

MANUAL METODOLÓGICO

EVALUACIÓN DE TECNOLOGÍAS EN SALUD

Manual metodológico para la elaboración de evaluaciones de efectividad clínica, seguridad y validez diagnóstica de tecnologías en salud



Instituto de Evaluación
Tecnológica en Salud®

Evidencia que promueve Confianza



Instituto de Evaluación
Tecnológica en Salud®

Evidencia que promueve Confianza

Manual metodológico para la
elaboración de evaluaciones de
efectividad clínica, seguridad y
validez diagnóstica de tecnologías en
salud

Febrero de 2022

Tabla de contenido

| | |
|--|-----------|
| Listado de siglas y abreviaturas..... | 9 |
| Glosario..... | 11 |
| Capítulo 1. Fase de preparación y conformación del grupo desarrollador | 31 |
| 1. Conformación del grupo desarrollador | 31 |
| 1.1 Funciones del grupo desarrollador..... | 32 |
| 1.2 Composición del grupo desarrollador | 32 |
| 1.3 Aspectos éticos y conflictos de interés | 35 |
| 2. Alcance de la evaluación..... | 35 |
| 3. Preguntas de evaluación | 36 |
| 3.1 Formulación de las preguntas de investigación | 36 |
| 3.2 Refinamiento de las preguntas de evaluación..... | 40 |
| 4. Selección del comparador..... | 40 |
| 4.1 Definición..... | 40 |
| 4.2 Características de un comparador..... | 40 |
| 4.3 Selección de comparadores | 41 |
| 4.4 Procedimiento para la selección de comparadores | 41 |
| 4.5 Comparadores por subpoblaciones de pacientes | 43 |
| 5. Selección y calificación de la importancia relativa de los desenlaces en salud... 43 | 43 |
| 5.1 Definición..... | 43 |
| 5.2 Criterios para la selección de desenlaces | 44 |
| 5.3 Gradación de los desenlaces | 47 |
| 5.4 Consideraciones adicionales: | 48 |
| 6. Protocolo de la revisión..... | 50 |
| Capítulo 2. La condición de salud y uso actual de la tecnología..... | 51 |
| Capítulo 3. La tecnología y sus características..... | 52 |
| Capítulo 4. Análisis de efectividad clínica y seguridad | 54 |
| 1. Definición de los criterios de elegibilidad de la evidencia | 55 |
| 1.1. Criterios de inclusión | 56 |
| 1.2. Criterios de exclusión..... | 58 |
| 2. Búsqueda de estudios..... | 59 |
| 2.1 Búsqueda en bases electrónicas de datos | 59 |
| 2.2 Métodos de búsqueda complementarios..... | 60 |
| 2.3 Actualización de una búsqueda sistemática de la literatura | 60 |
| 2.4 Gestión documental..... | 62 |
| 3. Tamización de referencias y selección de estudios..... | 62 |
| 4. Evaluación del riesgo de sesgos..... | 63 |
| 5. Extracción de datos | 64 |
| 6. Síntesis de la evidencia | 64 |
| 6.1 Consideraciones adicionales..... | 66 |
| 7. Análisis estadístico..... | 66 |
| 7.1 Metanálisis clásico | 66 |

| | | |
|--|---|------------|
| 7.2 | Metanálisis en red (NMA) | 67 |
| 7.3 | Adición de nuevos datos a las medidas combinadas del efecto | 68 |
| 8. | Valoración de la calidad y del cuerpo de la evidencia por cada desenlace..... | 69 |
| 9. | Determinación de valor umbral terapéutico o significancia clínica de los resultados..... | 70 |
| 9.1 | Proceso deliberativo para determinar umbral de valor terapéutico | 74 |
| 10. | Formulación de las conclusiones | 85 |
| 11. | Revisión por pares..... | 85 |
| | | |
| Capítulo 5. Guía para el reporte de protocolos (versión abreviada) e informes de evaluación rápida de tecnologías en salud | | 87 |
| Referencias bibliográficas..... | | 93 |
| Anexos | | 99 |
| Anexo 1. Registros sanitarios vigentes para las tecnologías de interés | | 99 |
| Anexo 2. Reportes de búsqueda de evidencia en bases electrónicas de datos..... | | 100 |
| Anexo 3. Diagrama PRISMA: flujo de la búsqueda, tamización y selección de estudios | | 101 |
| Anexo 4. Diagrama PRISMA para una actualización: flujo de la búsqueda, tamización y selección de estudios..... | | 102 |
| Anexo 5. Análisis del espectro eficacia-efectividad en estudios de intervenciones en salud | | 103 |
| Anexo 6. Herramientas para la evaluación de riesgo de sesgos por tipo de diseño.. | | 105 |
| Evaluación de riesgo de sesgos de revisiones sistemáticas de la literatura | | 105 |
| Evaluación de calidad de metanálisis de comparaciones indirectas o en red | | 106 |
| Evaluación de riesgo de sesgos en experimentos clínicos aleatorizados..... | | 109 |
| Evaluación de riesgo de sesgos en estudios de intervención no aleatorizados | | 110 |
| Anexo 7. Contenido del informe de efectividad y seguridad | | 121 |
| Anexo 8. Lista de chequeo para los reportes de evaluación de efectividad, seguridad o validez diagnóstica de tecnologías en salud | | 123 |

El Instituto de Evaluación Tecnológica en Salud - IETS, es una corporación sin ánimo de lucro, de participación mixta y de carácter privado, con patrimonio propio, creado según lo estipulado en la Ley 1438 de 2011. Su misión es contribuir al desarrollo de mejores políticas públicas y prácticas asistenciales en salud, mediante la producción de información basada en evidencia, a través de la evaluación de tecnologías en salud y guías de práctica clínica, con rigor técnico, independencia y participación. Los miembros de su asamblea son el Ministerio de Salud y Protección Social, el Ministerio de Ciencia, Tecnología e Innovación - Minciencias, el Instituto Nacional de Vigilancia de Medicamentos y Alimentos - INVIMA, el Instituto Nacional de Salud - INS, la Asociación Colombiana de Facultades de Medicina - ASCOFAME y la Asociación Colombiana de Sociedades Científicas.

Información general del manual

Autores

Estrada-Orozco Kelly, Médica cirujana, MSc Neurociencias y biología del comportamiento, MSc Epidemiología Clínica, Experta en mejoramiento continuo de la calidad, PhDc Salud Pública, PhDc Epidemiología clínica. Instituto de Evaluación Tecnológica en Salud - IETS.

Cortes-Muñoz Ani Julieth, Bacterióloga y laboratorista clínica, MSc Epidemiología. Instituto de Evaluación Tecnológica en Salud - IETS.

León-Guzmán Erika. Odontóloga. Especialista en Epidemiología. MSc en Epidemiología Clínica. Instituto de Evaluación Tecnológica en Salud IETS

Osorio-Arango Luz Karime. Bacterióloga y laboratorista clínico, especialista en epidemiología, MSc en salud pública. Instituto de Evaluación Tecnológica en Salud – IETS

Ospina-Lizarazo Nathalie Nutricionista Dietista MSc en Epidemiología Clínica. Instituto de Evaluación Tecnológica en Salud – IETS.

Pinilla Forero, Margoth Cristina. Ingeniería Biomédica, Esp. en Gerencia de la calidad y auditoría en salud, MSc en Gestión de tecnologías en salud, estudiante de especialización en evaluación económica en salud. Instituto de Evaluación Tecnológica en Salud – IETS.

Segura Sandino, Diana Marcela. Química farmacéutica, Esp. en Economía y Gestión de la salud, MS(c) en Farmacología, Instituto de Evaluación Tecnológica en Salud - IETS.

Sierra-Matamoros Fabio Alexander, Psicólogo, MSc en Epidemiología Clínica. Instituto de Evaluación Tecnológica en Salud - IETS.

Revisores

Herrera T. Ana Milena MD. Magister en epidemiología. PhD en patología. Instituto de Evaluación Tecnológica en Salud - IETS.

Mesa, Lorena. Enfermera, M.Sc. en Salud Pública, MSc en Economía de la Salud, Ph.D. en Salud Pública. Instituto de Evaluación Tecnológica en Salud

Sánchez, Guillermo. MD. MSc. Epidemiología Clínica PhD. Salud Pública. Director Centro Asociado Cochrane-FUCS. Coordinador Posgrados Epidemiología Clínica. Profesor Investigador Titular-FUCS.

Mejía Aurelio. Economista, MSc en Economía de la Salud. Director de Regulación. Ministerio de Comercio, Industria y Turismo MINCIT.

Fuentes de financiación

Instituto de Evaluación Tecnológica en Salud - IETS.

Conflictos de interés

Los autores de este manual declaran que no existe ningún tipo de conflicto financiero, intelectual, de pertenencia o familiar que afecte las recomendaciones incluidas en el presente manual.

Citación

Segunda versión

Este documento debe citarse de la siguiente manera:

Estrada-Orozco K, Cortés-Muñoz A, León E, Osorio K, Ospina-Lizarazo N, Pinilla M, Segura D, Sierra-Matamoros F. Manual metodológico para la elaboración de

evaluaciones de efectividad clínica, seguridad y validez diagnóstica de tecnologías en salud. Instituto de Evaluación Tecnológica en Salud – IETS. Bogotá D.C. 2022.

Derechos de autor

Los derechos de propiedad intelectual del contenido de este documento son de propiedad del Instituto de Evaluación Tecnológica en Salud - IETS. Lo anterior, sin perjuicio de los derechos morales y las citas y referencias bibliográficas enunciadas.

En consecuencia, constituirá violación a la normativa aplicable a los derechos de autor, y acarreará las sanciones civiles, comerciales y penales a que haya lugar, su modificación, copia, reproducción, fijación, transmisión, divulgación, publicación o similares, parcial o total, o el uso del contenido de este sin importar su propósito, sin que medie el consentimiento expreso y escrito del Instituto de Evaluación Tecnológica en Salud - IETS.

Correspondencia

Instituto de Evaluación Tecnológica en Salud - IETS
Carrera 49 A # 91-91
Bogotá, D.C., Colombia.
www.iets.org.co
contacto@iets.org.co

© Instituto de Evaluación Tecnológica en Salud, 2022

Listado de siglas y abreviaturas

| | |
|----------|---|
| ADMC | Análisis de decisión multicriterio |
| ANVISA | Agencia Nacional de Vigilancia Sanitaria |
| ATC | Anatomical, Therapeutic, Chemical Classification System |
| CADTH | Canadian Agency for Drugs and Technologies in Health |
| CENETEC | Centro Nacional de Excelencia Tecnológica en Salud |
| CENTRAL | Cochrane Central Register of Controlled Trials |
| CINAHL | Cumulative Index to Nursing and Allied Health Literature |
| CLIB | Biblioteca Cochrane |
| COMET | Core Outcome Measures in Effectiveness Trials |
| CUM | Código Único de Medicamentos |
| CUPS | Clasificación Única de Procedimientos en Salud |
| DARE | Database of Abstracts of Reviews of Effects |
| DeCS | Descriptores de Ciencias de la Salud |
| ECA | Ensayo Controlado Aleatorizado |
| EMA | European Medicines Agency |
| Emtree | Embase Subject Headings |
| EOC | Entidades Obligadas a Compensar |
| ENDS | Encuesta Nacional de Demografía y Salud |
| EPS | Entidades Promotoras de Salud |
| ETES | Evaluación de Tecnología Sanitaria |
| EUnetHTA | European network for health technology assessment |
| FDA | Food and Drug Administration |
| GLOBOCAN | Global Cancer Observatory |
| GRADE | Grading of Recommendations Assessment, Development and Evaluation |
| GIN | Guidelines International Network |
| GPC | Guía de Práctica Clínica |
| HR | Hazard Ratio |
| HTA | Health Technology Assessment |
| IARC | The International Agency for Research on Cancer |
| IC | Intervalo de Confianza |
| ICHOM | International Consortium for Health Outcomes Measurement |
| ICTRP | WHO International Clinical Trials Registry Platform |
| IECS | Instituto de Efectividad Clínica y Sanitaria |
| IETS | Instituto de Evaluación Tecnológica en Salud |

| | |
|----------|--|
| IHME | Institute for Health Metrics and Evaluation - University of Washington |
| INAHTA | International Network of Agencies for Health Technology Assessment |
| INS | Instituto Nacional de Salud |
| INVIMA | Instituto Nacional de Vigilancia de Medicamentos y Alimentos |
| ITC | Indirect Treatment Comparisons |
| JBI | Joanna Briggs Institute |
| LH | Likelihood Ratio |
| MeSH | Medical Subject Headings |
| MHRA | Medicines and Healthcare Products Regulatory Agency |
| MTC | Multiple Treatment Comparison |
| NA | No Aplica |
| NGC | National Guideline Clearinghouse |
| NNT | Número Necesario Para Tratar |
| NZGG | New Zealand Guidelines Group |
| OMS | Organización Mundial de la Salud |
| ONS | Observatorio Nacional de Salud |
| OR | Odds Ratio |
| POS | Plan Obligatorio de Salud |
| PRISMA | Preferred Reporting Items for Systematic Reviews and Meta-Analyses |
| QUADAS | Quality Assessment of Diagnostic Accuracy Studies |
| RAR | Reducción Absoluta del Riesgo |
| RoB | Risk of Bias |
| ROC | Receiver Operating Characteristic Curve |
| RR | Riesgo Relativo |
| RSL | Revisión Sistemática de la Literatura |
| SGSSS | Sistema General de Seguridad Social en Salud |
| SIGN | Scottish Intercollegiate Guidelines Network |
| SISPRO | Sistema integral de información de la Protección Social |
| SIVIGILA | Sistema de Vigilancia en Salud Pública |
| UPC | Unidad de Pago por Capitación |

Glosario

- Aceptabilidad: valoración general del paciente tras evaluar las características de un fármaco o tecnología y su preferencia con respecto a otros tratamientos disponibles (1).
- Actores interesados: hace referencia a las partes interesadas en una decisión de cobertura o financiación de una tecnología sanitaria (2).
- Análisis de decisión multicriterio (ADMC): se define como "un término general para describir una colección de enfoques formales que buscan tener en cuenta explícitamente múltiples criterios para ayudar a las personas o grupos a explorar decisiones importantes". Cualquier ADMC implica al menos 3 pasos: definir el problema de decisión, seleccionar los criterios que reflejan los valores relevantes y construir la matriz de desempeño (3).
- Aproximación GRADE: sistema para calificar la calidad del cuerpo de la evidencia en revisiones sistemáticas y otras síntesis de evidencia, como evaluaciones de tecnologías en salud y guías, y formular recomendaciones en salud. La aproximación GRADE ofrece un proceso transparente y estructurado para desarrollar y presentar los resúmenes de la evidencia, y la formulación de recomendaciones. Se puede utilizar para desarrollar guías de práctica clínica (GPC) y otras recomendaciones del área de la salud (p.ej. Salud pública, sistemas y políticas públicas y decisiones de cobertura en salud) (4).
- Base de datos administrativas: cualquiera de una amplia variedad de bancos de datos (a menudo informatizados) que recogen observaciones e información sobre un grupo de pacientes (p. ej. varones residentes en Gotemburgo), una enfermedad (p. ej. hipertensión), una intervención (p. ej. farmacoterapia antihipertensiva) u otros eventos o características. (5).
- Base de datos bibliográfica: fuente indexada, informatizada o impresa, de citas de artículos de revistas y otros documentos publicados. Las citas bibliográficas normalmente incluyen autor, título, fuente, resumen y/o información relacionada (incluyendo en algunos casos, el texto completo). Ejemplo de ello son MEDLINE y EMBASE (5).

- Biblioteca Cochrane (CLIB): colección de bases de datos en línea actualizadas regularmente que contienen la Base de datos Cochrane de revisiones sistemáticas (CDSR), el Registro Cochrane Central de Ensayos Controlados (CENTRAL), la Base de datos de resúmenes de revisiones de efectos, el Registro de metodología Cochrane, la Base de datos HTA, NHSEED, e información sobre la Colaboración Cochrane (6).
- Calidad metodológica: grado en que el diseño y la realización de un estudio probablemente hayan prevenido errores sistemáticos (sesgo). La variación de calidad puede explicar la variación en los resultados. Los estudios diseñados con rigor, presentan una mayor probabilidad de obtener resultados más próximos a la 'verdad'. (Véase también validez externa y validez) (5). Sin embargo, en algunas ocasiones se pueden encontrar estudios con alta calidad metodológica pero que distan de la realidad.
- Cegamiento: [En un ensayo controlado] es el proceso de evitar que las personas involucradas en un ensayo sepan a qué grupo de comparación pertenece un participante en particular. Los participantes, cuidadores, evaluadores de resultados y analistas son todos candidatos para ser cegados. El cegamiento de ciertos grupos no siempre es posible, por ejemplo, cirujanos en ensayos quirúrgicos. Los términos ciego, doble ciego y triple ciego son de uso común, pero no se usan de manera consistente y, por lo tanto, son ambiguos a menos que se enumeren las personas específicas que están cegadas (6).
- Certeza en la evidencia/ Calidad de la evidencia: refleja el grado de confianza en que las estimaciones del efecto son correctas. En el contexto de hacer recomendaciones, refleja el grado de confianza en que las estimaciones de un efecto son adecuadas para respaldar una decisión o recomendación en particular (7).
- CINAHL (Cumulative Index to Nursing and Allied Health Literature): base de datos electrónica que contiene bibliografía sobre enfermería y ciencias afines. Años de cobertura: 1981 hasta la actualidad
- Coeficiente de correlación: medida numérica entre -1 y 1 que expresa la fuerza de una asociación lineal observada entre dos variables. Expresada como r , el valor $r = 0$ indica, o bien la ausencia de una relación lineal o una posible relación no lineal entre las dos variables (5) .

- Coeficiente de probabilidad (Likelihood ratio): 1. cociente entre la probabilidad de obtener un resultado positivo (o negativo) en aquellas personas que tienen una determinada enfermedad y la citada probabilidad en personas sin la mencionada enfermedad. La razón de verosimilitud para una prueba con resultado positivo es sensibilidad/ (1 - especificidad). La razón de verosimilitud de una prueba con resultado negativo es (1 - sensibilidad) /especificidad. 2. índice estadístico que compara la bondad del ajuste de dos modelos relacionados, lo que permite contrastar hipótesis en un gran número de situaciones (5).
- Comorbilidad: presencia de enfermedades o condiciones coexistentes o adicionales a la que se está estudiando. También describe el efecto que estas condiciones coexistentes tienen sobre el individuo además de la condición primaria (5).
- Consenso: acuerdo en un comité por coincidencia de opiniones. Dicho acuerdo puede ser parcial, o incluso puede no alcanzarse, por lo que las opiniones de los comités suelen referirse a mínimos éticos, e incluir votos particulares con las opiniones discordantes. En el desarrollo del consenso surgen diversas formas de opinión de grupos en las que un grupo (o panel) de expertos interactúa para la evaluación de una intervención y la formulación de propuestas por votación u otro proceso, para llegar a un acuerdo general. Este proceso puede ser informal o formal y conlleva técnicas como la de grupo nominal y la Delphi. Sinónimo: métodos de consenso (5).
- Contraindicación: circunstancia, condición, síntoma o factor que aumenta el riesgo asociado con un procedimiento médico, medicamento o tratamiento. Una contraindicación se refiere a cualquier intervención considerada inapropiada o desaconsejada en función de factores únicos de la situación, como las posibles interacciones perjudiciales entre medicamentos o afecciones médicas que vuelven vulnerable a un individuo si se implementa (8).
- Cribado o tamización: servicio de salud pública en el cual se formula una pregunta u ofrece una prueba a miembros de una población definida, que no necesariamente perciben que están en peligro de sufrir una enfermedad o sus complicaciones, para identificar a aquellas personas que presentan más probabilidades de beneficiarse que de perjudicarse por la realización de otras pruebas o tratamientos. Sinónimos: tamización (5). En el ámbito de las

revisiones sistemáticas de la literatura, la tamización se refiere al proceso de lectura y selección de documentos por título y resumen.

- Curva de características operativas del receptor (ROC): gráfico que representa la relación entre la tasa de verdaderos positivos (sensibilidad) y la tasa de falsos positivos ("1 - especificidad"), como una función del nivel de corte del marcador de una enfermedad (o condición clínica). Las curvas ROC ayudan a demostrar cómo el incremento o disminución del punto de corte que define una prueba como positiva afecta al equilibrio entre la identificación correcta de los enfermos (verdaderos positivos) y la denominación incorrecta como positivos de los sanos (falsos positivos). El área bajo la curva ROC es una expresión del potencial diagnóstico de un marcador independiente de la población de pacientes y puede utilizarse para comparar uno o más marcadores (5).
- Curva SROC: gráfico bidimensional que en un metanálisis de pruebas diagnósticas, representa en un eje la sensibilidad, y en el otro (1-especificidad) para cada estudio individual y la curva que mejor se ajusta a ese conjunto de puntos (5).
- DARE (Database of Abstracts of Reviews of Effects): base de datos de revisiones sistemáticas cuya calidad ha sido evaluada sobre los efectos de las intervenciones de salud. (Véase también biblioteca Cochrane) (5).
- Datos del mundo real: datos obtenidos fuera del contexto de los ensayos controlados aleatorizados, que se generan durante la práctica clínica rutinaria. Esto incluye datos de estudios observacionales, retrospectivos o prospectivos, y registros observacionales (5).
- Declaración de conflicto de intereses: manifestación explícita realizada por las personas que participan en un informe o una revisión de la evidencia, sobre los intereses personales, económicos o de otra índole que podrían influir en los resultados del informe o en su interpretación (5).
- Desenlace: componente del estado clínico o funcional de un participante después de que se ha aplicado una intervención, que se utiliza para evaluar el beneficio o daño de una intervención (4). Un desenlace describe una medida válida del beneficio clínico debido al tratamiento: el impacto de la intervención en cómo se siente, funciona y sobrevive un paciente. Es clínicamente relevante, sensible (responde al cambio) y es aceptado y

utilizado tanto por los médicos como por los pacientes. Los desenlaces pueden ser un evento clínico (Por ejemplo: mortalidad), una combinación de varios eventos, una medida del estado clínico o calidad de vida relacionada con la salud (CVRS).

- Desenlace (clasificación de la importancia): al analizar la evidencia siguiendo las pautas GRADE, se clasifican los desenlaces en: críticos, los que son importantes, pero no críticos, y los de limitada importancia. Los desarrolladores de ETES pueden optar por calificar los resultados numéricamente en una escala de 1 a 9 (7 a 9, crítico; 4 a 6, importante; y 1 a 3, de importancia limitada) para distinguir entre las categorías de importancia). Debido a que los expertos, los médicos y los pacientes pueden tener diferentes valores y preferencias, los aportes realizados por los afectados por la decisión (pacientes) pueden fortalecer este proceso.
- Desenlaces subrogados: medida intermedia entre la intervención y el desenlace, y por ser un paso intermedio en el proceso causal, puede ser un indicador indirecto del desenlace definitivo, por ejemplo, la presión arterial no es directamente significativa para los pacientes, pero a menudo se usa como resultado en ensayos clínicos porque es un factor de riesgo de accidente cerebrovascular y ataques cardíacos. Los desenlaces subrogados son a menudo marcadores fisiológicos o bioquímicos que se pueden medir de forma relativamente rápida y fácil, y que se toman como predictores de resultados clínicos importantes. A menudo se usan cuando la observación de resultados clínicos requiere un seguimiento prolongado (6).
- Diferencia mínima clínicamente importante: el cambio clínico más pequeño que es importante para el paciente. Es una herramienta útil que permite a los clínicos interpretar adecuadamente los efectos de las intervenciones evaluadas en los estudios clínicos (9).
- Efectividad: significa que un efecto se detecta no bajo condiciones ideales sino bajo condiciones del mundo real (10). Es la medida en que una intervención específica, cuando se usa en circunstancias normales, hace lo que se pretende que haga. Los ensayos clínicos que evalúan la efectividad a veces se denominan ensayos pragmáticos o de manejo (6).
- Efectividad comparativa: comparación de los resultados del mundo real que resultan de opciones de tratamiento alternativas que están disponibles para una afección médica determinada. Sus elementos clave incluyen el estudio

de la efectividad (efecto en el mundo real), en lugar de la eficacia (efecto ideal), y las comparaciones entre estrategias alternativas (11).

- Efecto adverso: efecto no deseado o no intencionado de una intervención. (Véase también evento adverso y efecto secundario) (5).
- Efecto del tratamiento: efecto de un tratamiento (intervención) sobre los resultados, es decir, sólo atribuible al efecto de la intervención. Los investigadores tratan de estimar el verdadero efecto del tratamiento utilizando la diferencia entre los resultados observados en el grupo de tratamiento y el grupo control. (Véase también tamaño del efecto) (5).
- Efecto placebo: efecto sobre los resultados (mejora o empeoramiento) de un paciente que puede producirse debido a la expectativa de este (o proveedor) de que una determinada intervención tenga un efecto. El efecto placebo (también conocido como el efecto Hawthorne) es independiente del efecto verdadero (farmacológico, quirúrgico, etc.) de una intervención dada. Para controlar este factor, puede administrarse un placebo al grupo de control de un ensayo (5).
- Eficacia: la medida en que una intervención produce un resultado beneficioso en condiciones ideales (condiciones altamente controladas de experimentación). Los ensayos clínicos que evalúan la eficacia a veces se denominan ensayos explicativos y están restringidos a los participantes que cooperan completamente (6). Se utiliza para expresar que se puede observar un efecto biológico en condiciones ideales de estudio (10).
- EMBASE: base de datos biomédica y farmacológica (Excerpta Medical). Contiene un gran número de publicaciones europeas. Años de cobertura: 1947 hasta la actualidad (5).
- EMTREE: tesauro de vocabulario controlado de la base de datos EMBASE (5).
- Enmascaramiento: (véase cegamiento).
- Ensayo controlado aleatorio: experimento de dos o más intervenciones en el que se asigna a personas elegibles a las intervenciones mediante aleatorización. El uso de la aleatorización permite utilizar de forma válida una

variedad de métodos estadísticos para comparar los resultados de las intervenciones (5).

- Ensayos de equivalencia: ensayos que tienen como objetivo determinar si una intervención (típicamente nueva) es terapéuticamente similar a otro tratamiento (generalmente existente). Los ensayos de equivalencia son muy similares, excepto que la equivalencia se define como el efecto del tratamiento entre $-\Delta$ y $+\Delta$ (12) .
- Ensayos de fase 0, 1 ,2 ,3 ,4: Fase 0: estudio exploratorio que involucra una exposición humana muy limitada al medicamento, sin objetivos terapéuticos o de diagnóstico (por ejemplo, estudios de detección, estudios de microdosis). Fase 1: estudios que generalmente se realizan con voluntarios sanos y que enfatizan la seguridad. El objetivo es descubrir cuáles son los eventos adversos más frecuentes y graves del medicamento y, a menudo, cómo se metaboliza y excreta el medicamento. Fase 2: estudios que recopilan datos preliminares sobre la efectividad (si el medicamento funciona en personas que tienen una determinada enfermedad o condición). Por ejemplo, los participantes que reciben el medicamento pueden compararse con participantes similares que reciben un tratamiento diferente, generalmente una sustancia inactiva (llamada placebo) o un medicamento diferente. La seguridad continúa siendo evaluada y se estudian los eventos adversos a corto plazo. Fase 3: estudios que recopilan más información sobre seguridad y efectividad al analizar diferentes poblaciones y diferentes dosis y al usar el medicamento en combinación con otros medicamentos. Fase 4: estudios que se realizan después de que se haya aprobado un medicamento para su comercialización. Estos incluyen los requisitos de posventa y los estudios de compromiso requeridos o acordados por el patrocinador. Estos estudios recopilan información adicional sobre la seguridad, efectividad o uso óptimo de un medicamento (13).
- Ensayo de no inferioridad: buscan determinar si un nuevo tratamiento no es peor que un tratamiento de referencia en más de una cantidad aceptable. Debido a que la prueba de equivalencia exacta es imposible, se define un margen preestablecido de no inferioridad (Δ) para el efecto del tratamiento en un resultado primario del paciente (12).
- Ensayo de superioridad: busca mostrar que una nueva intervención de tratamiento (fármaco, psicoterapia) es superior a (mejor que) otro tratamiento (14).

- Ente gestor: hace referencia a la entidad, agente o solicitante interesado en la realización de la evaluación de tecnologías en salud, quien contrata una institución coordinadora del proceso (contratista) que pueden ser institutos, grupos de investigación, universidades, centros de investigación, sociedades científicas, etc.
- Estimación puntual: resultados (p. ej. media, diferencia de media ponderada, Odds ratio, riesgo relativo, diferencia de riesgos) obtenidos de una muestra (estudio o metanálisis) que se utilizan como la estimación de lo que ocurre en la realidad en la población de la cual se ha obtenido la muestra. Un intervalo de confianza es una medición de la incertidumbre (debida al azar) asociada a esa estimación. (Véase también intervalo de confianza) (5).
- Estrategia de búsqueda: combinación de fuentes, términos y límites utilizados en una búsqueda bibliográfica, con el fin de identificar información para revisiones sistemáticas o evaluaciones de tecnología sanitaria (5).
- Estudio primario: "Investigación original" en la que se recopilan datos. El término estudio primario a veces se usa para distinguirlo de un estudio secundario o integrativo (nuevo análisis de datos recopilados previamente), metaanálisis y otras formas de combinar estudios (como el análisis económico y el análisis de decisiones) (6).
- Evaluación de Tecnologías Sanitarias (ETES): proceso sistemático de valoración de las propiedades, los efectos y/o los impactos de una o más tecnologías sanitarias. Puede abordar tanto las consecuencias directas y deseadas de las tecnologías como las indirectas y no deseadas. Su objetivo principal es informar la toma de decisiones en atención sanitaria. Las ETES son realizadas por grupos multidisciplinares utilizando marcos analíticos explícitos basados en diversos métodos (5).
- Evaluación económica en salud: consiste en una comparación de los costos y beneficios en salud de dos o más tecnologías indicadas para abordar un problema de salud específico, es una herramienta para la toma de decisiones de los agentes del sistema de salud que se enmarca en el análisis microeconómico: se fundamenta en el análisis marginal en su definición económica (en donde se evalúa el aporte o el costo adicional de producir una unidad adicional de resultados) y contribuye al proceso de asignación eficiente de los recursos disponibles. En general, el objetivo de una

evaluación económica es ofrecer información relevante para los tomadores de decisiones, basada en la mejor evidencia, para promover una asignación eficiente de los recursos disponibles que permitan maximizar el bienestar de los usuarios del sistema de salud (15).

- Evento adverso: cualquier cambio nocivo, patológico o no intencionado en las funciones anatómicas, físicas o metabólicas evidente por signos físicos, síntomas y/o cambios microbiológicos producidos durante cualquier fase de un estudio clínico, estén o no relacionados con el tratamiento. Incluye la exacerbación de dolencias o eventos preexistentes, enfermedades intercurrentes, accidentes, la interacción con otros fármacos o el empeoramiento significativo de la enfermedad (5)
- Evidencia del mundo real: aquella obtenida de analizar los datos del mundo real (16) (ver datos del mundo real).
- Experto temático: persona que tiene un profundo conocimiento de un tema específico, puede elaborar argumentos y justificaciones sobre distintas proposiciones relacionadas con éste y puede discriminar entre argumentos útiles y robustos y argumentos fútiles. Para precisar con mayor detalle los atributos de un experto temático consulte el Manual de Participación y Deliberación- IETS 2014 (17).
- Fuerza de la recomendación: grado de confianza en que los efectos deseables de una intervención superan los efectos indeseables, o viceversa(4).
- Gráfico de embudo: visualización gráfica de alguna medida de precisión de estudio trazada contra el tamaño del efecto, que puede usarse para investigar si existe un vínculo entre el tamaño del estudio y el efecto del tratamiento. Una posible causa de una asociación observada es el sesgo de informe o publicación (6).
- Heterogeneidad: en el metanálisis, la heterogeneidad se refiere a la variabilidad o a las diferencias en las estimaciones de los efectos entre los estudios. En ocasiones, se distingue entre la "heterogeneidad estadística" (diferencias en los efectos declarados), la "heterogeneidad metodológica" (diferencias en el diseño de estudios) y la "heterogeneidad clínica" (diferencias entre estudios en las características fundamentales de participantes, intervenciones o mediciones de resultados). Las pruebas

estadísticas de heterogeneidad se utilizan para evaluar si la variabilidad observada en los resultados de las estimaciones del tamaño del efecto, son diferentes a las que podrían esperarse por azar. Sin embargo, estas pruebas tienen escasa potencia estadística (5).

- Indicación: síntoma clínico, factor de riesgo o circunstancia en la cual es apropiada la realización de una intervención sanitaria, de acuerdo a lo determinado o especificado, por ejemplo, en una guía de práctica clínica, en el estándar asistencial, por un órgano regulador o por otra fuente autorizada (5).
- Intervalo de confianza: representa el rango de incertidumbre sobre una estimación del efecto de un tratamiento. Se calcula a partir de las diferencias observadas en los resultados de los grupos de tratamiento y control, y del tamaño de la muestra de un estudio. El intervalo de confianza (IC) es el rango de valores por encima y por debajo de la estimación puntual que, con determinada probabilidad, incluye el valor del parámetro del efecto del tratamiento. Los investigadores normalmente utilizan IC del 90%, 95% o 99%. Por ejemplo, un IC del 95% indica que existe una probabilidad del 95% de que el IC calculado a partir de un estudio dado, incluya el parámetro del efecto de un tratamiento. (5).
- Involucramiento de actores: proceso iterativo de solicitar activamente el conocimiento, la experiencia, el juicio y los valores de actores identificados para representar intereses en un tema particular, con el doble propósito de crear un entendimiento compartido y de tomar decisiones relevantes, transparentes y efectivas (18).
- Literatura gris: documentos que no se publican en revistas o bases de datos formalmente publicados por medios académicos. Incluye actas de congresos que contiene los resúmenes de investigaciones presentadas en conferencias, tesis inéditas, etc.(6)
- Límites de confianza: los límites superior e inferior de un intervalo de confianza (6).
- MEDLINE: base de datos electrónica generada por la Biblioteca Nacional de Medicina de los Estados Unidos (*United States National Library of Medicine*). Años de cobertura: 1948 hasta la actualidad. (Véase también PubMed) (5).

- Medical Subject Headings MeSH (Descriptor de Ciencias de la Salud): es un vocabulario controlado de términos biomédicos que identifican el contenido de cada artículo en la base de datos MEDLINE (5).
- Metanálisis: uso de técnicas estadísticas en una revisión sistemática para integrar los resultados de los estudios incluidos (6). Se utiliza para proporcionar una estimación combinada del tamaño del efecto. Se recomienda que la técnica metanalítica se realice en el marco de una revisión sistemática de la literatura.
- Número necesario para tratar para beneficiarse (NNTb): estimación de cuántas personas necesitan recibir un tratamiento antes de que una persona experimente un resultado beneficioso. Por ejemplo, si necesita administrar un medicamento para la prevención del accidente cerebrovascular a 20 personas antes de que se prevenga un accidente cerebrovascular, el número necesario a tratar para beneficiarse de ese medicamento para la prevención del accidente cerebrovascular es 20. El NNTb se estima como el recíproco de la diferencia de riesgo absoluto (6). Se calcula como el inverso de la Reducción Absoluta del Riesgo (RAR) dado como $1/RAR$.
- Número necesario para tratar el daño (NNTd): estimación de cuántas personas necesitan recibir un tratamiento antes de que una persona experimente un resultado perjudicial (6).
- *Odds Ratio* (OR): es una medida de efecto comúnmente utilizada para comunicar los resultados de una investigación en salud que expresa el cociente entre la probabilidad de que algo ocurra (P) y la probabilidad de que no ocurra (1-P) (5).
- Patrón de oro: método, procedimiento o medición generalmente aceptado como el mejor de su clase, con respecto al cual deben compararse las intervenciones nuevas. Es especialmente importante en los estudios de exactitud (o validez) de las pruebas diagnósticas. Sinónimo: estándar de referencia (5).
- Perfil de evidencia GRADE: aproximación que permite presentar información clave acerca de todos los desenlaces relevantes para una pregunta de investigación. Este perfil presenta información acerca del cuerpo de la evidencia (p.ej. Número de estudios), los juicios acerca de la calidad de la evidencia, los resultados estadísticos claves y la calificación de la evidencia

para cada desenlace. Es útil para presentar la evidencia que soporta una recomendación en las guías de práctica clínica, pero también como un resumen de la evidencia para otros propósitos, como cuando los usuarios necesitan o quieren entender los juicios acerca de la calidad de la evidencia con mayor detalle (4).

- Precisión diagnóstica: capacidad de una prueba para discriminar entre la condición objetivo y la salud. Esta capacidad puede estimarse o cuantificarse a partir de las medidas del rendimiento operativo de la prueba como la sensibilidad y la especificidad, los valores predictivos, las razones de probabilidad, el área bajo la curva ROC, entre otros (19).
- Prueba de Mantel-Haenszel: prueba Ji-cuadrado para datos estratificados que contrasta la asociación controlando la posible confusión. En los metanálisis, la prueba Mantel-Haenszel se utiliza para analizar los datos estratificados (agrupados) por estudio (5).
- Prueba estadística: procedimiento o función matemática que se utiliza, entre otras cosas, para determinar si las diferencias en los resultados entre el grupo de tratamiento y el grupo control son lo suficientemente grandes como para concluir que son estadísticamente significativas. Entre las pruebas estadísticas más comunes se encuentran las siguientes: Pruebas F, t, Z y Ji-cuadrado. La elección de la prueba depende de las condiciones de un estudio, por ejemplo el tipo de variable de resultado utilizada, si los pacientes fueron seleccionados aleatoriamente de una población más grande, y si puede asumirse que los valores de los resultados de la población tienen una distribución normal u otro tipo de distribución (5)
- Pubmed: Es un sistema de búsqueda de la Biblioteca Nacional de Medicina de Estados Unidos que contiene citas de artículos biomédicos que datan desde 1947. Estas citas son de MEDLINE y otras publicaciones de ciencias de la vida. Pubmed contiene enlaces a muchos sitios web que ofrecen artículos en texto completo y otros recursos relacionados. Además de MEDLINE, Pubmed incluye otras bases de datos, como por ejemplo OLDMEDLINE, registros facilitados por la editorial y registros en curso, y se puede acceder a ella de forma gratuita en Internet. (Véase también MEDLINE) (5).
- Reducción Absoluta del Riesgo (RAR): sinónimo de diferencia del riesgo (DR), es la diferencia en el riesgo entre dos grupos. Por ejemplo, si un grupo

tiene un riesgo del 15% de contraer una enfermedad particular, y el otro tiene un riesgo de 10% de contraer la enfermedad, la diferencia del riesgo es 5 puntos porcentuales (4).

- Reducción Relativa del Riesgo (RRR): reducción proporcional en el riesgo en un grupo de tratamiento comparado con otro. Se calcula restando uno de la razón de riesgo. Si la razón de riesgo es 0,25, entonces la reducción relativa del riesgo es $1-0,25=0,75$ o 75%(4).
- Relevancia clínica: diferencia entre dos resultados de terapia que es lo suficientemente grande como para justificar el cambio del estándar de atención(20). Refleja el alcance del cambio, si el cambio hace una diferencia real en la vida de los sujetos, cuánto duran los efectos, la aceptabilidad del usuario, los resultados del análisis de costo-efectividad y la facilidad de implementación (21).
- Resultado secundario: Un resultado utilizado para evaluar los efectos adicionales de la intervención considerados *a priori* como menos importantes que los resultados primarios (6).
- Revisión de la literatura: resumen e interpretación de los hallazgos de investigación publicados en la literatura. Puede incluir revisiones cualitativas no estructuradas por autores individuales, así como diversos procedimientos sistemáticos y cuantitativos como el metanálisis. Sinónimo: descripción general (5).
- Revisión narrativa: descripción general de los estudios primarios que no se han identificado o analizado de forma sistemática (estandarizada y objetiva) (5).
- Revisión por pares: proceso por el cual los manuscritos enviados a revistas de salud, biomédicas y otras publicaciones de tipo científico se someten a una revisión por otros expertos con experiencia en el área (normalmente anónimos para el autor) a fin de determinar si ofrecen la calidad necesaria para ser publicados (5) o para verificar importancia de los informes de investigación (6).
- Revisión sistemática de la literatura (RSL): Forma de revisión estructurada que tienen como objetivo proporcionar una síntesis completa e imparcial de estudios relevantes en un solo documento, utilizando métodos rigurosos,

transparentes y reproducibles. Una revisión sistemática tiene como objetivo sintetizar y resumir el conocimiento existente. Intenta descubrir “todas” las pruebas pertinentes a una pregunta (22) . Para analizar y resumir los resultados de los estudios incluidos, pueden utilizarse optativamente métodos estadísticos (metanálisis). (Véase también revisión Cochrane) (5).

- Revisor: persona responsable de preparar y realizar una revisión sistemática. El término “revisor” también se utiliza para referirse a la persona que participa en la revisión por pares o a la que dirime las discrepancias entre evaluadores (5).
- Riesgo: probabilidad de que un individuo, libre de un desenlace y susceptible de él, lo desarrolle en un periodo determinado (23).
- Riesgo relativo: razón de riesgo (estadístico) en el grupo de intervención con respecto al riesgo en el grupo de control o entre los expuestos y los no expuestos. Un riesgo relativo de 1 indica que no hay diferencia entre los grupos comparados. Para efectos no deseados, un RR inferior a 1 indica que la intervención resultó eficaz para reducir el riesgo de ese resultado (5).
- Seguridad de una intervención: se refiere a los efectos adversos graves, como los que amenazan la vida, requieren o prolongan la hospitalización, provocan discapacidad permanente o causan defectos de nacimiento. Los efectos adversos indirectos, como los accidentes de tránsito, la violencia y las consecuencias perjudiciales del cambio de humor, también pueden ser graves (6).
- Sesgo: error sistemático o desviación en los resultados o inferencias de la validez. En los estudios de los efectos de la atención en salud, los principales tipos de sesgo se originan de diferencias sistemáticas en los grupos que se comparan (sesgo de selección), la atención provista, exposición a otros factores aparte de la intervención de interés (sesgo de desempeño), retiros/pérdidas o exclusiones de personas incluidas en un estudio (sesgo de deserción) o cómo se están evaluando los desenlaces (sesgo de detección). Las revisiones sistemáticas de estudios se pueden ver afectadas también particularmente por el sesgo de reporte, donde un subconjunto sesgado de todos los datos relevantes se encuentra disponible (4).
- Significancia estadística: en investigación, la significación estadística es una medida de la probabilidad de que la hipótesis nula sea cierta en comparación

con el nivel aceptable de incertidumbre con respecto a la respuesta verdadera (24).

- **Significancia clínica:** un resultado clínicamente significativo es aquel (por ejemplo, un efecto del tratamiento) que es lo suficientemente grande como para ser de importancia práctica para los pacientes y los proveedores de atención médica. Esto no es lo mismo que estadísticamente significativo. La evaluación de la importancia clínica tiene en cuenta factores como: el tamaño del efecto del tratamiento, la gravedad de la afección que se está tratando, los efectos secundarios del tratamiento y el costo. Por ejemplo, si el efecto estimado de un tratamiento para el acné fuera pequeño, pero estadísticamente significativo, pero el tratamiento fuera muy costoso y causara que muchos de los pacientes tratados sintieran náuseas, este no sería un resultado clínicamente significativo. Mostrar que un medicamento redujo la frecuencia cardíaca en un promedio de 1 latido por minuto tampoco sería clínicamente significativo (6).
- **Solidez de la evidencia:** magnitud, precisión y reproducibilidad del efecto de la intervención (incluye la magnitud del tamaño del efecto, la amplitud del intervalo de confianza, el valor p y la exclusión de efectos sin importancia clínica). En el caso de estudios no aleatorios, pueden tenerse en cuenta factores adicionales como la verosimilitud biológica, el gradiente biológico y la temporalidad de las asociaciones (5).
- **Tamaño del efecto:** El tamaño del efecto se refiere a la magnitud de un resultado. Existen varios métodos para calcularlo. El término tamaño del efecto puede referirse a tamaños del efecto no estandarizados (p. ej., la diferencia entre las medias de los grupos, el riesgo relativo o la razón de posibilidades) o tamaños del efecto estandarizados (como "correlación" o "d de Cohen"). Un tamaño de efecto grande significa que un hallazgo de investigación tiene un significado práctico, mientras que un tamaño de efecto pequeño indica aplicaciones prácticas limitadas (25).
- **Tecnología sanitaria:** cualquier intervención que pueda ser utilizada en la promoción de la salud, prevención, diagnóstico o tratamiento de una enfermedad, rehabilitación o cuidados prolongados. Se incluyen los medicamentos, los dispositivos, los procedimientos médicos y quirúrgicos, así como los sistemas organizativos dentro de los cuales se proporciona dicha atención sanitaria (5).

- Tecnología sanitaria emergente: tecnología que aún no ha sido adoptada por el sistema de salud (los fármacos suelen encontrarse en ensayos clínicos fase II o fase III, tal vez en pre-lanzamiento; los dispositivos están a 6 meses o menos de introducirse al mercado, ya se comercializan pero tienen una difusión menor al 10% o se comercializan en unos cuantos centros), o tecnología existente a la que se le han hecho cambios de indicación o uso (5).
- Umbral de valor terapéutico: para fines de este Manual, se entenderá como la medida de significancia clínica de una intervención, considerando otros elementos del contexto además de la acostumbrada significancia estadística.
- Validez del estudio: grado en que se garantizan las inferencias extraídas del estudio cuando se tienen en cuenta los métodos del estudio, la representatividad de la muestra del mismo y la naturaleza de la población de la cual se extrae validez interna, validez externa y aplicabilidad y generalizabilidad (5).
- Validez externa: grado en que los hallazgos obtenidos de una investigación realizada en determinadas circunstancias son generalizables a otras poblaciones distintas. En la medida en que las circunstancias de una determinada investigación (p. ej. características de pacientes o la manera de administrar el tratamiento) difieran de las circunstancias de interés, la validez externa de los resultados de esa investigación puede ser cuestionada (5).
- Validez interna: grado en que los resultados de un estudio representan con exactitud la relación causal entre una intervención y un resultado en las circunstancias particulares de ese estudio. La validez interna de un ensayo puede ser dudosa cuando determinados tipos de sesgo en el diseño o la realización de un ensayo podrían haber afectado los resultados, ocultando así la verdadera dirección, magnitud o certeza del efecto del tratamiento (5).
- Valor: describe la perspectiva individual que podría ser muy diferente en dos personas ante cierta condición (10).
- Valor social: es una forma de pensar acerca de cómo se asignan y utilizan los escasos recursos en una sociedad determinada; también es una forma de mantener la responsabilidad social en un entorno amenazado por la racionalización económica. En un contexto de salud, los conceptos relacionados incluyen la prescripción social y la personalización, que

representan un alejamiento de los modelos médicos que se centran en el tratamiento hacia una situación que promueve la independencia y el bienestar a través de enfoques preventivos para la atención médica. El valor social proporciona un marco legítimo para evaluar y comprender el impacto de las intervenciones a nivel individual y comunitario (26).

- Vocabulario controlado: sistema de términos en el que intervienen, por ejemplo, definiciones, estructura jerárquica y referencias cruzadas. Su objetivo es eliminar la ambigüedad del lenguaje natural al indexar los documentos. Ejemplos de vocabularios controlados son el MeSH, utilizado en MEDLINE, y Emtree, utilizado en EMBASE (5).

Acerca de este manual

La Red Internacional de Agencias de Evaluación de Tecnología en Salud (INAHTA-*International Network of Agencies for Health Technology Assessment*) y la Sociedad para la Evaluación Internacional de Tecnología en Salud (HTAi-*Health Technology Assessment International*) definen la Evaluación de Tecnologías en Salud (ETES) como un proceso multidisciplinario (médico, social, ético y económico) que utiliza métodos explícitos para determinar el valor de una tecnología de salud en diferentes puntos de su ciclo de vida (es decir, antes de la comercialización, durante la aprobación del mercado, después de la comercialización, hasta la desinversión de una tecnología de salud). El propósito es informar la toma de decisiones para promover un sistema de salud equitativo, eficiente y de alta calidad a través de un proceso formal, sistemático y transparente utilizando métodos estandarizados para considerar la mejor evidencia disponible (5).

Las dimensiones de valor para una tecnología de salud pueden evaluarse examinando las consecuencias intencionadas y no deseadas del uso de una tecnología de salud en comparación con las alternativas existentes. El valor general puede variar según la perspectiva tomada, las partes interesadas involucradas y el contexto de decisión (5).

La evaluación de las tecnologías en salud tiene planteado nueve dominios: El problema de la salud y el actual uso de la tecnología, descripción y características técnicas, seguridad, efectividad clínica, costos y evaluación económica, análisis ético, aspectos organizacionales, aspectos sociales y aspectos legales, los cuales son importantes para discriminar los cuatro primeros dominios como una evaluación rápida, y los nueve dominios como una evaluación completa.

Teniendo en cuenta lo anterior y debido a los diferentes avances en los métodos y herramientas para la síntesis y evaluación de evidencia científica de investigación en salud y en Colombia, surge la necesidad de contar con recursos acordes a estos, por lo que se realiza la actualización de este Manual. El IETS ha considerado pertinente este documento para el desarrollo de evaluaciones de efectividad clínica, seguridad y validez diagnóstica de tecnologías en salud, tras ocho años del lanzamiento de su versión original.

El propósito de este manual es brindar a los autores de las evaluaciones de tecnología en salud los insumos en cuanto a métodos y técnicas que se emplean en este tipo de investigación integrativa, para dar respuesta al dominio de Efectividad y Seguridad. Así mismo, se pretende dar los lineamientos generales que deben ser tenidos en cuenta para la realización de las evaluaciones de tecnología en salud y que serán verificadas por parte del IETS, de acuerdo con la resolución 586 de 2021; así como para otros procesos en los cuales es un insumo para la toma de decisiones de inclusión de tecnologías al Plan de Beneficios en Salud (PBS), para el procedimiento técnico-científico y participativo para la determinación de los servicios y tecnologías que no podrán ser financiados con recursos públicos asignados a la salud y se establecen otras disposiciones exclusiones, elaboración de lineamientos y protocolos, posicionamientos terapéuticos, evaluaciones de efectividad y seguridad en el marco de presupuestos máximos y en el marco de valor terapéutico. Las recomendaciones que aquí se presentan están orientadas a que las evaluaciones de efectividad clínica, seguridad y validez diagnóstica, sean procesos rigurosos, sistemáticos y transparentes.

Las pautas metodológicas que incorpora el manual aplican a la evaluación de cualquier tipo de tecnología en salud, incluyendo medicamentos, dispositivos, pruebas diagnósticas o procedimientos médicos.

En su mayoría, los principios y métodos contenidos en el manual parten del concepto de “investigación de efectividad comparada”, definida como la generación y síntesis de evidencia que compara los beneficios y daños de alternativas para prevenir, diagnosticar, tratar y monitorizar una condición clínica, o para mejorar la prestación del cuidado en salud (27).

El presente manual no tiene como finalidad profundizar en los métodos para el desarrollo de una revisión sistemática convencional de intervenciones o pruebas diagnósticas, para lo cual se sugiere la utilización de metodologías propuestas por el *Manual Cochrane de revisiones sistemáticas de intervenciones* o el Instituto Joanna Briggs. Este manual se ubica en la práctica de quienes son responsables o están interesados en la conducción de evaluaciones de tecnologías en salud, según el modelo adaptado de la *European Network for Health Technology Assessment (EUnetHTA)*.

En cuanto a la estructura del manual, este se compone de cinco capítulos. El primero de ellos describe la fase de preparación y conformación del grupo desarrollador de la evaluación, el alcance de la evaluación, la formulación de pregunta de evaluación y aspectos a tener en cuenta para la selección del

comparado y desenlaces. El segundo y tercer capítulo brindan orientaciones para el desarrollo de los dominios de la condición de salud y la tecnología, que definen una evaluación de tecnología en salud, con base en el modelo propuesto por *EUnetHTA*: la condición de salud y uso actual de la tecnología (capítulo dos), la tecnología y sus características (capítulo tres), y el análisis de efectividad clínica y seguridad (capítulo cuatro). Un quinto capítulo, proporciona una guía para el reporte de protocolos e informes de evaluación de efectividad clínica y seguridad de tecnologías en salud.

Capítulo 1. Fase de preparación y conformación del grupo desarrollador

Para esta fase, se presentarán los aspectos relacionados con la conformación del grupo desarrollador, la determinación del alcance y el planteamiento de la pregunta de investigación.

1. Conformación del grupo desarrollador

Los miembros del grupo desarrollador deben ser seleccionados por su conocimiento y experiencia; la calidad de este grupo determinará en parte la calidad de las evaluaciones de efectividad clínica, seguridad y validez diagnóstica de tecnologías en salud y es determinante en el éxito del trabajo del equipo.

Con el fin de asegurar que todas las áreas de conocimiento pertinentes estén representadas, toda la evidencia científica sea localizada y evaluada en forma crítica, se identifique y resuelvan los problemas prácticos de la evaluación de efectividad clínica, seguridad y validez diagnóstica, el grupo desarrollador debe ser multidisciplinario e incluir:

- Equipo técnico interdisciplinario: compuesto por profesionales clínicos, epidemiólogo, químico farmacéutico, ingeniero biomédico, proveedores e investigadores, del sector público, privado o voluntario, u otros proveedores independientes de atención, cuidado, apoyo o servicios (28).
- Actores de interés: son personas que utilizan las tecnologías, familiares y cuidadores, miembros del sector público y comunitario o voluntario con experiencia relevante. Los actores de interés son quienes defienden las perspectivas de las personas que usan las tecnologías, los cuidadores o el público, por lo tanto, deben estar dispuestos a reflejar las experiencias de una amplia gama de personas afectadas y no basarse únicamente en la experiencia personal. Son quienes se afectan con la decisión de inclusión de las tecnologías (28).

Todos los miembros del grupo desarrollador, incluidos los profesionales, proveedores, actores de interés y demás, tienen el mismo estatus, reconociendo la importancia de la experiencia que cada miembro aporta al grupo.

Se recomienda que los integrantes del grupo desarrollador no pertenezcan a las organizaciones con un interés comercial directo en intervenciones o servicios, debido a posibles conflictos de intereses; para garantizar este proceso es altamente recomendable que las evaluaciones de tecnología sean realizadas por agentes expertos independientes.

Es importante tener en cuenta que el ente gestor no hace parte del grupo desarrollador, sin embargo, la definición del alcance y los objetivos se realiza en conjunto con sus representantes, quienes realizan seguimiento a todo el proceso, así como la revisión y aprobación de los productos finales.

1.1 Funciones del grupo desarrollador

El grupo desarrollador está encargado de la definición de los siguientes aspectos dentro de la evaluación de efectividad clínica, seguridad y validez diagnóstica y terapéutica de la tecnología en salud (29):

- Formulación de los alcances y objetivos (en conjunto con representantes del ente gestor).
- Definición de las preguntas PICOT (Población, Intervención, Comparador, Outcome o desenlace, Tipo de estudio).
- Definición y priorización de los desenlaces con relevancia clínica
- Búsqueda de evidencia científica, desarrollo de revisiones sistemáticas de la literatura.
- Evaluación y síntesis de la evidencia.
- Procesos de participación (expertos u otros actores).
- Formulación de las conclusiones clínicas.
- Socialización de los resultados de las evaluaciones con los diferentes interesados.

1.2 Composición del grupo desarrollador

El número de personas que conforman el grupo desarrollador es variable y dependerá de la complejidad y extensión de la tecnología a evaluar. Sus miembros constituirán el grupo central de la evaluación de tecnologías y serán los encargados de la ejecución de las labores técnicas, logísticas y de construcción de los documentos de reporte y de soporte al finalizar el proceso. Se distinguen cuatro categorías de integrantes: el líder del grupo, expertos técnicos o metodológicos, profesionales clínicos o expertos temáticos y pacientes o cuidadores.

1.1.1. Líder del grupo: funciones y roles

El líder del grupo es quien guía la evaluación de tecnologías en términos de tarea (desarrollo) y proceso (cómo funciona el grupo). Ayuda a trabajar en colaboración, asegurando una contribución equilibrada de todos los miembros del equipo (30). El líder debe ser un individuo neutral, idealmente experto en el tema, que tenga conocimientos en el desarrollo de procesos basados en la evidencia y de trabajo en equipo.

1.1.2. Expertos técnicos o metodológicos

Los expertos técnicos o metodológicos son las personas encargadas de las labores técnicas centrales del desarrollo de la evaluación de efectividad clínica, seguridad y validez diagnóstica, con el apoyo de los profesionales clínicos. Estos expertos deberán contar con formación y experiencia en el campo de la epidemiología clínica, salud pública o investigación en salud, e idealmente deben tener experiencia previa en revisiones sistemáticas de la literatura o evaluación de tecnologías (30) (Ver tabla 1).

1.1.3. Profesionales clínicos

Los profesionales clínicos o expertos temáticos son profesionales con experiencia en la práctica clínica, que incluye a médicos generales y especialistas, enfermeras, odontólogos y otros profesionales de la salud, según el área o la condición de interés de la tecnología a evaluar. Deben representar la(s) perspectiva(s) de los profesionales de la salud involucrados en la atención de pacientes afectados por la enfermedad para la cual se utiliza la tecnología a evaluar. No es necesaria una experiencia en investigación extensa, sin embargo, es deseable que tengan conocimientos en medicina basada en la evidencia. Este grupo participa en la definición y priorización de los desenlaces en salud y la definición del umbral de valor terapéutico, que se considerarán para la evaluación de efectividad, seguridad, para lo cual, no se espera que representen las opiniones de sus asociaciones o sociedades científicas o académicas (30) (Ver Tabla 1). Estos profesionales son seleccionados y delegados por las asociaciones o sociedades científicas respectivas del tema a tratar.

1.1.4. Pacientes y cuidadores

Los pacientes y cuidadores son personas con experiencia directa en la condición de interés, ya sea como paciente para la cual la tecnología a evaluar es utilizada o forma parte de la población objeto como cuidador, miembro de la familia, funcionario

o miembro de una organización de pacientes, cuidadores o grupo de apoyo (profesionales o no profesionales). Si bien los pacientes pueden participar en diferentes momentos de la ETS, es necesario que se garantice su participación en la valoración de aceptabilidad de los desenlaces y en la definición de umbral de valor terapéutico en salud, que se considerarán para la evaluación de efectividad y seguridad. Deben estar dispuestos a reflejar no solo sus propias experiencias sino también las experiencias que ellos conozcan de otros pacientes (29) (Ver Tabla 1).

Tabla 1. Funciones del grupo desarrollador

| Miembro del grupo desarrollador | Funciones |
|--|---|
| Líder del grupo | <ul style="list-style-type: none"> * Proponer los miembros del grupo desarrollador *Estructurar el grupo desarrollador *Establecer las reglas del funcionamiento del grupo. *Planear las reuniones del grupo. *Garantizar que el grupo cuente con la información y recursos suficientes. *Establecer un clima de confianza y respeto mutuo entre los miembros. *Proveer a todos los miembros la oportunidad de participar en las discusiones y actividades. *Manejar la discusión de acuerdo con la agenda. *Estimular el debate y no presionar el acuerdo. *Resumir los puntos principales y las decisiones del debate. *Hacer seguimiento a los compromisos de los miembros del GD |
| Técnicos o metodológicos | <ul style="list-style-type: none"> *Formular las preguntas PICOT. *Realizar las revisiones sistemáticas de la literatura. *Realizar la revisión de la calidad metodológica. *Evaluar la calidad de la evidencia. *Analizar la evidencia y formular las conclusiones. |
| Profesionales clínicos | <ul style="list-style-type: none"> *Contribuir constructivamente a las reuniones, tener una buena comunicación y trabajo en equipo. *Ayudar en la formulación o validar las preguntas PICOT. *Proporcionar orientación al equipo técnico que realizan las revisiones sistemáticas y el análisis económico. *Revisar, leer o evaluar los estudios, información relevante, hacer comentarios constructivos y aportes correspondientes en las reuniones. *Apoyar la formulación de las conclusiones basadas en las revisiones de evidencia o por consenso cuando la evidencia es pobre o falta (junto con los demás miembros del grupo). |
| Pacientes y cuidadores | <ul style="list-style-type: none"> *Asegurar que las preguntas de revisión abarquen tanto los problemas del paciente como los profesionales. *Considerar hasta qué punto la evidencia publicada ha medido y tomado en cuenta medidas de resultado que los pacientes consideran importantes. |

| Miembro del grupo desarrollador | Funciones |
|---------------------------------|---|
| | *Asegurar que las conclusiones aborden los problemas y preocupaciones del paciente. |

Fuente: elaboración IETS

1.3 Aspectos éticos y conflictos de interés

Es requerido que antes de iniciar sus funciones, como parte del desarrollo metodológico y con el fin de garantizar la transparencia de los procesos, todos los integrantes del grupo desarrollador sean evaluados para determinar su potencial conflicto de interés y se decida de manera definitiva sus funciones en el proyecto. Específicamente, para este proceso es necesario que cada miembro del grupo desarrollador, tanto del equipo técnico interdisciplinario como los actores de interés, diligencien el formato establecido para ello, disponible en la página web del IETS en el siguiente enlace: <http://herramientas.iets.org.co/declaracion/home.aspx>. En este enlace, además de anexar una copia de su hoja de vida actualizada o formulario de solicitud en el caso de los actores de interés, se firma el acuerdo de confidencialidad (Por ejemplo, como lo describe la Guía de involucramiento de actores del IETS (31).

Los formatos diligenciados con los intereses declarados por los participantes de cada proyecto son analizados por un comité conformado para tal fin, el cual emite una calificación sobre los potenciales intereses en conflicto identificados, estableciendo si corresponden a conflictos inexistentes, probables o confirmados directos o indirectos y de acuerdo con la calificación define el alcance de la participación de los actores en el proceso.

2. Alcance de la evaluación

La primera fase en el desarrollo de una evaluación de efectividad y seguridad es la definición del alcance o enfoque del informe, específicamente, en términos de las preguntas que se abordarán. En este punto debe aclararse la perspectiva, puesto que definirá el alcance y permitirá asegurar que la evaluación de efectividad, seguridad, preferencias y valores del paciente, junto a otra evidencia recolectada, sea consistente con la requerida para el análisis económico y otros dominios de la evaluación. Un alcance definido proporciona un enfoque, ayuda al planteamiento de las preguntas de investigación, definir la metodología para dar respuesta a las mismas y asegura que las conclusiones de la evaluación permitan resolver las necesidades o problemas que motivan a la misma.

Definir el alcance incluye especificar el propósito de la evaluación, así como las intenciones o beneficios esperados con la misma. También la perspectiva con la que las recomendaciones derivadas de la misma serán generadas, la población objetivo de las recomendaciones, el contexto donde las recomendaciones serán aplicadas, las intervenciones que serán cubiertas en la evaluación y la justificación de las que no serán cubiertas, actores y recursos clave que se van a considerar en la evaluación, así como las referencias de documentos de interés que soportan las decisiones tomadas por el grupo desarrollador en la definición del alcance (32).

3. Preguntas de evaluación

El grupo desarrollador deberá acordar el alcance de la evaluación de efectividad y seguridad en conjunto con el ente gestor; para esto planteará la pregunta de evaluación en términos de población, intervención, comparadores y desenlaces como se presenta a continuación.

3.1 Formulación de las preguntas de investigación

Las preguntas de investigación deben plantearse mediante la estructura PICOT y presentarse de forma desagregada, tal como lo indica la Tabla 2 :

Tabla 2 Pregunta de investigación en estructura PICOT

| | |
|--|---|
| Población | Los pacientes que son candidatos al uso de las intervenciones bajo estudio. Subgrupos de interés (si aplica) |
| Intervenciones | Las tecnologías en salud de interés |
| Comparación | Las alternativas disponibles para el manejo de la condición de salud, que son comparables con las tecnologías de interés, o aquellas estrategias metodológicas empleadas como grupo de control |
| Desenlaces (del inglés outcomes) | <u>Efectividad clínica</u> Desenlaces críticos e importantes para la toma de decisiones <u>Seguridad</u> Desenlaces críticos e importantes para la toma de decisiones y diferencias mínimas clínicamente importantes |
| Tipo de estudio | Los diseños de investigación seleccionados para dar respuesta a la pregunta de evaluación |
| Consideraciones acerca de los comparadores y desenlaces no incluidos en esta pregunta | |

Fuente: elaboración IETS

Deben tenerse en cuenta las siguientes consideraciones para cada uno de los elementos de la pregunta PICOT.

- **Población:** debe especificar las características clínicas y sociodemográficas de la población en la que se hace la evaluación de la ETES y a la que se aplicarán posteriormente los resultados. Por lo tanto, debe incluir la condición de salud de interés, sospecha clínica (en el caso de evaluaciones de pruebas diagnósticas), estado, duración de la enfermedad (si aplica), falla a tratamientos previos, comorbilidades, cointervenciones, edad, sexo, raza, nivel educativo, estrato socioeconómico, localización geográfica, etc. Si la evaluación se realiza para distintos subgrupos de interés, también se deben especificar en esta sección. Si no se incluye algún grupo de interés en el que se pueda aplicar la intervención (por ejemplo, población pediátrica) debe presentarse la justificación.
- **Intervención:** son las tecnologías de interés objeto de la evaluación, pueden ser medicamentos, dispositivos, procedimientos quirúrgicos, etc. Deben presentarse sus características como dosis, concentraciones o formas farmacéuticas, sobre las cuales se presentarán los resultados de la evaluación. Así mismo, debe especificarse el propósito de uso de las tecnologías: prevención, tamización, diagnóstico, tratamiento, seguimiento, rehabilitación, cuidado paliativo u organizacional, y si su ámbito de uso es hospitalario o ambulatorio.

Para el caso de los medicamentos, debe especificarse la modalidad del tratamiento (monoterapia, terapia multimodal, tratamiento coadyuvante, terapia secuencial, etc.) junto con la línea terapéutica de interés, según sea el caso. En el escenario de tecnologías diagnósticas, debe precisarse si la prueba es única, en paralelo o en serie.

Además, debe justificarse el análisis de aquellas tecnologías que son imprescindibles para el uso de las tecnologías de interés, es decir, que son complementarias.

- **Comparadores:** son las intervenciones que tienen la misma indicación de las tecnologías de interés frente a las cuales se desea conocer, de modo comparativo, los beneficios, daños o utilidad. En todo caso los comparadores deben ser las tecnologías de referencia en la práctica clínica sugerida o GPC, manuales, protocolos o lineamientos disponibles en el país o a nivel internacional. En caso de que no se haya determinado un comparador de referencia o no se disponga de un comparador activo, es decir distinto al placebo, se deberá comparar con el estándar de cuidado actual.

En el caso de las evaluaciones de validez diagnóstica, deben seleccionarse, siempre que sea posible, patrones de referencia fuertes (por ejemplo, biopsia,

autopsia, cirugía, etc.). Si el patrón de referencia seleccionado no se ha validado aún como patrón de oro de la condición de salud de interés, deben presentarse los supuestos para su uso.

Para los comparadores deben presentarse las características de la tecnología como dosis, concentraciones o formas farmacéuticas o en caso de ser un procedimiento se debe realizar su descripción. Tanto los medicamentos como los procedimientos deben estar autorizados para Colombia. Tener en cuenta, además, las especificaciones del apartado *Selección del comparador*.

- **Desenlaces:** son los resultados en salud atribuibles al uso de las tecnologías.

Se deben seleccionar desenlaces tanto de efectividad como de seguridad: críticos e importantes, junto con su tiempo de ocurrencia. Los desenlaces deben ser relevantes para los pacientes y para los servicios y sistemas de salud, preferiblemente desenlaces finales, sin considerar si han sido o no reportados en estudios previos y definirse como variables, sin especificar si el desenlace se favorece o no por el efecto de la intervención (por ejemplo: mortalidad y no *aumento o reducción* en la mortalidad).

En el caso de evaluaciones de validez diagnóstica debe incluirse el desenlace de presencia o ausencia de la condición de salud de interés, así como la sensibilidad, efectividad y otras medidas del rendimiento operativo como *Likelihood Ratio* (LR), exactitud diagnóstica, valores predictivos y desenlaces de utilidad clínica como oportunidad en el inicio de terapia y desenlaces clínicos; adicionalmente, desenlaces de seguridad, cuando se conozca que el uso de la tecnología puede estar asociado a eventos adversos serios y frecuentes.

En el apartado *Selección y calificación de la importancia relativa de los desenlaces en salud*, se describen de manera detallada las recomendaciones para la selección de desenlaces, los casos en los que se pueden usar desenlaces subrogados y compuestos y la forma para hacer su gradación.

- **Tipo de estudios:** son los diseños epidemiológicos seleccionados para dar respuesta a la pregunta de investigación.

Se recomienda como primera instancia el uso de revisiones sistemáticas convencionales, preferiblemente de ensayos controlados aleatorizados; revisiones sistemáticas panorámicas (*overviews*, *umbrella* o globales); revisiones sistemáticas con metanálisis de comparaciones indirectas o

metanálisis en red, especialmente si no existe evidencia sobre la efectividad y seguridad de la tecnología frente a los comparadores propuestos en RSL de comparaciones directas o en estudios primarios cabeza a cabeza.

Si no se cuenta con revisiones sistemáticas, se pueden considerar estudios epidemiológicos individuales con los siguientes diseños: ensayos controlados aleatorizados (idealmente de fase III, con un enfoque pragmático “análisis por intención de tratar”), cohortes analíticas (preferiblemente prospectivas), casos-controles y aquellos estudios observacionales comparativos que estén bajo la denominación de “evidencia del mundo real”. Este aspecto será abordado con mayor detalle en el Capítulo 4, 1.1. Criterios de inclusión.

Para la formulación preliminar de las preguntas de evaluación que involucren tecnologías (intervención a evaluar), por ejemplo, en el caso de medicamentos, deben seguirse los siguientes pasos:

- a. Tener autorización de la autoridad competente o registro sanitario por lo que se debe consultar los registros sanitarios vigentes de la página del INVIMA para las tecnologías e indicaciones objeto de evaluación, registrando la fecha de corte en la que se realizó la consulta. No deben incluirse en la evaluación tecnologías sin el registro sanitario para la indicación de interés, a menos que forme parte del listado UNIRS (Usos No Incluidos en el Registro Sanitario), en estos casos específicos, se deberá consultar el listado actualizado del Ministerio de Salud y Protección Social. Así mismo, en caso de no identificar registros sanitarios en la base del INVIMA ni en el listado UNIRS, debe revisarse el listado de medicamentos vitales no disponibles de dicha entidad.
- b. Identificar los posibles comparadores mediante la revisión del Sistema de Clasificación Anatómica, Terapéutica, Química (ATC - del inglés *Anatomical, Therapeutic, Chemical Classification System*).
- c. Delimitar la población y los comparadores a través de la revisión de recomendaciones de guías de práctica clínica.
- d. Verificar que las tecnologías definidas como comparadores estén disponibles en el país sin tener en cuenta su estado de financiación con recursos públicos.

Ampliar la información de lo descrito anteriormente, con lo descrito en la sección *Selección del comparador*.

3.2 Refinamiento de las preguntas de evaluación

La pregunta de evaluación debe refinarse con el grupo desarrollador y mediante el involucramiento de actores relevantes como expertos técnicos o metodológicos, pacientes y cuidadores, profesionales clínicos, entre otros. La pregunta con estructura PICOT preliminar debe publicarse en la página web del IETS para comentarios (si aplica), y tras los ajustes que tengan lugar, debe procederse a su publicación como pregunta definitiva (33).

4. Selección del comparador

4.1 Definición

El comparador o comparadores en un estudio de evaluación de efectividad y seguridad son las intervenciones que tienen la misma indicación de la tecnología de interés a evaluar, frente a las cuales se desea conocer, de modo comparativo, los beneficios, daños o utilidad diagnóstica. Idealmente, los comparadores deben ser las tecnologías de referencia en la práctica clínica (estándares de atención) (34, 35).

4.2 Características de un comparador

Para definir el comparador más adecuado, se debe asegurar que cuente con las siguientes características (4):

- Es la terapia estándar o la(s) tecnología(s) más utilizada(s) en la práctica clínica o es la tecnología alternativa más relevante en términos de efectividad y seguridad o de exactitud diagnóstica en caso de las intervenciones de pruebas diagnósticas.
- Es actualmente aceptada en el ámbito clínico.
- Tiene la misma indicación que la intervención de interés.
- Tener autorización de la autoridad competente o registro sanitario.
- La inclusión de comparadores sin autorización por la autoridad competente en el país se admite exclusivamente en los casos en los que se usen como intervenciones puente para construir metaanálisis de comparaciones indirectas o metanálisis en red, en tales casos, solo servirán para poder evaluar las comparaciones de interés en la evaluación de tecnología y no se incluirán recomendaciones de la evaluación de efectividad y seguridad sobre las intervenciones no aprobadas.

En caso de que no se haya determinado un comparador de referencia o no se disponga de un comparador activo, es decir distinto al placebo, se deberá comparar con el cuidado estándar actual (35).

4.3 Selección de comparadores

Los juicios realizados para la selección del (los) comparador(es) deben ser incorporados de manera detallada en el informe de la evaluación de efectividad y seguridad. Esta sección también debe presentar toda la información sobre los atributos de las tecnologías seleccionadas como comparadores: tipo de dispositivo, medicamentos, técnica, procedimiento o terapia, mecanismo de acción, evidencia de efectividad y seguridad y también describir cómo la tecnología difiere de sus predecesores, y las diversas modificaciones actuales, especialmente si las diferencias afectan el rendimiento (36).

En caso de que el comparador sea un medicamento, se debe considerar la elección de la dosis adecuada. Esta debe ser una dosis que haya mostrado la mejor relación riesgo/beneficio y un efecto clínicamente relevante en la población de interés o haya sido establecida de acuerdo con la información contenida en la autorización de comercialización o guías de práctica clínica (36).

En el caso de las evaluaciones de exactitud diagnóstica deben seleccionarse, siempre que sea posible, estándares de referencia fuertes (por ejemplo, biopsia, autopsia, cirugía, etc.). Si el patrón de referencia seleccionado no se ha validado aún como patrón de oro de la condición de salud de interés, deben presentarse los supuestos para su uso.

Los medicamentos utilizados fuera de la indicación y uso aprobados “*off label*” solo pueden considerarse como tratamiento comparativo si su uso para la indicación en cuestión está suficientemente justificado por la investigación clínica, o descrito en guías de práctica clínica (GPC) o protocolos basados en la evidencia, también su uso debe ser aceptado como el tratamiento habitual en la práctica diaria (4).

El cuidado estándar puede ser considerado un comparador, siempre y cuando no exista una alternativa disponible con indicación similar a la intervención evaluada y que constituya el estándar de referencia en el país.

4.4 Procedimiento para la selección de comparadores

Para la selección de los comparadores se debe establecer un proceso deliberativo entre expertos clínicos y metodólogos, que propongan de manera preliminar los posibles comparadores y validarlos por el ente gestor con el fin de que cubran la necesidad de la evaluación. A partir de esta selección se sugiere seguir el siguiente proceso:

- a) Identificar (para el caso de evaluaciones de medicamentos) los posibles comparadores mediante la revisión de grupos farmacológicos (clasificación ATC), con base en la información del Centro Colaborador de la OMS para la metodología de estadísticas de medicamentos (http://www.whocc.no/atc_ddd_index/).
- b) Consultar los registros sanitarios vigentes y en trámite de renovación de la página del INVIMA para las tecnologías e indicación de interés, registrando la fecha de corte para la consulta. No deben incluirse en la evaluación tecnologías sin el registro sanitario para la indicación de interés, a menos que forme parte del listado UNIRS (Usos No Incluidos en el Registro Sanitario), en estos casos específicos, se deberá consultar el listado actualizado del Ministerio de Salud y Protección Social. Así mismo, en caso de no identificar registros sanitarios en la base del INVIMA ni en el listado UNIRS, debe revisarse el listado de medicamentos vitales no disponibles de dicha entidad.
- c) Consultar las recomendaciones de guías de práctica clínica recientes sobre la condición de interés. Se sugiere consultar las siguientes fuentes:
 - National Institute for Health and Care Excellence (NICE)
<https://www.nice.org.uk/>
 - American College of Physicians (ACP)
<https://www.acponline.org/>
 - National Guideline Clearinghouse (NGC)
<http://www.guideline.gov/index.aspx>
 - Guidelines International Network (GIN)
<http://www.g-i-n.net/gin>
 - New Zealand Guidelines Group (NZGG)
<http://www.health.govt.nz/>
 - Scottish Intercollegiate Guidelines Network (SIGN)
<http://www.sign.ac.uk/index.html>
 - Organización Mundial de la Salud (OMS)
<https://www.who.int/publications/guidelines/atoz/en/>
 - Ministerio de Salud y Protección Social - IETS

<http://www.iets.org.co/>

- GuíaSalud
<http://portal.guiasalud.es/web/guest/home>
 - Centro Nacional de Excelencia Tecnológica en Salud (CENETEC)
<http://www.cenetec.salud.gob.mx/>
- d) Consultar revisiones Cochrane, estudios primarios, bases de datos como UpToDate (<https://www.uptodate.com/home>).

4.5 Comparadores por subpoblaciones de pacientes

Si una tecnología tiene una amplia indicación terapéutica, que cubre el tratamiento de pacientes en diferentes líneas de tratamiento (por ejemplo, primera línea o segunda línea), diferentes grados de gravedad, diferentes etapas de una enfermedad (por ejemplo, cáncer temprano o cáncer metastásico) o características genéticas, es importante considerar que el comparador puede ser diferente para las subpoblaciones. Otras características clínicas específicas de los grupos de pacientes, como las comorbilidades también pueden ser importantes. En estos casos, puede ser necesario definir los comparadores por separado para cada subgrupo relevante dependiendo de la evidencia disponible (36).

5. Selección y calificación de la importancia relativa de los desenlaces en salud

5.1 Definición

Un desenlace clínico puede definirse como una forma de medir el impacto que un tratamiento tiene en la funcionalidad, supervivencia o en cómo se sienten los pacientes (37); el impacto puede reflejarse tanto en la mejoría como en el empeoramiento del estado de salud; así, un desenlace clínico puede ser un evento clínico como la mortalidad, una medición del estatus clínico o la calidad de vida relacionada con la salud (38, 39). También puede definirse el desenlace como un componente del estado clínico o funcional de un participante después de que se ha

aplicado una intervención, que se utiliza para evaluar la efectividad de una intervención y que permite definir la medida del efecto de esa intervención (4).

Los desenlaces clínicos deben reunir ciertas características: ser relevantes para la práctica clínica, medirse de manera válida y confiable, ser sensibles al cambio (por ejemplo, ser capaces de detectar cambios debidos al tratamiento o en el tiempo) y ser reconocidos por los médicos tratantes y pacientes (38).

5.2 Criterios para la selección de desenlaces

Para realizar una evaluación de tecnologías en salud, se deben incluir desenlaces tanto de efectividad (efectos deseados) como de seguridad (efectos indeseados). Teniendo en cuenta que puede existir una variabilidad importante en el número, tipo, forma, tiempo de medición e importancia de los desenlaces en los distintos estudios que evalúan las ETES, es necesario seleccionar y hacer una gradación o clasificación de la importancia de esos desenlaces. Para esto se presentan las siguientes recomendaciones:

- Seleccionar como primera opción desenlaces finales, que sean relevantes tanto para los pacientes como para los servicios y sistemas de salud; estos desenlaces están relacionados con el objetivo terapéutico final de la intervención, no solo con los resultados clínicos, por lo cual tienen más relevancia para los pacientes y para la priorización (38, 40). Dependiendo de la condición de salud, puede cambiar la elección de los desenlaces: en enfermedades que amenazan la vida los desenlaces deben ser finales como la mortalidad mientras que en condiciones que no la amenazan se pueden considerar desenlaces de morbilidad o funcionamiento (38).
- El tiempo de medición de la ocurrencia del desenlace debe ser el necesario para que pueda ocurrir el desenlace final. Sin embargo, en condiciones agudas sin consecuencias a largo plazo, pueden emplearse desenlaces evaluados a corto plazo; por ejemplo, si se evalúa la efectividad de una tecnología para la prevención de enfermedad diarreica aguda, los desenlaces de efectividad evaluados son la incidencia de episodios de diarrea y la incidencia de diarrea persistente. En cualquier caso, el tiempo de seguimiento debe declararse al elaborar y refinar la pregunta PICOT; si los estudios seleccionados, luego de realizar la búsqueda de la literatura, emplean menores tiempos de seguimiento, debe presentarse una discusión sobre cómo se afecta la estimación del efecto.

- Si se hace análisis de supervivencia, la supervivencia global debe ser el desenlace crítico.
- Si no se dispone de desenlaces finales, por ejemplo, cuando la tecnología en salud que se está evaluando es relativamente reciente, de manera que el tiempo de seguimiento de los estudios no es el suficiente, los eventos finales son poco frecuentes porque la población es pequeña o la enfermedad es rara o progresa de manera muy lenta, puede realizarse la evaluación a partir de desenlaces intermedios y subrogados (40). En este caso, se debe indicar y discutir hasta qué punto se puede sesgar el efecto de la intervención al evaluarse por medio de desenlaces intermedios.
- Cuando se utilicen desenlaces subrogados o intermedios ¹ se debe especificar el desenlace crítico y el correspondiente desenlace subrogado. Para el uso de los desenlaces subrogados se debe partir del concepto que, estos resultados no son importantes para el paciente, pero pueden estar correlacionados con los desenlaces importantes para los pacientes, por ejemplo, la hemoglobina glucosilada (HbA1c) que es una medida de control glucémico como subrogado de las complicaciones a largo plazo en pacientes con diabetes (40). De emplearse un desenlace subrogado en la revisión, este deberá estar adecuadamente validado, lo cual implica que se pueda establecer una relación entre el desenlace subrogado y el desenlace final con base en la plausibilidad biológica² y la evidencia empírica.
- No se deben usar desenlaces compuestos si existe algún desenlace singular que sea confiable para realizar la evaluación; si no existe o si es más confiable el desenlace compuesto, se puede incluir en la evaluación; en este caso, cada uno de los componentes del desenlace compuesto deben ser clínicamente relevantes para la condición de salud y se deben reportar también de manera individual cada uno de los componentes (39, 41).
- Si se incluyen desenlaces clínicos compuestos, los eventos que los constituyen deben cumplir las características generales de todo desenlace (mencionados anteriormente); su cambio debe ser homogéneo, es decir debe

¹ El desenlace intermedio es un desenlace asociado con el uso de la intervención tal como una medición de la función o síntomas, que está correlacionado con cambios en el desenlace final y cuya mejoría puede ser bien percibida por el paciente, aunque no lleve a una mejoría en la morbilidad o mortalidad. El desenlace subrogado se propone para substituir al desenlace clínico de interés y predice el beneficio clínico a partir de evidencia epidemiológica, patofisiológica, terapéutica u de otros tipos. Ejemplo de intermedio es la supervivencia libre de progresión; ejemplo del segundo es la presión sistólica. Un desenlace subrogado puede considerarse válido si es sensible, medible, interpretable y altamente preciso al predecir el desenlace clínicamente relevante.

² Debe existir evidencia que indique que el desenlace debe estar en la senda causal al desenlace clínico, patogénesis de la enfermedad y mecanismo de acción de la patología.

existir consistencia en el efecto a través de los distintos eventos al desagregarlos; su agrupación debe tener sentido desde el punto de vista clínico y biológico (38). Se debe precisar en el caso en que los eventos pueden ser repetitivos en el tiempo si se registra solo la primera ocurrencia o todas las posibles ocurrencias durante el tiempo de observación del estudio (41).

- Si se realiza un metanálisis a partir de desenlaces compuestos, la definición del desenlace debe ser similar a través de los distintos estudios, es decir, debe incluir los mismos componentes; si varían, se puede realizar el metanálisis a partir de los componentes individuales, pero debe considerarse el efecto del tamaño de la muestra en las estimaciones del efecto (41).
- Los desenlaces compuestos pueden ser usados para evaluar también la seguridad de una tecnología (41).
- En el reporte de resultados de desenlaces compuestos se debe presentar el resultado para cada componente, independientemente de si alcanzó o no significancia estadística. Si alguno muestra un efecto negativo del tratamiento (el cual puede ocultarse por un beneficio mayor en los demás componentes) no se debe concluir que el tratamiento tiene un efecto clínicamente relevante en el desenlace compuesto como un todo (41); esto debe considerarse en la discusión de los resultados.
- Los denominados Resultados Percibidos por los Pacientes (*Patient Reported Outcomes* en la bibliografía anglosajona) como, por ejemplo, la calidad de vida relacionada con la salud, síntomas de la enfermedad o satisfacción con el tratamiento, pueden incluirse en la evaluación; en este caso debe indicarse cuál fue el instrumento con el que fueron medidos (refiriendo si es específico para la condición de salud e incluyendo la bibliografía que sustenta su confiabilidad y validez). Además, debe establecerse de manera a priori cómo se define un cambio clínicamente importante, es decir a partir de qué valor se define si hay si hay anormalidad (o presencia de una condición) y cuál cambio (cuántos puntos de incremento o disminución) debe considerarse para determinar diferencias clínicamente significativas entre grupos (intervención versus comparador) o entre los momentos antes y después de la intervención.
- Además de desenlaces de efectividad la evaluación de tecnologías en salud debe incluir desenlaces de seguridad, como efectos o eventos adversos, para determinar el balance beneficio-riesgo; su inclusión asegura la ponderación

informada, relacionada con la población pero también individual de beneficio y daño (39).

- Se debe especificar el grado de severidad y seriedad de los eventos y efectos adversos (42), que se incluirán como desenlaces en la evaluación de las tecnologías de salud (por ejemplo, como desenlaces de seguridad pueden incluirse eventos adversos serios o eventos adversos grado 3 y 4).
- Dado que los experimentos clínicos fueron diseñados utilizando como desenlace principal alguna medición de efectividad, el tamaño de muestra y el tiempo de seguimiento para la evaluación de eventos adversos puede llevar a una subestimación de estos, la búsqueda de desenlaces de seguridad debe complementarse con otras fuentes, tales como estudios de cohortes, datos de farmacovigilancia, etc.
- Si algún desenlace dicotómico fue obtenido a partir de un puntaje o resultado cuantitativo, el punto de corte empleado, para dicotomizarlo, debe ser definido de manera a priori y debe justificarse de acuerdo con la literatura.
- El listado de los desenlaces deberá construirse independientemente de la dirección, la significancia estadística y clínica y la calidad de la evidencia disponible. No deberán elegirse sobre la base de la magnitud del efecto anticipada u observada o porque es probable que hayan sido incluidos en los estudios que serán revisados (43).

5.3 Gradación de los desenlaces

- Para hacer la gradación o clasificación de la importancia de los desenlaces se recomienda seguir la metodología propuesta por el *GRADE Working Group* (desenlaces críticos, importantes y no importantes). La importancia relativa otorgada a los desenlaces en salud deberá reflejar la perspectiva de los afectados por la futura decisión, que informará la revisión: la de los pacientes y la del sistema de salud. Lo anterior implica que, para hacer esta clasificación, deberá consultarse la opinión de los expertos técnicos o metodológicos, profesionales clínicos, pacientes y cuidadores, con el propósito de capturar las preferencias de dichos actores.
- En las enfermedades que amenazan la vida, los desenlaces críticos serán la mortalidad o la sobrevida a un tiempo determinado, más los eventos adversos serios, otros desenlaces asociados con la morbilidad y la calidad

de vida relacionada con la salud pueden ser desenlaces importantes. La gradación de estos como críticos debe ser justificada en el informe.

- Para hacer la clasificación de la importancia de los desenlaces se deberá solicitar a los expertos técnicos o metodológicos, profesionales clínicos, pacientes y cuidadores que valoren de forma individual la importancia relativa de cada uno de los desenlaces identificados, tanto de efectividad como de seguridad, en una escala de 9 puntos:

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|
| De importancia limitada para la toma de decisiones (no incluidos en el perfil de evidencia) | | | Importantes, mas no críticos para la toma de decisiones (incluidos en el perfil de evidencia) | | | Críticos para la toma de decisiones (incluidos en el perfil de evidencia) | | |

- Una vez realizada la calificación individual se calculará la mediana; si está entre 7 y 9 se clasificará el desenlace como crítico, si está entre 4 y 6 se clasificará como importante, pero si está entre 1 y 3 se clasificará como no importante. Se debe reportar tanto la gradación, como el puntaje (mediana) obtenido para cada desenlace (por ejemplo, la supervivencia libre de progresión fue clasificada como un desenlace importante; la mediana fue 6; las votaciones individuales del grupo desarrollador fueron 8, 7, 5, 4, 6, 5, 6).
- La gradación de los desenlaces debe reflejar la importancia que tiene cada uno para medir el efecto de la intervención; por ejemplo, la seguridad, medida como presencia de eventos adversos grados 3 o 4, puede ser más importante que el grado de funcionalidad alcanzado, medido por medio de una escala, para el grupo desarrollador de la evaluación.
- Al reportar los perfiles de evidencia, se deben incluir en las tablas de evidencia, todos los desenlaces críticos o importantes, aunque estos no hayan sido evaluados en los estudios incluidos en la revisión. Este es un resultado importante para el tomador de decisiones.

5.4 Consideraciones adicionales:

- La identificación de desenlaces puede realizarse a partir de la revisión de la literatura, consulta a expertos o consulta en las siguientes fuentes (43):

- La batería de desenlaces en salud propuesta por el *International Consortium for Health Outcomes Measurement* (ICHOM), disponible de forma gratuita en <http://www.ichom.org/medical-conditions/>
 - La base de datos de la iniciativa COMET (*Core Outcome Measures in Effectiveness Trials*), disponible de forma gratuita en: <http://www.comet-initiative.org>
- El grupo desarrollador de la revisión sistemática deberá considerar la posibilidad que la importancia de un desenlace (p.ej., un efecto adverso serio) podrá hacerse evidente solo después de que se revise la evidencia o se lleve a cabo el análisis. En estos casos, se debe explicitar en el apartado de diferencias entre el protocolo y la revisión, la inclusión de estos desenlaces y su importancia relativa.
- Los desenlaces considerados críticos para la toma de decisiones (calificados de 7 a 9) constituirán la evidencia determinante en la conclusión acerca de los efectos de la tecnología, en la recomendación y en la calidad global de la evidencia. Solo en aquellos casos donde no exista información para los desenlaces críticos, se emplearán los desenlaces importantes, pero no críticos. Sin embargo, en el contexto de evaluaciones realizadas por terceros, por ejemplo en el proceso de inclusión de nuevas tecnologías para su financiamiento en presupuestos máximos, el grupo desarrollador de la evaluación debe presentar los resultados de desenlaces tanto críticos como no críticos; de esta manera se garantiza que la evaluación disponga de toda la evidencia requerida, en el caso en que la validación que hace el IETS determine una clasificación diferente para alguno de los desenlaces.
- En la revisión de evaluaciones de tecnologías llevadas a cabo por terceros se debe realizar una validación de los desenlaces. Esta consiste en la verificación de los posibles desenlaces que permitan analizar la efectividad y seguridad de la tecnología, así como su gradación, para lo cual se consultarán las fuentes mencionadas por el elaborador de la evaluación, además de guías de práctica clínica y una consulta independiente a expertos. De manera independiente se realizará la gradación de los desenlaces por un grupo de expertos temáticos. Como pueden existir discrepancias entre las dos graduaciones (la proporcionada por el desarrollador y la realizada por el grupo evaluador) se requiere que la evaluación de efectividad y seguridad incluya la evidencia correspondiente tanto a desenlaces críticos además de los importantes; si bien algunos procesos, como el de exclusión de tecnologías de la financiación de recursos públicos asignados a la salud, se basa en desenlaces críticos, reportar también la evidencia correspondiente a desenlaces importantes, permitirá al

IETS contar con toda la información en caso de que alguno sea clasificado en una categoría diferente al realizar la validación de los resultados. Sin embargo, si se requiere, se puede realizar una comunicación entre los grupos para resolver posibles discrepancias en la determinación de la importancia de los desenlaces.

6. Protocolo de la revisión

Las Revisiones Sistemáticas de Literatura (RSL) proporcionan en general, la mejor evidencia para responder preguntas de investigación por su metodología rigurosa y preespecificada. Por tanto, es necesario preparar antes de comenzar la RSL un protocolo que describa el método del desarrollo de la revisión, el cual debe estar disponible para todos los actores interesados (44).

El protocolo de la revisión tendrá que ser remitido al correo tecnica@iets.org.co para su publicación correspondiente en la página web del IETS: <http://www.iets.org.co/>

El contenido del protocolo de la revisión se presenta a continuación:

- Título de la evaluación
- Grupo desarrollador (autores, afiliación y función dentro de la evaluación)
- Entidad que solicita el estudio
- Fuentes de financiación
- Declaración de conflictos de interés
- Declaración de independencia editorial
- Derechos de autor
- Citación
- Correspondencia
- Tabla de contenido
- Listado de siglas y abreviaturas
- Objetivo y alcance de la evaluación
- Descripción de la condición
- Descripción de la intervención
- Pregunta de investigación en estructura PICOT
- Criterios de elegibilidad de los estudios
- Metodología de búsqueda (bases de datos, términos, estrategias, límites)
- Metodología para la síntesis de la evidencia
- Metodología para la evaluación de la calidad
- Referencias bibliográficas

Capítulo 2. La condición de salud y uso actual de la tecnología

La descripción de la condición de salud de interés debe contener los siguientes aspectos: definición, epidemiología (población que afecta, distribución y frecuencia), factores de riesgo, historia natural, fisiopatología, sintomatología, diagnóstico, tratamiento, curso clínico, pronóstico y carga de enfermedad. Para la caracterización epidemiológica de la condición de salud, debe documentarse su prevalencia o incidencia a partir de publicaciones y fuentes de información preferiblemente nacionales como estudios de base poblacional, encuestas nacionales de salud, Análisis de Situación en Salud (ASIS), estimaciones de carga de enfermedad o cifras e indicadores de salud. El reporte debe mencionar la ausencia de información, cuando no sea posible identificar datos específicos de frecuencia nacional, regional o local para Colombia de la condición de salud de interés.

Así mismo, es fundamental describir claramente lo relacionado al uso actual de la tecnología de interés para la condición de salud objeto, teniendo en cuenta los aspectos y características propias del contexto colombiano.

Se recomienda realizar la consulta de las siguientes fuentes, según aplique, de acuerdo con la condición de salud de interés:

- Sistema Integral de Información de la Protección Social (SISPRO): Análisis de Situación de Salud - ASIS <http://rssvr2.sispro.gov.co/reportesAsis/>, <http://rssvr2.sispro.gov.co/reportesasis2/>
- Sistema de Vigilancia en Salud Pública (SIVIGILA)
- Observatorio Nacional de Salud (ONS)
- Otros sistemas de información que administre el INS
- Cuenta de Alto Costo
- Encuesta Nacional de Demografía y Salud – ENDS
- Institute for Health Metrics and Evaluation - IHME (University of Washington)
- GLOBOCAN (The International Agency for Research on Cancer - IARC)
- Orphanet

Tener en cuenta las preguntas descritas en la sección cuatro de la herramienta “Estándares de reporte IETS para evaluaciones rápidas de tecnologías en salud” (Ver capítulo 5).

Capítulo 3. La tecnología y sus características

El término Tecnologías hace referencia a medicamentos, procedimientos, o dispositivos médicos. Mediante la descripción de la tecnología a evaluar, se busca dar un contexto general acerca de las características principales de esta, su utilización, indicación y nivel de atención donde se emplea, así como el estado de comercialización en el país. Según corresponda, deben incluirse los siguientes elementos:

- Medicamentos: nombre de la tecnología (en Denominación Común Internacional), y descripción del código ATC , grupo terapéutico al que pertenece, código CUM, estado de registro sanitario, información sobre el titular de registro sanitario, fase de desarrollo en la que encuentra, presentación, nombre comercial del medicamento, mecanismo de acción e indicación, forma farmacéutica, concentración, condiciones especiales de administración, dosis, contraindicaciones, precauciones, interacciones, reacciones adversas y alertas sanitarias, contexto y nivel de atención donde se usa el medicamento, equipos y suministros para su administración. Adicionalmente se elabora un consolidado de la información de los registros sanitarios en el país que se reporta como anexo.
- Dispositivos médicos, pruebas diagnósticas y reactivos de diagnóstico *in vitro*: nombre de las tecnologías, descripción de la clasificación según riesgo de acuerdo a registro sanitario, clasificación por familia de DM (no se cuenta con estándar semántico para DM), estado de registro sanitario, información sobre el titular de registro sanitario, fase de desarrollo de la tecnología, presentación, nombre comercial, indicaciones, condiciones especiales y precauciones de uso, , incidentes y eventos adversos, alertas sanitarias, contexto y nivel de atención donde se emplea la tecnología, equipos, suministros y capacidad del talento humano necesario para utilizar la tecnología, así como las necesidades de la infraestructura
- Procedimientos: nombre del procedimiento, el código CUPS, (no aplica información de registro sanitario ni fase de desarrollo), indicación, proceso de aplicación, contraindicaciones, precauciones y riesgos asociados al procedimiento, contexto y nivel de atención donde se realiza el procedimiento, equipos, suministros e infraestructura necesaria para su realización.

Así mismo, para todas las tecnologías, deben señalarse las fuentes de financiación (UPC, presupuesto máximo, fondos o programas especiales, recobros o gastos de bolsillo), para dispositivos médicos, se debe tener en cuenta el procedimiento al que se encuentra asociado.

En relación con este dominio de la evaluación, el grupo desarrollador deberá dar respuesta a cada una de las preguntas descritas en la sección cinco y anexos de la herramienta “Estándares de reporte IETS para evaluaciones rápidas de tecnologías en salud” (Ver Capítulo 5, sección 5).

Para abordar las preguntas propuestas en la herramienta, se recomienda realizar la consulta de las siguientes fuentes:

- Sistema de trámites en línea - Instituto Nacional de Vigilancia de Medicamentos y Alimentos (Invima)
- WHO - Uppsala Monitoring Centre
- Food and Drug Administration - FDA
- European Medicines Agency - EMA
- Agencia Española de Medicamentos y Productos Sanitarios - AEMPS
- Medicines and Healthcare Products Regulatory Agency - MHRA
- Agencia Nacional de Vigilancia Sanitaria de Brasil – ANVISA
- Agencia Salud de Canadá “HC-SC”
- Agencia Sanitaria de Australia “TGA”
- Agencia Nacional de seguridad de medicamentos y productos de salud “ANSM” de Francia
- Medicamentos a un clic – Ministerio de Salud y Protección social
- Bases de datos especializadas como Up to date, Lexicomp, entre otras.
- Bases de datos especializadas nacionales e internacionales de farmacovigilancia y tecnovigilancia.
- Otras agencias sanitarias del mundo, dependiendo de la tecnología.
- Guías de práctica clínica nacionales o internacionales.
- Protocolos o rutas de atención en salud.
- Fichas técnicas y manuales de dispositivos médicos y reactivos
- Buscadores libres.
- Resolución vigente "Tecnologías y Servicios financiación con recursos de la UPC".
- Resolución vigente que establece la Clasificación Única de Procedimientos en Salud CUPS.

Capítulo 4. Análisis de efectividad clínica y seguridad

El análisis de efectividad y seguridad parte de una revisión sistemática de la literatura de intervenciones. Una vez definida la pregunta de investigación, se iniciará la búsqueda en bases de datos especializadas utilizando vocabularios controlados y términos libres. Se considera que una buena búsqueda debe incluir al menos tres de las principales bases de datos MEDLINE, EMBASE, y la librería COCHRANE; otras fuentes pueden ser consultadas de acuerdo con el diseño de estudio y características específicas de las tecnologías.

Una vez realizada la búsqueda, se debe realizar la selección de los estudios, iniciando con la tamización de los títulos y resúmenes que cumplan con los criterios de selección preespecificados. Posterior a esta fase, se completará la selección por texto completo. Este proceso debe ser llevado en duplicado, para minimizar el riesgo de selección de los artículos para la revisión, cualquier desacuerdo deberá ser resuelto por consenso o con ayuda de un tercer revisor. Todos los estudios excluidos en la fase de revisión en texto completo deberán ser presentados con la respectiva justificación para la exclusión. Los resultados de este proceso, deben ser presentados utilizando el diagrama de flujo propuesto en la declaración *PRISMA* (45).

Una vez se complete la selección, los estudios deberán ser evaluados en su calidad metodológica, acudiendo a las herramientas validadas para este fin de acuerdo con el diseño de estudio: para ensayos clínicos controlados se utilizará la herramienta de riesgo de sesgos de la colaboración Cochrane en su segunda versión (ROB 2) (46), ROBIS para revisiones sistemáticas, QUADAS 2 para estudios de precisión diagnóstica, entre otros. Es indispensable que los resultados de este proceso se presenten como anexos del documento del informe de la evaluación de tecnología.

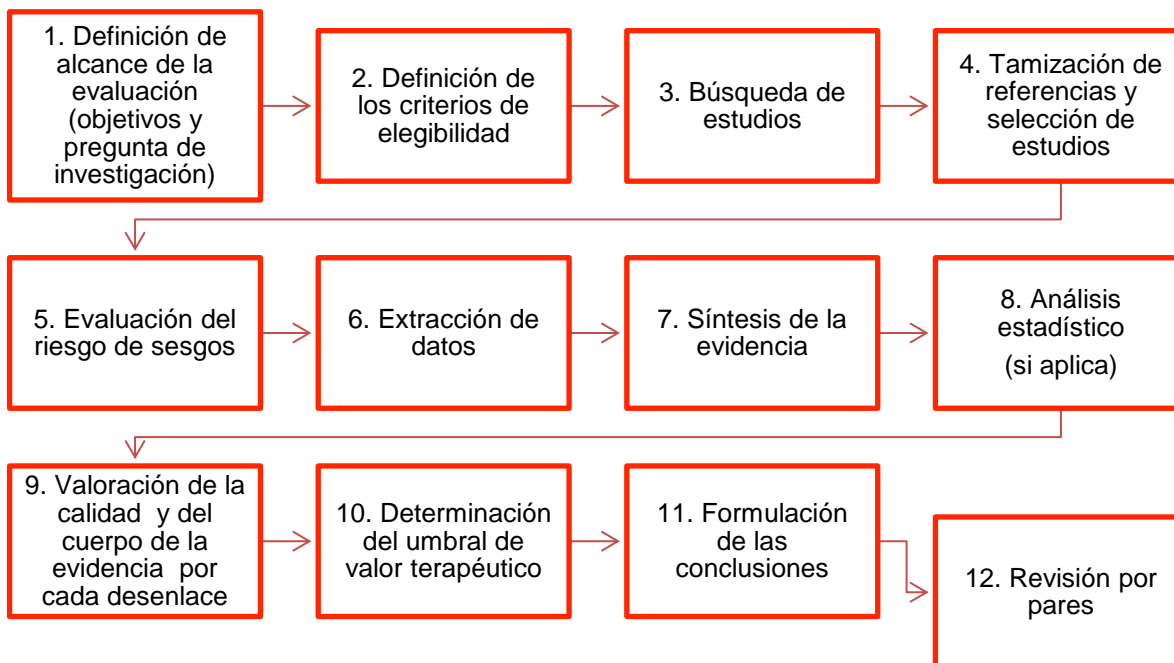
El proceso de extracción de información debe incluir de manera general los datos de identificación de los estudios, las características de la población, intervención, comparador, desenlaces de interés y los tiempos de seguimiento que son claves para el análisis de la información. Es necesario que este proceso se cumpla en duplicado.

La síntesis de evidencia debe realizarse utilizando las tablas que se recomiendan en este manual, la evaluación de la calidad global de la evidencia debe cumplir las recomendaciones de la aproximación GRADE, la calificación de la certeza en la evidencia encontrada deberá ser incorporada en el análisis e interpretación de los

resultados. Para más detalles consultar los manuales de calificación de la evidencia (47, 48)

Los pasos para el análisis de efectividad clínica y seguridad se presentan en el siguiente diagrama:

Figura 1 Pasos para el análisis de efectividad clínica y seguridad



Fuente: elaboración IETS

Este dominio de la evaluación deberá reportarse según lo recomendado en la sección seis de la herramienta “Estándares de reporte IETS para evaluaciones rápidas de tecnologías en salud” (Ver capítulo 5).

Dado que en el capítulo anterior se revisó el primer ítem correspondiente a la definición de alcance de la evaluación, se procederá a describir los siguientes a este:

1. Definición de los criterios de elegibilidad de la evidencia

A partir de las preguntas PICOT se definen los criterios de elegibilidad de la evidencia, en términos de criterios de inclusión y exclusión.

1.1. Criterios de inclusión

Los criterios de inclusión son definidos por el perfil de la población, subgrupos de interés, tecnologías a evaluar, comparadores, desenlaces, tiempo (si aplica) y tipo de estudios que debe contener la evaluación. En caso de considerar relevante analizar la efectividad y seguridad o validez diagnóstica de las tecnologías de interés en subgrupos específicos de la población, estos deben predefinirse como variables.

Los criterios que se presentan a continuación son orientativos. En todo caso, los criterios adoptados para cada evaluación deben ser explícitos y debidamente justificados.

- a) Criterios de inclusión de estudios relacionados con el formato, idioma y fecha de publicación.
 - El estudio debe estar disponible como publicación completa. Los estudios publicados únicamente en formato de resumen no son considerados, porque la información reportada es incompleta para evaluar su calidad metodológica. Además, los resultados de los estudios pueden cambiar significativamente entre la presentación inicial en un evento y la publicación final en un medio impreso.
 - Estudios en inglés o en español, independientemente de su estado de publicación (publicado, en prensa o literatura gris). Debe especificarse si se considera relevante incluir estudios publicados en un idioma diferente a los mencionados.
 - Se recomienda limitar la búsqueda de revisiones sistemáticas a aquellas publicadas en los últimos cinco años. Este rango de fecha de publicación se respalda en evidencia empírica la cual estima que una revisión sistemática tiene una sobrevida media libre de signos para requerir una actualización de 5,5 años (IC95 %: 4,6-7,6) (49). En este punto debe justificarse si se considera pertinente emplear un rango de fecha de publicación diferente.
- b) Criterios de inclusión de estudios relacionados con el diseño, la calidad metodológica y de reporte.

Se recomienda el uso de revisiones sistemáticas convencionales, compuestas por estudios primarios (preferiblemente ensayos controlados aleatorizados) y revisiones sistemáticas panorámicas (*overviews*, *umbrella* o globales), que evalúen la efectividad clínica comparativa de las tecnologías de interés (50). Se considerará

válido el uso de revisiones sistemáticas con metanálisis de comparaciones indirectas o metanálisis en red, siempre y cuando exista una justificación detallada para su uso, es decir, cuando no exista evidencia sobre la efectividad y seguridad de la tecnología frente a los comparadores propuestos en RSL de comparaciones directas o en estudios primarios cabeza a cabeza.

Actualmente, se considera que las RSL aportan información clave para la toma de decisiones en salud. Sus resultados dan cuenta de los efectos (beneficios y daños) de las intervenciones en una población. En comparación con estudios primarios, ofrecen ventajas como la minimización de los sesgos, tienen mayor poder estadístico, reducen el error aleatorio y proporcionan perspectivas complementarias de las opciones bajo comparación (48, 51, 52).

Para definir que un estudio corresponde a una revisión sistemática, se recomienda verificar los criterios propuestos por *Manual Cochrane de revisiones sistemáticas de intervenciones* (53):

El estudio debe tener:

- Objetivos claramente establecidos.
- Criterios de elegibilidad de estudios previamente definidos.
- Metodología explícita y reproducible.
- Una búsqueda sistemática que identifique todos los estudios que puedan cumplir los criterios de elegibilidad.
- Una evaluación de la validez de los resultados de los estudios incluidos, por ejemplo, mediante la evaluación del riesgo de sesgos.
- Una presentación sistemática y una síntesis de las características y resultados de los estudios incluidos.

Además de los anteriores criterios, se debe considerar que, si la revisión tiene solamente un autor o se buscó solamente en una base de datos, no debería ser reconocida como una revisión sistemática (54).

Ante la ausencia de revisiones sistemáticas, se sugiere la inclusión de estudios epidemiológicos individuales con los siguientes diseños: ensayos controlados aleatorizados (idealmente, de fase III, con un enfoque pragmático “análisis por intención de tratar”), cohortes analíticas (preferiblemente prospectivas), casos-contrroles y aquellos estudios observacionales comparativos que estén bajo la denominación de “evidencia del mundo real” (16). En este último escenario, se recomienda privilegiar el uso de ensayos controlados aleatorizados. La posibilidad

de buscar estudios primarios, ante la ausencia de estudios integrativos, debe redactarse de manera explícita en el protocolo de la revisión.

Se sugiere que todo estudio epidemiológico sin un grupo de control concurrente³ (un grupo de control observado por los investigadores al mismo tiempo que el grupo de tratamiento) sea excluido de la revisión, salvo que, no exista ninguna alternativa de manejo comparable con la tecnología principal bajo estudio.

Según el objetivo y alcance de la revisión, para determinar la elegibilidad de los estudios deberá aplicarse la herramienta “Análisis del espectro eficacia-efectividad en estudios de intervenciones en salud” (56), la cual tiene como propósito determinar si cualquier tipo de estudio epidemiológico estuvo más orientado a evaluar la eficacia (estudios explicativos) o la efectividad (estudios pragmáticos) de una intervención en salud Anexo 5.

En el caso que la tecnología objeto de la evaluación sea una prueba diagnóstica, también se privilegiará la inclusión de revisiones sistemáticas. Ante su ausencia, se recomienda el uso de estudios de validez diagnóstica tipo corte transversal o cohorte. Se desaconseja el empleo de estudios de casos y controles (57). Debe considerarse la relevancia de incluir, de forma complementaria, estudios de concordancia entre pruebas diagnósticas; esto es, cuando se tenga como objetivo determinar si una prueba diferente a la habitual tiene una utilidad diagnóstica similar que las haga intercambiables en su uso (58). Otros diseños también pueden ser incluidos especialmente cuando se busque demostrar resultados sobre la utilidad de la prueba, en este caso, experimentos clínicos de pruebas o algoritmos diagnósticos pueden ser necesarios.

1.2. Criterios de exclusión

Los criterios de exclusión deben definir las características que obligan a descartar de la evaluación a aquellos sujetos (población), estudios o desenlaces que cumplen los criterios de inclusión, pero que no pueden admitirse por diferentes motivos. Los criterios de exclusión no deben plantearse como una negación de los criterios de inclusión.

³ Los estudios en mención carecen de una hipótesis refutable. Dado su diseño, les es imposible evitar los sesgos en la selección, reclutamiento e inclusión de los individuos de observación, por cuanto no disponen de mecanismos de muestreo o de asignación aleatoria. En este mismo sentido, al no existir un grupo de control concurrente con el cual sea posible ejercer una comparación, se imposibilita el establecimiento de relaciones causales 55. Echeverry R. J, Pardo T. R. Diseños de investigación clínica. Estudios observacionales: Reporte de casos y serie de casos. Revista Colombiana de Cardiología. 1999;7(1):15-20.

Deben precisarse aquellas situaciones médicas, éticas, de conveniencia, adherencia u otras que obliguen a excluir a la población que cumple los criterios de inclusión, por considerar que introducen sesgos en la evaluación.

2. Búsqueda de estudios

Con el objetivo de identificar evidencia relevante en relación con las preguntas de evaluación, debe llevarse a cabo una búsqueda sistemática y exhaustiva. Todo el proceso debe acogerse a los estándares de calidad internacional utilizados en RSL (48). Las búsquedas deben llevarse a cabo por personal entrenado.

2.1 Búsqueda en bases electrónicas de datos

Para identificar publicaciones indexadas, deben consultarse las siguientes fuentes:

- MEDLINE, de ser posible incluir los repositorios In-Process & Other Non-Indexed Citations y Daily Update (plataforma Ovid o PubMed).
- EMBASE (Elsevier).
- Cochrane Database of Systematic Reviews (plataforma Wiley).
- Database of Abstracts of Reviews of Effects - DARE (plataforma Wiley).
- LILACS (Biblioteca Virtual en Salud - BVS, interfaz iAHx).
- Health Technology Assessment Database (Wiley)
- Red Española de Agencias de Evaluación de Tecnologías Sanitarias y Prestaciones del Sistema Nacional de Salud
- Base Regional de Informes de Evaluación de Tecnologías en Salud de las Américas (BRISA)

En el caso de medicamentos y dispositivos médicos, debe tenerse en cuenta la información referente a vigilancia poscomercialización descrita en el capítulo 3 del presente manual para documentar, en caso de ser posible, la proporción de eventos adversos asociados al uso de las tecnologías de interés.

Si se requiere la búsqueda de ECA, deben consultarse además de las anteriores, en *Cochrane Central Register of Controlled Trials - CENTRAL* (plataforma Ovid) y *Clinical Trials*. En caso de ser necesario, también debe incluirse la identificación de publicaciones a través del motor de búsqueda *Google*.

Para la realización de las búsquedas, inicialmente se identifican los términos clave en lenguaje natural a partir de las preguntas en formato PICOT. Posteriormente, se diseña una estrategia de búsqueda genérica, compuesta por vocabulario controlado

explotado como términos (MeSH (Medical Subject Headings), DeCS (Descriptores en Ciencias de la Salud) y Emtree (Embase Subject Headings) y por lenguaje libre, considerando sinónimos, abreviaturas, acrónimos, variaciones ortográficas y plurales. La estrategia se complementa con identificadores de campo, truncadores, operadores de proximidad y operadores booleanos, y se limita empleando filtros validados por tipo de estudio (59): de alta especificidad para RSL y de alta sensibilidad para estudios primarios. Esta estrategia se valida mediante consulta virtual con expertos temáticos (60) y se adapta para las diferentes bases de datos.

La búsqueda de revisiones sistemáticas debe limitarse a literatura publicada en los últimos cinco años. Cuando se trate de búsquedas *de novo* de estudios primarios, estas deben realizarse sin restricción en la fecha de publicación y de idioma.

Para el reporte de la búsqueda de estudios en bases electrónicas de datos, se recomienda usar el formato que se presenta en el Anexo 2, y el diagrama de flujo propuesto en la declaración *PRISMA* (45) que se presenta en el Anexo 3.

2.2 Métodos de búsqueda complementarios

Además de buscar en las bases de datos mencionadas, se sugiere el uso de métodos de búsqueda complementarios, tales como:

- Herramienta “*Similar articles*” de *PubMed*:
<https://www.ncbi.nlm.nih.gov/pubmed/>
- Herramienta “*Matrix of Evidence*” de *Epistemonikos*:
<https://www.epistemonikos.org/en/>
- *Google* (como fuente de literatura gris)
- Consulta a expertos temáticos, productores y comercializadores de las tecnologías evaluadas acerca de la disponibilidad de estudios adicionales publicados o no publicados que cumplan con los criterios de elegibilidad (de inclusión y de no exclusión) específicos para la evaluación (33).
- Búsqueda manual “en bola de nieve” mediante la revisión de las listas de referencias bibliográficas de los estudios seleccionados.
- Búsqueda de información en las bases de datos especializadas sugeridas por los expertos temáticos.

2.3 Actualización de una búsqueda sistemática de la literatura

Con el propósito de actualizar una revisión sistemática, se recomienda tener en cuenta los siguientes pasos:

- Si se requiere realizar la actualización de la búsqueda de literatura de una RSL, es importante contar con las estrategias de búsqueda utilizadas, las bases en que se realizó dicha búsqueda y las fechas en las que se realizó esta revisión; si esta información está disponible puede ser utilizada para llevar a cabo la nueva búsqueda, en caso contrario, es necesario revisar los elementos de la pregunta de investigación para construir una nueva estrategia de búsqueda.
- Si se decide realizar la actualización de la búsqueda es necesario:
 - Verificar si los términos controlados (DeCS, MeSH y Emtree), aún están vigentes; si no es así, es necesario actualizarlos en la estrategia de búsqueda y construir nuevos algoritmos.
 - La búsqueda inicial es el insumo para elaborar la estrategia de búsqueda para la actualización: si la pregunta de investigación debe incluir algún elemento adicional (por ejemplo, una nueva intervención), esta debe incluirse en la estrategia para actualizar la búsqueda.
 - Se sugiere realizar las búsquedas en las bases que se tuvieron en cuenta originalmente. Si el grupo desarrollador considera necesaria realizar la búsqueda en bases adicionales, se puede realizar reportando la completitud de la estrategia e información necesaria para su replicación.
 - De acuerdo con la base de datos, verifique la vigencia de los filtros, operadores de proximidad, truncadores y otras herramientas de búsqueda que se hayan usado en la búsqueda inicial, comprobando si han sido modificados o requieren actualizarse.
 - El tiempo de las búsquedas puede ser desde la fecha de la búsqueda inicial, no obstante, este tiempo se puede modificar a criterio de los revisores y justificar su decisión.
 - Es importante describir y justificar todos los cambios que se hayan hecho en el proceso de actualización.
- Consultar la opinión de los expertos temáticos sobre si las conclusiones de la revisión sistemática se mantienen validas en el tiempo actual, y si ellos están o no al tanto de cualquier nuevo estudio primario (adicional a los incluidos en la revisión sistemática) que pueda modificar las conclusiones de la revisión (61).

Para el reporte de la actualización de la búsqueda de estudios en bases electrónicas de datos, se recomienda usar el diagrama de flujo propuesto en la declaración *PRISMA* (45), que se presenta en el Anexo 4.

2.4 Gestión documental

Para cada búsqueda en bases de datos electrónicas debe generarse un reporte, garantizando su reproducibilidad y transparencia. Las estrategias de búsqueda y sus resultados deben almacenarse en formato electrónico.

Los resultados de las búsquedas deben incluirse en una base de datos, preferiblemente en un programa de manejo de referencias bibliográficas. Las publicaciones duplicadas deben eliminarse.

3. Tamización de referencias y selección de estudios

Las referencias deben ser tamizadas por dos revisores de forma independiente, revisando los títulos y resúmenes frente a los criterios de elegibilidad predefinidos. Previamente debe realizarse una prueba piloto para asegurar la consistencia en la aplicación de los criterios de elegibilidad.

En caso de duda sobre el cumplimiento de los criterios de elegibilidad, debe revisarse el texto completo del estudio para orientar la decisión. Deben describirse las estrategias empleadas para resolver los desacuerdos entre los pares revisores (consenso, consulta con un tercer revisor, etc.).

A partir del grupo de referencias preseleccionadas, se realiza la selección de estudios; para esto, un revisor con el apoyo de un experto temático verifica que el estudio cumpla los criterios de elegibilidad (de inclusión y de no exclusión) específicos para las preguntas de evaluación (población, tecnologías de interés, comparadores y desenlaces) mediante la revisión de cada publicación en texto completo.

Los estudios no recuperados en texto completo al momento de la publicación del reporte de evaluación deben clasificarse como estudios excluidos.

Es altamente recomendable que el proceso de selección se realice con ayuda de herramientas de gestión de referencias como Endnote®, Mendeley®, Rayyan®, con el fin de asegurar la trazabilidad del proceso.

4. Evaluación del riesgo de sesgos

El sesgo es un error sistemático o una desviación de la verdad, en los resultados o inferencias (1), dicho error puede sobreestimar o subestimar el efecto real de la intervención. Por tanto, la evaluación del riesgo de sesgos de los estudios incluidos en la revisión sistemática es determinante para conocer la aproximación real del efecto.

De otro lado, la evaluación de la calidad metodológica determina la apreciación crítica de los estudios incluidos y el grado en el cual los autores realizaron su investigación con los más altos estándares posibles (1). Existen escenarios en los cuales los efectos de un estudio pueden tener ausencia de sesgos a pesar de tener algunos defectos metodológicos (1), o tener altos estándares de calidad metodológica y riesgo de sesgo alto.

Este manual recomienda la evaluación del riesgo de sesgo de los estudios incluidos en las revisiones sistemáticas. Para tal fin se recomiendan las siguientes herramientas, las cuales deben ser elegidas según el diseño de los estudios incluidos:

1. Revisiones sistemáticas de literatura: ROBIS (62).
2. Revisiones sistemáticas con ITC (*Indirect Treatment Comparison, en español comparación de tratamiento indirecta*) o MTC (*Mixed Treatment Comparison, en español comparación de tratamientos múltiples*): IQWiG (63).
3. Revisiones sistemáticas con ITC (*Indirect Treatment Comparison/Network Meta-Analysis*) de ISPOR (64).
4. Ensayo controlado aleatorizado: Riesgo de sesgos de Cochrane RoB (1) preferiblemente en su segunda versión RoB 2 (65).
5. Estudios no aleatorizados de intervenciones: ROBINS-I (66).
6. Estudios primarios de exactitud diagnóstica: QUADAS-2 (57).
7. Estudios de cohortes y casos y controles JBI (67) o New Castle- Ottawa (47).

En este proceso es recomendable se tengan en cuenta aspectos como, la aplicación de las herramientas por parte de un experto metodólogo, la realización de la lectura completa del estudio objeto. Anexo 6.

5. Extracción de datos

Idealmente, se deberá construir una matriz estandarizada, que deberá ser piloteada antes de iniciar el ejercicio de extracción. Los datos deben ser extraídos a partir de los estudios con menor riesgo de sesgos. Los datos que deben ser extraídos incluyen por ejemplo: datos de identificación del estudio como autor, año de publicación, identificadores de indexación como DOI u otro sistema, diseño del estudio, objetivo, número de participantes, características de la población como comorbilidades, edad, sexo y otras si se anticipa una posible modificación del efecto a partir de ellas, datos sobre la intervención como vía de administración, dosis, frecuencia, uso de cointervenciones u otras características que puedan ser modificadoras del efecto, datos sobre los desenlaces de interés. Se considera pertinente incluir más de un estudio si la información contenida en cada publicación es complementaria en términos de comparaciones y desenlaces.

En caso de no contar con una medida resumen del efecto y sí con estudios primarios no combinables, debe presentarse un rango de los efectos con base en lo reportado por dichos estudios.

Los resultados deben extraerse tal y como son provistos por los autores en los estudios seleccionados. Cuando sea necesario, debe contactarse a los autores de los estudios para obtener información sobre hallazgos no reportados o incompletos. En caso de identificar múltiples publicaciones de un mismo estudio, debe tomarse en cuenta el reporte con fecha de publicación más reciente.

6. Síntesis de la evidencia

Se recomienda que la información clave acerca de los desenlaces relevantes para cada pregunta de investigación, se presente mediante perfiles de evidencia *GRADE*. El formato estándar para reportar el perfil de evidencia deberá incluir para cada comparación objeto de la revisión, los siguientes elementos (4):

- La identificación de los desenlaces críticos y de los desenlaces importantes para la toma de decisiones.
- El número de participantes y el número de estudios.
- El diseño de los estudios.
- Los juicios acerca de cada uno de los factores que puedan influenciar la calidad de la evidencia (riesgos de sesgos, inconsistencia, evidencia indirecta, imprecisión, sesgo de publicación, entre otros aspectos).

- Los resultados estadísticos clave en términos de los estimadores de asociación (riesgo relativo RR, *Odds ratio OR*, *hazard ratio HR*, diferencia de medias DM o diferencia estandarizada de medias DEM) y sus correspondientes medidas de incertidumbre (intervalos de confianza o intervalos de credibilidad). Se puede recurrir a la presentación de resultados en términos de estimadores de impacto (número necesario a tratar, número necesario a dañar, etc.) como datos adicionales. No se recomienda incluir resultados exclusivamente de valores p de pruebas de hipótesis estadísticas.

En caso de usar como evidencia metanálisis en red, las estimaciones del efecto deberán acompañarse de los valores del área bajo la curva del *ranking* acumulativo (valores *SUCRA*). Cuando se trate de metanálisis de indicadores de desempeño operativo de pruebas diagnósticas, se recomienda el uso de las curvas *ROC* resumen.

- Los efectos absolutos anticipados (*GRADEpro* los calcula automáticamente).
- La calificación de la calidad de la evidencia para cada desenlace.
- La calificación de la importancia de cada desenlace.
- La calificación de la calidad global del cuerpo de la evidencia (para cada comparación).
- Notas al pie de página, para proveer las explicaciones de los juicios acerca de la calidad de la evidencia.

Además del perfil de evidencia *GRADE*, se deberá presentar un resumen en texto, en lenguaje plano, que dé cuenta del análisis en conjunto de la significancia estadística y la significancia clínica de los hallazgos, y de la direccionalidad de los efectos (es decir, si favorece a la tecnología de interés o a su comparador). En dicha síntesis narrativa, se desaconseja incorporar juicios acerca de la magnitud o el tamaño de los efectos (por ejemplo: “el medicamento presenta un gran efecto benéfico”).

Para la creación de los perfiles de evidencia *GRADE* se sugiere emplear la herramienta *GRADEpro*, disponible de forma gratuita en el siguiente enlace:

<http://gdt.guidelinedevelopment.org/app/>

Para mayores detalles sobre la metodología de construcción de los perfiles de evidencia *GRADE*, se recomienda consultar la versión en español del manual *GRADE* (4), disponible en línea:

<http://gdt.guidelinedevelopment.org/app/handbook/translations/es/handbook.html#h.1hmsyys>

Los tutoriales acerca del uso del sistema *GRADE* pueden ser vistos en: cebgrade.mcmaster.ca

6.1 Consideraciones adicionales

- Las revisiones sistemáticas que consideren múltiples comparaciones (por ejemplo, un medicamento innovador frente a todas sus alternativas terapéuticas empleadas para el manejo de la condición de salud) van a requerir la presentación de un perfil de evidencia por cada comparación. Para cada comparación de tecnologías, todos los desenlaces se deben reportar juntos en el mismo perfil de evidencia.
- Es probable que un solo estudio incluido en la revisión no proporcione evidencia para todos los desenlaces relevantes, por ende, el perfil de evidencia puede estar soportado a partir de múltiples estudios (idealmente, revisiones sistemáticas) que sean complementarios entre sí para responder a la pregunta de investigación.

7. Análisis estadístico

A continuación, se presentan las metodologías para el desarrollo de los metanálisis *de novo*. Deben presentarse las razones por las cuales se decide realizar o no un metanálisis y justificar la selección del modelo (efectos fijos, efectos aleatorios, enfoque frecuentista o bayesiano, etc.).

7.1 Metanálisis clásico

Debe contemplarse la opción de un metanálisis en caso de no identificar una medida combinada del efecto en las revisiones seleccionadas y de disponer de al menos dos estudios primarios similares. De forma previa al análisis, debe explorarse la presencia de heterogeneidad clínica, metodológica y estadística, y si existe reporte insuficiente de desenlaces o variación en su definición operativa.

En caso de tomar la decisión de agrupar los estudios en un metanálisis, la estimación promedio del efecto debe obtenerse a través de un metanálisis, ponderando por el tamaño muestral (prueba estadística de Mantel - Haenszel). La significancia estadística del efecto debe evaluarse mediante la prueba estadística Z.

El análisis de la heterogeneidad estadística debe considerar el valor de la prueba de inconsistencia (I^2), el valor p de la prueba de ji-cuadrado y una inspección visual de la consistencia del efecto a través de los estudios con base en un *forest plot* (48). En caso de identificar heterogeneidad importante, debe explorarse la fuente de la misma mediante herramientas como: estimaciones de modelos de efectos fijos *versus* efectos aleatorios, análisis de subgrupos o meta-regresión. Las fuentes de heterogeneidad pueden ser clínica, metodológica o estadística, por lo que las características de la población, intervención, diseños y calidad metodológica deben ser consideradas.

En todos los casos, el análisis de subgrupos debe realizarse sobre las variables predefinidas en los criterios de inclusión. Debe realizarse un análisis de sensibilidad para estudiar la influencia del riesgo de sesgo en los resultados del metanálisis. La presencia de sesgo de publicación debe evaluarse mediante un gráfico de embudo y la prueba estadística de Egger (48). Debe reportarse el nivel de confiabilidad, de significancia estadística y el programa empleado para el análisis estadístico.

Para mayor información sobre los aspectos anteriormente señalados, se sugiere revisar el manual Cochrane de revisiones sistemáticas de intervenciones (48). Para información específica sobre el desarrollo de metanálisis de estudios de validez diagnóstica, se sugiere revisar el *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (68).

7.2 Metanálisis en red (NMA)

Existen situaciones en las que no siempre se dispone de estudios de comparaciones directas (cabeza a cabeza) para evaluar y comparar la efectividad entre las tecnologías en salud, por ejemplo:

- Estudios de posicionamiento terapéutico en los que el objetivo es establecer un ordenamiento entre moléculas de una categoría farmacológica en los cuales no siempre hay estudios que comparen todos los posibles pares de moléculas.
- Guías de práctica clínica sobre tratamientos para los que existen múltiples opciones terapéuticas y se carece de estudios para comparaciones entre ellas o algunas de ellas para analizar los criterios sobre efectos deseados, efectos indeseados, balance de efectos y certeza en la evidencia.
- Estudios de efectividad y seguridad para grupos terapéuticos para lo que se requiere de múltiples comparaciones.

En estos casos debe analizarse la posibilidad de realizar un metanálisis en red. Esta técnica permite comparar tres o más intervenciones de manera simultánea combinando tanto evidencia directa como evidencia indirecta, lo cual permite a su vez comparar, por medio de una medida relativa, cada pareja de intervenciones; así mismo permite generar un ordenamiento o *ranking* de las mismas (69, 70). En el Manual Cochrane puede revisarse la fundamentación teórica para la realización de este tipo de estudios (69).

Si se considera esta posibilidad, en este apartado de la evaluación de efectividad y seguridad se debe justificar la necesidad de realización del metanálisis en red, se debe presentar un diagrama de red en el que se describa y sintetice la información disponible (número de experimentos clínicos, comparaciones directas realizadas, comparaciones que se estimarán de manera indirecta, si la red utilizará un comparador común, etc.) por cada desenlace evaluado, y se debe presentar el juicio sobre el cumplimiento del supuesto de transitividad.

Luego debe indicarse cuál fue la aproximación utilizada para hacer las estimaciones (frecuentista o bayesiana), el método empleado para evaluar la consistencia o inconsistencia entre estimaciones directas e indirectas y el método empleado para generar el *ranking* entre intervenciones. Se debe indicar el software o lenguaje de programación empleado y anexar el código y la base de datos empleada.

La evaluación de la certeza en la evidencia en este caso debe realizarse de acuerdo con las indicaciones propuestas por el grupo GRADE para la evaluación de la certeza (71, 72), incoherencia (73), imprecisión (74, 75), usando contextos mínimamente contextualizados (76) o parcialmente contextualizados (77) y usar las tablas de resumen de hallazgos propuestas en (78).

7.3 Adición de nuevos datos a las medidas combinadas del efecto

Cuando se identifiquen estudios primarios adicionales a los incluidos en una revisión seleccionada para la síntesis de evidencia que cumplan los criterios de elegibilidad (de inclusión y de no exclusión) de la revisión y se disponga de un metanálisis clásico (con comparaciones directas entre las tecnologías de interés), debe evaluarse la posibilidad de incluir los datos de estos nuevos estudios primarios en la estimación combinada del efecto, independientemente de la magnitud, dirección y precisión de los nuevos datos. Para esto, debe evaluarse si los nuevos estudios son combinables con el grupo de estudios previamente agrupados, es decir, si cumplen los criterios de elegibilidad del estudio original y no generan importante heterogeneidad (clínica, metodológica o estadística).

8. Valoración de la calidad y del cuerpo de la evidencia por cada desenlace

Para mayor información sobre la metodología, se sugiere revisar los manuales de *GRADE Working Group* (52), *Manual Cochrane para revisiones sistemáticas de intervenciones* (79) y Singh y cols (80).

El método para llevar a cabo la evaluación de la calidad de la evidencia de las RSL se basa en la propuesta del *GRADE Working Group* (81). Por tanto, todas las RSL realizadas deben incluir este método.

La calidad de la evidencia presenta el grado de confianza en que el estimativo del efecto es correcto (5). Dicha estimación de la calidad se basa en la calificación del cuerpo de la evidencia para cada uno de los desenlaces de manera independiente. Por lo tanto, la calidad puede variar entre los desenlaces (5).

Esta aproximación clasifica la calidad en uno de cuatro grados (alto, moderado, bajo y muy bajo). Para la determinación de estos grados, se tiene en cuenta la evaluación del tipo de diseño de los estudios incluidos y el análisis de los factores que determinan la calidad de la evidencia (5). A continuación, se presentan dichos factores:

Factores que disminuyen la calidad de la evidencia estudios aleatorizados:

- Limitaciones del estudio (riesgo de sesgo de ensayos controlados aleatorizados o estudios observacionales).
- Inconsistencia de los resultados.
- Evidencia indirecta.
- Imprecisión.
- Sesgo de publicación.

Factores que aumentan la calidad de la evidencia de los estudios observacionales y estudios no aleatorizados:

- Efecto de gran magnitud.
- Efecto de los potenciales factores de confusión residual.
- Gradiente dosis respuesta.

Factores que disminuyen la calidad de la evidencia de los estudios de precisión diagnóstica:

- Diseño del estudio.
- Riesgo de sesgo (limitaciones en el diseño y la ejecución del estudio).
- Evidencia indirecta.
- Inconsistencia importante en los resultados de los estudios.
- Evidencia imprecisa.
- Alta probabilidad de sesgo de publicación.
- Subir por dosis efecto, posible sesgo de grandes efectos residuales y factores de confusión.

La información obtenida de este proceso, es decir la calidad y los juicios que la sustentan (riesgos de sesgos, inconsistencia, evidencia indirecta, imprecisión, sesgo de publicación, etc.), se deben presentar en el perfil de evidencia GRADE para cada desenlace, según lo descrito previamente en este manual.

Para determinar la calidad global de la evidencia se debe tener en cuenta que:

- Si la calidad de la evidencia es la misma para todos los desenlaces críticos incluidos en la evaluación, esta se considera la calidad global de la evidencia.
- Si la calidad de la evidencia difiere entre los desenlaces críticos se asumirá como calidad global, la menor calidad de la evidencia determinada para cualquier desenlace (5).

Para mayor información sobre la metodología, se sugiere revisar el manual de *GRADE Working Group* (5).

9. Determinación de valor umbral terapéutico o significancia clínica de los resultados

En general, el alcance del análisis de ETS relacionado con el paciente, cuidadores y otros actores puede ser muy amplio, es por esto, que se deben destacar aquellos temas que son de particular relevancia para la tecnología bajo evaluación.

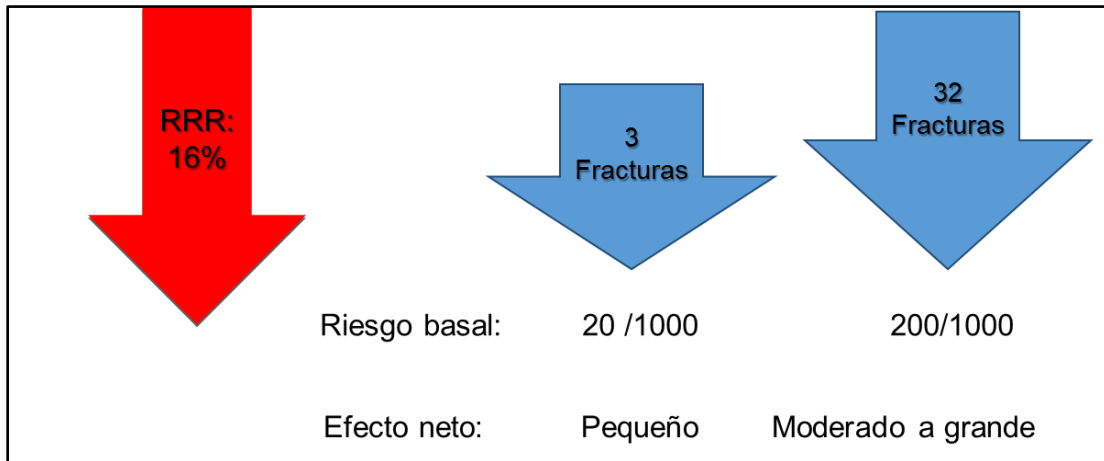
Se propone para poder juzgar qué asuntos son relevantes para una tecnología dada, llegar a un cálculo de umbral de valor terapéutico para cada uno de los desenlaces analizados. Esta definición parte de la reflexión del grupo GRADE, cuando se califica la certeza de la evidencia para un desenlace, se califica cuán

seguro se está de que el verdadero efecto reside en un rango particular o a un lado de un umbral (7, 82).

El enfoque GRADE postula cada vez más la importancia del efecto de la intervención, presentado como el riesgo o la diferencia en el efecto, como números absolutos (por ejemplo, 5 muertes menos por cada 100) (83), en lugar de usar efectos relativos que a menudo pueden ser engañosos. Por ejemplo, considere un Riesgo Relativo (RR) de 0,84 IC 95% (0,65 a 0,90), con una Reducción del Riesgo Relativo del 16% en las fracturas de cadera en adultos mayores. Si por un lado, el riesgo basal de fracturas de cadera es 20/1000 durante 1 año, la reducción relativa de riesgo 0,84 se traduciría en 3 fracturas (un RR de 0.84 equivale a una reducción relativa del riesgo de 16%; 20/1000 es 2%; entonces $0,16 \times 0,02 \times 1000 = 3,2$) menos por 1000 tratados, lo que la mayoría consideraría un efecto pequeño (83).

Por otro lado, si el riesgo basal es mayor por ejemplo 200/1000, muchos considerarían que la reducción absoluta resultante de 32 fracturas (un RR de 0.84 equivale a una reducción relativa del riesgo de 16%; 200/1000 es 20%; entonces $0,16 \times 0,20 \times 1000 = 32$) por 1000 es un efecto de moderado a grande (83), ver Figura 2.

Figura 2. Cambios en la apreciación del desenlace disminución de fracturas según Reducción de Riesgo Relativo (RRR) y Reducción de Riesgo Absoluto dependiendo del riesgo basal



Fuente: elaboración IETS

El grupo GRADE indica que la elección de un umbral (o rango) puede estar totalmente contextualizada (basada en la consideración de todos los resultados críticos) o parcialmente contextualizada (basado en el valor del resultado individual); sin embargo, para el desarrollo metodológico del presente manual se elegirán umbrales de valor para cada desenlace (83).

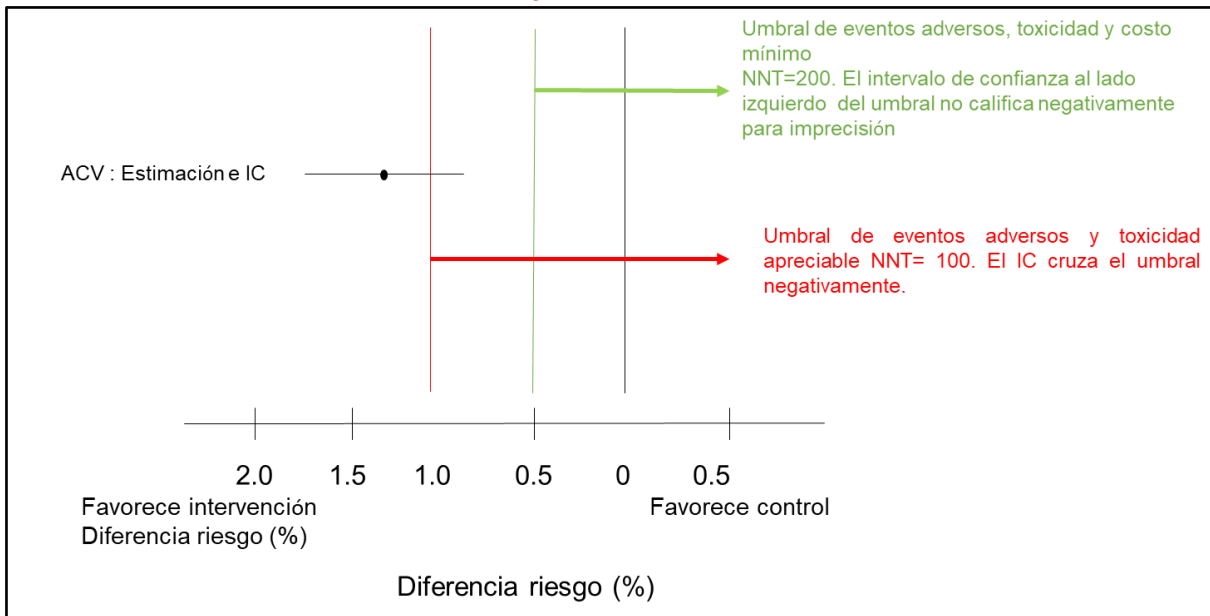
Para hacer la definición de los umbrales, se pueden considerar diferencias absolutas de riesgo, números necesarios a tratar (o para hacer daño) o efectos absolutos como los presentados en GRADE, luego se calcula la diferencia de riesgos o el número necesario a tratar o el efecto absoluto.

Al decidir los umbrales, se necesita ser consciente del riesgo de malinterpretar un resultado con un intervalo de confianza que incluye "1" (para efectos relativos) o "0" (para efectos absolutos) (83).

Para entender mejor este concepto se presenta a continuación el ejemplo descrito por el grupo GRADE:

Los autores presentan una revisión sistemática hipotética de ensayos clínicos aleatorizados de una intervención para prevenir accidentes cerebrovasculares mayores (ACV) que produce una estimación agrupada de una reducción absoluta de 1.3%, con un intervalo de confianza (IC 95%: 0,6 a 2,0 %). Por lo tanto, al llevarlo a números absolutos, se debe tratar a 77 (100 / 1,3) pacientes durante un año para prevenir un solo accidente cerebrovascular mayor, el IC del 95% alrededor del Número Necesario para Tratar (NNT) de 50 a 167, nos dice que mientras 77 es el mejor estimador, es posible que se necesiten tratar tan solo 50 o tantos como 167 personas para prevenir un solo accidente cerebrovascular (82).

Figura 3. Valoración de la certeza de que el verdadero efecto se encuentra en un rango particular



Fuente: Guyatt et al. Modificado grupo GRADE 2012 (82).

Por ejemplo, por criterio clínico y de pacientes (suponiendo que la intervención es un medicamento sin efectos adversos graves, con inconvenientes mínimos y es equitativo, factible y aceptable para administrar), se puede definir un umbral de la intervención para reducir los accidentes cerebrovasculares en tan solo un 0,5% (línea media vertical verde en la Figura 3) que corresponde a un NNT de 200. El IC completo (0.6% a 2.0%) alrededor del efecto sobre la reducción del accidente cerebrovascular se encuentra a la izquierda de la decisión clínica del umbral de 0.5% y por lo tanto, se excluiría un beneficio más pequeño que el umbral, y así la precisión de la evidencia es suficiente (83).

Se establece un umbral, 0.5%, que define el valor terapéutico y permitirá recomendar o no la intervención.

Si en el escenario se adicionan eventos serios de toxicidad, podría modificar la anterior recomendación, a menos que, la reducción absoluta del ACV esté en al menos 1% (NNT= 100 – Línea vertical roja de la Figura 3). En estas circunstancias, la precisión es insuficiente (porque el límite superior del IC95% de la diferencia de riesgos excede el 1%, o el límite superior del NNT es mayor a 100) para considerar que la intervención tiene valor terapéutico, por lo tanto, no se recomendaría su uso. El IC abarca los efectos más pequeños del tratamiento que este umbral, por lo que no excluye un beneficio absoluto apreciablemente inferior al 1% (83).

Particularmente en este ejemplo, la certeza en el tamaño del efecto para el desenlace de reducción del ACV no cambia, pero como el umbral para cambios en el tratamiento (como la magnitud de la salud no deseada de los efectos) aumenta, la certeza de que las consecuencias deseables supere las consecuencias indeseables disminuye (83).

Si se conoce la frecuencia del desenlace en el grupo control (por ejemplo, con el uso de la intervención actual), se puede determinar el umbral en términos de efecto absoluto, como se presenta en el sistema GRADE. Por ejemplo, el número de eventos cardiovasculares mayores en pacientes con falla cardíaca y FEVI reducida, es 5%, es decir 50 por cada 1000 pacientes; el grupo de expertos considera que, para considerar superior a la intervención nueva, esa frecuencia debe ser por lo menos de 10 menos por cada 1000 pacientes. Si el estudio que compara las dos intervenciones arroja un RR de 0.93 (IC95% 0,87 a 0,96), el efecto absoluto será de 3.5 eventos menos (2 menos a 6,5 menos), con lo cual no se cumple el umbral.

La lógica del ejemplo anterior se aplica a la identificación del **umbral de valor terapéutico**: dadas las consecuencias indeseables de una intervención, qué tan seguros están los actores (especialistas clínicos, pacientes) que los beneficios para la salud se encuentran por encima de un umbral que hace que valga la pena administrar esa intervención. En otras palabras, el umbral de valor terapéutico refleja la medida de significancia clínica de una intervención, considerando otros elementos del contexto además de la acostumbrada significancia estadística.

9.1 Proceso deliberativo para determinar umbral de valor terapéutico

A continuación, se describen los aspectos principales en el proceso de deliberación, para la determinación del umbral de valor terapéutico:

9.1.1 Participantes y convocatoria

La selección de los participantes debe estar basada en un mapeo de actores adecuado a la condición e intervención a evaluar. Se espera que se incluyan al menos los siguientes actores: profesionales de la salud, pacientes y/o cuidadores principales.

El proceso de convocatoria debe garantizar transparencia y representatividad de los actores e incluir el suministro de información previa a los participantes sobre el tema a deliberar en términos de una traducción de la evidencia científica a lenguaje cotidiano para que todos los actores, expertos y no expertos, dispongan de la

información adecuada para el proceso de deliberación (84), para esto se sugiere proveer un resumen de la traducción de evidencia previo a la deliberación de modo que se asegure que la información relevante será tomada en cuenta por los actores participantes, así se garantiza que su participación sea informada, lo que además permite niveles comunes de información por parte de los actores (52).

Se debe solicitar a los actores que participan en la deliberación, la declaración de los posibles conflictos de intereses y el consentimiento verbal para grabar la sesión. En este mismo sentido, es fundamental que se tenga la autorización del tratamiento de datos personales por parte del IETS, como elemento de criterio para la auditabilidad del involucramiento de actores (Ver Guía de Involucramiento de actores (31)).

El proceso de convocatoria de los participantes debe sistematizarse y presentarse como parte del informe de participación (Ver Guía de Involucramiento de actores).

9.1.2 Contexto de toma de decisiones: carga de enfermedad y alternativas de tratamiento.

Se realizará una presentación de la condición de salud objeto del uso de la tecnología a evaluar, haciendo énfasis en el riesgo basal, como contextualización para todos los participantes. Se describirá la tecnología ya sea medicamento o dispositivo médico (vías de administración, dosis, requerimientos de seguimiento médico, equipos y suministros, capacidad de talento humano, entre otros) bajo evaluación y de sus comparadores.

9.1.3 Métodos para obtención de preferencias de los participantes.

En términos generales, la evidencia no reporta cuál es la técnica más adecuada para la captar e incluir las preferencias de los participantes en una ETES (85), si bien pueden existir diversos métodos para esto, se sugiere usar técnicas que garanticen la deliberación y permitan la votación como el caso del grupo nominal, la técnica Delphi o los diálogos deliberativos, esto dependerá de la condición o intervención a valorar y la finalidad que se trace en grupo desarrollador en términos alcance del involucramiento de actores.

Una vez seleccionada la técnica, con su respectiva justificación, durante la sesión se presentarán los hallazgos de la efectividad, eficacia y seguridad de la tecnología objeto de la evaluación, mediante los hallazgos de las tablas GRADE. Para calcular

el umbral, se seleccionan solamente los desenlaces que presentan significancia estadística.

Se debe exponer claramente el incremento/reducción del número de pacientes con el desenlace del comparador en términos absolutos como se indica al inicio de este capítulo para cada uno de los desenlaces (efectividad/seguridad) de manera transparente. Se sugiere utilizar material como pictogramas para un mejor entendimiento por parte de los pacientes, en términos de número de pacientes o eventos evitados tanto de la efectividad de los tratamientos como de la severidad de eventos adversos, ver Figura 7, Figura 9 y Figura 11; **Error! No se encuentra el origen de la referencia..**

En la figura 4 se muestra un resultado de un metanálisis en el cual se comparan dos antibióticos para la curación de tejidos blandos infectados, la estimación del efecto ponderado favorece el uso de linezolid comparada con vancomicina (RR:1.09, IC 95%: 1.03 a 1.16) ¿Cómo se mostraría en términos de pacientes curados con las dos terapias?

Figura 4. Valoración tabla GRADE de desenlace de efectividad

Autor(es): Grupo Desarrollador

Pregunta: Linezolid comparado con Vancomicina para curación clínica de infecciones de tejidos blandos

Bibliografía: YueJ, DongBR, YangM, ChenX, WuT, LiuGJ. Linezolid versus vancomycin for skin and so) tissue

infections. Cochrane Database of Systematic Reviews 2016, Issue 1. Art. No.: CD008056. DOI:

10.1002/14651858.CD008056.pub3.

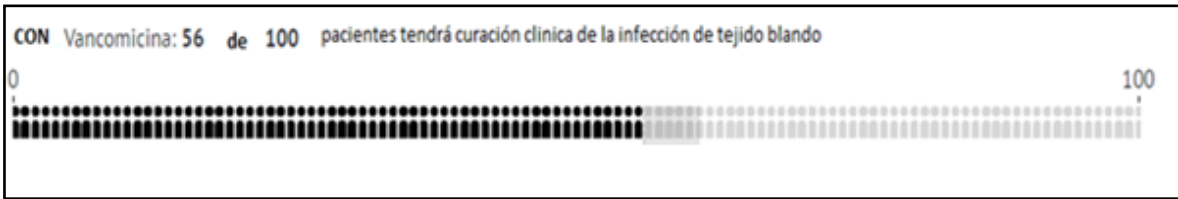
| No de estudios | Diseño de estudio | Riesgo de sesgo | Medición certeza | | | | No de pacientes | | Efecto | | Certeza | Importancia |
|--|--------------------|-----------------|------------------|---------------------|-------------|-----------------------|-----------------|------------------|-----------------------|-------------------|-----------|-------------|
| | | | Inconsistencia | Evidencia indirecta | Imprecisión | Otras consideraciones | Linezolid | Vancomicina | Relativo (95% CI) | Absoluto (95% CI) | | |
| Curación clínica (seguimiento: mediana 10 días ; evaluado con : Proporción de pacientes con curación clínica). | | | | | | | | | | | | |
| 9 | ensayos aleatorios | no es serio | no es serio | no es serio | no es serio | ninguno | | 838/1498 (55.9%) | RR 1.09 (1.03 a 1.16) | | ⊕⊕⊕⊕ ALTA | CRÍTICO |

Fuente: Yue et al., 2016

Primero se mostrará la información de número de pacientes con el desenlace del comparador es decir vancomicina.

Se puede decir que de cada 100 pacientes con vancomicina se curan 56.

Figura 5 Curación clínica de la infección de tejido blando con Vancomicina



Fuente: YueJ, DongBR, YangM, ChenX, WuT, LiuGJ., 2016

¿Cuántos pacientes más deben tener curación clínica de la infección para que la Linezolid sea considerada superior?

Las preguntas por realizarse son:

- ¿Son X número de pacientes (más curados) importantes desde el punto de vista clínico?
- ¿Qué tan relevante es el tamaño del efecto?
 - ✓ Efecto grande
 - ✓ Efecto moderado
 - ✓ Efecto pequeño (importante)
 - ✓ Efecto pequeño (no importante)
 - ✓ Sin efecto
- ¿Qué consideraciones clínicas adicionales se deben tener en cuenta, vías de administración, dosis, etc.?
- Ahora bien, ¿Qué pasa si X pacientes curados no son suficientes, desde el punto de vista clínico, para garantizar superioridad de Linezolid sobre Vancomicina? ¿Cuál sería el número de pacientes mínimo para garantizar la superioridad de Linezolid?

Este valor se debe negociar y decidir con el grupo participante mediante votación.

9.1.4 Proceso para el consenso de participantes.

Se deben evaluar cada uno de los desenlaces que la literatura reporta con significancia estadística. Una vez se haga el ejercicio descrito en el apartado anterior para cada desenlace y se ha obtenido umbral de valor el grupo debe estar de acuerdo (someter a votación).

Si hay varias propuestas, se elegirá la mediana de dichos valores y se someterá a votación.

El moderador es el encargado de la coordinación de la discusión en torno a las ideas presentadas, se debe determinar la votación para lograr consenso de la siguiente manera (se debe hacer para cada desenlace):

1. Presentación del valor del umbral terapéutico del desenlace
2. Votación
3. Resultados de la votación
4. Replanteamiento del nuevo valor de umbral terapéutico
5. Nueva votación

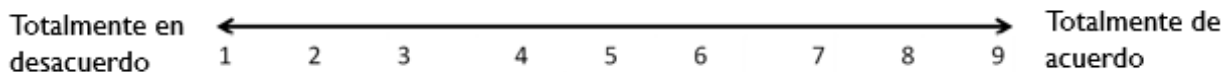
En cada ronda de discusión, los participantes de forma individual analizan y califican mediante una votación *on-line* anonimizada, utilizando una escala Likert entre uno (1) y nueve (9) donde 1 es “Totalmente en desacuerdo” y 9 “Totalmente de acuerdo” (Figura 6).

Posteriormente, se debe realizar el análisis de la votación presentando los resultados obtenidos al grupo de participantes, analizados en términos de porcentaje y mediana con un intervalo de confianza (IC) del 95%. Las posibilidades de resultado se interpretan de la siguiente manera:

- Los puntajes entre 7 y 9 de más del 80% de los participantes y/o una mediana de 8 IC95% entre 7 y 9 (se considera válido pasar alguna de las 2 opciones): se aprueba el valor del umbral.
- Los puntajes entre 1 y 3 de más del 80% de los expertos y/o una mediana de 2 IC95% entre 1 y 3 (se considera válido pasar alguna de las 2 opciones): no se aprueba el valor del umbral.
- En los casos en que no se alcanza alguna de las 2 condiciones anteriores se replantea el valor del umbral, mediante una discusión en una segunda ronda con el grupo de participantes, quienes vuelven a realizar la votación hasta lograr consenso.

Los resultados de las votaciones se deben procesar, analizar y presentar a los participantes, y reportar en el documento que se entrega al IETS.

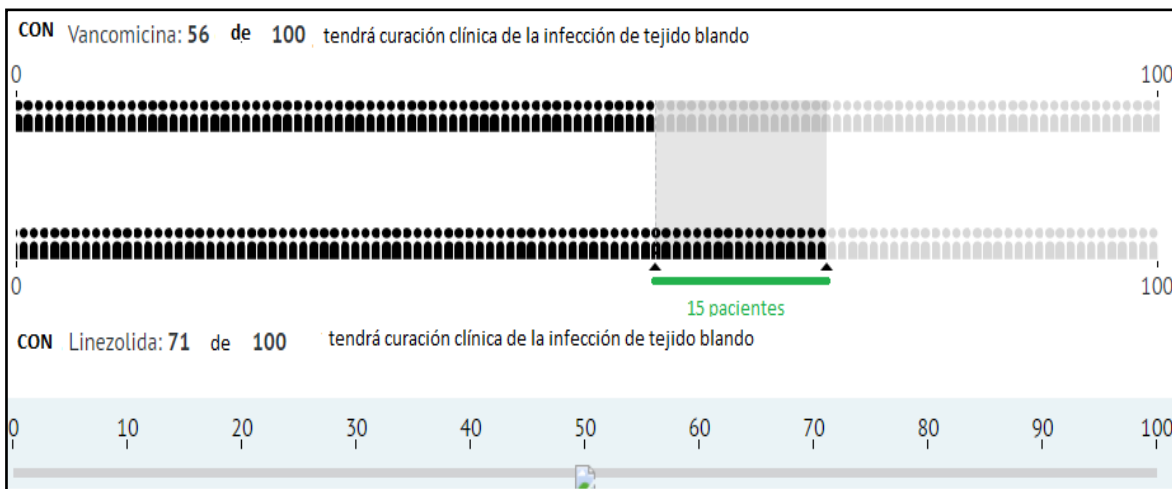
Figura 6. Escala usada para el proceso de votación



Análisis de los resultados

Supuesto: El consenso considera que para que Linezolid sea superior en términos de efectividad, debe curar mínimo 15 pacientes más que vancomicina, es decir 71 pacientes de cada 100.

Figura 7 Curación clínica de la infección de tejido blando con vancomicina vs linezolid



Fuente: Yue et al. Herramienta GradePro., 2016

Es decir un RAR 15,22% IC 95% (11,9% a 18,59%) y un NNT: 7 IC 95% (5 a 8)

Es decir, que para el grupo debería haber una diferencia absoluta del riesgo del 15% (Efecto absoluto 15 pacientes más por cada 100 pacientes).

A continuación, se presentan los valores verdaderos del estudio clínico:

Primero se enseña la tabla GRADE con todos sus valores

Figura 8 Valoración tabla GRADE de desenlace de efectividad

Autor(es): Grupo Desarrollador

Pregunta: Linezolid comparado con Vancomicina para curación clínica de infecciones de tejidos blandos

Bibliografía: YueJ, DongBR, YangM, ChenX, WuT, LiuGJ. Linezolid versus vancomycin for skin and so) tissue infections. Cochrane Database of Systematic Reviews 2016, Issue 1. Art. No.: CD008056. DOI: 10.1002/14651858.CD008056.pub3.

| Nº de estudios | Diseño de estudio | Riesgo de sesgo | Medición certeza | | | | Nº de pacientes | | Efecto | | Certeza | Importancia |
|--|--------------------|-----------------|------------------|---------------------|-------------|-----------------------|------------------|------------------|-----------------------|-----------------------------------|-----------|-------------|
| | | | Inconsistencia | Evidencia indirecta | Imprecisión | Otras consideraciones | Linezolid | Vancomicina | Relativo (95% CI) | Absoluto (95% CI) | | |
| Curación clínica (seguimiento: mediana 10 días ; evaluado con : Proporción de pacientes con curación clínica). | | | | | | | | | | | | |
| 9 | ensayos aleatorios | no es serio | no es serio | no es serio | no es serio | ninguno | 991/1616 (61.3%) | 838/1498 (55.9%) | RR 1.09 (1.03 a 1.16) | 5 más por 100 (de 2 más a 9 más) | ⊕⊕⊕⊕ ALTA | CRÍTICO |

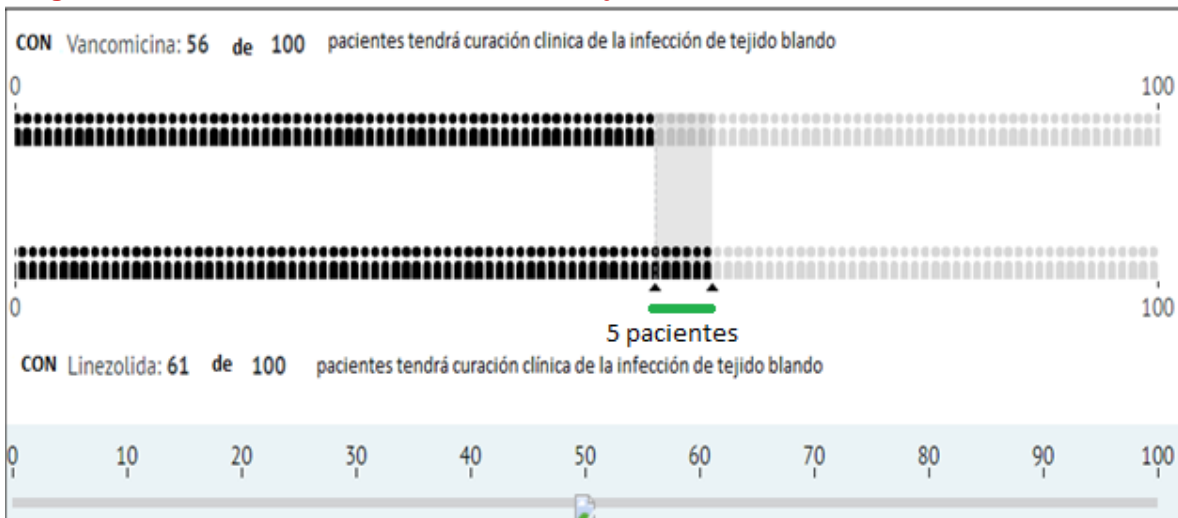
Fuente: Yue et al., 2016

Valor de Reducción Absoluta del Riesgo (RAR es de: 5,38% IC 95% (1,94% a 8,84%) y Valor de NNT: 20 IC 95%: (11 a 52)

Se puede decir que de cada 100 pacientes tratados con Linezolid se curan 61 y de cada 100 pacientes tratados con Vancomicina se curan 56. Si se calcula el NNT esto nos da aproximadamente 20, es decir se deben tratar a 20 pacientes con Linezolid para curar un paciente de infección. Ahora, al pensarlo en términos de 100 pacientes, serían 5 pacientes (un paciente de cada 20).

La tabla indica que en términos absolutos hay curación clínica de 5 pacientes de cada 100 (entre 2 mínimo y 9 máximo) más con Linezolid, en comparación con Vancomicina (basados en datos de 3114 pacientes de 9 estudios), lo cual se muestra de la siguiente manera:

Figura 9 Curación clínica de la infección de tejido blando con Vancomicina vs. Linezolid



Fuente: Yue et al. Herramienta GradePro, 2016

Análisis 1: Con el umbral del valor elegido por el grupo: 15 pacientes más de cada 100

RAR: 15,22% IC 95% (11,9% a 18,59%)

NNT: 7 IC 95% (5 a 8)

- ¿Se curan por lo menos 15 pacientes más de cada 100 con la intervención (Linezolid) en comparación con vancomicina?

No, solamente se curan 5 de cada 100, luego no se alcanza el valor para definir superioridad.

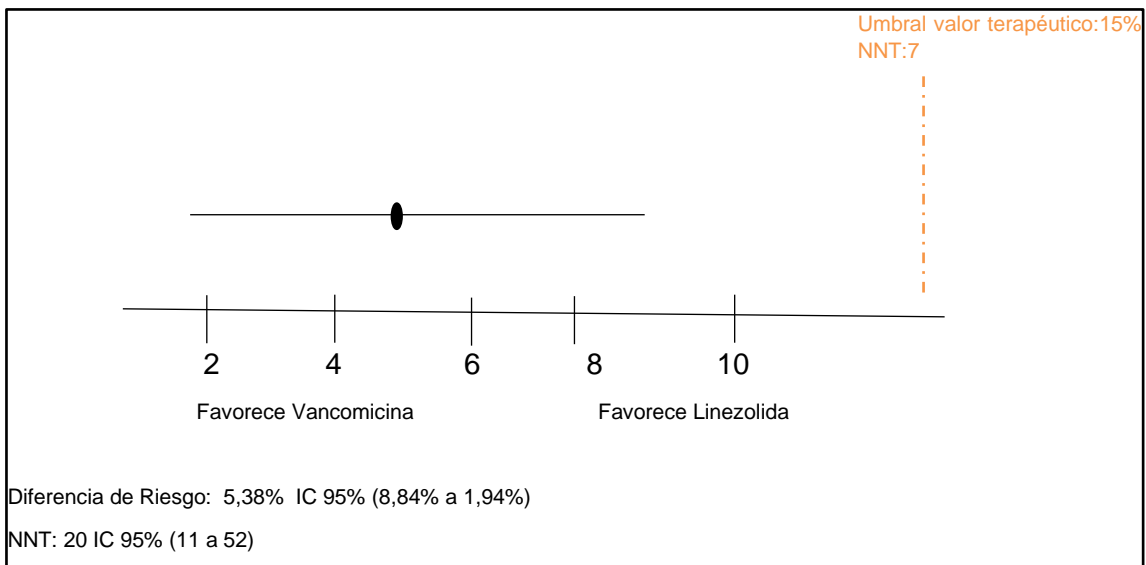
- ¿De cada 7 pacientes que trato con Linezolid se cura por lo menos uno?

No, debo tratar 20 pacientes para por lo menos curar 1 caso, luego no se alcanza el valor para definir superioridad.

- ¿Se curan un 15% más los pacientes tratados con Linezolid comparados con los tratados por vancomicina?

No, la diferencia de riesgo entre los dos medicamentos es del 5%, luego no se alcanza el valor para definir superioridad.

Figura 10 Umbral valor terapéutico vancomicina vs. linezolid



Fuente: elaboración IETS

Análisis 2: Con un umbral del valor supuesto elegido por el grupo: 4 pacientes más de cada 100

Ahora, si el grupo hubiera considerado que un umbral de 4.

RAR: 4%

NNT: 24

- ¿Se curan por lo menos 4 pacientes más de cada 100 con la intervención (Linezolid) en comparación con vancomicina?

Si, se curan 5 de cada 100, el valor de 4 luego se alcanza el valor para definir superioridad.

- ¿De cada 24 pacientes que trato con Linezolid se cura por lo menos uno?













Si, debo tratar 20 pacientes por lo menos curar 1 caso, luego se alcanza el valor para definir superioridad.




- ¿Se curan un 4% más los pacientes tratados con Linezolid comparados con los tratados por vancomicina?

Si, la diferencia de riesgo entre los dos medicamentos es del 5% luego se alcanza el valor para definir superioridad.

Recordar: Para los desenlaces de seguridad, deben presentarse de manera detallada (cuando aplique incluso la severidad de estos). Ver Figura 12.

Figura 11. Valoración de la severidad de los desenlaces de los tratamientos

| Atributo | Nivel 1 | Nivel 2 | Nivel 3 |
|---------------------------------|---|--|--|
| Náusea y vómito | Leve (1 vez en 24 horas)  | Moderado (2-5 veces en 24 horas)  | Severo (Más de 6 veces en 24 horas)  |
| Tiempo sin progresión del tumor | Alto (14 meses)  | Moderado (11 meses)  | Bajo (7 meses)  |
| Efectos en la piel | Ninguno  | Leve (1/10 partes del cuerpo)  | Moderado (Más de 1/3 del cuerpo)  |
| Diarrea | Leve (1-4 veces al día)  | Moderado (4-6 veces al día)  | Severo (Más de 7 veces día)  |

| Atributo | Nivel 1 | | Nivel 2 | | Nivel 3 | |
|-------------------|-----------------------------------|---|---|--|---|---|
| Cansancio /Fatiga | Leve (Poca influencia en las AVD) |  | Moderado (Influencia importante en las AVD) |  | Severo (Influencia considerable en las AVD) |  |

AVD: Actividades de la Vida Diaria

Fuente: Mülbacher y Bethge, 2015

Siguiendo el ejemplo anterior, (linezolid comparada con vancomicina), se va a evaluar el desenlace de seguridad: náusea severa.

Las preguntas por realizarse son:

- ¿Son X pacientes (más con náusea) importantes desde el punto de vista clínico?
- ¿Cambia la respuesta si fuera con náuseas leves (1 vez en 24 horas)?

Se realiza la discusión.

Se presenta al grupo, que la vancomicina produce dos pacientes con episodios de náuseas por cada 100 pacientes, hasta cuántos pacientes más se espera tener para concluir que linezolid es menos segura

Supuesto: El consenso considera que para que linezolid sea inferior en términos de seguridad, debe tener cinco pacientes más con náuseas que vancomicina, es decir 7 pacientes de cada 100.

RAR: 5% IC 95% (3,2% a 6,7%)

NNH: 20 IC 95% (15 a 31)

Es decir, que para el grupo debería haber una diferencia absoluta del riesgo del 5% (Efecto absoluto 5 pacientes más por cada 100)

A continuación, se presentan los valores verdaderos del estudio clínico:

Figura 12. Valoración tabla GRADE de desenlace de seguridad

| Certainty assessment | | | | | | | Nº de pacientes | | Efecto | | Certainty | Importancia |
|-------------------------------------|--------------------|-----------------|----------------|---------------------|-------------|-----------------------|-----------------|----------------|-----------------------|----------------------------------|-----------|-------------|
| Nº de estudios | Diseño de estudio | Riesgo de sesgo | Inconsistencia | Evidencia indirecta | Imprecisión | Otras consideraciones | Linezolid | Vancomicina | Relativo (95% CI) | Absoluto (95% CI) | | |
| Náusea (seguimiento: media 10 días) | | | | | | | | | | | | |
| 2 | ensayos aleatorios | no es serio | no es serio | no es serio | no es serio | ninguno | 58/1129 (5.1%) | 23/1103 (2.1%) | RR 2.45 (1.52 a 3.94) | 3 más por 100 (de 1 más a 6 más) | ⊕⊕⊕⊕ ALTA | IMPORTANTE |

CI: Intervalo de confianza ; RR: Razón de riesgo

Fuente: Yue et al 2016.

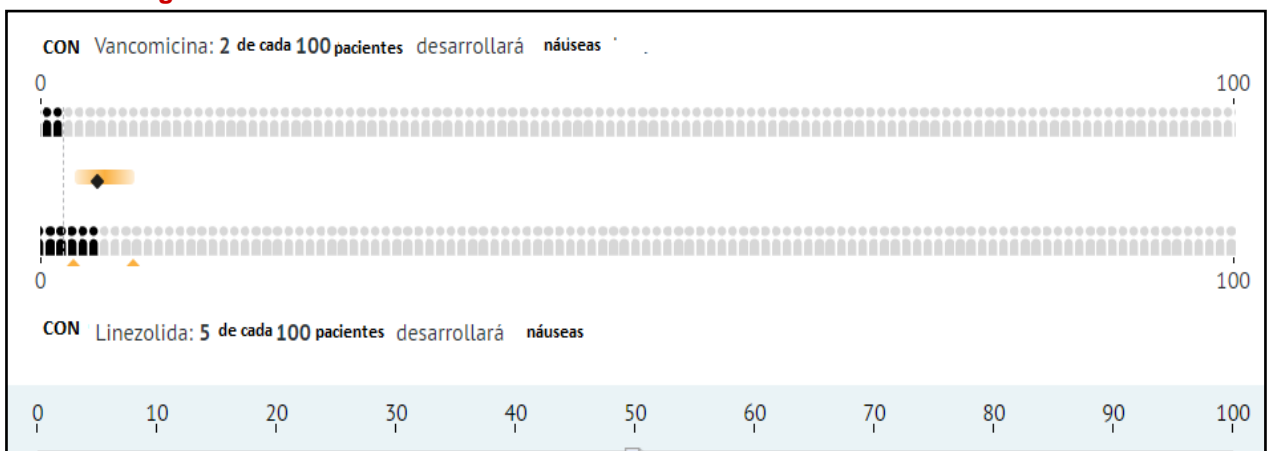
RAR: 3% IC 95% (1,4% a 4,5%)

NNH: 33 IC 95% (22 a 69)

Se puede decir que de cada 100 pacientes tratados con linezolid presentan náusea 5 pacientes y de cada 100 pacientes tratados con Vancomicina, solamente 2.

La tabla indica que en términos absolutos 3 pacientes más de cada 100, tendrán náusea con linezolid, en comparación con vancomicina lo cual se muestra de la siguiente manera:

Figura 13. Frecuencia de náuseas con uso de linezolid vs. vancomicina



Fuente: Yue et al. Herramienta GradePro, 2016

Análisis 1: Con el umbral del valor elegido por el grupo: 5 pacientes más de cada 100

- ¿Las náuseas ocurren en por lo menos 5 pacientes más de cada 100 con la intervención (linezolid) en comparación con vancomicina?

No, solamente las náuseas ocurren en 3 de cada 100 luego no se alcanza el valor del intervalo de confianza para definir superioridad.

- ¿De cada 20 pacientes que trato con linezolid presenta náuseas por lo menos uno?

No, debo tratar 33 pacientes para por lo menos 1 caso presente náuseas, luego no se alcanza el valor para definir superioridad.

- ¿Tienen náusea un 5% más los pacientes tratados con linezolid comparados con los tratados por vancomicina?

No, la diferencia de riesgo entre los dos medicamentos es del 3%, luego no se alcanza el valor para definir superioridad.

Tabla 3. Presentación de resultados del valor umbral terapéutico

| Desenlace | Medida de asociación | Valor de umbral terapéutico |
|-------------------------------|-------------------------------------|---|
| Curación clínica | RR: 1.09 IC 95% (1,03 a 1,16) | 15 pacientes más que el comparador por cada 100 pacientes |
| Náuseas | RR: 2.45 IC 95% (1,52 a 3,94) | 5 pacientes más que el comparador por cada 100 pacientes |
| Colesterol total | DM: 0.54 mg/dL IC 95% (0,04 a 1,04) | 2 mg /dL |
| Número de Kilogramos perdidos | DM: -7.43 IC 95% (-5 a -13) | 10 Kg |

RR: Riesgo Relativo, IC: Intervalo de Confianza, DM: Diferencia de Medias

Fuente: elaboración IETS

10. Formulación de las conclusiones

El grupo desarrollador deberá presentar las conclusiones de la evaluación en términos de:

- Efectividad clínica o validez diagnóstica
- Seguridad
- Significancia clínica de los resultados

- Calidad de la evidencia

Las conclusiones deberán formularse con base en los hallazgos para los desenlaces críticos e importantes en la toma de decisiones, según lo descrito en el apartado de síntesis de la evidencia del presente manual.

Deberán dar cuenta de si las tecnologías de interés frente a sus comparadores tienen menor, similar o mayor efectividad, seguridad o validez diagnóstica, o precisar que no se identificó evidencia o la evidencia disponible es no concluyente. En todos los casos, debe especificarse la población, la comparación y el desenlace. Deberá incluirse la calidad de la evidencia por desenlace y la calidad global del cuerpo de la evidencia.

11. Revisión por pares

Una vez termina la evaluación, se sugiere que la misma sea valorada en su calidad metodológica por un grupo revisor, que debe estar conformado por expertos temáticos y expertos metodológicos.

Se sugiere establecer controles de calidad empleando estrategias como:

- Replicación de la estrategia de búsqueda.
- Verificación y auditoria de los estudios excluidos y de las razones de exclusión.
- Auditoría de la combinación metodológica y estadística de resultados de los estudios en caso de realizar metanálisis de combinaciones directas, múltiples e indirectas o *network* metanálisis.
- Auditoría de la evaluación de la calidad metodológica de los estudios incluidos.
- Auditoría de la evaluación de la calidad global de la evidencia.
- Verificación de las conclusiones y recomendaciones sobre la tecnología.

Capítulo 5. Guía para el reporte de protocolos (versión abreviada) e informes de evaluación rápida de tecnologías en salud

Para la presentación del protocolo, el grupo desarrollador deberá verificar que se incluyan todos los ítems descritos en la siguiente lista de chequeo:

Tabla 4. Lista de chequeo para la presentación de protocolo por el grupo desarrollador

| Sección | Contenidos | Verificación (cumple, no cumple) |
|-----------|--|----------------------------------|
| Portada | Título de la evaluación. | |
| | Mes y año de publicación. | |
| Protocolo | Objetivo y alcance de la evaluación. | |
| | Pregunta de investigación en estructura PICOT. | |
| | Criterios de elegibilidad de los estudios. | |
| | Preguntas de consulta pública. | |
| | Referencias bibliográficas. | |

Fuente: elaboración IETS

Nota: en todo caso, los protocolos deberán hacer referencia explícita al uso de los métodos propuestos en el presente manual.

Para la presentación del reporte final, el grupo desarrollador deberá verificar que se incluyan todos los ítems descritos en la herramienta “Estándares de reporte IETS para informes de evaluación rápida de tecnologías en salud” (86).

Esta lista de chequeo pretende servir como un instrumento que apoya a verificar en qué medida, un informe de evaluación rápida de tecnología en salud incorpora los contenidos relevantes, esto como una primera aproximación a la calidad metodológica del informe. La herramienta citada fue construida con base en: la declaración *PRISMA* (45), la herramienta *AMSTAR* (87), el Listado de comprobación para informes de Evaluación de Tecnologías Sanitarias de *INAHTA* (88), los principios propuestos por Drummond y cols. en su documento “*Key principles for the improved conduct of health technology assessments for resource allocation decisions*” (89), una adaptación del documento “Core Model” de *EUnetHTA* (90) y la herramienta *GRADE Evidence to Decision (EtD) frameworks* (32).

Tabla 5. Estándares de reporte IETS para informes de evaluación rápida de tecnologías en salud

| Sección | Contenidos | Verificación (cumple, no cumple o no aplica) |
|--|--|--|
| Portada | Título de la evaluación. | |
| | Número de identificación de la evaluación. | |
| | Mes y año de publicación. | |
| 1. Información general del proyecto | Tipo de evaluación. | |
| | Grupo desarrollador (autores, afiliación y función dentro de la evaluación). | |
| | Revisión por pares. | |
| | Agradecimientos. | |
| | Dirección técnica del proyecto. | |
| | Entidad que solicita el estudio. | |
| | Fuentes de financiación. | |
| | Declaración de conflictos de interés. | |
| | Declaración de independencia editorial. | |
| | Derechos de autor. | |
| | Citación. | |
| | Correspondencia. | |
| | Tabla de contenido. | |
| | Listado de siglas y abreviaturas. | |
| 2. Resumen para el tomador de decisiones | Título de la evaluación. | |
| | Autores. | |
| | Objetivo y alcance. | |
| | Pregunta de evaluación. | |
| | Condición de interés | |
| | ¿Cuál es la tecnología evaluada? | |
| | ¿Para qué uso fue evaluada la tecnología? | |
| | ¿Cómo fue elaborada esta evaluación de tecnología? | |
| | Resumen de evidencia | |
| | Resultados | |
| | ¿Qué tan importantes son los efectos deseados? | |
| | <ul style="list-style-type: none"> ▪ Juicio del panel de valoración* ▪ Conclusiones a partir de la evidencia científica. | |
| | ¿Qué tan importantes son los efectos indeseados? | |
| | <ul style="list-style-type: none"> ▪ Juicio del panel de valoración* ▪ Conclusiones a partir de la evidencia científica. | |
| | ¿Cuál es la confianza global en la evidencia sobre los efectos? | |
| | <ul style="list-style-type: none"> ▪ Juicio del panel de valoración* ▪ Conclusiones a partir de la evidencia científica. | |
| | ¿El balance entre los efectos deseados e indeseados favorece a la intervención o a la comparación? | |
| | <ul style="list-style-type: none"> ▪ Juicio del panel de valoración* ▪ Conclusiones a partir de la evidencia científica. | |
| | <ul style="list-style-type: none"> ▪ Juicio del panel de valoración* ▪ Conclusiones a partir de la evidencia científica. | |

| Sección | Contenidos | Verificación (cumple, no cumple o no aplica) |
|--|---|---|
| | <p>¿Cuál es la evidencia científica clave?</p> <ul style="list-style-type: none"> ▪ Perfiles de evidencia <i>GRADE</i>. <p>Recomendaciones sobre la tecnología Justificación de la recomendación. Consideraciones a tener en cuenta para el análisis de la recomendación. Referencias bibliográficas de los estudios seleccionados para la síntesis de la evidencia. Declaración de conflictos de interés.</p> | |
| <p>3. Objetivo y alcance de la evaluación</p> | <p>Objetivo y alcance. Proceso de formulación y refinamiento de la pregunta de investigación Pregunta de investigación en estructura PICOT (en formato de cuadro)</p> | |
| <p>4. La condición de salud y uso actual de la tecnología</p> | <p>Metodología. Hallazgos sobre los elementos de evaluación:</p> <ul style="list-style-type: none"> ▪ ¿Para qué condiciones de salud y con qué propósito se utiliza la tecnología objeto de la evaluación? ▪ ¿Cuál es la condición de salud objeto de la evaluación? ▪ ¿Cuál es la población elegible para el uso de la tecnología bajo evaluación? ▪ ¿Cuál es la historia natural de la condición de salud objeto de la evaluación? ▪ ¿Cuáles son los síntomas de la condición de salud objeto de la evaluación? ▪ ¿Cuál es la carga de la enfermedad asociada con la condición de salud objeto de la evaluación? ▪ ¿Cuáles son los desenlaces a los que se dirige la tecnología objeto de la evaluación? ▪ ¿Qué tanto se utiliza actualmente la tecnología objeto de la evaluación? ▪ ¿Cuáles son las otras alternativas típicas o comunes para la tecnología objeto de la evaluación? ▪ ¿Cómo se diagnostica actualmente la condición de salud objeto de la evaluación? (este ítem solamente aplica para las evaluaciones de pruebas diagnósticas). ▪ ¿Quién decide qué población es elegible para el uso de la tecnología objeto de la evaluación? ▪ ¿Es la tecnología objeto de la evaluación: un nuevo modo de atención, un complemento, una modificación o el reemplazo de un modo estándar de atención? | |
| <p>5. La tecnología y sus características</p> | <p>Metodología. Hallazgos sobre los elementos de evaluación:</p> | |

| Sección | Contenidos | Verificación (cumple, no cumple o no aplica) |
|---|---|---|
| | <ul style="list-style-type: none"> ▪ ¿Qué es y cómo se clasifica la tecnología objeto de la evaluación y sus comparadores? ▪ ¿Cuál es el código de identificación según el estándar semántico vigente para el país, o cuál es el código único de procedimientos en salud de la tecnología objeto de la evaluación? ▪ ¿Cuál es el estatus de autorización para la comercialización de la tecnología objeto de la evaluación? ▪ ¿Quién fabrica la tecnología objeto de la evaluación? ▪ ¿Cuál es la fase de desarrollo de la tecnología objeto de la evaluación? ▪ ¿Cuál es la presentación y nombre comercial de la tecnología objeto de la evaluación? ▪ ¿Cuál es el mecanismo de acción de la tecnología objeto de la evaluación? ▪ ¿Cuál es la forma farmacéutica y la concentración de la tecnología objeto de la evaluación? (este ítem solo aplica para la evaluación de medicamentos). ▪ ¿Cuál es la forma de administración o aplicación de la tecnología objeto de la evaluación? ▪ ¿Cuál es la dosificación para el uso de la tecnología objeto de la evaluación? (este ítem solo aplica para la evaluación de medicamentos). ▪ ¿Cuáles son las contraindicaciones de la tecnología objeto de la evaluación? ▪ ¿Cuáles son las precauciones de uso de la tecnología objeto de la evaluación? ▪ ¿En qué contexto y en qué nivel de atención es utilizada la tecnología objeto de la evaluación? ▪ ¿Qué equipos o suministros se necesitan para usar la tecnología objeto de la evaluación? <p>Nota: las anteriores preguntas deberán desarrollarse en el contexto específico del uso de la tecnología para la condición de salud de interés de la evaluación.</p> | |
| 6. Análisis de efectividad clínica y seguridad | Presentación del protocolo. | |
| | Resumen de la metodología (según lo definido en el presente manual). | |
| | Diferencias entre el protocolo y la revisión. <ul style="list-style-type: none"> ▪ ¿Qué diferencias se encontraron en el desarrollo de la revisión respecto a lo planteado en el protocolo? | |
| | Hallazgos: <ul style="list-style-type: none"> ▪ Diagrama de flujo <i>PRISMA</i>. | |

| Sección | Contenidos | Verificación (cumple, no cumple o no aplica) |
|---|--|---|
| | <ul style="list-style-type: none"> ▪ Perfiles de evidencia <i>GRADE</i> (para cada comparación de interés). ▪ Síntesis de la evidencia en lenguaje plano. ▪ Resultados de umbral de valor terapéutico ▪ Cuadro resumen de las alertas identificadas en la vigilancia post-comercialización. | |
| 7. Conclusiones | <ul style="list-style-type: none"> ▪ Efectividad clínica ▪ Seguridad ▪ Balance entre beneficios y riesgos ▪ Calidad de la evidencia ▪ Umbral de valor terapéutico (si aplica)** | |
| 8. Recomendación | <p>Metodología para la valoración*</p> <p>Recomendación*</p> | |
| 9. Consideraciones a tener en cuenta para el análisis de la recomendación | <p>Aspectos técnicos, éticos, legales, organizacionales, sociales, entre otros, que el grupo desarrollador de la evaluación considere pertinentes para el análisis de la recomendación generada por el panel de valoración de la tecnología.</p> | |
| 10. Limitaciones de la evaluación | <p>Declaración de potenciales sesgos.</p> | |
| Referencias bibliográficas | <p>Listado completo de referencias bibliográficas en estilo <i>Vancouver</i>.</p> | |
| Anexos | <ul style="list-style-type: none"> ▪ Análisis de comentarios al protocolo de la revisión. ▪ Metodología y hallazgos para el dominio de efectividad clínica y seguridad: <ul style="list-style-type: none"> - Criterios de elegibilidad de los estudios. - Métodos para la búsqueda de información. - Bitácoras de búsqueda en bases electrónicas de datos. - Métodos para la selección de los estudios. - Listado de los estudios incluidos (estilo <i>Vancouver</i>). - Listado de los estudios excluidos (estilo <i>Vancouver</i>) y las razones para su exclusión. - Métodos para la recolección de datos. - Características de los estudios incluidos (solo aplica cuando se presentan metanálisis desarrollados <i>de novo</i>). - Métodos para la valoración del riesgo de sesgos. - Riesgo de sesgos de los estudios incluidos. - Métodos para la síntesis de los hallazgos. - Figuras de la red de intervenciones (solo aplica cuando se presentan metanálisis de comparaciones indirectas o metanálisis en red). | |

| Sección | Contenidos | Verificación (cumple, no cumple o no aplica) |
|---------|---|---|
| | <ul style="list-style-type: none"> - Figuras de bosque (solo aplica cuando se presentan metanálisis desarrollados <i>de novo</i>). ▪ Registros sanitarios de la tecnología objeto de la evaluación. ▪ Listas de chequeo correspondientes | |

* La presentación de esta información estará a cargo únicamente del IETS.

**Siempre debe realizarse en informes en los que el objetivo es determinar el valor terapéutico de una tecnología (por ejemplo, en términos de su inferioridad, no diferencias o superioridad frente a un comparador) como en el caso de las evaluaciones de tecnologías candidatas a financiación con presupuestos máximos.

Fuente: Adaptado de PRISMA, herramienta AMSTAR, Listado de comprobación para informes de Evaluación de Tecnologías Sanitarias de INAHTA, “*Key principles for the improved conduct of health technology assessments for resource allocation decisions*”, “Core Model” de EUnetHTA, y la herramienta GRADE Evidence to Decision (EtD) frameworks.

Referencias bibliográficas

1. Sánchez AMB, Roces LV. Efectividad, seguridad y aceptabilidad del tratamiento con dimetil fumarato en pacientes con esclerosis múltiple recurrente-remitente en el Hospital Universitario Central de Asturias (Oviedo) que iniciaron tratamiento en 2015. *FarmaJournal*. 2017;2(2):39-50.
2. A Pichon-Riviere, N Soto, F Augustovski, LS-C. SG-M. Involucramiento de los actores relevantes en el proceso de evaluación de tecnología sanitarias. Segundo Foro Latinoamericano de Políticas de Evaluación de Tecnologías Sanitarias ed. Lima, Perú 2017.
3. Baltussen R, Marsh K, Thokala P, Diaby V, Castro H, Cleemput I, et al. Multicriteria Decision Analysis to Support Health Technology Assessment Agencies: Benefits, Limitations, and the Way Forward. *Value in Health*. 2019;22(11):1283-8.
4. Schünemann H, Brozek J, Guyatt G, Oxman A. Manual GRADE para calificar la calidad de la evidencia y la fuerza de la recomendación (1a Ed. en español). P.A. Orrego & M.X. Rojas (Trans.) Mar 2017. Publicación Original: <http://gdt.guidelinedevelopment.org/app/handbook/handbook.html2013>.
5. INAHTA Hea. Health Technology Assessment International (HTAi) "HTA Glossary.net." [Available from: [http://htaglossary.net/INAHTA+\(es\)](http://htaglossary.net/INAHTA+(es))].
6. Community C. Glossary 2020 [Available from: <https://community.cochrane.org/glossary>].
7. Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, et al. The GRADE Working Group clarifies the construct of certainty of evidence. *Journal of clinical epidemiology*. 2017;87:4-13.
8. Sanzone M. Contraindication. In: Kreutzer JS, DeLuca J, Caplan B, editors. *Encyclopedia of Clinical Neuropsychology*. New York, NY: Springer New York; 2011. p. 701-2.
9. Donado Gómez JH, Jiménez Cotes EA, Cardona LM. Diferencia mínima clínicamente importante en algunas escalas de medición utilizadas en enfermedades musculoesqueléticas. *Revista Colombiana de Reumatología*. 2014;21(3):125-32.
10. Porzolt F, Rocha NG, Toledo-Arruda AC, Thomaz TG, Moraes C, Bessa-Guerra TR, et al. Efficacy and effectiveness trials have different goals, use different tools, and generate different messages. *Pragmat Obs Res*. 2015;6:47-54.
11. Majors R, Carberry K, Ransaw TS. The international handbook of black community mental health. *The international handbook of black community mental health*. 2020.
12. Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG, Consort Group ft. Reporting of Noninferiority and Equivalence Randomized Trials: Extension of the CONSORT 2010 Statement. *JAMA*. 2012;308(24):2594-604.
13. ClinicalTrials.gov. Description: Phase [Available from: <https://www.clinicaltrials.gov/ct2/help/glossary/phase>].
14. Wang B, Wang H, Tu XM, Feng C. Comparisons of Superiority, Non-inferiority, and Equivalence Trials. *Shanghai Arch Psychiatry*. 2017;29(6):385-8.
15. Moreno M MA, Casto H. . Manual para la elaboración de evaluaciones económicas en salud. . 2014.
16. Berger ML, Sox H, Willke RJ, Brixner DL, Eichler HG, Goettsch W, et al. Good Practices for Real-World Data Studies of Treatment and/or Comparative Effectiveness: Recommendations from the Joint ISPOR-ISPE Special Task Force on Real-World Evidence in Health Care Decision Making. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*. 2017;20(8):1003-8.
17. Instituto de Evaluación Tecnológica en Salud. Manual metodológico de procesos participativos. Bogotá DC., 2014.

18. Oortwijn W, Jansen M, Baltussen R. Evidence-informed deliberative processes. A practical guide for HTA agencies to enhance legitimate decision-making. Version 1.0. Nijmegen: Radboud University Medical Centre. Radboud Institute for Health Sciences. 2019.
19. Simundic A-M. Measures of Diagnostic Accuracy: Basic Definitions. *Med Biol Sci.* 2008;22.
20. Mariani AW, Pêgo-Fernandes PM. Statistical significance and clinical significance. *Sao Paulo Medical Journal.* 2014;132:71-2.
21. Fethney J. Statistical and clinical significance, and how to use confidence intervals to help interpret both. *Australian Critical Care.* 2010;23(2):93-7.
22. Institute JB. JBI Manual for Evidence Synthesis. Aromataris E MZE, editor2020.
23. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic research: principles and quantitative methods:* John Wiley & Sons; 1982.
24. Tenny S AI. *Statistical Significance Treasure Island (FL): StatPearls Publishing; 2022* [Available from: <https://www.ncbi.nlm.nih.gov/books/NBK459346/>].
25. Aarts S, van den Akker M, Winkens B. The importance of effect sizes. *Eur J Gen Pract.* 2014;20(1):61-4.
26. *Frontiers. Research Topic: Social Value for Health: Understanding, Framing and Capturing the Wider Psycho-social Outcomes Of Health Interventions and Programmes.*
27. Institute of Medicine. *Initial national priorities for comparative effectiveness research.* Washington, DC: National Academies Press; 2009.
28. Excellence NifHC. *Developing NICE guidelines: the manual: National Institute for Health and Care Excellence (NICE); 2015.*
29. Carrasquilla G, Pulido A, De la Hoz A, Mieth K, Muñoz O, Guerrero R. *Guía Metodológica para la elaboración de guías de práctica clínica con evaluación económica en el Sistema General de Seguridad Social en Salud Colombiano. . Segunda edición, Versión completa final. ed. Bogotá2014*
30. Health Nif, Clinical E. *The guidelines manual. Process and methods. 2012.*
31. IETS. *Guía para el involucramiento de actores en evaluaciones de tecnologías en salud2020.*
32. Alonso-Coello P, Schunemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. *GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. Bmj.* 2016;353:i2016.
33. Instituto de Evaluación Tecnológica en Salud. *Manual metodológico de procesos participativos. Bogotá DC.,2014.*
34. *Presentación: prevención y control de infecciones y nuevo coronavirus (COVID-19): precauciones estándares y uso de equipos de protección personal* [Available from: <https://www.paho.org/es/documentos/presentacion-prevencion-control-infecciones-nuevo-coronavirus-covid-19-precauciones>].
35. *Assessment ENfHT. Comparators & comparisons. Criteria for the choice of the most appropriate comparator(s) Summary of current policies and best practice recommendations. 2015.*
36. *Assessment ENfHT, Model HC. for Rapid Relative Effectiveness Assessment of Pharmaceuticals. Version 3.0. 2013 [cited 2015-10-30].*
37. Group. BDW. *Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther.* 2001;69(3):89-95.
38. *European Network for Health Technology Assessment (EunetHta). Guideline Endpoints used for Relative Effectiveness Assessment: Clinical Endpoints. Adapted version (2015) [Internet]: Eunethta; 2015 [cited 2020 Marzo 09]. Available from: <https://eunethta.eu/methodology-guidelines/>.*

39. Institute for Quality and Efficiency in Health Care (IQWiG). General Methods. Version 5.0 of 10 July 2017. Colonia (Alemania): IQWiG; 2017 [cited 2020 Marzo 09]. Available from: <https://www.iqwig.de/en/methods/methods-paper.3020.html>.
40. European network for health technology assessment (EUnetHTA). Guideline Endpoints used in Relative Effectiveness Assessment: Surrogate Endpoints. Adapted version (2015). [Internet]: EUnetHTA; 2015 [cited 2020 Marzo 09]. Available from: <https://eunetha.eu/methodology-guidelines/>.
41. European network for health technology assessment (EUnetHTA). Guideline Endpoints used for Relative Effectiveness Assessment Composite endpoints. Adapted version (2015). [Internet]: EUnetHTA; 2015 [cited 2020 Marzo 09]. Available from: <https://eunetha.eu/methodology-guidelines/>.
42. European network for health technology assessment (EUnetHTA). Guideline Endpoints used in Relative Effectiveness Assessment SAFETY. Adapted version (2015). [Internet]: EUnetHTA; 2015 [cited 2020 Marzo 09]. Available from: <https://eunetha.eu/methodology-guidelines/>.
43. Higgins JPT, Cochrane Collaboration. Cochrane handbook for systematic reviews of interventions. Second edition. ed. Hoboken, NJ: Wiley-Blackwell; 2020. pages cm p.
44. PRISMA. Planning a systematic review? Think protocols <http://prisma-statement.org/Protocols/WhyProtocols2015> [updated Jan 2015].
45. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol*. 2009;62(10):e1-34.
46. Group CBM. RoB 2: A revised Cochrane risk-of-bias tool for randomized trials 2020 [cited 2020. Available from: <https://methods.cochrane.org/bias/resources/rob-2-revised-cochrane-risk-bias-tool-randomized-trials>.
47. Wells GA, Wells G, Shea B, Shea B, O'Connell D, Peterson J, et al., editors. The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomised Studies in Meta-Analyses 2014.
48. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions 5.1.0 [updated March 2011] 2011.
49. Shojania KG, Sampson M, Ansari MT, Ji J, Garritty C, Rader T, et al. Updating Systematic Reviews. Rockville MD 2007 Sep.
50. Lavis JN, Oxman AD, Grimshaw J, Johansen M, Boyko JA, Lewin S, et al. SUPPORT Tools for evidence-informed health Policymaking (STP) 7: Finding systematic reviews. *Health research policy and systems*. 2009;7 Suppl 1:S7.
51. de Helsinki D. Principios éticos para las investigaciones médicas en seres humanos. Adoptada por la. 2004;18:20-3.
52. Lavis JN, Boyko JA, Oxman AD, Lewin S, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP) 14: Organising and using policy dialogues to support evidence-informed policymaking. *Health research policy and systems*. 2009;7(1):S14.
53. Pavy JJ, Denekamp J, Letschert J, Littbrand B, Mornex F, Bernier J, et al. Late effects toxicity scoring: The soma scale. *International Journal of Radiation Oncology • Biology • Physics*. 1995;31(5):1043-7.
54. Puljak L. If there is only one author or only one database was searched, a study should not be called a systematic review. *J Clin Epidemiol*. 2017;91:4-5.
55. Echeverry R. J, Pardo T. R. Diseños de investigación clínica. Estudios observacionales: Reporte de casos y serie de casos. *Revista Colombiana de Cardiología*. 1999;7(1):15-20.

56. Díaz MH, Cañón L, Pinzón C. Análisis del espectro eficacia-efectividad en estudios de intervenciones en salud. Bogotá, D.C.: Instituto de Evaluación Tecnológica en Salud - IETS; 2018.
57. Whiting PF, Rutjes AS, Westwood ME, et al. Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*. 2011;155(8):529-36.
58. Cortés-Reyes É, Rubio-Romero JA, Gaitán-Duarte H. Métodos estadísticos de evaluación de la concordancia y la reproducibilidad de pruebas diagnósticas. *Revista Colombiana de Obstetricia y Ginecología*. 2010;61(3):247-55.
59. Glasziou PP, Shepperd S, Brassey J. Can we rely on the best trial? A comparison of individual trials and systematic reviews. *BMC medical research methodology*. 2010;10(1):23.
60. Instituto de Evaluación Tecnológica en Salud - IETS. Manual de procesos participativos. Bogotá D.C.2014.
61. Shekelle PG, Motala A, Johnsen B, Newberry SJ. Assessment of a method to detect signals for updating systematic reviews. *Systematic reviews*. 2014;3:13.
62. Whiting P, Savovic J, Higgins JP, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol*. 2016;69:225-34.
63. Schmetz A, Emich H. The IQWiG Checklist for Indirect Comparison and Network Meta-Analyses. *The Evidence Forum: Institute for Quality and Efficiency in Healthcare (IQWiG)*; 2017.
64. Jansen JP, Trikalinos T, Cappelleri JC, Daw J, Andes S, Eldessouki R, et al. Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*. 2014;17(2):157-73.
65. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *Bmj*. 2019;366:l4898.
66. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *Bmj*. 2016;355:i4919.
67. Institute JB. JBI's critical appraisal tools assist in assessing the trustworthiness, relevance and results of published papers The University of Adelaide.
68. Pathela P, Zahid Hasan K, Roy E, Huq F, Kasem Siddique A, Bradley Sack R. Diarrheal illness in a cohort of children 0–2 years of age in rural Bangladesh: I. Incidence and risk factors. *Acta Paediatrica*. 2006;95(4):430-7.
69. Chaimani A, Caldwell D, Li T, Higgins J, Salanti G. Undertaking network metaanalyses. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions version 62 (updated February 2021)*: Cochrane; 2021.
70. Rochwerg B, Brignardello-Petersen R, Guyatt G. Network meta-analysis in health care decision making. *Med J Aust*. 2018;209:151-4.
71. Brignardello-Petersen R, Bonner A, Alexander PE, Siemieniuk RA, Furukawa TA, Rochwerg B, et al. Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis. *J Clin Epidemiol*.2018:36-44.
72. Brignardello-Petersen R, Bonner A, Alexander PE, Siemieniuk RA, Furukawa TA, Rochwerg B, et al. Corrigendum to "Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis" [*J Clin Epidemiol* 2018;93:36-44]. *J Clin Epidemiol*. 2018;98:162.

73. Brignardello-Petersen R, Mustafa RA, Siemieniuk RAC, Murad MH, Agoritsas T, Izcovich A, et al. GRADE approach to rate the certainty from a network meta-analysis: addressing incoherence. *J Clin Epidemiol.* 2019;108:77-85.
74. Brignardello-Petersen R, Guyatt GH, Mustafa RA, Chu DK, Hultcrantz M, Schünemann HJ, et al. GRADE guidelines 33: Addressing imprecision in a network meta-analysis. *J Clin Epidemiol.* 2021;139:49-56.
75. Brignardello-Petersen R, Murad MH, Walter SD, McLeod S, Carrasco-Labra A, Rochweg B, et al. GRADE approach to rate the certainty from a network meta-analysis: avoiding spurious judgments of imprecision in sparse networks. *J Clin Epidemiol.* 2019;105:60-7.
76. Brignardello-Petersen RA-O, Florez ID, Izcovich A, Santesso N, Hazlewood G, Alhazanni W, et al. GRADE approach to drawing conclusions from a network meta-analysis using a minimally contextualised framework. *Bmj.* 2020;371:m3900.
77. Brignardello-Petersen RA-O, Izcovich A, Rochweg B, Florez ID, Hazlewood G, Alhazanni W, et al. GRADE approach to drawing conclusions from a network meta-analysis using a partially contextualised framework. *Bmj.* 2020;371:m3907.
78. Yepes-Nuñez JJ, Li SA, Guyatt G, Jack SM, Brozek JL, Beyene J, et al. Development of the summary of findings table for network meta-analysis. *J Clin Epidemiol.* 2019;115:1-13.
79. Schünemann HJ HJ, Vist GE, Glasziou P, Akl EA, Skoetz N, Guyatt GH. Chapter 14: Completing 'Summary of findings' tables and grading the certainty of the evidence. . In: Higgins JPT TJ, Chandler J, Cumpston M, Li T, Page MJ, Welch VA editor. *Cochrane Handbook for Systematic Reviews of Interventions version 60 (updated July 2019):* Cochrane; 2019.
80. Singh S, Chang SM, Matchar DB, Bass EB. Chapter 7: grading a body of evidence on diagnostic tests. *Journal of general internal medicine.* 2012;27 Suppl 1:S47-55.
81. Al-Tawfiq JA, Perl TM. Middle East respiratory syndrome coronavirus in healthcare settings. *Curr Opin Infect Dis.* 2015;28(4):392-6.
82. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence—imprecision. *Journal of clinical epidemiology.* 2011;64(12):1283-93.
83. Santesso N, Glenton C, Dahm P, Garner P, Akl EA, Alper B, et al. GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions. *Journal of Clinical Epidemiology.* 2020;119:126-35.
84. Boyko JA, Lavis JN, Abelson J, Dobbins M, Carter N. Deliberative dialogues as a mechanism for knowledge translation and exchange in health systems decision-making. *Social science & medicine.* 2012;75(11):1938-45.
85. Bayón J, Gutierrez-Iglesias M, Vasco L, Gutiérrez-Ibarluzea I. Síntesis de información relevante de apoyo a los MCDA (análisis de decisión multicriterio) para la toma de decisiones 2019.
86. Díaz MH, Cañón L, Pinzón C. Estándares de reporte IETS para evaluaciones rápidas de tecnologías en salud. Bogotá, D.C.: Instituto de Evaluación Tecnológica en Salud - IETS; 2018.
87. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol.* 2009;62(10):1013-20.
88. INAHTA. Listado de comprobación para informes de Evaluación de Tecnologías Sanitarias 2007 [Available from: http://www.inahta.org/wp-content/uploads/2014/04/INAHTA_HTA_Checklist_Español.pdf.
89. Drummond MF, Schwartz JS, Jonsson B, Luce BR, Neumann PJ, Siebert U, et al. Key principles for the improved conduct of health technology assessments for resource

allocation decisions. *International journal of technology assessment in health care*. 2008;24(3):244-58; discussion 362-8.

90. EUnetHTA Joint Action 2 WP. HTA Core Model® version 2.0. 2013.

Anexos

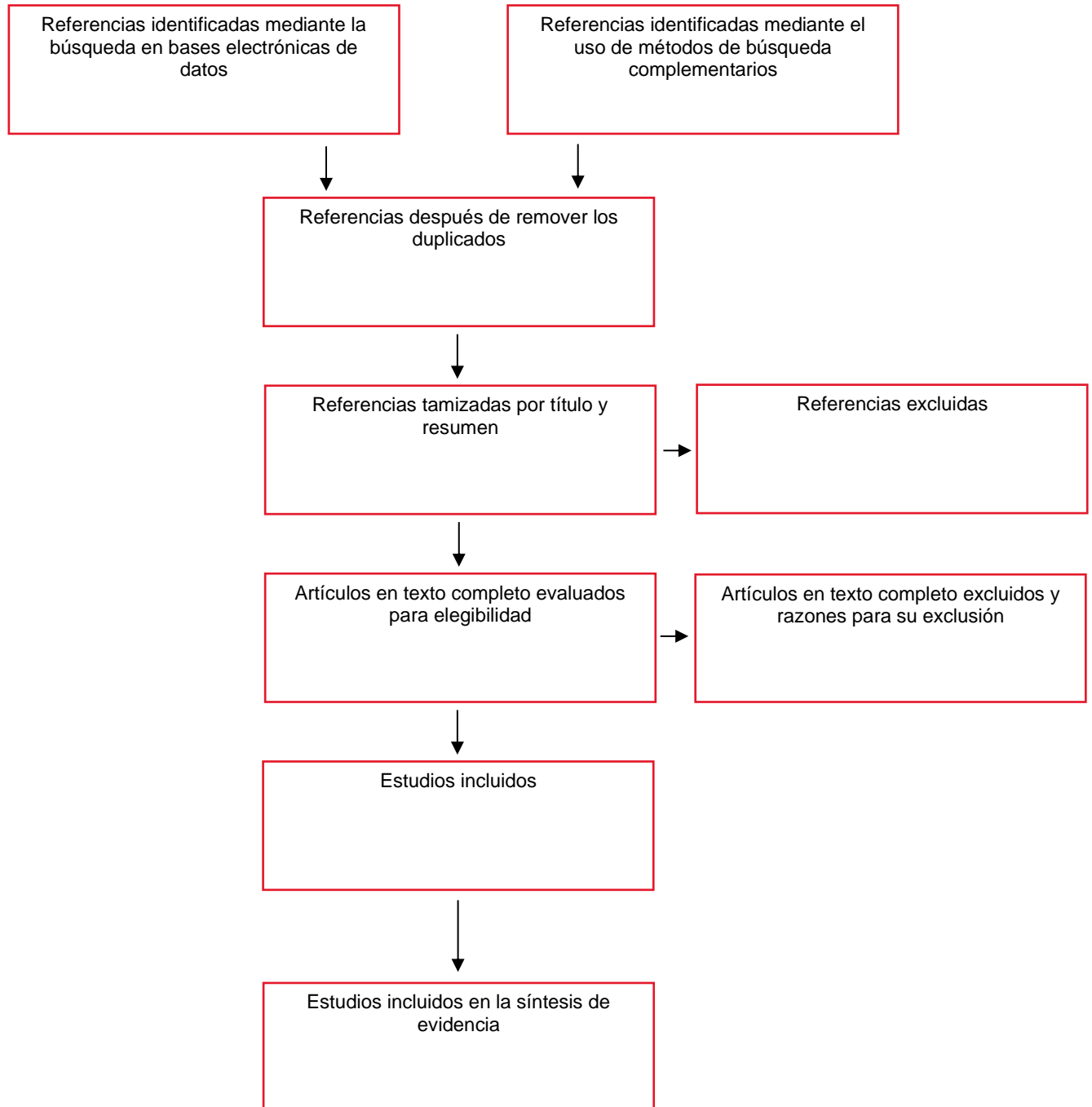
Anexo 1. Registros sanitarios vigentes para las tecnologías de interés

| Núm. | Registro sanitario | Nombre comercial | Consideraciones farmacéuticas | Indicación | Titular registro |
|------|--------------------|------------------|-------------------------------|------------|------------------|
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |

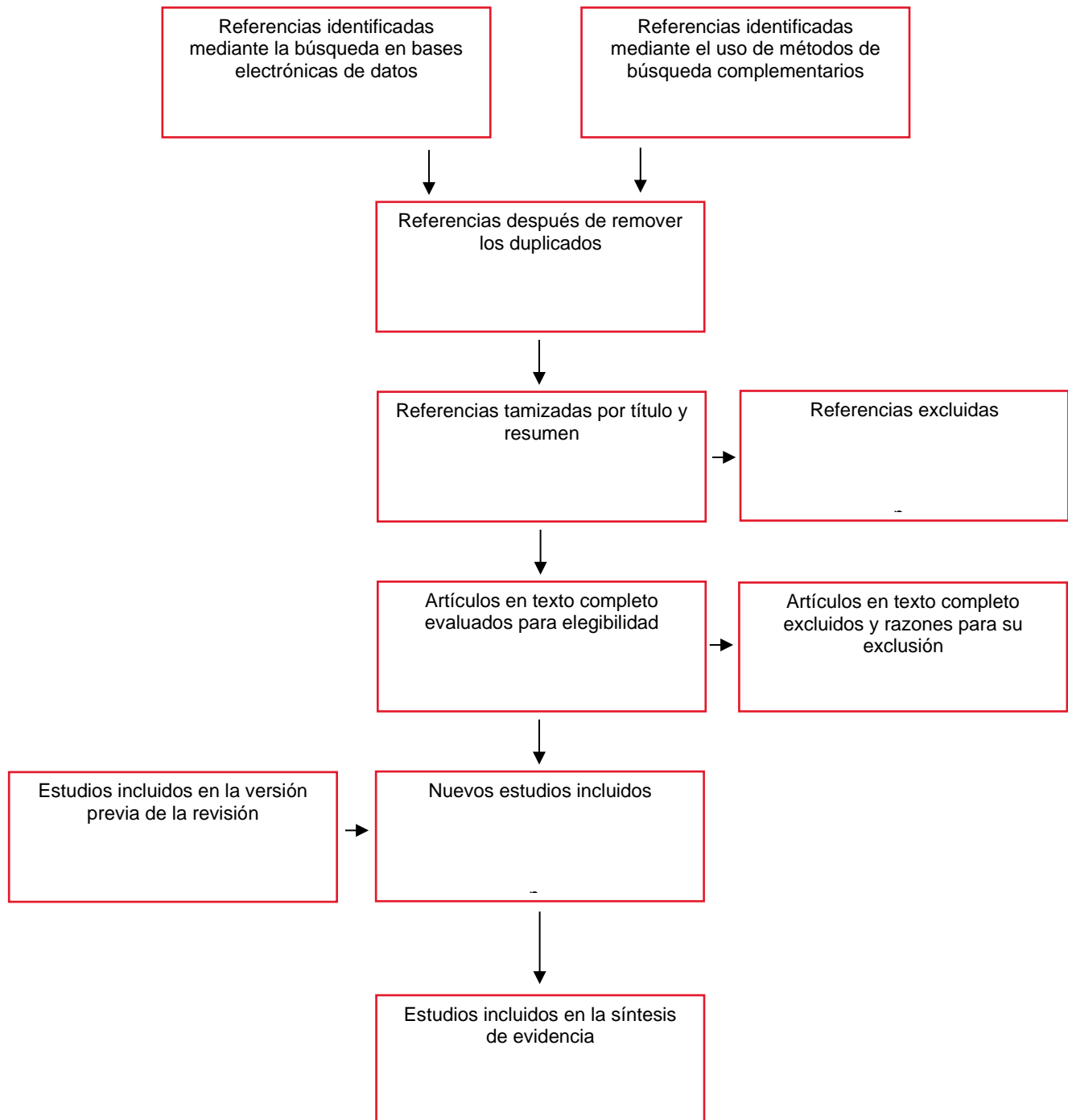
Anexo 2. Reportes de búsqueda de evidencia en bases electrónicas de datos

| | |
|----------------------------|--|
| Tipo de búsqueda | |
| Base de datos | |
| Plataforma | |
| Fecha de búsqueda | |
| Rango de fecha de búsqueda | |
| Restricciones de lenguaje | |
| Otros límites | |
| Estrategia de búsqueda | |
| Referencias identificadas | |

Anexo 3. Diagrama PRISMA: flujo de la búsqueda, tamización y selección de estudios



Anexo 4. Diagrama PRISMA para una actualización: flujo de la búsqueda, tamización y selección de estudios



Anexo 5. Análisis del espectro eficacia-efectividad en estudios de intervenciones en salud

Análisis del espectro eficacia-efectividad en estudios de intervenciones en salud (56)

La herramienta propuesta no es una lista de chequeo ni una escala (no otorga puntajes), es una evaluación basada en dominios, en la cual la valoración es aplicada separadamente por cada criterio relevante.

Estudio:

Dominio N° 1: las características de los participantes.

Aspectos para considerar: la rigurosidad de los criterios de elegibilidad de los participantes, en términos de las variables sociodemográficas (por ejemplo, edad, sexo, etc.), la gravedad de la condición de salud, y la presencia de comorbilidades.

| | | | | |
|---|---|---|--|--|
| | | | | |
| Énfasis fuerte en la eficacia (condiciones controladas) | Énfasis probablemente fuerte en la eficacia (condiciones controladas) | Énfasis balanceado entre la eficacia y la efectividad | Énfasis probablemente fuerte en la efectividad (condiciones de la atención habitual) | Énfasis fuerte en la efectividad (condiciones de la atención habitual) |

Dominio N° 2: el escenario de realización del estudio.

Aspectos para considerar: el número de centros involucrados en el estudio, el ámbito y el nivel de atención médica.

| | | | | |
|---|---|---|--|--|
| | | | | |
| Énfasis fuerte en la eficacia (condiciones controladas) | Énfasis probablemente fuerte en la eficacia (condiciones controladas) | Énfasis balanceado entre la eficacia y la efectividad | Énfasis probablemente fuerte en la efectividad (condiciones de la atención habitual) | Énfasis fuerte en la efectividad (condiciones de la atención habitual) |

Dominio N° 3: la relevancia clínica de la intervención y de sus comparadores.

Aspectos para considerar: la evaluación en el estudio de aquellas intervenciones consideradas las mejores y actuales prácticas de manejo para una determinada condición de salud. En el caso de los metanálisis de comparaciones indirectas y metanálisis en red, es necesario determinar cuáles fueron las comparaciones de interés en el análisis, pues es posible que al momento de configurar la red de tratamientos al comienzo de la revisión, solo existan estudios con un comparador común que no sea de relevancia clínica (por ejemplo, placebo).

| | | | | |
|---|---|---|--|--|
| Énfasis fuerte en la eficacia (condiciones controladas) | Énfasis probablemente fuerte en la eficacia (condiciones controladas) | Énfasis balanceado entre la eficacia y la efectividad | Énfasis probablemente fuerte en la efectividad (condiciones de la atención habitual) | Énfasis fuerte en la efectividad (condiciones de la atención habitual) |
|---|---|---|--|--|

Dominio N° 4: la flexibilidad en el uso de las intervenciones.

Aspectos para considerar: las características de administración de las intervenciones evaluadas (dosis, duración del tratamiento, estrategias de seguimiento, protocolos de adherencia a las intervenciones, etc.) y el uso de cointervenciones.

| | | | | |
|---|---|---|--|--|
| Énfasis fuerte en la eficacia (condiciones controladas) | Énfasis probablemente fuerte en la eficacia (condiciones controladas) | Énfasis balanceado entre la eficacia y la efectividad | Énfasis probablemente fuerte en la efectividad (condiciones de la atención habitual) | Énfasis fuerte en la efectividad (condiciones de la atención habitual) |
|---|---|---|--|--|

Dominio N° 5: otros aspectos relacionados con el diseño y la conducción del estudio.

Aspectos para considerar: estrategias de reclutamiento de los participantes, la duración del estudio, y el uso de análisis por intención de tratar para los desenlaces de interés.

| | | | | |
|---|---|---|--|--|
| Énfasis fuerte en la eficacia (condiciones controladas) | Énfasis probablemente fuerte en la eficacia (condiciones controladas) | Énfasis balanceado entre la eficacia y la efectividad | Énfasis probablemente fuerte en la efectividad (condiciones de la atención habitual) | Énfasis fuerte en la efectividad (condiciones de la atención habitual) |
|---|---|---|--|--|

¿Cuál es el énfasis del estudio?

| | | | | |
|---|---|---|--|--|
| Énfasis fuerte en la eficacia (condiciones controladas) | Énfasis probablemente fuerte en la eficacia (condiciones controladas) | Énfasis balanceado entre la eficacia y la efectividad | Énfasis probablemente fuerte en la efectividad (condiciones de la atención habitual) | Énfasis fuerte en la efectividad (condiciones de la atención habitual) |
|---|---|---|--|--|

Anexo 6. Herramientas para la evaluación de riesgo de sesgos por tipo de diseño.

Evaluación de riesgo de sesgos de revisiones sistemáticas de la literatura

Se realizó por medio de la herramienta *Risk of Bias of Systematic Reviews (RoBIS)* cuyos dominios son:

- Dominio 1: preocupaciones respecto a la especificación de los criterios de elegibilidad de los estudios.
- Dominio 2: preocupaciones respecto a los métodos usados para identificar o seleccionar los estudios.
- Dominio 3: preocupaciones respecto a los métodos usados para recolectar los datos y evaluar los estudios.
- Dominio 4: preocupaciones respecto a la síntesis y resultados.
- Revisión: riesgo de sesgos en la revisión.




En cada uno de los dominios el juicio puede ser: bajo riesgo de sesgo= ; riesgo de sesgo poco claro= , alto riesgo de sesgo = . En el archivo XXXX se puede ver la evaluación detallada de cada uno de los dominios.

Tabla 6. Evaluación de riesgo de sesgo de revisiones sistemáticas de la literatura

| Desenlace | Dominio 1 | Dominio 2 | Dominio 3 | Dominio 4 | Revisión |
|-------------------------|--|---|---|---|---|
| Autor, año (referencia) |  ^a |  |  |  |  |
| Autor, año (referencia) |  |  |  |  |  |
| Autor, año (referencia) |  |  |  |  |  |

Notas

^a No se mencionan los límites de búsqueda...

Ingrese en notas al pie las razones que llevaron a la clasificación “riesgo de sesgo poco claro” o “alto riesgo de sesgo”.

Evaluación de calidad de metanálisis de comparaciones indirectas o en red

Tabla 7. Evaluación de riesgo de sesgo de metanálisis de comparaciones indirectas o en red
 Escoger la herramienta que vaya a emplear (IQWiG o ISPOR) y eliminar el otro formato.

Valoración de la calidad metodológica de revisiones sistemáticas con metanálisis de comparaciones indirectas o metanálisis en red. Herramienta adaptada al español a partir de: (8) (IQWiG)

| Estudio: referencia | | |
|---------------------|---|-----------------|
| | Criterio | Juicio |
| 1. | ¿Se ha establecido <i>a priori</i> la pregunta? <ul style="list-style-type: none"> ▪ Descripción clara de la pregunta. ▪ Transferencia a una hipótesis estadística. ▪ Explicación de las desviaciones a partir del plan establecido originalmente. | Sí |
| 2. | ¿Se ha explicado suficientemente la justificación para el uso de una comparación indirecta? | No ^a |
| 3. | ¿Se ha explicado suficientemente la selección de un comparador común en lugar de una comparación directa? | xxx |
| 4. | ¿Se ha realizado una revisión sistemática y exhaustiva de la literatura, y esta se ha descrito en detalle? <ul style="list-style-type: none"> ▪ ¿Para la intervención de interés primario? ▪ ¿Para el comparador común? | xxx |
| 5. | ¿Se han usado y descrito criterios de inclusión y exclusión inicialmente definidos? | xxx |
| 6. | ¿Hubo un reporte completo de todos los datos relevantes de los estudios? <ul style="list-style-type: none"> ▪ Características de todos los estudios incluidos. ▪ Evaluación de todos los estudios incluidos. ▪ Gráficos de la red, descripción de la geometría de la red. ▪ Para todos los desenlaces relevantes, comparaciones y subgrupos: <ul style="list-style-type: none"> - Resultados individuales de todos los estudios (las estimaciones del efecto y sus correspondientes intervalos de confianza). - Estimaciones del efecto e intervalos de confianza a partir de los metanálisis en parejas. | xxx |
| 7. | ¿Se han investigado los supuestos principales, y se han tratado adecuadamente los resultados de esta investigación? <ul style="list-style-type: none"> ▪ Similitud (transitividad). ▪ Homogeneidad. ▪ Consistencia (coherencia). | xxx |
| 8. | ¿Se han usado herramientas estadísticas adecuadas, y ellas se han descrito con suficiente detalle? <ul style="list-style-type: none"> ▪ Uso de comparaciones indirectas ajustadas. ▪ Tratamiento de estudios con múltiples grupos. ▪ Detalles técnicos (especialmente, cuando se usan modelos bayesianos). ▪ Código de programación. | xxx |

Valoración de la calidad metodológica de revisiones sistemáticas con metanálisis de comparaciones indirectas o metanálisis en red. Herramienta adaptada al español a partir de: (8) (IQWiG)

| | |
|--|-----|
| Estudio: referencia | |
| <ul style="list-style-type: none"> ▪ Análisis de sensibilidad. | |
| 9. ¿Se han descrito las limitaciones y estas se han discutido suficientemente? | xxx |
| <ul style="list-style-type: none"> ▪ Calidad y exhaustividad de la base de datos. ▪ Incertidumbres metodológicas, análisis de sensibilidad. ▪ Conflictos con los supuestos principales. | |
| Valoración global de la calidad metodológica: | xxx |

Notas

^a Comentario que argumenta el juicio

Valoración de la calidad metodológica de revisiones sistemáticas con metanálisis de comparaciones indirectas o metanálisis en red. Herramienta adaptada al español a partir de: (6,7) (ISPOR)

| | Criterio | Autor, año | Autor, año |
|---------------------|---|------------|-----------------|
| Introducción | ¿La justificación y los objetivos del estudio se presentan con claridad? | Sí | No ^a |
| Métodos | ¿La sección de métodos incluye lo siguiente?: -Descripción de los criterios de elegibilidad. -Fuentes de información. -Estrategias de búsqueda. -Proceso de selección de los estudios. -Extracción de datos. -Evaluación de la calidad de los estudios incluidos. | | |
| | ¿Las medidas de resultado son descritas? | | |
| | ¿Existe una descripción de los métodos para el análisis y síntesis de la evidencia? ¿Los métodos descritos incluyen lo siguiente?: - Descripción de los métodos de análisis/modelos. - Manejo de sesgos potenciales /inconsistencia. - Marco de análisis. | | |
| | ¿Los análisis de sensibilidad son presentados? | | |
| Resultados | ¿Los resultados incluyen un resumen de los estudios incluidos en la red de evidencia? ¿Se incluyen datos de estudios individuales? | | |

Valoración de la calidad metodológica de revisiones sistemáticas con metanálisis de comparaciones indirectas o metanálisis en red. Herramienta adaptada al español a partir de: (6,7) (ISPOR)

| | | | |
|-----------------------|---|-------------|-------------|
| | ¿Se incluye la red de estudios? | | |
| | ¿El estudio describe una evaluación del ajuste del modelo? | | |
| | ¿Los modelos competidores son comparables? | | |
| | ¿Los resultados de la síntesis de evidencia (ITC/MTC) son presentados claramente? | | |
| | ¿Los análisis de sensibilidad/de escenarios son presentados? | | |
| Discusión | ¿La discusión incluye lo siguiente?: - Descripción/resumen de los hallazgos principales - Validez interna del análisis - Validez externa - Implicaciones de los resultados para la población blanco | | |
| Calidad global | | /10 Alta | /10 Baja |

Notas

^a Comentario que argumenta el juicio

Evaluación de riesgo de sesgos en experimentos clínicos aleatorizados

Se realizó por medio de la herramienta *Risk of Bias* v2.0 (RoBIS) cuyos dominios son:

- Dominio 1: sesgo derivado del proceso de aleatorización.
- Dominio 2: sesgo debido a desviaciones de las intervenciones previstas.
- Dominio 3: sesgo debido a la falta de datos de resultados.
- Dominio 4: sesgo en la medición del resultado.
- Dominio 5: sesgo en la selección del resultado informado.










En cada uno de los dominios el juicio puede ser: bajo riesgo de sesgo= ; riesgo de sesgo poco claro= , alto riesgo de sesgo = . En el archivo XXXX se puede ver la evaluación detallada de cada uno de los dominios.

Tabla 8. Evaluación de riesgo de sesgo de experimentos clínicos aleatorizados

| Desenlace | Dominio 1 | Dominio 2 | Dominio 3 | Dominio 4 | Dominio 5 | Global |
|--------------------------------|---|---|---|-----------|-----------|--|
| Autor, año (referencia) | | | | | | |
| Nombre del desenlace |  |  |  | | |  ^a |
| Nombre del desenlace |  |  | | | | |
| Autor, año (referencia) | | | | | | |
| Nombre del desenlace | | | | | | |
| Nombre del desenlace | | | | | | |
| Nombre del desenlace | | | | | | |

Notas

^a No se realizó aleatorización...

Diligencie la justificación de la valoración de riesgo global, de manera resumida, preferiblemente numerada consecutivamente.

Evaluación de riesgo de sesgos en estudios de intervención no aleatorizados

Tabla 9. Evaluación de riesgo de sesgo en estudios de intervención no aleatorizados

| Sesgo | Autor, año (ref) | Autor, año (ref) | Autor, año (ref) |
|---|--------------------|--------------------|--------------------|
| Sesgo debido a confusión | Elija un elemento. | Elija un elemento. | Elija un elemento. |
| Sesgo en la selección de los participantes en el estudio | Elija un elemento. | Elija un elemento. | Elija un elemento. |
| Sesgo en la clasificación de las intervenciones | Elija un elemento. | Elija un elemento. | Elija un elemento. |
| Sesgos debido a desviaciones de las intervenciones propuestas | Elija un elemento. | Elija un elemento. | Elija un elemento. |
| Sesgo debido a datos perdidos | Elija un elemento. | Elija un elemento. | Elija un elemento. |
| Sesgo en la medición de los desenlaces | Elija un elemento. | Elija un elemento. | Elija un elemento. |
| Sesgo en la selección de los resultados reportados | Elija un elemento. | Elija un elemento. | Elija un elemento. |
| Sesgo global | Elija un elemento. | Elija un elemento. | Elija un elemento. |

6.4 ROBINS - I: evaluación del riesgo de sesgos de estudios no aleatorizados de intervenciones (66)

1. Especificar la pregunta de revisión

| | |
|---------------------------|--|
| Participantes | |
| Intervención experimental | |
| Comparador | |
| Desenlaces | |

2. Enumerar los dominios de confusión relevantes para todos o la mayoría de los estudios

3. Enumere co-intervenciones que podrían ser diferentes entre los grupos de intervención y que podrían tener un impacto en los resultados

4. Especifique un ensayo clínico aleatorio específico para el estudio

| | |
|---------------------------|--|
| Diseño | Individually randomized / Cluster randomized / Matched (e.g. cross-over) |
| Participantes | |
| Intervención experimental | |
| Comparador | |

5. ¿Es su objetivo para este estudio...?

- Para evaluar el efecto de la asignación a la intervención
- Para evaluar el efecto de comenzar y adherirse a la intervención

Especificar el desenlace

Especifique qué resultado está siendo evaluado para determinar el riesgo de sesgo (típicamente de entre los asignados a la tabla Resumen de hallazgos). Especifique si se trata de un beneficio o daño de la intervención.

Especificar el resultado numérico que se está evaluando

En el caso de que se presenten múltiples análisis alternativos, especifique el resultado numérico (por ejemplo, RR = 1,52 (IC del 95%: 0,83 a 2,77) y / o una referencia (por ejemplo, una tabla, figura o párrafo) que defina de manera única el resultado que se está evaluando.

Consideración preliminar de los factores de confusión

Completar una fila para cada dominio de confusión importante (i) enumerado en el protocolo de revisión; Y (ii) relevantes para el establecimiento de este estudio en particular, o que los autores del estudio identificaron como potencialmente importantes.

Los dominios de confusión "importantes" son aquellos para los cuales, en el contexto de este estudio, se espera que el ajuste conduzca a un cambio clínicamente importante en el efecto estimado de la intervención. "Validez" se refiere a si la variable o variables de confusión miden completamente el dominio, mientras que "confiabilidad" se refiere a la precisión de la medida (más error de medición significa menos confiabilidad).

| (i) dominios de confusión enumerados en el protocolo de revisión | | | | |
|---|-------------------------|---|--|---|
| Dominio de confusión | Variable (s) medida (s) | ¿Existe evidencia de que el control de esta variable era innecesario? | ¿Es el dominio de confusión medido válida y fiablemente por esta variable (o estas variables)? | OPCIONAL: ¿El fracaso para ajustarse a esta variable (solo) se espera que favorezca la intervención experimental o el comparador? |
| | | | Sí / No / Sin información | A favor del brazo experimental / A favor del comparador / No información |
| | | | | |
| | | | | |

| | | | | |
|--|--|--|--|--|
| | | | | |
|--|--|--|--|--|

(ii) Dominios de confusión adicionales relevantes para el establecimiento y desarrollo de este estudio en particular, o que los autores del estudio identificaron como importantes

| Dominio de confusión | Variable (s) medida (s) | ¿Existe evidencia de que el control de esta variable era innecesario? | ¿Es el dominio de confusión medido válida y fiablemente por esta variable (o estas variables)? | OPCIONAL: ¿El fracaso para ajustarse a esta variable (solo) se espera que favorezca la intervención experimental o el comparador? |
|----------------------|-------------------------|---|--|---|
| | | | Sí / No / Sin información | A favor del brazo experimental /A favor del comparador / No información |
| | | | | |
| | | | | |
| | | | | |

*En el contexto de un estudio en particular, se puede demostrar que las variables no son factores de confusión y por lo tanto no se incluyen en el análisis: (a) si no son predictivos del resultado; (B) si no son predictivos de la intervención; O (c) porque el ajuste no hace o diferencia mínima al efecto estimado del parámetro primario. Nótese que "ninguna asociación estadísticamente significativa" no es lo mismo que "no predictivo".

Examen preliminar de las co-intervenciones

Completar una fila para cada co-intervención importante (i) enumerada en el protocolo de revisión; Y (ii) relevantes para el establecimiento de este estudio en particular, o que los autores del estudio identificaron como importantes.

Las co-intervenciones "importantes" son aquellas para las cuales, en el contexto de este estudio, se espera que el ajuste conduzca a un cambio clínicamente importante en el efecto estimado de la intervención.

| (i) Co-intervenciones enumeradas en el protocolo de revisión | | |
|---|---|---|
| Co-intervención | ¿Existe evidencia de que el control de esta co-intervención fuera innecesario (por ejemplo, porque no se administró)? | ¿Es probable que la presencia de esta co-intervención favorezca los resultados en la intervención experimental o en el comparador |
| | | A favor del brazo de experimentación / A favor del comparador / No información |
| | | A favor del brazo de experimentación / A favor del comparador / No información |
| | | A favor del brazo de experimentación / A favor del comparador / No información |
| | | A favor del brazo de experimentación / A favor del comparador / No información |

| (ii) Co-intervenciones adicionales relevantes para el establecimiento de este estudio en particular, o que los autores del estudio identificaron como importantes | | |
|--|---|---|
| Co-intervenciones | ¿Existe evidencia de que el control de esta co-intervención fuera innecesario (por ejemplo, porque no se administró)? | ¿Es probable que la presencia de esta co-intervención favorezca los resultados en la intervención experimental o en el comparador |
| | | A favor del brazo de experimentación / A favor del comparador / No información |
| | | A favor del brazo de experimentación / A favor del comparador / No información |
| | | A favor del brazo de experimentación / A favor del comparador / No información |

| | | |
|--|--|--|
| | | A favor del brazo de experimentación / A favor del comparador / No información |
|--|--|--|

Evaluación de Riesgo de sesgo

Las respuestas subrayadas en verde son marcadores potenciales de bajo riesgo de sesgo, y las respuestas en rojo son marcadores potenciales para un riesgo de sesgo. Cuando las preguntas se refieren sólo al post de señalización a otras preguntas, no se utiliza el formato.

| Sesgo debido a confusión | |
|--|--------------------|
| <p>1.1. ¿Existe potencial efecto de confusión en la estimación del efecto de la intervención en este estudio? Si N / PN a 1,1: el estudio se puede considerar que tiene un bajo riesgo de sesgo debido a la confusión y no más preguntas se necesitan para ser considerados?</p> | Elija un elemento. |
| Si respondió SI/PROBABLEMENTE SI en 1.1: determine si hay necesidad de evaluar confusión que varía en el tiempo: | |
| <p>1.2. ¿El análisis se basó en dividir el tiempo de seguimiento de los participantes según la intervención recibida? Si N / PN, responda a las preguntas relacionadas con la confusión de la línea de base (1.4 a 1.6) Si Y / PY, pase a la pregunta 1.3?</p> | Elija un elemento. |
| <p>1.3. ¿Los cambios o interrupciones de la intervención probablemente estarían relacionadas con factores pronósticos del resultado? Si N / PN, responda a las preguntas relacionadas con la confusión de la línea de base (1.4 a 1.6) Si Y / PY, responde a las preguntas relacionadas tanto con la confusión inicial como con la confusión que varía en el tiempo (1.7 y 1.8)</p> | Elija un elemento. |
| Preguntas relacionadas con confusión en la línea de base | |
| <p>1.4. ¿Utilizaron los autores un método de análisis apropiado que controlaba todos los dominios de confusión importantes?</p> | Elija un elemento. |
| <p>1.5. Si Y/PY a 1,4: ¿Los dominios de la confusión fueron controlados por estrategias válidas y replicables por las variables disponibles en este estudio?</p> | Elija un elemento. |
| <p>1.6. ¿Los autores controlaron las variables posteriores a la intervención que podrían haber sido afectadas por la intervención?</p> | Elija un elemento. |
| Preguntas relacionadas con confusión en la línea de base y confusión que varía en el tiempo | |

| | |
|---|--------------------|
| 1.7. ¿Utilizaron los autores un método de análisis apropiado que controlaba todos los dominios de confusión importantes y los factores de confusión variables en el tiempo? | Elija un elemento. |
| 1.8. Si Y / PY a 1,7: ¿Los dominios de confusión fueron controlados por estrategias de medición válidas y replicables por las variables disponibles en este estudio?? | Elija un elemento. |
| Juicio de riesgo de sesgo | Elija un elemento. |
| Opcional: ¿Cuál es la dirección prevista del sesgo debido a la confusión? | Elija un elemento. |
| Sesgo en la selección de los participantes en el estudio | |
| 2.1. ¿La selección de los participantes en el estudio (o en el análisis) se basó en las características de los participantes observadas después del inicio de la intervención? Si N / PN a 2.1: vaya a 2.4? | Elija un elemento. |
| 2.2. Si Y / PY a 2.1: ¿Las variables post-intervención que influyeron en la selección probablemente estarían asociadas con la intervención?? | Elija un elemento. |
| 2.3. ¿Las variables post-intervención que influyeron en la selección probablemente estarían influenciadas por el resultado o una causa del resultado?? | Elija un elemento. |
| 2.4. ¿El inicio del seguimiento y el inicio de la intervención coinciden para la mayoría de los participantes? | Elija un elemento. |
| 2.5. Si Y / PY a 2.2 y 2.3, o N / PN a 2.4: ¿Se usaron técnicas de ajuste que probablemente corregirían la presencia de sesgos de selección? | Elija un elemento. |
| Juicio de riesgo de sesgo | Elija un elemento. |
| Opcional: ¿Cuál es la dirección predicha del sesgo debido a la selección de los participantes en el estudio? | Elija un elemento. |
| Sesgo en la clasificación de las intervenciones | |
| 3.1. ¿Se definieron claramente los grupos de intervención? | Elija un elemento. |
| 3.2. ¿La información utilizada para definir grupos de intervención se registró al inicio de la intervención? | Elija un elemento. |
| 3.3. ¿Podría la clasificación del estado de intervención haber sido afectada por el conocimiento del resultado o riesgo del resultado? | Elija un elemento. |
| Juicio de riesgo de sesgo | Elija un elemento. |
| Opcional: ¿Cuál es la dirección prevista del sesgo debido a la clasificación de las intervenciones? | Elija un elemento. |
| Sesgos debido a desviaciones de las intervenciones propuestas | |

| | | |
|--|-----------------|----|
| Si su objetivo para este estudio es evaluar el efecto de la asignación de la intervención, responda las preguntas 4.1 y 4.2 | | |
| 4.1. ¿Hubo desviaciones de la intervención prevista más allá de lo que se esperaría en la práctica habitual?? | Elija elemento. | un |
| 4.2. Si Y / PY a 4.1: ¿Fueron estas desviaciones de la intervención deseada desequilibradas entre los grupos y probablemente afectaron el resultado? | Elija elemento. | un |
| Si su objetivo para este estudio es evaluar el efecto de iniciar y adherirse a la intervención, responda las preguntas 4.3 a 4.6 | | |
| 4.3. ¿Hubo co-intervenciones importantes equilibradas entre los grupos de intervención?? | Elija elemento. | un |
| 4.4. ¿Se implementó la intervención con éxito para la mayoría de los participantes? | Elija elemento. | un |
| 4.5. ¿Los participantes del estudio se han adherido al régimen de intervención asignado? | Elija elemento. | un |
| 4.6. Si N / PN a 4,3, 4,4 o 4,5: ¿Se utilizó un análisis apropiado para estimar el efecto del inicio y la adherencia a la intervención? | Elija elemento. | un |
| Juicio de riesgo de sesgo | Elija elemento. | un |
| Opcional: ¿Cuál es la dirección predicha del sesgo debido a las desviaciones de las intervenciones previstas? | Elija elemento. | un |
| Sesgo debido a datos perdidos | | |
| 5.1. ¿Se disponía de datos sobre los resultados para todos o casi todos los participantes? | Elija elemento. | un |
| 5.2. ¿Se excluyeron los participantes por falta de datos sobre el estado de intervención? | Elija elemento. | un |
| 5.3. ¿Se excluyeron los participantes debido a la falta de datos sobre otras variables necesarias para el análisis? | Elija elemento. | un |
| 5.4. Si PN / N a 5.1, o Y / PY a 5.2 o 5.3: ¿La proporción de participantes y las razones para la falta de datos son similares a través de las intervenciones? | Elija elemento. | un |
| 5.5. Si PN / N a 5.1 o Y / PY a 5.2 o 5.3: ¿Hay evidencia de que los resultados fueran robustos a la presencia de datos faltantes? | Elija elemento. | un |
| Juicio de riesgo de sesgo | Elija elemento. | un |
| Opcional: ¿Cuál es la dirección prevista del sesgo debido a la falta de datos? | Elija elemento. | un |
| Sesgo en la medición de los desenlaces | | |
| 6.1. ¿Podría la medición de los resultados haber sido influenciada por el conocimiento de la intervención recibida?? | Elija elemento. | un |
| 6.2. ¿Los evaluadores de resultados eran conscientes de la intervención recibida por los participantes del estudio?? | Elija elemento. | un |

| | | |
|--|-----------------|----|
| 6.3. ¿Fueron comparables los métodos de evaluación de resultados entre los grupos de intervención? | Elija elemento. | un |
| 6.4. ¿Hubo algún error sistemático en la medición del resultado relacionado con la intervención recibida? | Elija elemento. | un |
| Juicio de riesgo de sesgo | Elija elemento. | un |
| Opcional: ¿Cuál es la dirección prevista del sesgo debido a la medición de los resultados? | Elija elemento. | un |
| Sesgo en la selección de los resultados reportados | | |
| Es probable que la estimación del efecto notificado sea seleccionada, sobre la base de los resultados obtenidos, a partir de: | Elija elemento. | un |
| 7.1. ¿...medidas de resultados múltiples dentro del dominio de resultados? | | |
| 7.2. ¿... análisis múltiples de la relación intervención-resultado? | Elija elemento. | un |
| 7.3. ¿... diferentes subgrupos?? | Elija elemento. | un |
| Juicio de riesgo de sesgo | Elija elemento. | un |
| Opcional: ¿Cuál es la dirección predicha de sesgo debido a la selección del resultado reportado? | Elija elemento. | un |
| Sesgo global | | |
| Juicio de riesgo de sesgo | Elija elemento. | un |
| Opcional: ¿Cuál es la dirección prevista global del sesgo para este resultado? | Elija elemento. | un |

Nota: traducción libre realizada por funcionarios de la Subdirección de Evaluación de Tecnologías en Salud, Instituto de Evaluación Tecnológica en Salud - IETS.

6.5 Calidad de los estudios primarios de validez diagnóstica incluidos en la evaluación (herramienta QUADAS-2).

| Dominio | | Criterio | Autor año |
|------------------------|--------------------|--|-----------|
| Selección de pacientes | A. Riesgo de sesgo | ¿Se realizó una muestra consecutiva o aleatoria de los pacientes reclutados? | |
| | | ¿Se evitó un diseño de casos y controles? | |
| | | ¿El estudio evitó exclusiones inadecuadas? | |
| | | ¿Podría la selección de los pacientes haber introducido un sesgo? | |

| Dominio | | Criterio | Autor año |
|------------------------------|---------------------|---|-----------|
| | B. Aplicabilidad | ¿Existe preocupación que los pacientes incluidos no correspondan a la pregunta de investigación? | |
| Prueba índice | A. Riesgo de sesgo | ¿Fueron los resultados de la prueba índice interpretados sin conocimiento de los resultados del estándar de referencia? | |
| | | Si se usó un umbral, ¿éste fue especificado previamente? | |
| | | ¿La conducción o interpretación de la prueba índice podría haber introducido un sesgo? | |
| | B. Aplicabilidad | ¿Existe preocupación acerca que la prueba índice, su conducción o interpretación no correspondan con la pregunta de investigación? | |
| Estándar de referencia | A. Riesgo de sesgo | ¿Es probable que el estándar de referencia clasifique correctamente la condición de interés? | |
| | | ¿Fueron los resultados del estándar de referencia interpretados sin conocimiento de los resultados de la prueba índice? | |
| | | ¿La conducción o interpretación del estándar de referencia podría haber introducido un sesgo? | |
| | B. Aplicabilidad | ¿Existe preocupación acerca que la condición de interés definida por el estándar de referencia no corresponda a la pregunta de investigación? | |
| Flujo de pacientes y tiempos | A. Riesgo de sesgos | ¿Hubo un intervalo de tiempo adecuado entre la prueba índice y el estándar de referencia? | |
| | | ¿Todos los pacientes recibieron el estándar de referencia? | |
| | | ¿Los pacientes recibieron el mismo estándar de referencia? | |
| | | ¿Fueron incluidos todos los pacientes en el análisis? | |
| | | ¿Podría el flujo de pacientes haber introducido un sesgo? | |

Nota: traducción libre realizada por funcionarios de la Subdirección de Evaluación de Tecnologías en Salud, Instituto de Evaluación Tecnológica en Salud - IETS.

Anexo 7. Contenido del informe de efectividad y seguridad

Título del informe: en el formato Evaluación de efectividad y seguridad de [intervenciones] comparadas con [comparadores] para [condición de salud].

Preliminares: nominador de la tecnología, autores de la evaluación, revisores, fuente de financiamiento, declaración de intereses, correspondencia.

Tabla de contenido.

Índice de tablas.

Índice de figuras.

Lista de abreviaturas y siglas.

Resumen ejecutivo: estructurado con los apartados introducción, objetivo, métodos, resultados y conclusiones; no debe ser mayor a 100 palabras.

Introducción.

1. Descripción de la condición de salud.
2. Descripción de la(s) tecnologías en salud evaluadas: mecanismo de acción, dosis, vía de administración, eventos adversos, registro sanitario INVIMA
3. Comparadores.
4. Pregunta de evaluación.
5. Metodología para la revisión de efectividad y seguridad
 - 5.1. Criterios de elegibilidad
 - 5.2. Búsqueda de evidencia, tamización de referencias, selección de documentos.
 - 5.3. Evaluación de riesgo de sesgos.
 - 5.4. Síntesis de la información y evaluación de la certeza en la evidencia.
6. Resultados
 - 6.1. Búsqueda de evidencia, tamización de referencias, selección de documentos.
 - 6.2. Evaluación de riesgo de sesgos.
 - 6.3. Síntesis de la información y evaluación de la certeza en la evidencia.
7. Discusión y conclusiones
8. Referencias bibliográficas

Anexos

1. Registros sanitarios
2. Estrategias de búsqueda

3. Diagramas de flujo PRISMA
4. Lista de estudios incluidos
5. Lista de estudios excluidos y razones para la exclusión
6. Evaluación de riesgo de sesgos
7. Características de los estudios incluidos
8. Perfiles de evidencia GRADE
9. Códigos para la realización de metanálisis

Anexo 8. Lista de chequeo para los reportes de evaluación de efectividad, seguridad o validez diagnóstica de tecnologías en salud

Estos criterios de verificación se desarrollaron con base en las herramientas AMSTAR, INAHTA, los principios propuestos por Drummond y cols. en el documento "*Key principles for the improved conduct of health technology assessments for resource allocation decisions*" del año 2008, y el manual metodológico para la elaboración de evaluaciones de efectividad, seguridad y validez diagnóstica de tecnologías en salud del IETS.

Pretende servir como una herramienta que ayuda a verificar en qué medida un informe de efectividad, seguridad o validez diagnóstica de tecnologías en salud responde a los criterios más relevantes sobre su calidad metodológica y de reporte. El referente para la elaboración de las evaluaciones de efectividad y seguridad es el **Manual metodológico para la elaboración de evaluaciones de efectividad, seguridad y validez diagnóstica de tecnologías en salud**, disponible en la página web del IETS www.iets.org.co

Fecha:

Autor:

Título del reporte:

Revisó:

| Documentos para entregar | ¿Entrega? (Sí/No) |
|---|-------------------|
| Reporte de efectividad y seguridad en la plantilla establecida, con todas las tablas y anexos diligenciados | |
| Artículos incluidos en la revisión | |

| N° | Ítem | Evaluación (Cumple, no cumple, cumple parcialmente, no aplica) | Pág. | Observación |
|---|---|---|------|-------------|
| Introducción | | | | |
| 1 | ¿Se identifican los autores del reporte (grupo desarrollador)? | | | |
| 2 | ¿Participan en la evaluación todos los actores clave de acuerdo con la tecnología y condición evaluadas (expertos en metodología, expertos temáticos y pacientes o sus representantes)? | | | |
| 3 | ¿Se declara si el reporte fue sometido a revisión externa por pares? | | | |
| 4 | ¿Se describieron las fuentes de financiación? | | | |
| 5 | ¿Se declaró la existencia de conflictos de interés? | | | |
| 6 | ¿Se indicaron los datos de contacto para solicitar más información? | | | |
| 7 | ¿Se presenta un resumen ejecutivo incluyendo introducción, objetivo, metodología, resultados y conclusiones? | | | |
| 8 | ¿Se presenta una introducción, incluyendo la síntesis del problema a investigar, justificación, alcance y objetivo de la evaluación? | | | |
| 9 | ¿Se presenta una descripción de la condición de salud en la que se incluya su definición, epidemiología, factores de riesgo, historia natural, curso clínico, sintomatología, pronóstico y carga de la enfermedad? | | | |
| 10 | ¿Se presenta una descripción de la tecnología en salud en la que se incluya su definición, código de identificación para el país, estatus de autorización, mecanismo de acción, forma farmacéutica y concentración, forma de administración, dosificación, contraindicaciones, precauciones y contexto y nivel de atención en el que utiliza? | | | |
| Pregunta de evaluación y criterios de elegibilidad | | | | |
| 11 | ¿La pregunta de investigación está formulada en estructura PICOT? | | | |
| 12 | ¿Se menciona la condición de salud, características socio-demográficas y clínicas o subgrupos que permitan identificar de manera precisa la población de estudio? | | | |
| 13 | ¿En la intervención se describe la modalidad del tratamiento, dosis o forma farmacéutica? | | | |
| 14 | ¿La tecnología comparadora está financiada con recursos públicos asignados a la salud? | | | |

| N° | Ítem | Evaluación (Cumple, no cumple, cumple parcialmente, no aplica) | Pág. | Observación |
|--------------------|---|---|------|-------------|
| 15 | El comparador muestra evidencia y soporte que es la terapia estándar o la(s) tecnología(s) más utilizada(s) en la práctica clínica o es la tecnología alternativa más relevante en términos de efectividad y seguridad o de exactitud diagnóstica en caso de las intervenciones de pruebas diagnósticas y es actualmente aceptada en el ámbito clínico. | | | |
| 16 | Se menciona el tiempo de seguimiento para la evaluación de los desenlaces | | | |
| 17 | ¿Se realiza una consulta con expertos temáticos, representantes de sociedades científicas, pacientes y cuidadores para el refinamiento de la pregunta de evaluación? | | | |
| 18 | ¿Se presentan los ajustes realizados a la pregunta de investigación luego de la reunión para su refinamiento? | | | |
| 19 | ¿Se presenta una gradación de los desenlaces usando la propuesta GRADE? | | | |
| 20 | ¿Se presentan los resultados de la votación para hacer la gradación de los desenlaces? | | | |
| 21 | Si proponen desenlaces subrogados o intermedios ¿se justifica su uso? | | | |
| 22 | ¿Se definen con claridad y precisión los criterios de elegibilidad de la evidencia (de inclusión y exclusión) en términos de población, subgrupos, tecnologías de interés, comparadores, desenlaces y tiempo? | | | |
| 23 | ¿Se publicó la pregunta con estructura PICOT de manera preliminar en la página web del IETS, tras el envío del protocolo del ente gestor y se tuvo en cuenta los ajustes (si aplica) para su publicación como pregunta definitiva? | | | |
| Metodología | | | | |
| 24 | ¿La evaluación es realizada de acuerdo con un protocolo definido a priori por el grupo desarrollador, revisado por pares y que incluyó: pregunta de evaluación en estructura PICOT, criterios de elegibilidad de la evidencia y métodos de búsqueda de evidencia? | | | |
| 25 | ¿La descripción metodológica de la búsqueda de evidencia incluye las bases de datos, términos, | | | |

| N° | Ítem | Evaluación (Cumple, no cumple, cumple parcialmente, no aplica) | Pág. | Observación |
|-------------------|--|---|------|-------------|
| | límites de fecha e idioma, filtros por tipo de estudio? | | | |
| 26 | ¿La descripción metodológica de la estrategia de búsqueda incluye métodos de búsqueda complementarios: consulta con expertos temáticos, productores y comercializadores de las tecnologías evaluadas y de sus comparadores, búsqueda manual "en bola de nieve", revisión de información sobre vigilancia poscomercialización, etc? | | | |
| 27 | ¿La descripción metodológica del proceso de búsqueda menciona si se realizó actualización de la búsqueda de estudios primarios para las revisiones sistemáticas seleccionadas? | | | |
| 28 | ¿La descripción metodológica menciona se realizó la tamización de manera duplicada e independiente, junto con el método empleado para resolver desacuerdos entre los pares revisores? | | | |
| 29 | ¿La descripción metodológica menciona las herramientas empleadas para la evaluación de calidad de cada uno de los estudios incluidos? | | | |
| 30 | ¿Se evalúa la calidad de la evidencia de acuerdo con lo especificado en la metodología GRADE? | | | |
| 31 | ¿Se describen los métodos de extracción de datos y síntesis de la evidencia? | | | |
| 32 | ¿Se realiza un control de calidad sobre la extracción de datos? | | | |
| Resultados | | | | |
| 33 | ¿Se presentan los reportes de la búsqueda de evidencia en bases de datos electrónicas, que contengan información sobre bases de datos empleadas, con las fechas, filtros y términos de búsqueda de tal manera que se pueda replicar la búsqueda de la literatura? | | | |
| 34 | ¿Existe consistencia entre estos elementos y los presentados en la pregunta 25? | | | |
| 35 | ¿Se presentaron los resultados de la búsqueda de evidencia, tamización de referencias y selección de estudios mediante el diagrama de flujo PRISMA? | | | |

| N° | Ítem | Evaluación (Cumple, no cumple, cumple parcialmente, no aplica) | Pág. | Observación |
|---------------------------------|---|---|------|-------------|
| 36 | ¿Se presenta el listado de referencias de los estudios excluidos junto con las razones de la exclusión? | | | |
| 37 | ¿Se describen los resultados obtenidos por medio de la búsqueda complementaria? | | | |
| 38 | ¿El reporte de los resultados incluye los siguientes apartados: búsqueda de estudios, tamización de referencias y selección de estudios, valoración de la calidad de la evidencia, descripción de estudios y síntesis de la evidencia de efectividad y seguridad? | | | |
| 39 | ¿Se presentan las tablas con las características de todos los estudios incluidos? | | | |
| 40 | ¿Cuándo no se encontraron estudios integrativos, se hizo la búsqueda de estudios primarios para responder la pregunta de evaluación? | | | |
| 41 | ¿Se presenta la evaluación de calidad para cada uno de los estudios incluidos? | | | |
| 42 | ¿Se presentan las razones cuando se disminuyó la calidad de los estudios incluidos? | | | |
| 43 | ¿Se presentan las tablas GRADE con las notas explicativas cuando la calidad de la evidencia es considerada moderada, baja o muy baja? | | | |
| 44 | ¿Se presentan las tablas GRADE con las medidas de efecto relativas y con las medidas de frecuencia o tendencia central para cada uno de los grupos? | | | |
| 45 | ¿Las tablas GRADE incluyen todos los desenlaces mencionados en la PICOT , aunque no hayan estudios que los evalúen? | | | |
| 46 | ¿Se presenta la búsqueda de eventos adversos en otras fuentes como los reportes de agencias de tecnologías en salud? | | | |
| 47 | Si usaron desenlaces compuestos ¿presentan los resultados para el desenlace global y para cada uno de sus componentes? | | | |
| 48 | ¿Se mencionan desviaciones a partir del protocolo original de la revisión? | | | |
| Discusión y conclusiones | | | | |
| 49 | ¿La discusión contiene los siguientes aspectos: interpretación de los resultados, similitudes y diferencias con la literatura, limitaciones y riesgo de sesgos en la ETES? | | | |

| N° | Ítem | Evaluación (Cumple, no cumple, cumple parcialmente, no aplica) | Pág. | Observación |
|--|--|---|------|-------------|
| 50 | ¿Se presentaron las conclusiones de la evaluación de acuerdo con lo especificado en el manual metodológico IETS? | | | |
| Si la evaluación implicó la realización de un metanálisis de comparaciones directas | | | | |
| 51 | ¿Se indica el método empleado para evaluar la heterogeneidad? | | | |
| 52 | ¿Se menciona el método empleado para hacer la estimación (efectos fijos o aleatorios)? | | | |
| 53 | ¿Se menciona la estrategia empleada para evaluación de posible sesgo de publicación? | | | |
| Si la evaluación implicó la realización de un metanálisis de comparaciones indirectas o en red | | | | |
| 54 | ¿Se presenta una justificación para la realización de comparaciones indirectas o en red? | | | |
| 55 | ¿Se presenta una justificación para la elección del comparador común de la red? | | | |
| 56 | ¿Se presenta la red de evidencia y se describe su estructura? | | | |
| 57 | ¿Se presentan los resultados obtenidos para cada uno de los estudios incluidos en la red? | | | |
| 58 | ¿Se describe la manera en la que se realizó la evaluación del supuesto de transitividad? | | | |
| 59 | ¿Se describe la manera en la que se realizó la evaluación del supuesto de coherencia o consistencia? | | | |
| 60 | ¿Se presentan los resultados tanto de las comparaciones directas, como de las indirectas o en red? | | | |
| 61 | ¿Se presenta el ranking de los tratamientos, ya sea en forma tabular o gráfica? | | | |
| 62 | ¿Se emplea alguna de las aproximaciones GRADE para evaluar la calidad de la evidencia del NMA? | | | |
| Concepto final: | | | | |
| Observaciones: | | | | |



Instituto de Evaluación
Tecnológica en Salud®

Evidencia que promueve Confianza



www.iets.org.co



Carrera 49 a No. 91 - 91
Bogotá, D.C., Colombia



(+571) 3770100



contacto@iets.org.co



[@ietscolombia](https://twitter.com/ietscolombia)



[ietscolombia](https://www.instagram.com/ietscolombia)