reportes de casos.

Los estudios no aleatorizados tienen mayor riesgo de sesgos, en especial sesgos de confusión, de selección, de detección y de desgaste.

Para las evaluaciones de tecnología se sugiere la herramienta de Newcastle–Ottawa para determinar la validez de los estudios de tipo cohorte o casos y controles.

7. EVALUACIÓN DEL CONJUNTO DE LA EVIDENCIA

María Ximena Rojas Reyes, MSc, PhD
Hernando G. Gaitán Duarte, MD, MSc

7.1 ¿Qué es el conjunto de la evidencia y cómo evaluarlo?

El grupo de estudios seleccionados como la “mejor evidencia disponible” para resolver las preguntas planteadas en la ET se conoce como “cuerpo de la evidencia”. Cuando este se evalúa se deben considerar los siguientes aspectos: el diseño epidemiológico, cómo se ejecutaron los estudios, qué tan directa es la evidencia obtenida respecto a la tecnología en evaluación, el número de estudios que cumplieron con los criterios de selección, el tamaño del efecto, la homogeneidad y consistencia de sus resultados, y el riesgo de sesgos. Estos criterios determinan la calidad global de la evidencia que soporta los resultados de la evaluación¹²³.

Hay varias aproximaciones para evaluar la calidad de la evidencia, entre ellas: la del National Institute for Health and Care Excellence (NICE)¹²⁴, la de los miembros de la tarea de fuerza en servicios comunitarios preventivos (Task Force)¹²⁵ y la del grupo de trabajo “Grading of Recommendations Assessment, Development and Evaluation (GRADE)”¹²⁶. La aproximación propuesta por GRADE es novedosa ya que busca evaluar, no un estudio de manera aislada, sino el conjunto de estudios que evalúan los efectos

de una intervención. Lo hace de una manera rigurosa, sistemática y transparente; por lo tanto, permite a los grupos desarrolladores de guías y elaboradores de informes de la ET, hacer una evaluación de la “calidad” del cuerpo de la evidencia en el desarrollo de una RS o de la confianza que podemos tener en qué tan “correcta” es la estimación del efecto de la tecnología en la condición, en un informe de evaluación tecnológica que soportará la toma de decisiones clínicas o en salud pública relacionadas con la tecnología¹²⁶. Es transparente ya que permite conocer los argumentos por los cuales los autores de la RS hacen los juicios de valor, de una u otra manera, para calificar la evidencia. Por otra parte, el concepto de calidad prioriza la implicación en investigación futura y el concepto de confianza prioriza la confianza en que el efecto estimado se aproxime actualmente al efecto real¹²⁷.

La aproximación GRADE es la más utilizada actualmente en organizaciones dedicadas al desarrollo de GPC como, por ejemplo, NICE, la Organización Mundial de la Salud (OMS), Scottish Intercollegiate Guidelines Network (SIGN) y es la más recientemente recomendada por las asociaciones de agencias de la ET, como European Network for Health Technology Assessment - EunetHTA, entre otras¹²⁸. En esta sección revisaremos los pasos sugeridos para hacer la evaluación de la calidad de la evidencia bajo la aproximación GRADE.

- a) Selección de los resultados en salud a considerar en la evaluación de tecnologías
- b) Calificación de la calidad del cuerpo de la evidencia identificado para la ET

c) Resumen del cuerpo de la evidencia relacionada con la ET

A continuación, describiremos estos pasos con más detalle.

7.2 ¿Cómo selecciono y califico los desenlaces relevantes para la evaluación de tecnologías?

Para iniciar se sugiere escoger los desenlaces que serán considerados en la ET. Se sugiere seleccionar los 4 a 5 desenlaces más relevantes para la toma de decisiones, con base en las preguntas PICO planteadas en el protocolo en la definición del alcance de la ET (ver sección 2.2). Se recomienda que la selección sea hecha por expertos temáticos y el grupo de metodólogos que realiza la ET. Es muy importante considerar el punto de vista de los pacientes, el de los tomadores de decisiones y, en el caso de las ET, de los pagadores. Se deben considerar tanto los resultados de beneficio o efectividad como los de riesgo o daño. Se sugiere incluir los desenlaces más importantes, aún si no hay evidencia que los haya evaluado, ya que será parte del informe de la ET en cuanto a la incertidumbre que hay en ese aspecto⁴⁴. Sobre cada uno de los desenlaces escogidos se valorará la

calidad del cuerpo de la evidencia para definir así la “confianza en la estimación del efecto” encontrado. Dado que es posible que una RS no haya evaluado todos los resultados que interesan a una ET, podría ser posible tener que recurrir a más de una RS⁴⁴.

Posteriormente, se deberá calificar la importancia relativa para cada desenlace. Para esto se utilizará una escala ordinal de 9 puntos que categoriza en tres estratos: 1 a 3 que califica como no importante al resultado, 4 a 6 como importante, y 7 a 9 como crítico (ver sección 2.3). Es posible que varios desenlaces tengan la misma calificación. Se sugiere incluir todos los resultados críticos y evaluar muy bien si es conveniente incluir alguno de los importantes, como también no incluir resultados intermedios o complejos⁴⁴ (Cuadro 7.1).

Cuadro 7.1 Criterios de calificación de los desenlaces

1 – 3	No importante (No incluida en la tabla de SoF)
4 – 6	Importante, mas no crítico para toma de decisiones (Incluida en la tabla de SoF)
7 – 9	Crítico para la toma de decisiones (Incluida en la tabla de SoF)

7.3 ¿Cómo juzgar la calidad del cuerpo de la evidencia?

Para los desenlaces de efectividad y seguridad de las intervenciones, la aproximación GRADE considera que los ensayos clínicos aleatorios (ECA) proveen la evidencia de más alta calidad; sin embargo, la confianza en la estimación del efecto del cuerpo de la evidencia que proviene de los ECA puede verse afectada por cinco aspectos: a) limitaciones en el diseño y la conducción del estudio que los lleven a un alto riesgo de sesgo; b) inconsistencia en los resultados de los estudios (diferencias que los hacen realmente heterogéneos); c) cuando los estudios proporcionan evidencia indirecta, es decir que la pregunta que se aborda en la ET es diferente a la que la evidencia disponible presenta en cuanto a la población, la tecnología en evaluación, el comparador o el desenlace evaluado; d) imprecisión en los resultados que se presenta cuando los estudios tienen bajo poder (los estudios incluyen relativamente pocos pacientes) o se presentan pocos eventos, lo que lleva a que el intervalo de confianza del estimador de efecto (relativo o absoluto) tenga límites amplios alrededor del efecto; y e) sesgo de publicación que lleva a subestimación o sobrestimación del efecto derivada de una publicación selectiva de estudios que muestran resultados positivos.

Ante la presencia de alguno de ellos (o de varios), disminuirá la “confianza o certeza” en el efecto estimado a partir del cuerpo de la evidencia o la calidad de la evidencia para el desenlace en evaluación. De esta manera la confianza en la estimación del efecto o la calidad a partir de los ECA se podrá calificar como alta, moderada, baja o muy baja¹²⁷.

En cuanto a los estudios no aleatorizados, la metodología GRADE los considera en principio como estudios con baja calidad metodológica, debido al alto riesgo de sesgos que tienen en: a) la selección de los sujetos, b) el control de los factores de confusión y c) el control en el análisis de los factores pronósticos, además de los sesgos que pueden afectar los ECA (ver sección 6.2). Sin embargo, provee algunos criterios para incrementar la calidad del cuerpo de la evidencia si el grupo de estudios muestra una importante fuerza de asociación, se muestra una substancial relación dosis-efecto y la persistencia de la asociación a pesar de la presencia de sesgos que tengan un efecto hacia el no rechazo de la hipótesis nula. Volveremos sobre estos conceptos más adelante¹²⁷.

A continuación, se explican en detalle las características a evaluar en cada uno de los estudios que aportan al cuerpo de la evidencia en cada uno de estos cinco aspectos.

7.3.1 Limitaciones del diseño (riesgo de sesgos)

Este aspecto se evalúa siguiendo los criterios para evaluar el riesgo de sesgos en los ECA¹²⁹. Se considera que hay limitaciones cuando hay:

- a) Un método de asignación no aleatorio
- b) Falta de ocultamiento de la asignación
- c) Falta de cegamiento (particularmente si los desenlaces son subjetivos o blandos)
- d) Pérdidas importantes durante el seguimiento, no están balanceadas entre los grupos o falta de adherencia en el análisis acorde al principio de intención de tratar

- e) Suspensión temprana de un estudio para beneficio
- f) Reporte selectivo de eventos: los investigadores omiten reportar desenlaces que habían sido planteados en el protocolo o han sido medidos (típicamente aquellos para los que no observaron efecto alguno)

Una vez evaluadas estas características en cada uno de los estudios, los evaluadores deben hacer una evaluación global de los estudios incluidos en las categorías: no serio; serio o muy serio. Por ejemplo, Moja y colaboradores (2012) realizaron una RS sobre el uso del trastuzumab en pacientes con cáncer de seno temprano¹³⁰. Incluyeron 8 ECA, todos los cuales tuvieron alto riesgo de sesgo de desempeño y detección por ser estudios abiertos; tres de ellos tuvieron alto riesgo de sesgo de reporte selectivo, y en otros tres no hubo claridad sobre este riesgo. Además, no fue claro en seis de ellos el método de ocultamiento. Con base en esto la calificación del riesgo de sesgos en el cuerpo de la evidencia podría ser como serio por falta de enmascaramiento en todos los estudios y reporte selectivo en el 40% de los estudios. Los autores de la RS, sin embargo, consideraron como bajo el riesgo de sesgos. Esta discrepancia ejemplifica una de las bondades de la metodología GRADE que permite conocer de manera transparente las razones que tuvieron los autores de la RS para calificar un criterio, y permite a los autores de la ET concordar o discrepar con la calificación y dar las razones para hacerlo. Para mayores detalles de este dominio se recomienda la lectura del capítulo 5 (ver sección 5.1) en la sección de ensayos clínicos aleatorizados.

7.3.2 Inconsistencia

Cuando los resultados de la estimación relativa del efecto del tratamiento (RR, HR, OR) difieren ampliamente entre los diferentes estudios, se denomina inconsistencia heterogeneidad o variabilidad de resultados¹³¹.

La inconsistencia puede surgir de diferencias en la población, la intervención o la medición de los desenlaces.

El grado de inconsistencia se evalúa por medio de:

- a) La visualización de cómo se distribuyen los resultados en la gráfica de efectos (*forest plot* en inglés)
- b) Ausencia de superposición de los intervalos de confianza
- c) Valor de la p en el valor del test Chi 2 es menor al 10%
- d) El valor del estadístico I^2 . Se considera baja heterogeneidad si el valor de $I^2 < 40\%$, importante heterogeneidad si el valor de I^2 es del 40 al 70% y heterogeneidad sustancial si el $I^2 > 70\%$

Para evaluar la inconsistencia se sugiere revisar si la fuente de heterogeneidad está originada en diferencias de respuesta en algún subgrupo de población (mayor o menor susceptibilidad al tratamiento), en la administración de la tecnología en evaluación (mayor dosis y efectos más grandes con dosis más altas) o en la medición de los resultados (disminución del efecto del tratamiento con el paso del tiempo o diferencias en el método de medición). Esta indagación se hace mediante el análisis de subgrupos. De esta manera,

en situaciones donde hay heterogeneidad importante o sustancial es pertinente presentar la síntesis de evidencia por subgrupos de población, de intervención o cómo midieron los desenlaces y no solamente de manera global. Por lo tanto, este análisis de subgrupos se debe buscar al evaluar la evidencia presentada por la RS seleccionada.

Si, a pesar de tomar en cuenta las posibles causas de heterogeneidad, esta no se puede explicar, significa que existen reales diferencias entre los estudios en la estimación del efecto real del tratamiento; es decir que hay inconsistencia. Cuando existe inconsistencia y los investigadores no logran detectar una explicación para ello, se debe bajar uno o dos niveles la calificación de la calidad de la evidencia, dependiendo de la magnitud de la inconsistencia en los resultados y del impacto de la inconsistencia en la evaluación del efecto de la intervención. Por ejemplo, no es lo mismo que la mitad de los estudios muestren un efecto significativo a favor de la intervención y la otra mitad no muestre beneficio con un I^2 del 75%, a que el 80% de los estudios muestre beneficio y solo el 20% no lo muestren con un valor I^2 del 50%. En el primer caso, se bajaría dos niveles, y en el segundo, solo un nivel.

Como ejemplo Singh y colaboradores¹³² publicaron una RS sobre el efecto de los medicamentos biológicos como único tratamiento en pacientes con artritis reumatoide resistentes al metotrexate (MTX) y los fármacos antirreumáticos modificadores de la enfermedad (FARME). Ellos incluyen la comparación de los medicamentos biológicos contra placebo en cuanto a la mejoría en un 50% (ACR 50) de la inflamación articular o en la evaluación que hace el paciente o el médico de la enfermedad, en la mejoría de la discapacidad o en el nivel sérico de

los reactantes de fase aguda, como resultado de efectividad. Incluyen en el metaanálisis 9 ECA, que hacen la comparación “cabeza a cabeza” entre los medicamentos y un placebo, e informan un OR de 6,39 (IC 95%: 3,76 – 10,85) con un valor para el estadístico I^2 del 58%, es decir hay importante heterogeneidad entre los resultados. Los autores hacen un análisis de subgrupos por tipo de biológicos, por lo que se hace la comparación entre medicamentos biológicos que actúan contra el Factor de necrosis tumoral (FNT), biológicos que actúan por otro mecanismo de acción, denominados en la RS como “no TNF” y el tofacitinb un bloqueador en la inhibición de varias de las isoformas de la proteína quinasa Janus. Al hacer este análisis informan:

- a) Comparación TNF vs placebo, que incluye un total de 7 ECA, con un OR de 5,43 (IC 95% 3,09-9,52), con un I^2 del 59%
- b) Comparación No TNF vs placebo con 2 ECA un OR de 13,20 (IC 95%: 5,13 – 32,99) con un I^2 de 0%
- c) Comparación Tofacitinob vs placebo que incluye un total de 2 estudios un OR de 5,63 (IC 95%: 2,0 – 15,45) con un I^2 del 65%

Los autores de la revisión para la comparación de Biológicos vs placebo disminuyen un nivel la calificación de inconsistencia (seria). Esta calificación se basa en que, a pesar de tener un valor de I^2 del 58%, la inconsistencia es parcialmente explicada en las diferencias en el tratamiento. Para la comparación TNF vs placebo los autores disminuyen un nivel la calificación de inconsistencia (seria) por el I^2 de 59%, sin embargo, dado que no hay explicación para esta heterogeneidad en los resultados, se podría considerar

como inconsistencia muy seria. De nuevo se resalta la importancia de proveer en el texto las razones que justifican el juicio de valores, en especial cuando no hay acuerdo en la evaluación del criterio

Los conceptos antes descritos solo aplican a las revisiones sistemáticas de comparaciones directas, ya que los conceptos de heterogeneidad e inconsistencia utilizados en las comparaciones indirectas tienen otras implicaciones¹³³(ver ejemplo: Cuadro 7.3).

7.3.3 Evidencia indirecta

El término evidencia indirecta involucra situaciones en las cuales se encuentran estudios o revisiones sistemáticas de estudios que evalúan la efectividad y seguridad de tecnologías, que contestan parcialmente la pregunta pico que soporta el alcance de la ET, pero pueden diferir en alguno de los componentes, por ejemplo, la población, la comparación directa de la tecnología en evaluación y el comparador o los resultados evaluados¹³⁴.

Miremos cada una de estas situaciones en detalle. En el caso de la población, se refiere a la aplicabilidad de los resultados de los estudios a la población objeto de la ET. Por ejemplo, se podrían estar evaluando como problema de política pública los beneficios y riesgos del uso de la vacuna del papiloma humano en adolescentes menores de 9 a 13 años sin antecedente de inicio de la vida sexual¹³⁵. Sin embargo, los estudios han sido realizados en mujeres mayores de 15 a 25 años con vida sexual activa^{136, 137}. De esta manera, la evidencia que soporta el uso de la vacuna en la población objeto (adolescentes de 9 a 13 años) es indirecta, ya que proviene del uso de la vacuna en una población diferente.

Respecto a la evaluación de la tecnología y el comparador, vale la pena recordar que cuando se está comparando una tecnología para prevenir una enfermedad (vacuna para el virus del papiloma humano (VPH) contra placebo) o para tratar una enfermedad (medicamento nuevo contra el tratamiento estándar, como ejemplo el uso de trastuzumab más quimioterapia estándar (tratamiento A) versus quimioterapia estándar sola (tratamiento B) para cáncer de mama), lo ideal es encontrar revisiones sistemáticas que hayan incluido ECA en los que se hace la comparación directa (llamadas cabeza a cabeza) de la tecnología a evaluar (tratamiento A) con el comparador (tratamiento B). Sin embargo, no siempre existen ECA que hayan hecho la comparación "cabeza a cabeza", por lo tanto, tenemos que recurrir a evidencia indirecta.

La comparación indirecta se presenta cuando encontramos dos estudios que comparan los dos medicamentos contra un tercero pero no de manera directa; por ejemplo, en uno la intervención A fue comparada con una intervención C (placebo), y en otro la intervención B también fue comparada con la intervención C (placebo). Estos estudios permiten hacer una aproximación al efecto de A con respecto a B de manera indirecta. En la actualidad hay metodologías robustas para hacer la síntesis de la evidencia a partir de comparaciones indirectas exclusivamente, ya sea mediante la utilización de un solo comparador intermedio (evidencia indirecta simple) o mediante el uso de múltiples comparadores intermedios (evidencia indirecta compleja)¹³⁸. Este tipo de evidencia nos da menor confianza en la estimación del efecto real de la comparación, que la evidencia que aportaría una comparación cabeza a cabeza entre A y B, en especial por el riesgo de una sobrestimación de la precisión en la estimación del efecto¹³⁹. Los metaanálisis en red,

que toman en cuenta tanto la evidencia directa como la indirecta, son una alternativa a considerar en la actualidad, ya que recopilan toda la evidencia en la comparación de dos tecnologías¹³⁹.

Como ejemplo de comparación indirecta tenemos el metaanálisis realizado para evaluar diferentes medicamentos sistémicos en el tratamiento de la psoriasis moderada a severa¹⁴⁰. En él se comparan tratamientos biológicos contra tratamientos convencionales a través de la diferencia de riesgos (DR) de mejoría del 75% en el índice de severidad del área de psoriasis (PASI 75). Aunque para algunas comparaciones hubo evidencia directa e indirecta (infiximab versus metrotexate), para muchas de ellas solo hubo evidencia indirecta (ej., infiximab versus adalimumab). En este último caso, ambos medicamentos fueron comparados con placebo en múltiples estudios y la diferencia de riesgos se tomó del estimador agrupado de los estudios que compararon el infiximab con placebo (seis estudios DR agrupada: 0,76 I²: 0%) y de los estudios que compararon adalimumab con placebo (cuatro estudios DR agrupada: 0,61 I²: 24%), para una diferencia de riesgos estimada de manera indirecta de 0,15 (IC 95%: 0,07–0,21).

En cuanto a los resultados, se podría tener evidencia indirecta cuando se toma información de un desenlace intermedio o en desenlaces relacionados con el tiempo si se tiene información de un resultado a un tiempo corto, pero el resultado más importante se debe medir a un tiempo mayor. Como ejemplo tenemos el metaanálisis que evalúa el efecto de la idursulfasa en síndrome de Hunter¹⁴¹ en el cual los resultados más importantes para el paciente son el cambio en la talla, en el peso y la edad a la muerte. Sin embargo, los

autores solo pudieron encontrar un ECA que evaluó el medicamento contra placebo, pero midieron la prueba de caminata de seis minutos, que es un desenlace importante, pero podría no ser crítico y concluir el beneficio del tratamiento a partir de este resultado.

Se deberán tomar en cuenta los tres criterios antes mencionados para disminuir 1 o 2 puntos la confianza en la estimación del efecto del cuerpo de la evidencia disponible. Por ejemplo, si la población de los estudios originales difiere de la población objeto de la ET, las comparaciones son indirectas y los resultados son intermedios, podríamos bajar dos niveles la calificación de la calidad.

7.3.4 Imprecisión

La imprecisión se evalúa observando el intervalo de confianza del estimador de efecto utilizado; si el intervalo de confianza, es amplio el estimador es impreciso¹⁴². El intervalo de confianza nos indica la probabilidad (usualmente 95%) de que el verdadero valor del estimador se encuentre entre un rango de valores en la población (el límite inferior y el límite superior del IC)⁷¹ (ver sección 4.5).

No hay un solo criterio para evaluar la imprecisión, más bien es un juicio de valores respecto a: el beneficio (efectividad) y el riesgo (efectos adversos) que implica el uso de la tecnología, la importancia de los resultados evaluados y el mínimo valor que aceptamos para considerar que la tecnología es útil. Las causas de imprecisión son: a) los estudios incluidos incorporaron relativamente pocos pacientes y b) los eventos son poco frecuentes, aunque el tamaño de muestra es importante.

Para desenlaces binarios se ha sugerido considerar el óptimo tamaño de la muestra OTM (más conocida por su nombre en inglés *optimal information size* OIS) para evaluar si el cuerpo de la evidencia tiene suficiente precisión. Si el número de sujetos incluidos en los estudios de la revisión es menor que el óptimo tamaño de muestra que se requeriría en un solo ECA —con un poder adecuado para poder encontrar una reducción esperada mínima en el desenlace dicotómico entre los grupos a evaluar— la evidencia obtenida se debe considerar imprecisa. Este cálculo también se puede basar en la reducción relativa del riesgo esperada (RRR) de tener malos resultados en la condición o el incremento relativo del riesgo (IRR) de tener efectos adversos con la tecnología, teniendo como base la proporción esperada de eventos en el grupo control (ver sección 4.5). Esta diferencia deberá ser fijada a priori por el grupo que desarrolla la ET con base en el conocimiento de la condición y el efecto mínimo esperado de la tecnología en la población objetivo de la tecnología. Este concepto también aplica para desenlaces cuantitativos, aunque en este escenario se tendría en cuenta la diferencia esperada absoluta, en unidades de medición, entre las alternativas de comparación. Como regla se ha sugerido que si se tiene un número total de eventos menor a 300 con una RRR del 30% se podría tener un óptimo tamaño de la muestra menor al deseable.

Por otra parte, cuando la frecuencia de los eventos es muy baja (menor al 1%), se sugiere considerar no solo el IC del 95% del estimador de efecto relativo (el RR), sino también el estimado del efecto absoluto (la RAR), ya que puede dar más información sobre la verdadera precisión del efecto estimado de la tecnología. Es importante que los eventos hayan sido seguidos por el tiempo en el que se suelen presentar los eventos

evaluados y no antes. En estas condiciones se sugiere evaluar si el tamaño de muestra incluye como mínimo 2000 e, idealmente, 4000 sujetos.

Como ejemplo tomaremos de nuevo la RS de Singh y colaboradores ¹³² en la que evaluó el efecto de los biológicos en pacientes con pobre respuesta al MTX o los medicamentos FARME. Volviendo al ejemplo de la comparación de biológicos como grupo contra placebo para el ACR 50. Se encuentra que se ingresaron en los 9 ECA un total 2204 pacientes y se tuvo un total de 455 eventos a evaluar (400 en el grupo de biológicos y 50 en el grupo de placebo). Por ende, no se considera hay imprecisión ya que cumple con los criterios para considerar hay precisión. Por otra parte, al analizar el grupo de tofacitinb contra placebo se encuentra que los 2 ECA ingresaron un total de 805 pacientes, con un total de 238 eventos y un OR de 5,62 (IC 95% 2,02–15,05) por lo que se considera hay importante imprecisión para esta comparación en este resultado (ACR 50), lo que lleva a otorgar la calificación de imprecisión seria.

Como criterio para bajar el nivel de confianza en la estimación en el desarrollo de la ET se sugiere:

Verificar si el efecto estimado supera el umbral mínimo esperado de beneficio (planteado a priori por el grupo desarrollador de la ET) para recomendar la tecnología. Si cruza el límite de recomendar y no recomendar (beneficio y no beneficio) se deberá bajar el nivel de la confianza en el efecto.

Si el estimador no cruza el umbral, verificar que se cumpla con un tamaño óptimo de la muestra, o si la tasa de eventos es baja (< 1%), verificar si se incluyeron al menos 2000 sujetos. Si se cumplen estos

criterios, se sugiere considerar que se tiene adecuada precisión. Si no se cumplen se debe bajar el nivel de confianza por imprecisión.

Cuando se hacen recomendaciones sobre el uso de la tecnología se deberá tomar en cuenta la precisión del estimador del efecto (relativo y absoluto, si hay pocos eventos) tanto en el “beneficio apreciable” como en el “daño apreciable”. Si en ambos hay imprecisión, se deberá bajar dos niveles la confianza por imprecisión (ver ejemplo: Cuadro 7.3).

7.3.5 Sesgo de publicación

El sesgo de publicación es una subestimación o sobrestimación del efecto subyacente (daño o beneficio), que ocurre de forma sistemática y se debe a la publicación selectiva de estudios¹⁴³. Es decir, algunos estudios realizados no son publicados (típicamente aquellos que no muestran efecto benéfico alguno), pues es menos probable que las revistas acepten para publicación estudios que no muestran resultados positivos. Se aconseja considerar que hay sesgos de publicación si el cuerpo de la evidencia está soportado por estudios pequeños y más aún si estos han sido patrocinados por la industria. Desde el punto de vista visual, se puede sospechar riesgo de sesgos de publicación si se observa asimetría en la distribución de los estudios en la gráfica de embudo (más conocida por su nombre en inglés *funnel plot*) o si el test estadístico de asimetría es significativo. Los autores del grupo GRADE reconocen que la calificación de este dominio tiene aún dificultades; por ejemplo, no es estimable si hay menos de 10 estudios, o la gráfica de embudo no toma en cuenta que el efecto estimado de los estudios incluidos está afectado por alto riesgo de sesgos; por lo tanto, se sugiere bajar la confianza en la estimación como máximo en un nivel por este sesgo.

Al final de la evaluación del cuerpo de la evidencia para cada desenlace, se tendrá el resultado sobre la confianza en las estimaciones, que puede ser “alta”, “moderada”, “baja” o “muy baja”¹²⁶. El Cuadro 7.2 resume el significado de cada uno de estos resultados y sus implicaciones en la toma de decisiones.

Cuadro 7.2 Calificaciones de la calidad en la evidencia según la aproximación GRADE

Calificación	Definiciones	Implicaciones
Alta	La investigación provee un muy buen indicador del efecto probable. La probabilidad de que el efecto sea sustancialmente diferente es muy baja.	La evidencia es una excelente base para la toma de decisiones sobre la implementación o no de la intervención (uso de la tecnología). Es probable que no se requiera evaluación de impacto ni monitoreo del impacto si la intervención es implementada.
Moderada	La investigación provee un buen indicador del efecto probable. La probabilidad de que el efecto sea sustancialmente diferente es moderada.	La evidencia es una buena base para la toma de decisiones sobre la implementación o no de una intervención. Es probable que se requiera monitoreo del impacto y podría justificarse una evaluación del impacto si la intervención es implementada.
Baja	La investigación provee alguna indicación del efecto probable; sin embargo, la probabilidad de que el efecto sea sustancialmente diferente es alta.	La evidencia provee alguna base para la toma de decisiones sobre la implementación o no de una intervención. Sin embargo, se justifica realizar una evaluación del impacto, si la intervención es implementada.
Muy baja	La investigación no provee una indicación confiable sobre el efecto probable. La probabilidad de que el efecto sea sustancialmente diferente es muy alta.	La evidencia no es una buena base para la toma de decisiones sobre la implementación o no de una intervención. Preferiblemente la intervención no debe ser implementada o, si lo es, se requiere realizar una evaluación de su impacto.

Cómo evaluar la evidencia sobre intervenciones a partir de estudios de comparaciones indirectas o metaanálisis en red es un aspecto que, al momento de escribir este manual, se encuentra en desarrollo; por lo tanto, solo es pertinente mencionar que ya hay publicaciones que abordan este tema en la literatura

internacional y los desarrolladores de la ET ya pueden consultar los casos que la evidencia encontrada provenga de este tipo de estudios ^{144, 145}.

A continuación, se ilustra el proceso de evaluación de la calidad de la evidencia obtenida como ejemplo

de la evaluación de la efectividad y seguridad de la vacuna del virus del papiloma humano a partir de la RS de la literatura de *Rambout y colaboradores*¹⁰⁶ que, de acuerdo con la evaluación que se realizó, tiene un puntaje de 9/11 en el instrumento AMSTAR (ver Cuadro 5.2).

Cuadro 7.3 Ejemplo de aplicación de evaluación de la calidad del cuerpo de la evidencia

Primero: Definición la pregunta clínica y en formato PICO

¿Se debe administrar la vacuna profiláctica contra el virus del papiloma humano (VPH) a las mujeres sin inicio de relaciones sexuales para reducir la incidencia de infección por VPH y lesiones cervicales precancerosas en comparación con la no vacunación?

Población: mujeres sin relaciones sexuales ni infección por VPH

Intervención: vacuna contra el VPH

Comparación: no vacunar

Desenlaces: lesiones precancerosas del cérvix o infección clínica por VPH (condiloma)

Segundo: Calificación de los resultados como críticos, importantes o no importantes desde el punto de vista de los pacientes y de los expertos en el tema:

Críticos: frecuencia de lesiones cervicales de alto grado (lesión intraepitelial escamosa de alto grado o neoplasia intraepitelial cervical de grado 2 o 3). Eventos adversos serios y muerte.

Importantes: infección persistente del VPH, lesiones cervicales de bajo grado (lesión intraepitelial escamosa de bajo grado o neoplasia intraepitelial cervical grado 1), lesiones genitales externas.

Tercero: Identificación de la evidencia para cada desenlace de interés:

Una vez cumplidos los pasos de búsqueda, apreciación y síntesis de la evidencia descritos en los capítulos anteriores se toma el cuerpo de la evidencia para cada desenlace y sobre este se hace la evaluación (habrá que evaluar la evidencia para cada uno de los definidos como críticos e importantes).

Este ejemplo comenzará por la evaluación del efecto del desenlace sobre "lesiones cervicales de alto grado".

Tomaremos como cuerpo de la evidencia para este desenlace la identificada por los autores de la RS: Rambout L, Hopkins L, Hutton B, Fergusson D. Prophylactic vaccination against human papillomavirus infection and disease in women: a systematic review of randomized controlled trials. CMAJ. 2007 ; 177(5):469-79106.

Cuarto: Definir el efecto de la intervención a partir de los estimadores de efecto relativo o efecto absoluto presentados por los autores.

Para esto definiremos el riesgo asumido (grupo de control = número de sujetos con el evento/ sujetos no expuestos a la intervención) y el riesgo correspondiente (grupo de intervención= número de sujetos con el evento/sujetos expuestos a la intervención). Los autores presentan los siguientes datos en el análisis de intención de tratar:

Riesgo en el grupo de control: $(277/18170) = 1,5\%$

Riesgo en el grupo con la vacuna: $(142/18096) = 0,08\%$

OR de Peto (IC 95%) = 0,52 (0,43 – 0,63)

Reducción absoluta del riesgo: 0,73% NNT = 135

En este mismo punto es importante definir si la población a la que se van a extrapolar los resultados tiene un riesgo diferente a la de la población incluida en los estudios. Por ejemplo, podríamos determinar que nuestra población tiene un riesgo mayor de tener VPH al momento de la vacuna por ser población adolescente sexualmente activa o de no tener la infección por pertenecer a una comunidad religiosa o con alta frecuencia de monogamia absoluta. Esta información sobre posibles mayores o menores riesgos puede obtenerse de la literatura médica o de consensos de expertos.

En este ejercicio, asumimos que la población de interés tiene los mismos riesgos que la población estudiada y, por tanto, no se calculará el efecto para diferentes poblaciones de alto o de más bajo riesgo.

Quinto: Evaluar las limitaciones del cuerpo de la evidencia para cada desenlace

Así, para el desenlace “lesiones intraepiteliales de alto grado”.

Riesgo de sesgos en los estudios: No serio (All 6 studies were of high methodologic quality (score of 5/5 on the Jadad scale 12). Adequate allocation concealment was reported in most instances. The studies reported results according to per-protocol and modified intention-to-treat assessments, and the per-protocol assessments were most common y used for the primary analysis).

Inconsistencia entre los resultados de los estudios: No serio (La figura 1 muestra consistencia a través de los resultados de los estudios para este desenlace, todos los 12 fueron = 0%)

La imprecisión de los resultados: no serio (Peto odds ratio was 0,14 [95% [CI] 0,09–0,21]). Si se toma en cuenta la regla de oro para condiciones en las cuales la frecuencia del evento es menor al 1%, como es en el desarrollo de LIE de alto grado, de tomar al menos 2.000 o, idealmente, 4.000 sujetos, el estudio cumple ya que incluyó a 36.000 sujetos.

Valoración de evidencia indirecta: los estudios han sido realizados en mujeres con vida sexual activa

Presencia de sesgo de publicación: no definido

Sexto: Determinar la calidad final del cuerpo de la evidencia para el desenlace a partir de los criterios anteriores.

De acuerdo con los criterios anteriores, la calidad de la evidencia para el efecto de la vacuna en mujeres sin relaciones sexuales, sobre la frecuencia de lesiones cervicales de alto grado de malignidad es de moderada calidad. (La investigación provee un buen indicador del efecto probable. La probabilidad de que el efecto sea sustancialmente diferente es moderada).

7.4 ¿Cómo generar una tabla de resumen de hallazgos a partir de esta evaluación?

Es pertinente que el grupo desarrollador de la ET, desde el protocolo, defina la necesidad de realizar una tabla por cada subgrupo de interés (por población o por intervención) que se incluya en la ET y qué desenlaces serán evaluados para el subgrupo en cuestión (preferiblemente no más de cinco).

Para presentar la síntesis de la evidencia encontrada para cada subgrupo y la valoración que se hace de la calidad de esta para cada desenlace, el GRADE Working Group (GWG) ofrece varias alternativas para presentar el resumen de la valoración de la calidad del cuerpo de la evidencia que soporta el informe de la ET. Sin embargo, estas se podrían resumir en dos alternativas: el formato de perfil de evidencia GRADE y en el formato de resumen de los hallazgos (SoF, por sus siglas en inglés).

7.4.1 El perfil de evidencia GRADE

Los resultados de la evaluación de la evidencia se presentan en un formato que incluye: el desenlace a considerar, el número y tipo de estudios que aportan información, los criterios de calificación antes mencionados y las razones que justifican el juicio emitido, el riesgo absoluto para el grupo de la tecnología en evaluación (riesgo correspondiente) y del grupo de comparación (riesgo asumido), el estimador de efecto relativo (con su respectivo IC 95%) y la calificación del cuerpo de la evidencia. Es el formato sugerido para el personal que revisa el informe desde el punto de vista técnico.

7.4.2 Resumen de hallazgos

Este formato contiene el resultado evaluado, el estimador de efecto relativo (con el intervalo de confianza respectivo), el efecto absoluto para el grupo de la tecnología en evaluación (riesgo correspondiente) y para el grupo de comparación (riesgo asumido) presentado sobre 1000 sujetos tratados o también puede ser presentada la diferencia de riesgo absoluta y el número necesario para tratar (NNT) o número necesario para daño (NND), el número de participantes (opcional) y la calificación de la calidad de la evidencia. Este formato está diseñado para el tomador de decisiones.

El programa para el desarrollo de las tablas se accede desde la página del GRADE pro (disponible en: <https://grade.pro.org/>).

Para las personas que quieran profundizar sobre este proceso se recomienda la lectura del capítulo 11 del Cochrane Handbook for Systematic Reviews of Interventions⁸³.

7.5 ¿Cómo juzgar la calidad del cuerpo de la evidencia sobre estudios no aleatorios que evalúan intervenciones?

Como se ha planteado anteriormente, en los casos en que el cuerpo de la evidencia para un desenlace de seguridad de una intervención provenga de estudios observacionales o no aleatorios, es necesario tener en cuenta otros aspectos en la evaluación de su calidad. En tal caso, la aproximación GRADE propone iniciar la evaluación considerando que estos estudios proveen baja calidad de evidencia (lo contrario de

lo que pasa cuando la evidencia proviene de ECA) y que la confianza en la evidencia que estos estudios representan puede aumentarse o subir si incluyen las siguientes características: a) la asociación fuerte entre la intervención y el evento, b) la magnitud del efecto es grande, c) el efecto de asociación entre la intervención y el desenlace persiste a pesar de que el factor de

confusión favorece la hipótesis nula; d) es claro el gradiente dosis-respuesta¹⁴⁶. El Cuadro 7.4 presenta los aspectos a considerar sobre cada uno de estos criterios.

Cuadro 7.4 Criterios para aumentar la calificación de calidad de la evidencia en estudios no aleatorios que evalúan intervenciones

Asociación Fuerte:

Cuando estudios observacionales metodológicamente fuertes producen estimaciones de la magnitud del efecto de un tratamiento o una exposición, y estas estimaciones son grandes o muy grandes y consistentes, podemos confiar tranquilamente en los resultados. En esas situaciones, un diseño débil del estudio no es la explicación probable de todo el daño o beneficio aparente, incluso cuando estudios observacionales pueden dar un sobrestimado del efecto verdadero. Cuanto más grande sea la magnitud del efecto, más fuerte se vuelve la evidencia.

Magnitud del efecto

RR grande >2 o <0.5 (basado en evidencia consistente de al menos 2 estudios, sin factores de confusión): Aumentar la calificación de la evidencia 1 nivel.

RR muy grande >5 o <0.2 (basado en evidencia directa sin amenazas importantes de su validez): Aumentar la calificación de la evidencia 2 niveles.

Efectos de todos los factores de confusión

Puede ocurrir que todos los factores de confusión de estudios observacionales y estudios aleatorizados trabajen para reducir el efecto demostrado, o para incrementar el efecto, si este no fue observado.

Por ejemplo, si solo los pacientes más enfermos reciben una intervención experimental o alguna exposición, y sin embargo les va bien, es probable que el efecto verdadero de la intervención o la exposición sea mayor que el que sugieren los datos.

Relación dosis-respuesta:

La presencia de un gradiente dosis-respuesta puede aumentar nuestra confianza en el hallazgo de estudios observacionales y subsecuentemente aumentar la calidad de la evidencia.

Solo se puede elevar la calificación de calidad de la evidencia de estudios que no tengan amenazas contra su validez (que no se les haya bajado su calificación de calidad por cualquier razón).

7.6 Conceptos clave

La “calidad de la evidencia” se refiere a la certeza que se puede tener sobre las estimaciones de efecto reportadas por los estudios que conforman el “cuerpo de la evidencia”.

La calidad se evalúa para cada desenlace prioritario (crítico o importante) según hayan sido clasificados por expertos temáticos.

La certeza en la evidencia encontrada para cada desenlace es mayor cuando los estudios que lo reportan están libres de sesgo, las estimaciones son precisas y basadas en un número suficiente de eventos y sujetos, los resultados son consistentes a través de los estudios y provienen de investigaciones que han evaluado en forma directa las mismas intervenciones (o tecnologías en este caso), en la misma población de interés para el ET y medido el desenlace de interés que plantea la pregunta de la ET.

Esta evaluación de evidencia debe ser reportada en los formatos dispuestos para la toma de decisiones o para conocer los criterios para la calificación según el público objetivo del informe.

BIBLIOGRAFÍA

1. Kristiansen, F. B., & Sigmund, H. Health Technology Assessment Handbook. Copenhagen: Danish Centre for Health Technology Assessment. 2007
2. Canadian Institutes of Health Research. Knowledge translation strategy 2004—2009: Innovation in action. 2004. Disponible en: http://www.cihr-irsc.gc.ca/e/documents/kt_strategy_2004-2009_e.pdf. Tomado en diciembre de 2016
3. Cochrane Public Health Group (s.f). Knowledge Translation. Disponible en: <http://ph.cochrane.org/knowledge-translation> . Tomado en diciembre de 2016
4. Jewell EJ, Abate F, eds. 2001. The New Oxford American Dictionary. New York: Oxford Univ. Press
5. Real Academia Española. Tomado en mayo de 2014: Disponible en: <http://lema.rae.es/drae/?val=evidencia> . Tomado en diciembre de 2016
6. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ. 1996 13;312(7023):71-2
7. Mulrow CD, Lohr KN. Proof and policy from medical research evidence. J Health Polit Policy Law. 2001;26(2):249-66
8. Danner M, Hummel JM, Volz F, van Manen JG, Wiegard B, Dintsios CM, et al. Integrating patients' views into health technology assessment: Analytic hierarchy process (AHP) as a method to elicit patient preferences. Int J Technol Assess Health Care. 2011;27(4):369-75
9. Van de Bovenkamp HM, Zuiderent-Jerak T. An empirical study of patient participation in guideline development: exploring the potential for articulating patient knowledge in evidence-based epistemic settings. Health Expect. 2015 ;18(5):942-55. doi: 10.1111/hex.12067
10. Utens CM, van der Weijden T, Joore MA, Dirksen CD. The use of research evidence on patient preferences in pharmaceutical coverage decisions and clinical practice guideline development: exploratory study into current state of play and potential barriers. BMC Health Serv Res. 2014; 14:540. doi: 10.1186/s12913-014-0540-2
11. Brooker AS, Carcone S, Witteman W, Krahn M. Integrating Quantitative Preference-Related Evidence into HTA: The Case of Ventilation for COPD. Toronto Health economics and Technology Assessment Collaborative. Disponible en: <https://smdm.confex.com/smdm/2011ch/webprogram/Paper6593.html>. Tomado en diciembre de 2016
12. Real Academia Española. Tomado en mayo de 2014: Disponible en: <http://dle.rae.es/?id=LXrOqrN> . Tomado en diciembre de 2016

- 13.** Oyebode O, Patrick H, Walker A, Campbell B, Powell J. The ghost in the Machine? The value of expert advice in the production of evidence-based guidance: a mixed methods study of the nice interventional procedures programme. *Int J Technol Assess Health Care*. 2016 Mar 22:1-8
- 14.** Brownson RC, Fielding JE, Maylahn CM. Evidence-based public health: a fundamental concept for public health practice. *Annual Review of Public Health*. 2009;30:175-201
- 15.** EUnetHTA. (2013). JA WP5 - Relative Effectiveness Assessment of Pharmaceuticals. In European Network for Health Technology Assessment. Disponible en: http://www.eunetha.eu/sites/5026.fedimbo.belgium.be/files/Model%20for%20Rapid%20REA%20of%20pharmaceuticals_final_20130311_reduced.pdf Tomado en diciembre de 2016
- 16.** Resolución WHA60.29 de la Asamblea Mundial de la Salud, mayo de 2007. Disponible en: http://www.who.int/medical_devices/policies/resolution_wha60_r29-sp.pdf . Tomado en diciembre de 2016
- 17.** INAHTA Health Technology Assessment (HTA) Glossary. Disponible en: <http://htaglossary.net/tecnolog%C3%ADa+sanitaria+%28n.f.%29> . Tomado en diciembre de 2016
- 18.** Fett M. Technology, Health and Health Care. Occasional Papers: Health Financing Series. Volume 5. Fett. Commonwealth Department of Health and Aged Care. 2000. ISBN 0 642 41581 1. Disponible en: [http://www.health.gov.au/internet/main/publishing.nsf/Content/DA8177ED1A80D332CA257BF0001B08EE/\\$File/ocpahfsv5.pdf](http://www.health.gov.au/internet/main/publishing.nsf/Content/DA8177ED1A80D332CA257BF0001B08EE/$File/ocpahfsv5.pdf) Tomado en diciembre de 2016
- 19.** Definition of Medical Technology. Washington lifescience. Disponible en: http://www.washingtonlifescience.com/industry/definition_medtech.htm . Tomado en diciembre de 2016
- 20.** Fletcher R Fletcher S, Wagner E Chapter 4 Frequency en Fletcher R Fletcher S, Wagner E. *Clinical Epidemiology. The essentials*. Williams & Wilkins 3er Ed. 1996
- 21.** Noorani, H,Z., Husereau, D.R., Boudreau, R. & Skidmore, B. Priority setting for health technology assessments: a systematic review of current practical approaches. *Int J Technol Assess Health Care*. 2007;23(3):310-315
- 22.** Del Llano, J. E., & Oliva, J. Cost effective medicine and evidence based medicine: impact on clinical decision making. *Med Clin (Barcelona)*, 2000; 114(3), 34-41
- 23.** Dehnavieh R, Noori Hekmat S, Ghasemi S, Mirshekari N. The vulnerable aspects of application of "Health Technology Assessment". *Int J Technol Assess Health Care*. 2015;31(3):197-8. doi: 10.1017/S0266462315000288

- 24.** Mobinizadeh M, Raeissi P, Nasiripour AA, Olyaeemanesh A, Tabibi SJ. A model for priority setting of health technology assessment: the experience of AHP-TOPSIS combination approach. *Daru*. 2016 ;24 :10. doi : 10.1186/s40199-016-0148-7.
- 25.** EUnetHTA (2013). Chapter 2 : en. EUnetHTA Joint Action 2, Work Package 8. HTA Core Model ® version 3.0 (PDF); 2016. Disponible en www.htacoremodel.info/BrowseModel.aspx. Tomado en diciembre de 2016
- 26.** Real Academia Española. (2001). Beneficio. Disponible en: <http://dle.rae.es/?id=5LctDVj> Tomado en diciembre 2016 Real Academia Española. (2001). Riesgo. Disponible en: <http://dle.rae.es/?id=WT8tAMI> Tomado en diciembre de 2016
- 27.** Real Academia Española. (2001). Riesgo. Disponible en: <http://dle.rae.es/?id=WT8tAMI> . Tomado en diciembre de 2016
- 28.** General Methods (benefit assessment).). In Institute for Quality and Efficiency in Health Care. 2013. Disponible en: <https://www.iqwig.de/en/methods/methods-paper.3020.html> . Tomado en diciembre de 2016.
- 29.** Fletcher, R., Fletcher, S. & Wagner, E. Chapter 7: Treatment. En Fletcher, R., Fletcher, S. & Wagner, E (Eds.). *Clinical Epidemiology. The essentials*. Williams & Wilkins 3er Ed. 1996
- 30.** Fletcher, R., Fletcher, S. & Wagner, E. Chapter 5: Risk. En Fletcher, R., Fletcher, S. & Wagner, E. (Eds.). *Clinical Epidemiology. The essentials*. Williams & Wilkins 3er Ed. 1996
- 31.** Saslow D, Solomon D, Lawson HW, Killackey M, Kulasingam SL, Cain J et al. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. *Am J Clin Pathol*. 2012;137(4):516-42. doi: 10.1309/AJCPTGD94EVRJCG
- 32.** Oeffinger KC, Fontham ET, Etzioni R, Herzig A, Michaelson JS, Shih YC, et al. Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. *JAMA*. 2015;314(15):1599-614. doi: 10.1001/jama.2015.12783
- 33.** Gaitán, H.G., Reveiz, L. & Farquhar C. Laparoscopy for the management of acute lower abdominal pain in women of childbearing age. *Cochrane Database Syst Rev*. 2011 May 22;(5):CD007683. doi: 10.1002/14651858.CD007683
- 34.** Straus, S., Glasziou, P., Richardson, W. & Haynes, B. Chapter 5: Diagnosis and Screening. En Straus, S., Glasziou, P., Richardson, W. & Haynes, B. (Eds.). *Evidence Based Medicine. How to practice and teach it*. Churchill Livingstone 4th ed 2011

- 35.** Dirksen CD, Utens CM, Joore MA, van Barneveld TA, Boer B, Dreesens, et al. Integrating evidence on patient preferences in healthcare policy decisions: protocol of the patient-VIP study. *Implement Sci.* 2013 ;8:64. doi: 10.1186/1748-5908-8-6
- 36.** Boivin A, Currie K, Fervers B, Gracia J, James M, Marshall C, et al Patient and public involvement in clinical guidelines: international experiences and future perspectives. *Qual Saf Health Care.* 2010;19(5):e22. doi: 10.1136/qshc.2009.034835
- 37.** Krahn M, Naglie G. The next step in guideline development: incorporating patient preferences. *JAMA.* 2008;300(4):436-8. doi: 10.1001/jama.300.4.436
- 38.** Gold MRS JA, Russel LB, Weinstein MC. In: Cost-effectiveness in health and medicine. Weinstein MC, Russel LB, Gold JA, editor. Oxford University; 1996
- 39.** Dilla T, Lizan L, Paz S, Garrido P, Avendaño C, Cruz-Hernández Do new cancer drugs offer good value for money? The perspectives of oncologists, health care policy makers, patients, and the general population. *Patient Preference Adherence.* 2015 Dec 18;10:1-7. doi: 10.2147/PPA.S93760
- 40.** Elder L, Paul R. El Arte de Formular Preguntas Esenciales. Foundation for Critical Thinking. 2002 Disponible en: [http://www.criticalthinking.org/resources/PDF/SP-AskingQuestions .pdf](http://www.criticalthinking.org/resources/PDF/SP-AskingQuestions.pdf). Tomado en diciembre de 2016
- 41.** Straus, S., Glasziou, P., Richardson, W. & Haynes, B. Chapter 1: Asking answerable clinical questions. En Straus, S., Glasziou, P., Richardson, W. & Haynes, B. (Eds.). *Evidence Based Medicine. How to practice and teach it.* Churchill Livingstone 4th ed. 2011
- 42.** Haynes R. Forming research questions. En Haynes, R., Sacket, D., Guyatt, G,H. & Tugwell, P.,(Eds.) *Clinical Epidemiology: How to do Clinical Practice Research.* Philadelphia, PA. Lippincott Williams & Wilkins. 3ed. 2006
- 43.** EUnetHTA (2013) Guideline Endpoints used for relative effectiveness assessment of Pharmaceuticals: Clinical Endpoints. Disponible en: <https://5026.fedimbo.belgium.be/outputs/methodological-guideline-rea-pharmaceuticals-clinical-endpoints> . Tomado en diciembre de 2016
- 44.** Guyatt, G. H., Oxman, A. D., Kunz, R., Atkins, D., Brozek, J., & Vist, G. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *Journal of Clinical Epidemiology.* (2011); 64(4), 395-400
- 45.** Davies, L., Drummond, M., & Papanikolaou, P. Prioritizing investments in health technology assessment. Can we assess potential value for money? *Int J Technol Assess Health Care.* 2000);16(1), 73-91

- 46.** Grupo de evaluación de tecnologías y políticas en salud (GETS) Instituto de Investigaciones Clínicas, Universidad Nacional de Colombia. "Manual metodológico para la elaboración y adaptación de guías de práctica clínica basadas en la evidencia". Bogotá, Editorial Universidad Nacional de Colombia. 2010
- 47.** Kristensen, F. B., & Sigmund, H. (Eds.). Literature searches. Chapter 3 en. Kristensen, F. B., & Sigmund, H Health Technology Assessment Handbook. Copenhagen: Danish Centre for Health Technology Assessment. 2007
- 48.** Lefebvre, C., Manheimer, E. & Glanville, J). Chapter 6: Searching for studies. En Higgins, J.P.T. & Green, S. (Eds). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. The Cochrane Collaboration. 2011. Disponible en: <http://handbook.cochrane.org/> . Tomado en diciembre de 2016
- 49.** Golder, S., & Loke, Y. K.). The performance of adverse effects search filters in MEDLINE and EMBASE. Health Info Libr Journal, 2012; 29, 141-151
- 50.** Golder S, McIntosh HM, Loke Y. Identifying systematic reviews of the adverse effects of health care interventions. BMC Med Res Methodol. 2006;6:22
- 51.** Rojas MX. Lozano JM. Capítulo 20: Revisiones sistemáticas y meta-análisis de la literatura. En: Ruiz A- Gómez C. Epidemiología Clínica: Investigación Aplicada 2ª Edición. Bogotá, Editorial Panamericana. 2015
- 52.** Muñoz, N., Kjaer, S. K., Sigurdsson, K., Iversen, O. E., Hernandez-Avila, M., & Wheeler, C. M.. Impact of human papi lomavirus (HPV)-6/11/16/18 vaccine on all HPV-associated genital diseases in young women. J Natl Cancer Inst, 2010; 102(5), 325-339
- 53.** Giordano SH, Temin S, Kirshner JJ, Chandarlapaty S, Crews JR, Davidson NE, et al Systemic therapy for patients with advanced human epidermal growth factor receptor 2-positive breast cancer: American Society of Clinical Oncology. Clinical Practice Guideline. J Clin Oncol. 2014;32(19):2078-99
- 54.** Kleimbaum, D., Kupper, L. & Morgensten, H. Chapter 6: Measures of Disease Frequency: Incidence. En Kleimbaum D, Kupper L, Morgensten H (Eds.). Epidemiological Research. Van Nostrand Reinhold 1er Ed. 1982
- 55.** Guyatt, G. H., Oxman, A. D., Santesso, N., Helfand, M., Vist, G., & Kunz, R. GRADE guidelines: 12. Preparing summary of findings tables-binary outcomes. Journal of Clinical Epidemiology. 2013;66(2), 158-172
- 56.** Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis part II: multivariate data analysis-an introduction to concepts and methods. Br J Cancer. 2003;89(3):431-6

- 57.** Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part I: basic concepts and first analyses. *Br J Cancer*. 2003;89(2):232-8
- 58.** Duszyńska W, Rosenthal VD, Szczęsny A, Woźnica E, Ulfik K, Ostrowska E, ET AL, Urinary tract infections in intensive care unit patients - a single-centre, 3-year observational study according to the INICC project. *Anaesthesiol Intensive Ther*. 2016;48(1):1-6. doi: 10.5603/AIT.2016.0001
- 59.** Kleimbaum, D., Kupper, L. & Morgensten, H. (1982). Chapter 3: Types of Epidemiological Research. En Kleimbaum D, Kupper L, Morgensten H (Eds.). *Epidemiological Research*. Van Nostrand Reinhold 1er Ed.
- 60.** Deeks, J.J., Higgins, J.P.T. & G Altman, D.G, Chapter 9: Analysing data and undertaking meta-analyses. En Higgins, J.P.T. & Green S (Eds). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. 2011. The Cochrane Collaboration. Disponible en: <http://handbook.cochrane.org/> . Tomado en diciembre de 2106
- 61.** Deeks, J. J. (2002). Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine*, 21(11), 575-600.
- 62.** Gordis, L. Capítulo 11: Estimación del riesgo ¿existe alguna asociación?. En Gordis, L. (Ed.). *Epidemiología*. Elsevier. 3ª ed. 2005
- 63.** Szklo, M. & Nieto, J. Chapter 3: Measuring associations between exposures and outcomes. En Szklo, M. & Nieto, J. (Eds.). *Epidemiology beyond the basics*. Jones & Bartlett Learning. 2007
- 64.** Baldur-Felskov, B., Dehlendorff, C., Munk, C., & Kjaer, S. K. (2014). Early impact of human papillomavirus vaccination on cervical neoplasia - nationwide follow-up of young Danish women. *J Natl Cancer Inst*, 106(3). doi:10.1093/jnci/djt460
- 65.** Zhang J, Yu KF. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA*. 1998 Nov 18;280(19):1690-1
- 66.** Gordis, L. Capítulo 12: Más sobre la estimación de riesgo: Estimación de las probabilidades de prevención. En Gordis, L. (Ed.). *Epidemiología*. Elsevier. 3ª ed. 2005
- 67.** Sacket, D., Haynes, B., Guyatt, G. & Tugwell, P. Capítulo 7: Decidir la mejor terapéutica. En Sacket, D., Haynes, B., Guyatt, G. & Tugwell, P. (Eds.) *Epidemiología clínica ciencia básica para la medicina clínica*. Ed Panamericana 2a Ed. 1991
- 68.** Straus, S., Glasziou, P., Richardson, W. & Haynes, B. Chapter 4: Therapy. En Straus, S., Glasziou, P., Richardson, W. & Haynes, B. (Eds.). *Evidence Based Medicine. How to practice and teach it*. Churchill Livingstone 4th ed. 2011
- 69.** Kleimbaum, D., Kupper, L. & Morgensten, H. (1982). Chapter 8: Measures of potential impact and summary of the measures. En Kleimbaum D, Kupper L, Morgensten H (Eds.). *Epidemiological Research*. Van Nostrand Reinhold 1er Ed.

- 70.** Medina, D. M., Valencia, A., de Velasquez, A., Huang, L. M., Prymula, R., & García-Sicilia, J. Safety and immunogenicity of the HPV-16/18 AS04-adjuvanted vaccine: a randomized, controlled trial in adolescent girls. *Journal of Adolescent Health*, 2010;46(5), 414-421
- 71.** Gardner, M. & Altman, D. Chapter 2: Estimation rather than hypothesis testing: confidence intervals rather than p values. En Gardner, M. & Altman, D. (Eds.) *Statistics with Confidence*. British Medical Journal 1st Ed. 1989
- 72.** Dobson, A. J. Calculating sample size. *Trans. Menzies Foundation*, 7, 75-79. 1984
- 73.** Higgins, J.P.T., Altman, D.G. & Sterne, J.A.C. Chapter 8: Assessing risk of bias in included studies. En Higgins, J.P.T. & Green, S. (Eds.). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. The Cochrane Collaboration. 2011. Disponible en: <http://handbook.cochrane.org/> . Tomado en diciembre de 2106
- 74.** Mulrow, C. D. The medical review article: state of the science. *Annals of Internal Medicine: Journal*. 1987;106(3): 485-488
- 75.** Greenhalgh, T. (1997). Assessing the methodological quality of published papers. *British Medical Journal*. 1997; 315(7103): 305-308
- 76.** Fletcher, R., Fletcher, S. & Wagner, E. (1996). Chapter 11: Cause. En Fletcher, R., Fletcher, S. & Wagner, E. (Eds.). *Clinical Epidemiology. The essentials*. Williams & Wilkins 3er Ed.
- 77.** Oxford Centre for Evidence-based Medicine – Levels of Evidence. (2009). In Centre for Evidence Based Medicine. Disponible en: <http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/> . Tomado en diciembre de 2106
- 78.** Guyatt, G. H., Sackett, D. L., & Cook, D. J. Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? *Evidence-Based Medicine Worki. Journal of the American Medical Association*. 1994; 271(1), 59-63
- 79.** Whitlock, E. P., Lin, J. S., Chou, R., Shekelle, P., & Robinson, K. A. (2008). Using existing systematic reviews in complex systematic reviews. *Annals of Internal Medicine: Journal*. 2008; 20(148), 776-802
- 80.** De Morton NA. The PEDro scale is a valid measure of the methodological quality of clinical trials: a demographic study. *Aust J Physiother*. 2009;55(2):129-33
- 81.** Equator network. Enhancing the QUALity and Transparency Of health Research. Disponible en: <http://www.equator-network.org/> . Tomado en diciembre de 2016

- 82.** Sordo, L., Indave, B. I., Barrio, G., Degenhardt, L., de la Fuente, L., & Bravo, M. J.. Cocaine use and risk of stroke: A systematic review. *Drug and Alcohol Dependence, Journal*. 2014;142(1). doi:10.1016/j.drugalcdep.2014.06.041
- 83.** Schünemann, H., Oxman, A., Higgins, J., Vist, G., Glasziou, P. & Guyat, G. (2011) Chapter 11: Presenting results and 'Summary of findings' tables. En Higgins, J.P.T. & Green, S. (Eds.). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. The Cochrane Collaboration. Disponible en: <http://handbook.cochrane.org/> . Tomado en diciembre de 2016
- 84.** Schulz, K. F., Chalmers, I., Hayes, R. J., & Altman, D. G. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of the American Medical Association*. 1995; 273(5), 408-412
- 85.** Jadad AR, Moore RA, Carroll D et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control. Clin. Trials* 1996; 17: 1 – 12
- 86.** Chung JH, Kang DH, Jo JK, Lee SW Assessing the quality of randomized controlled trials published in the *Journal of Korean Medical Science* from 1986 to 2011. *J Korean Med Sci*. 2012;27(9):973-80. doi: 10.3346/jkms.2012.27.9.973
- 87.** Jüni, P., Witschi, A., Bloch, R., & Egger, M. The hazards of scoring the quality of clinical trials for meta-analysis. *Journal of the American Medical Association*, 1999;282(11), 1054-1060
- 88.** Guyatt, G. H., Oxman, A. D., Vist, G., Kunz, R., Brozek, J., & Alonso-Coello, P.. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *Journal of Clinical Epidemiology*. 2011;64(4), 407-415
- 89.** Altman, G. G., & Bland, J. M. Statistics notes. Treatment allocation in controlled trials: why randomise? *British Medical Journal*, 1999;318(7192), 1209
- 90.** Viera, A. J., & Bangdiwala, S. I.. Eliminating bias in randomized controlled trials: importance of allocation concealment and masking. *Family Medicine Journal*, 2007;39(2), 132-137
- 91.** Devereaux, P. J., Manns, B. J., Ghali, W. A., Quan, H., Lacchetti, C., Montori, V. M., & Bhandari, M. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *Journal of the American Medical Association*, 2001;285(15), 2000-2003
- 92.** Probst P, Grummich K, Heger P, Zschke S, Knebel P, Ulrich A, et al Blinding in randomized controlled trials in general and abdominal surgery: protocol for a systematic review and empirical study. *Syst Rev*. 2016 ;5:48. doi: 10.1186/s13643-016-0226-4

- 93.** Berkman ND, Santaguida PL, Viswanathan M, Morton SC. The Empirical Evidence of Bias in Trials Measuring Treatment Differences. Methods Research Report. (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290-2007-10056-1.) AHRQ Publication No. 14-EHC050-EF. Rockville, MD: Agency for Healthcare Research and Quality 2014. Disponible en: https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0076773/pdf/PubMedHealth_PMH0076773.pdf . Tomado en diciembre 2016
- 94.** Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials. *Am J Public Health.* 2004 ;94(3):416-22
- 95.** Hills M, Armitage P. The two-period cross-over clinical trial. *Br J Clin Pharmacol.* 1979;8(1):7-20
- 96.** EUnetHTA Guideline. Endpoints used for relative effectiveness assessment of pharmaceuticals. Composite endpoints. February 2013 Disponible en: <https://eunethta.fedimbo.belgium.be/sites/5026.fedimbo.belgium.be/files/Composite%20endpoints.pdf> . Tomado en diciembre de 2016
- 97.** Fernandes RM, van der Lee JH, Offringa M. A systematic review of the reporting of Data Monitoring Committees' roles, interim analysis and early termination in pediatric clinical trials. *BMC Pediatr.* 2009 ;9:77. doi: 10.1186/1471-2431-9-77
- 98.** Tharmanathan P, Calvert M, Hampton J, Freemantle N. The use of interim data and Data Monitoring Committee recommendations in randomized controlled trial reports: frequency, implications and potential sources of bias. *BMC Med Res Methodol.* 2008 8:12. doi: 10.1186/1471-2288-8-12
- 99.** Oxman, A. D., Cook, D. J., & Guyatt, G. H. Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. *Journal of the American Medical Association,* 1994;272(17), 1367-1371
- 100.** Mulrow, C. D. Rationale for systematic reviews. *British Medical Journal.* 1994; 309, 597-599
- 101.** Green, S., Higgins, J.P.T., Alderson, P., Clarke, M., Mulrow, C.D. & Oxman, A.D. Chapter 1: Introduction. En Higgins, J.P.T. & Green, S. (Eds.). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0.* 2011. Disponible en: <http://handbook.cochrane.org/> . Tomado en diciembre de 2016
- 102.** Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N. D., & Hamel, C. D. (2007). Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology,* 7, 10

- 103.** Walton, D. M., Carroll, L. J., Kasch, H., Sterling, M., Verhagen, A. P., Macdermid, J.C., & Gross, A.. An Overview of Systematic Reviews on Prognostic Factors in Neck Pain: Results from the International Collaboration on Neck Pain (ICON) Project. *The Open Orthopaedics Journal*. 2013;7, 494-505
- 104.** Shojania KG, Sampson M, Ansari MT, Ji J, Garritty C, Rader T, Moher D. Updating Systematic Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2007 Sep.
- 105.** Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. *Ann Intern Med*. 2007;147(4):224-33
- 106.** Rambout L, Hopkins L, Hutton B, Fergusson D. Prophylactic vaccination against human papillomavirus infection and disease in women: a systematic review of randomized controlled trials. *CMAJ*. 2007 ;177(5):469-79
- 107.** Friedman ML, Furberg CurtD, deMets DL. *Fundamentals of clinical trials 3rd Ed* Springer, New York Inc 1998. ISBN: 0-387-98586-7
- 108.** Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. . Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;7 (27)
- 109.** Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ*. 1996;312:1215–18
- 110.** Mackay, F. J. Post-marketing studies: the work of the drug safety research unit. *Drug Safety*. 1998;19, 343-359
- 111.** Meyboom, R. B., Lindquist, M., & Egberts, A. G.). An ABC of drug-related problems. *Drug Safety*. 2000; 22, 415-423
- 112.** EUnerHTA Endpoints used in relative effectiveness assessment of pharmaceuticals. *Safety* 2013 tomado en junio de 2016. Disponible en: <http://www.eunetha.eu/outputs/methodological-guideline-rea-pharmaceuticals-safety> . Tomado en diciembre de 2016
- 113.** Marciante, K. D., Durda, J. P., Heckbert, S. R., Lumley, T., Rice, K., McKnight, B., & Psaty, B. M. Cerivastatin, genetic variants, and the risk of rhabdomyolysis. *Pharmacogenetics and Genomics*. 2011;21(5), 280-288
- 114.** Nordqvist, S. R., Fidler, W. J., Woodruff, J. M., & Lewis, J. L. Clear cell adenocarcinoma of the cervix and vagina. A clinicopathologic study of 21 cases with and without a history of maternal ingestion of estrogens. *American Cancer Society*. 1976;37(2), 858-871
- 115.** Loke, Y.K., Price. D. & Herxheimer, A. Cochrane Adverse Effects Methods Group. En Higgins, J.P.T. & Green, S. (Eds.). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. 2011. Disponible en: <http://handbook.cochrane.org/> . Tomado en diciembre de 2016
- 116.** Reeves, B.C., Deeks, J.J., Higgins, J.P.T. & Wells, G.A. Chapter 13 Including non-randomized

studies. En Higgins, J.P.T. & Green, S. (Eds.).
Cochrane Handbook for Systematic Reviews of
Interventions Version 5.1.0. 2011. Disponible en:
<http://handbook.cochrane.org/> . Tomado en
diciembre de 2016

117. Wells GA, Shea B, O'Connell D, Peterson J,
Welch V, Losos M, et al The Newcastle-
Ottawa Scale (NOS) for assessing the quality
of nonrandomised studies in meta-analyses.
Disponible en: [http://www.ohri.ca/programs/
clinical_epidemiology/oxford.asp](http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp) . Tomado el 30
de agosto de 2104

118. Reisch, J., Tyson, J. E., & Mize, S. G. Aid to the
evaluation of therapeutic studies. *Pediatrics*.
1989;84, 815-827

119. 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 12;355:i4919.
doi: 10.1136/bmj.i4919

120. Savović J, Weeks L, Sterne JA, Turner L, Altman
DG, Moher y colaboradores Evaluation of the
Cochrane Collaboration's tool for assessing the
risk of bias in randomized trials: focus groups,
online survey, proposed recommendations and
their implementation. *Syst Rev*. 2014;3:37. doi:
10.1186/2046-4053-3-37

121. Armijo-Olivo S, Ospina M, da Costa BR, Egger
M, Saltaji H, Fuentes J, cols. Poor reliability
between Cochrane reviewers and blinded
external reviewers when applying the Cochrane
risk of bias tool in physical therapy trials. *PLoS*

One. 2014;9(5):e96920. doi: 10.1371/journal.
pone.0096920

122. Schwarz, T. F., Huang, L. M., Medina, D. M.,
Valencia, A., Lin, T. Y., & Behre, U. Four-year
follow-up of the immunogenicity and safety of
the HPV-16/18 AS04-adjuvanted vaccine when
administered to adolescent girls aged 10-14 years.
Journal of Adolescent Health. 2012. 50(2), 187-
194

123. Velasco-Garrido, M., & Busse, R. Health
technology assessment: An introduction to
objectives, role of evidence, and structure in
Europe. In *European Observatory on Health
Systems and Policies*. . 2005. Disponible en:
[http://www.euro.who.int/
__data/assets/pdf_
file/0018/90432/E87866.pdf?ua=1](http://www.euro.who.int/__data/assets/pdf_file/0018/90432/E87866.pdf?ua=1) . Tomado
diciembre de 2016

124. NICE. *Methods for the development of NICE
public health guidance*. 2. London: National
Institute for Health and Clinical Excellence; 2012
3rd Ed. Disponible en: [https://www.nice.org.uk/
process/pmg4/chapter/introduction](https://www.nice.org.uk/process/pmg4/chapter/introduction). Tomado en
diciembre de 2106

125. Briss, P. A., Pappaioanou, M., Fielding, J.,
Wright-De Agüero, L., Truman, B. I., Hopkins,
D. P., & Harris, J. R. Developing an evidence-
based Guide to Community Preventive Services--
methods. The Task Force on Community Preventive
Services. *American Journal of Preventive Medicine*.
2000;18(1s), 35-43

- 126.** GRADE Working Group [Atkins D et al.] Grading quality of evidence and strength of recommendations. *British Medical Journal*. 2004; 328:1490–1494
- 127.** Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J. et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol*. 2011;64(4):401–406. doi: 10.1016/j.jclinepi.2010.07.015
- 128.** GRADE. Website. Disponible en: <http://www.gradeworkinggroup.org/> . Tomado en junio de 2016
- 129.** Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias).*J Clin Epidemiol*. 2011;64(4):407-15
- 130.** Moja L, Tagliabue L, Balduzzi S, Parmelli E, Pistotti V, Guarneri V, D'Amico R. Trastuzumab containing regimens for early breast cancer. *Cochrane Database of Systematic Reviews* 2012, Issue 4. Art. No.: CD006243. DOI: 10.1002/14651858.CD006243.pub2
- 131.** Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, GRADE guidelines: 7. Rating the quality of evidence--inconsistency. *J Clin Epidemiol*. 2011;64(12):1294-302. doi: 10.1016/j.jclinepi.2011.03.017
- 132.** Singh JA, Hossain A, Tanjong Ghogomu E, Mudano AS, Tugwell P, Wells GA. Biologic or tofacitinib monotherapy for rheumatoid arthritis in people with traditional disease-modifying anti-rheumatic drug (DMARD) failure: a Cochrane Systematic Review and network meta-analysis (NMA).*Cochrane Database Syst Rev*. 2016 Nov 17;11:CD012437
- 133.** Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JP. Evaluating the quality of evidence from a network meta-analysis. *PLoS One*. 2014;9(7):e99682. doi: 10.1371/journal.pone.0099682
- 134.** Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence--indirectness. *J Clin Epidemiol*. 2011;64(12):1303-10. doi: 10.1016/j.jclinepi.2011.04.014
- 135.** WHO. Weekly Epidemiological Record, no. 43, 24 October 2014. Disponible en: <http://www.who.int/wer/2014/wer8943.pdf?ua=1> Tomado en diciembre de 2016
- 136.** FUTURE II Study Group. Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions. *N Engl J Med*. 2007;356(19):1915-27
- 137.** Lehtinen M, Paavonen J, Wheeler CM, Jaisamrarn U, Garland SM, Castellsagué X, et al. Overall efficacy of HPV-16/18 AS04-adjuvanted vaccine against grade 3 or greater cervical intraepithelial neoplasia: 4-year end-of-study analysis of the randomised, double-blind PATRICIA trial. *Lancet Oncol*. 2012;13(1):89-99. doi: 10.1016/S1470-2045(11)70286-8

- 138.** Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R, et al. Indirect comparisons of competing interventions. *Health Technol Assess* 2005;9(26)
- 139.** Hoaglin DC, Hawkins N, Jansen JP, Scott DA, Itzler R, Cappelleri JC Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. *Value Health*. 2011;14(4):429-37. doi: 10.1016/j.jval.2011.01.011
- 140.** Schmitt J, Rosumeck S, Thomaschewski G, Sporbeck B, Haufe E, Nast A. Efficacy and safety of systemic treatments for moderate-to-severe psoriasis: meta-analysis of randomized controlled trials. *Br J Dermatol*. 2014;170(2):274-303. doi: 10.1111/bjd.12663
- 141.** Da Silva EM, Strufaldi MW, Andriolo RB, Silva LA. Enzyme replacement therapy with idursulfase for mucopolysaccharidosis type II (Hunter syndrome). *Cochrane Database Syst Rev*. 2016 Feb 5;2:CD008185. doi: 10.1002/14651858.CD008185.pub4
- 142.** Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence--imprecision. *J Clin Epidemiol*. 2011;64(12):1283-93. doi: 10.1016/j.jclinepi.2011.01.012
- 143.** Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence--publication bias. *J Clin Epidemiol*. 2011;64(12):1277-82. doi: 10.1016/j.jclinepi.2011.01.011
- 144.** Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JP. Evaluating the quality of evidence from a network meta-analysis. *PLoS One*. 2014 Jul 3;9(7):e99682. doi: 10.1371/journal.pone.0099682
- 145.** Puhan MA, Schünemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh JA, A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ*. 2014;349:g5630. doi: 10.1136/bmj.g5630
- 146.** Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol*. 2011;64(12):1311-6. doi: 10.1016/j.jclinepi.2011.06.004

ANEXOS

Anexo 1. Instrumento de evaluación del riesgo de sesgo de los ECA

Anexo 2. Instrumentos AMSTAR

Anexo 3. Actualización de la búsqueda de ECA para una revisión sistemática

Anexo 4. Escala de Newcastle-Ottawa para evaluar la calidad de los estudios de cohorte

Anexo 5. Escala de Newcastle-Ottawa para evaluar la calidad de los estudios de casos y controles

Anexo 1.

Instrumento de evaluación del riesgo de sesgo de los ECA



BÚSQUEDA, EVALUACIÓN Y SÍNTESIS DE LA EVIDENCIA QUE SOPORTA LAS EVALUACIONES DE TECNOLOGÍA. MANUAL METODOLÓGICO

Anexo 1. Instrumento de evaluación del riesgo de sesgo en ECA

Dominio	Alto riesgo	Bajo riesgo	No claro	Autor/año publicación			Autor/año publicación		
				Clasificación del estudio			Clasificación del estudio		
				alto	bajo	no claro	alto	bajo	no claro
1. Generación de la secuencia aleatoria (búsqueda de posibles sesgos de selección)	Métodos de forma alternativa basados en la fecha de cumpleaños, el número del caso, etc.	Estudios que tengan métodos como lanzar una moneda, lista de números aleatorios o generados por computador.	Afirmación de uso de técnicas de asignación aleatoria o diseño aleatorizado insuficiente para saber la técnica utilizada que dificulten su clasificación en "alto riesgo" o "bajo riesgo" de sesgos.						
2. Asignación y ocultamiento (búsqueda de posibles sesgos de selección)	Asignación basada en el día de admisión o el número del caso. Igualmente, los métodos que permiten conocer el grupo al cual son asignadas las intervenciones previo a su implementación (ej., sobres visibles, alternación o rotación o números pares o impares).	Implementación estricta de una secuencia aleatoria sin conocimiento previo de las asignaciones de la intervención (ej., sobres sellados no traslúcidos, frascos de medicamentos de apariencia idéntica).	Información insuficiente acerca de la asignación aleatoria para permitir el juzgamiento de "bajo riesgo" o "alto riesgo", como cuando no se describe el método o está descrito pero de forma no detallada.						
3.1 Enmascaramiento de los participantes y el personal (búsqueda de posibles sesgos de desempeño)	Cegamiento ausente o incompleto, y el desenlace tiene un alta probabilidad de estar influenciado por la falta de cegamiento del personal y los participantes. Cegamiento de solo algunas personas involucradas en el ensayo o se intentó el cegamiento pero no fue exitoso y hay alto riesgo que de los desenlaces sean influenciados por el conocimiento de la intervención	Cegamiento incompleto o ausente, pero los autores de la revisión consideran que el desenlace es bajamente influenciado por la falta del mismo. Cegamiento del personal y participantes involucrados en el estudio y es poco probable que haya sido descubierto.	Información insuficiente acerca del enmascaramiento de los participantes y el personal para permitir el juzgamiento de "bajo riesgo" o "alto riesgo", como cuando no se describe el método o está descrito pero de forma no detallada.						

<p>3.2 Cegamiento de la evaluación del desenlace (búsqueda de posibles sesgos de selección)</p>	<p>Cegamiento ausente o incompleto, y el desenlace tiene una alta probabilidad de estar influenciado por la falta de cegamiento de los evaluadores del desenlace. Cegamiento de solo algunas personas involucradas en la evaluación del desenlace en el ensayo o se intentó el cegamiento pero no fue exitoso y hay alto riesgo que de los desenlaces sean influenciados por el conocimiento de la intervención.</p>	<p>Cegamiento incompleto o ausente, pero los autores de la revisión consideran que el desenlace es bajamente influenciado por la falta del mismo. Cegamiento del personal involucrado en el estudio (evaluadores, investigadores) y es poco probable que haya sido descubierto.</p>	<p>Información insuficiente acerca del cegamiento de la evaluación para permitir el juzgamiento de "bajo riesgo" o "alto riesgo", como cuando no se describe el método o está descrito pero de forma no detallada.</p>						
<p>3.3 Cegamiento de la evaluación del desenlace (búsqueda de sesgos de detección)</p>	<p>Cegamiento ausente o incompleto, y el desenlace tiene una alta probabilidad de estar influenciado por la falta de cegamiento. Cegamiento de algunas personas involucradas en el ensayo o se intentó el cegamiento pero no fue exitoso y hay alto riesgo que de los desenlaces sean influenciados por el conocimiento de la intervención.</p>	<p>Cegamiento incompleto o ausente, pero los autores de la revisión consideran que el desenlace es bajamente influenciado por la falta del mismo. Cegamiento del personal involucrado en el estudio (evaluadores, investigadores) y es poco probable que haya sido descubierto.</p>	<p>Información insuficiente acerca de la evaluación del desenlace para permitir el juzgamiento de "bajo riesgo" o "alto riesgo", como cuando no se describe el método o está descrito pero de forma no detallada.</p>						
<p>4. Datos de desenlace incompletos (búsqueda de los posibles sesgos de desempeño debido a la cantidad, naturaleza y manejo de los datos de desenlace incompletos)</p>	<p>a. Es probable que la causa de la pérdida de datos de algunos resultados se encuentren relacionadas con el desenlace final. Para los datos de desenlaces dicotómicos, la proporción de desenlaces faltantes en comparación con las tasas de eventos observadas no es suficiente para tener un impacto clínico relevante en el estimador del efecto de la intervención; b. Para los datos continuos, el tamaño del efecto entre los desenlaces perdidos induce sesgos clínicamente relevantes en el tamaño del efecto observado; c. El análisis por protocolo ha sido realizado con una desviación grande de lo que inicialmente se había elegido en la aleatorización de los tratamientos; d. Imputación de datos potencialmente inapropiada.</p>	<p>a. No se presentaron datos faltantes; b. Es poco probable que la causa de los datos faltantes en los desenlaces se encuentre relacionada con el verdadero desenlace (datos de supervivencia, etc.); c. Los datos perdidos se encuentran balanceados en números entre los grupos de intervención y control, y las razones para su pérdida son similares; d. Para los datos de desenlaces dicotómico, la proporción de desenlaces faltantes en comparación con las tasas de eventos observadas no es suficiente para tener un impacto clínico relevante en el estimador del efecto de la intervención; e. Para los datos continuos, el tamaño del efecto entre los desenlaces perdidos no induce sesgos clínicamente relevantes en el tamaño del efecto observado; f. Los datos perdidos han sido imputados utilizando métodos apropiados.</p>	<p>Información insuficiente acerca de las pérdidas/exclusiones para permitir el juzgamiento de "bajo riesgo" o "alto riesgo", como cuando no se menciona el número de aleatorizados, o no hay razones que expliquen los datos perdidos.</p>						
<p>5. Reporte selectivo (búsqueda de sesgos de reporte)</p>	<p>a. No fueron reportados todos los desenlaces primarios preespecificados; b. Uno o más de los desenlaces de interés son reportados de forma incompleta; c. El estudio no reporta un desenlace clave para la condición analizada.</p>	<p>Está claro que el reporte publicado incluye todos los desenlaces esperados e inicialmente contemplados en la metodología.</p>	<p>Información insuficiente acerca del reporte selectivo como para clasificar en "alto riesgo" o "bajo riesgo" de sesgos.</p>						
<p>6. Otros sesgos (búsqueda de posibles sesgos no cubiertos en las otras secciones)</p>	<p>El estudio tiene sesgos relacionados con el diseño específico de investigación utilizado.</p>	<p>El estudio parece estar libre de otras fuentes de sesgo.</p>	<p>No hay suficiente información para evaluar si existen otras fuentes de sesgos o hay insuficiente evidencia de que un problema identificado pueda llegar a introducir sesgo.</p>						

Anexo 2.

Instrumentos AMSTAR



Anexo 2: Instrumento AMSTAR para evaluación de revisiones sistemáticas en evaluación de tecnologías.

AUTOR-AÑO

REFERENCIA

EVALUADOR (INICIALES)

	Si	No	No es claro
<p>1. ¿Fue un diseño “a priori”? Criterio: la pregunta de investigación y los criterios de inclusión deben ser establecidos antes de iniciar la revisión. Criterio: Es necesario que haga referencia a que un protocolo, la aprobación ética o los objetivos de investigación fueron publicados previamente para marcar "sí".</p>			
<p>2. ¿Hubo duplicación de la selección de los estudios y de la extracción de datos? Criterio: deben existir al menos dos extractores independientes de los datos y un procedimiento para llegar a consenso en caso de que existan desacuerdos.</p>			
<p>3. ¿Se realizó una búsqueda amplia de la literatura? Criterio: la búsqueda debe realizarse en al menos dos fuentes electrónicas. El informe debe señalar los años que abarcó la búsqueda y las bases de datos utilizadas (ej., EMBASE, MEDLINE, Lilacs). Las palabras clave y términos MeSH deben estar explicitados y cuando sea posible, debe estar disponible la estrategia de búsqueda. Las búsquedas deben ser complementadas con consulta a registros especializados o expertos en el campo de estudio, y por la revisión de las listas de referencias en los estudios encontrados.</p>			
<p>4. ¿Se utilizó el estado de la publicación como criterio de inclusión (ejemplo, literatura gris)? Criterio: los autores deberían declarar que ellos buscaron trabajos sin fijarse en el estado de publicación, el idioma del artículo, etc.</p>			
<p>5. ¿Se provee una lista de los estudios incluidos y excluidos? Criterio: debe entregarse una lista de los estudios incluidos y excluidos.</p>			
<p>6. ¿Se entregan las características de los estudios? Criterio: debe incluirse en forma completa, en tablas, la información esencial de los estudios originales, tal como el tipo de participantes, las intervenciones y los desenlaces evaluados en cada estudio.</p>			
<p>7. ¿Se evaluó y documentó la calidad científica de los estudios? Criterio: se deben describir los métodos de evaluación a priori.</p>			
<p>8. ¿Se utilizó adecuadamente la calidad de los estudios en la formulación de conclusiones? Criterio: los resultados del rigor metodológico y calidad científica deberían considerarse en el análisis y las conclusiones de la revisión, y declararse explícitamente en la formulación de recomendaciones.</p>			
<p>9. ¿Fueron apropiados los métodos para combinar los hallazgos de los estudios? Criterio: para los resultados agrupados, se debe realizar un test para evaluar su homogeneidad (ejemplo, prueba de Chi cuadrado para homogeneidad, I²), para asegurar que los estudios son combinables. La heterogeneidad en los resultados de los estudios primarios puede explicarse por distintas razones: - Por efecto del azar (para ello se utilizan los “test de homogeneidad”, que evalúan si las diferencias en los efectos observados entre los distintos estudios difieren mas allá de lo que uno pudiera esperar simplemente por azar) - Por diferencias en los diseños de investigación, o - Por variaciones en los componentes básicos del estudio: población (ejemplo, poblaciones de edades distintas), intervención (ejemplo, el fármaco se usó por un tiempo o a una dosis distinta entre un estudio y otro), o en la forma de medir los resultados. La revisión debería analizar cada uno de estos factores. Si hay heterogeneidad, debe utilizarse un modelo de efecto aleatorio y/o debe considerarse la pertinencia clínica de combinarlos (ejemplo, ¿es prudente (razonable) combinar? O sea, ¿no son peras y manzanas?).</p>			
<p>10. ¿Se evaluó la probabilidad de sesgo de publicación? Criterio: una evaluación del sesgo de publicación debe incluir una combinación de apoyos gráficos (tales como “funnel plot”) y/o pruebas estadísticas (ejemplo, test de regresión de Egger). Si bien esto es lo óptimo, sobre todo si uno sospecha que podría haber un cúmulo importante de estudios no publicados en el tema, a los que no se pudo acceder con la búsqueda; en la práctica son pocas las revisiones que lo evalúan formalmente.</p>			

<p>11. ¿Fueron declarados los conflictos de interés?</p> <p>Criterio: las fuentes de ayuda económica deben estar claramente reconocidas, tanto en la revisión sistemática como en los estudios incluidos en ella.</p> <p>Es importante verificar también que las conclusiones de los autores deriven directamente de los resultados de la revisión, y que no se planteen como ciertas asociaciones que no hayan sido debidamente demostradas, ni que, existiendo suficiente evidencia como para sostenerlas, la magnitud de las mismas sea distorsionada (exagerada) por un uso abusivo o poco preciso del lenguaje.</p>			
<p>PUNTAJE</p>			
<p>Tomado de : Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N. D., & Hamel, C. D. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Medical Research Methodology. 2007; Feb 15 7, 10</p>			

Anexo 3.

Actualización de la búsqueda de ECA para una revisión sistemática



ANEXO 3. Actualización de la búsqueda de ECA para una revisión sistemática

Miguel Díaz, MSc

Hernando G. Gaitán Duarte, MD, MSc

La literatura y experiencia con la actualización de la búsqueda de ECA para una RSL muestran que este proceso demanda importantes recursos y tiempo (Shojania 2007). En la mayoría de casos es imposible reproducir las estrategias de búsqueda de la RSL original, bien sea por un reporte incompleto de las estrategias o por no tener acceso a las bases de datos para llevar a cabo la actualización. Por ejemplo, muchas RSL Cochrane emplean EMBASE mediante la plataforma Ovid. En Colombia ninguna universidad posee este recurso. Por otra parte, se ha demostrado que la Colaboración Cochrane, con toda su infraestructura, actualiza solo el 20% de sus RSL cada dos años (Takwoingi 2013), y solo el 9% han derivado en un cambio en las conclusiones de la revisión (French 2005). Sin embargo, esto es motivo de controversia ya que también se ha mencionado que, a los dos años de publicación, hasta un 23% están desactualizadas por no haber incorporado nueva evidencia que cambia de manera sustancial las conclusiones sobre la efectividad y seguridad de las terapias (Shojania 2007).

A pesar de que se ha propuesto, iniciar con la búsqueda de RSL publicadas en los últimos 5 años, hoy en día, a la luz de la disponibilidad de las nuevas tecnologías de la información, la cantidad de investigación que se realiza en el desarrollo de nuevos medicamentos y, dadas las limitaciones metodológicas de los métodos tradicionales de actualización de las RS, hay autores que proponen llevar a cabo revisiones sistemáticas vivas (Living SR) para poder incorporar las más reciente evidencia de manera rápida a las RS (Elliot 2014).

De esta manera, la propuesta para este modelo es hacer una “actualización rápida”. Sugerimos actualizar las RS si tienen más de tres años de publicación en un intento por conciliar las dos posiciones antes descritas. La búsqueda de nuevos ECA se hará mediante la combinación de las siguientes técnicas:

Búsqueda, evaluación y síntesis de la evidencia de efectividad y seguridad en evaluaciones de tecnología sanitarias. Manual metodológico.

- » A partir de los tres ECA de más reciente publicación y con mayor tamaño de muestra en la RSL identificada, buscar nuevos ECA mediante la herramienta “Related citations in PubMed”. Existen estudios que dan cuenta de la eficiencia de esta estrategia (Tsertsvadze 2011 y Shojania 2007).
- » Búsqueda en CENTRAL y WHO International Clinical Trials Registry Platform ICTRP portal, limitada entre la última fecha de búsqueda reportada por la RSL y la fecha actual. Dado que Wiley es de libre acceso e integra CENTRAL y la base de datos de revisiones sistemáticas Cochrane (CDSR), la búsqueda en CENTRAL se realizaría mediante esta plataforma, empleando la misma estrategia de búsqueda diseñada para identificar RSL en CDSR (esto ahorraría tiempo considerable).
- » Consultar a los expertos temáticos sobre la existencia de nuevos ECA, adicionales a los incluidos en la RSL.
- » Consultar la opinión de los expertos temáticos sobre si las conclusiones de la RSL se mantienen válidas en el tiempo actual, y si ellos están o no al tanto de cualquier nuevo ECA que pueda cambiar las conclusiones de la revisión (Shekelle y colegas).
- » Búsqueda de alertas sobre la seguridad de las tecnologías en la página web de la FDA.

En caso de tratarse de una RSL con comparaciones indirectas (ITC o MTC) se propone una búsqueda en CENTRAL y WHO International Clinical Trials Registry Platform ICTRP portal, limitada entre la última fecha de búsqueda reportada por la RSL y la fecha actual, con el objetivo de descartar la existencia de ECA “cabeza a cabeza” para las comparaciones de interés.

La elegibilidad de los nuevos ECA se determinaría con base en los criterios de inclusión y exclusión de la RSL que es objeto de la actualización.

En caso de no identificar una RSL publicada en el periodo de los últimos 5 años, se propone ampliar la búsqueda a los últimos 10 años. Las RSL identificadas en este periodo requerirían una actualización completa de las búsquedas; esto es, una búsqueda en MEDLINE, EMBASE, CENTRAL, WHO International Clinical Trials Registry Platform ICTRP portal y al menos otro método de búsqueda complementario.

Referencias bibliográficas

- French SD, McDonald S, McKenzie JE, Green SE. Investing in updating: how do conclusions change when Cochrane systematic reviews are updated? *BMC medical research methodology*. 2005;5:33.
- Shekelle PG, Motala A, Johnsen B, Newberry SJ. Assessment of a method to detect signals for updating systematic reviews. *Systematic reviews*. 2014;3:13.
- Shojania KG, Sampson M, Ansari MT, et al. *Updating Systematic Reviews*. Rockville (MD): Agency for Healthcare Research and Quality (US); 2007 Sep. (Technical Reviews, No. 16.) Available from: <http://www.ncbi.nlm.nih.gov/books/NBK44099/>
- Takwoingi Y, Hopewell S, Tovey D, Sutton AJ. A multicomponent decision tool for prioritising the updating of systematic reviews. *BMJ (Clinical research ed)*. 2013;347:f7191.
- Tsertsvadze A, Maglione M, Chou R, Garritty C, Coleman C, Lux L, et al. Updating comparative effectiveness reviews: current efforts in AHRQ's Effective Health Care Program. *Journal of clinical epidemiology*. 2011 Nov;64(11):1208-15.
- Elliott JH, Turner T, Clavisi O, Thomas J, Higgins JP, Mavergames C, Gruen RL. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. *PLoS Med*. 2014;11(2):e1001603. doi: 10.1371

Anexo 4.

Escala de Newcastle-Ottawa para evaluar la calidad de los estudios de cohorte



Anexo 4. Escala de Newcastle-Ottawa para evaluar la calidad de los estudios de cohorte

Nota: Un estudio puede ser calificado como máximo con una estrella por cada ítem en las categorías de selección y resultados, y un máximo de dos estrellas para la categoría de comparabilidad.

Selección

1) Representatividad de la cohorte expuesta

- a) Verdaderamente representativa del promedio _____ (describir) en la comunidad*
- b) Algo representativa del promedio _____ en la comunidad*
- c) Grupo seleccionado de usuarios (ej., enfermeras, voluntarios)
- d) No hay descripción de la representatividad de la cohorte

2) Selección de la cohorte de no expuestos

- a) Extraída de la misma comunidad que la cohorte expuesta*
- b) Extraída de un origen diferente
- c) No hay descripción de dónde se derivó la cohorte no expuesta

3) Evaluación de la exposición

- a) Registros seguros (ej., registros quirúrgicos)*
- b) Entrevista estructurada*
- c) Autorreporte escrito
- d) No hay descripción

4) Demonstración de que el desenlace de interés no estaba presente al inicio del estudio

- a) Sí*
- b) No

Comparabilidad

1) Comparabilidad de la cohorte con base en el diseño o el análisis

- a) El estudio controla por _____ (colocar el factor más importante a controlar)*
- b) El estudio controla por cualquier otro factor adicional* (Estos criterios podrían ser modificados para indicar el control específico para un segundo factor importante)

Resultado

1) Evaluación del resultado

- a) Evaluación independiente enmascarada*
- b) Revisión de registros*
- c) Autorreporte
- d) No hay descripción

2) ¿Hubo un seguimiento lo suficiente largo para que los resultados ocurrieran?

- a) Sí (selección de un periodo de seguimiento adecuado para el resultado de interés)*
- b) No

3) Suficiencia en el seguimiento de las cohortes

- a) Seguimiento completo. Todos los sujetos fueron tomados en cuenta*
- b) Es improbable que las pérdidas hayan podido introducir sesgos. Bajo número de pérdidas ____ % en el seguimiento (seleccione un ____ % aceptable de pérdidas) o se provee una descripción de las pérdidas
- c) Tasa de seguimiento < ____% (seleccione un ____ % mínimo aceptable de seguimiento) o no hay descripción de las pérdidas
- d) No se mencionan

Tomado de: Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Disponible en: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. Tomado el 30 de junio de 2016

Anexo 5.

Escala de Newcastle-Ottawa para evaluar la calidad de los estudios de casos y controles



Anexo 5. Escala de Newcastle-Ottawa para evaluar la calidad de los estudios de casos y controles

Nota: Un estudio puede ser calificado como máximo con una estrella por cada ítem en las categorías de selección y exposición, y un máximo de dos estrellas para la categoría de comparabilidad.

Selección

- 1) ¿Es la definición de caso adecuada?
 - a) Sí, con una validación independiente*
 - b) Sí, con vinculación a registros o basada en autorreportes
 - c) No hay descripción
- 2) Representatividad de los casos
 - a) Consecutivos o serie de casos obviamente representativa*
 - b) Potencial sesgo de selección o no se describe
- 3) Selección de controles
 - a) Controles de la comunidad*
 - b) Controles hospitalarios
 - c) No hay descripción
- 4) Definición de los controles
 - a) No hay historia de enfermedad (resultado final)*
 - b) No hay descripción de la fuente

Comparabilidad

- 1) Comparabilidad de los casos y los controles con base en el diseño y el análisis
 - a) El estudio controla por _____ (Seleccione el factor más importante a controlar.)*
 - b) El estudio controla por cualquier otro factor adicional* (Estos criterios podrían ser modificados para indicar el control específico por un segundo factor importante.)

Exposición

- 1) Evaluación de la exposición
 - a) Registro seguro (ej., registros quirúrgicos)*
 - b) Entrevista estructurada enmascarada de estatus de casos /control*
 - c) Entrevista no enmascarada de estatus de casos /control
 - d) Autorreporte escrito o historia clínica únicamente
 - e) No hay descripción
- 2) Mismo método de búsqueda de la exposición en los casos y los controles
 - a) Sí*
 - b) No
- 3) Tasa de no respondedores
 - a) La misma tasa para ambos grupos*
 - b) No se describen no respondedores
 - c) Diferente tasa y no se describe a dónde pertenecen

Tomado de: Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Disponible en: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. Tomado el 30 de junio de 2016

