



4.16 Anexo a la sección 4.5 – Ejemplos de herramientas de evaluación de calidad



Forma de evidencia	Ejemplos de herramientas para evaluar calidad
Tipos de evidencia para los cuales existen herramientas de evaluación de calidad	
Análítica de datos	<p>ROBINS-I (riskofbias.info) para estudios observacionales, como los que examinan las asociaciones entre factores específicos (incluyendo intervenciones) y desenlaces específicos, donde hay riesgo de sesgo proveniente de:</p> <ul style="list-style-type: none"> • variables de confusión (donde la relación observada entre un factor y un desenlace difiere de la relación real debido a uno o más factores adicionales que no son considerados) • selección de participantes en el estudio • clasificación de la(s) intervención(es) • desviaciones de la intervención pretendida • información faltante • medición de desenlaces • selección del resultado reportado
Evaluación	<p>Risk of Bias (RoB) 2 (riskofbias.info) para ensayos controlados aleatorizados, donde el riesgo de variables de confusión es menor, pero donde hay riesgo de sesgos que surge de algunas de las mismas fuentes (aunque menos) mencionadas previamente:</p> <ul style="list-style-type: none"> • proceso de aleatorización • desviaciones de las intervenciones pretendidas • información faltante (del desenlace) • medición de desenlaces • selección del resultado reportado
Investigación de comportamiento/ de implementación	Vea otras filas para tipos de estudios relevantes o síntesis
Investigación cualitativa	<p>Lista de verificación de evaluación crítica de JBI para investigación cualitativa (bit.ly/31Lsib1), donde entran en juego consideraciones muy distintas, como:</p> <ul style="list-style-type: none"> • congruencia entre la metodología de investigación y la pregunta de investigación, métodos de recolección de datos, representación y analítica de datos, e interpretación de resultados, así como también entre la perspectiva filosófica declarada y la metodología • reflexividad de parte del investigador, como las declaraciones que ubican al investigador cultural y teóricamente, y abordan la influencia del investigador en la investigación y viceversa • representación de los participantes del estudio y sus voces • flujo de conclusiones del análisis e interpretación de la información
Síntesis de evidencia	<p>Vea anteriormente los tipos relevantes de estudios considerados en las síntesis de evidencia</p> <p><i>A MeaSurement Tool to Assess systematic Reviews</i> (AMSTAR; amstar.ca) para la calidad de las síntesis de evidencia, donde el riesgo de sesgo puede surgir de:</p> <ul style="list-style-type: none"> • identificación de todos los estudios potencialmente relevantes a través de una búsqueda exhaustiva de la literatura publicada y la literatura gris y sin restricciones de idioma • selección de todos los estudios que abordan la pregunta de investigación usando criterios explícitos sobre diseños de estudios y sobre participantes, intervenciones/ factores, comparadores y desenlaces, y con al menos dos revisores aplicando los criterios • evaluación de calidad y extracción de datos de todos los estudios incluidos • síntesis de hallazgos de todos los estudios incluidos <p>Observe que hay dos versiones de AMSTAR: 1) la versión original que puede aplicarse en todos los tipos de síntesis, aunque eliminando algunos criterios del numerador y del denominador; 2) una segunda versión de AMSTAR que es más específicamente relevante para síntesis de ensayos controlados aleatorizados</p> <p><i>Grading of Recommendations, Assessment, Development and Evaluation</i> (GRADE; bit.ly/3C9pMrx) para la certeza de la evidencia de los resultados de una intervención, con:</p> <ul style="list-style-type: none"> • disminución de la certeza por riesgo de sesgos (con evidencia de ensayos controlados aleatorizados que empiezan con alta certeza y evidencia de estudios observacionales que empiezan con baja calidad y luego son ajustados con base en RoB2 o ROBINS-I), imprecisión (p. ej. Uno o dos estudios pequeños), inconsistencia (p. ej. Dos estudios que muestran hallazgos muy diferentes), carácter indirecto (p. ej. Desenlaces subrogados o entornos atípicos estudiados), y sesgo de publicación (p. ej. Más común con estudios observacionales debido a la falta de registros de estudios o con estudios financiados por la industria debido a los incentivos comerciales de publicar estudios con resultados positivos) • incremento de la certeza en los casos de gran magnitud del efecto, el gradiente dosis-respuesta, y cuando todas las variables confusoras residuales disminuirían la magnitud del efecto <p>GRADE CERQual (cerqual.org) para la certeza de la evidencia de representaciones cualitativas de un fenómeno de interés, con:</p> <ul style="list-style-type: none"> • disminución de la certeza debido a limitaciones metodológicas (dado que los problemas en la manera en la que fueron diseñados o reportados los estudios se identificaron usando una herramienta de evaluación crítica como JBI, descrita anteriormente), relevancia (dado que el contexto donde los estudios primarios fueron realizados es sustancialmente diferente del contexto de la pregunta de la síntesis), coherencia (dado que parte de la información contradice los resultados o es ambigua), e idoneidad (dado que la información no es lo suficientemente rica o solo proviene de un pequeño número de estudios o participantes)



Evaluación de tecnologías / análisis de costo-efectividad	Lista de verificación de <i>International Network of Agencies for Health Technology Assessment (INAHTA)</i> (bit.ly/2YJVMVK) para la calidad de evaluaciones de tecnologías, donde dos de las 14 preguntas abordan el enfoque para sintetizar la evidencia (con indicaciones similares a las de AMSTAR) y otra pregunta aborda si la evaluación fue contextualizada a través de un análisis de costo-efectividad (con información de costos locales – nacionales o subnacionales), y con consideración de las implicaciones locales legales, éticas y sociales
	Lista de verificación de Drummond de análisis de costo-efectividad (bit.ly/3FbnB8R), y para evaluaciones económicas en general, con preguntas sobre el diseño del estudio, la recolección de datos, y el análisis e interpretación de resultados
	Lista de verificación de Philips para análisis de costo-efectividad que incluye un componente de modelamiento analítica de decisiones (bit.ly/3FcWBGc) con preguntas sobre la estructura del modelo (p. ej. Justificación explícita, supuestos justificados y horizonte temporal adecuado), la información usada (p. ej. Probabilidades basales de estudios observacionales, efectos del tratamiento de ensayos controlados aleatorizados, y evaluaciones de los cuatro tipos de incertidumbre, a saber la estructura del modelo, los pasos metodológicos seguidos, la heterogeneidad en la población estudiada, y los parámetros usados), y la consistencia (interna y externa) – también está la herramienta complementaria TRUST para evaluar incertidumbres en modelos analíticos de decisiones (bit.ly/3quFSKp)



Guías	Herramienta AGREE II (bit.ly/30qyFAB) para evaluar el desarrollo, el reporte y la evaluación (o valoración de calidad) de guías, que usa 23 ítems agrupados en seis dominios, cada uno de los cuales recibe una calificación independiente: <ul style="list-style-type: none"> • alcance y propósito descrito • involucramiento de actores interesados (ciudadanos/pacientes y profesionales) • rigor del desarrollo (con síntesis de evidencia usadas como insumo, un proceso robusto de desarrollo de recomendaciones, y recomendaciones vinculadas a la evidencia que las respalda) • claridad de presentación • aplicabilidad • independencia editorial (en relación con el financiador y los conflictos de interés de los miembros del panel)
	GRADE (bit.ly/3C9pMrx) para evaluar la fuerza de las recomendaciones, donde se tienen cuatro consideraciones claves: <ul style="list-style-type: none"> • balance entre desenlaces deseables e indeseables (compensaciones), teniendo en cuenta los mejores estimadores de la magnitud de los efectos en desenlaces deseables e indeseables, y la importancia de esos desenlaces (valores y preferencias típicos estimados) • confianza en la magnitud de los estimadores de efectos de las intervenciones en desenlaces importantes (ver GRADE en una fila previa) • confianza en valores y preferencias y en su variabilidad en uso de recursos



Tipos de evidencia para los cuales no existen aún herramientas de evaluación de calidad

Modelamiento	No existen aún herramientas ampliamente aceptadas para la mayoría de tipos de modelos, sin embargo, hay algunas preguntas generales que se pueden hacer sobre los modelos (como las mencionadas anteriormente en la lista de verificación de Philips), como: <ul style="list-style-type: none"> • estructura del modelo (p. ej. Justificación explícita, supuestos justificados, y horizonte temporal adecuado) • información usada (p. ej. Probabilidades basales de estudios observacionales, efectos de intervención de una gama de fuentes*, y evaluaciones de cuatro tipos de incertidumbre, a saber, la estructura del modelo, los pasos metodológicos seguidos, la heterogeneidad de la población estudiada, y los parámetros usados) • consistencia (interna y externa) • disponibilidad del software o la herramienta, de manera que pueda ser evaluado por otros <p>*Uno de los desafíos con COVID-19 fue que los diseños de los estudios que típicamente capturaban los efectos de la intervención, como los ensayos controlados aleatorizados, eran ética o logísticamente difíciles y/o tomaban tiempo para completarse, de manera que otros diseños de estudios debían ser usados y se debía recurrir a la opinión de expertos (y hay enfoques que permiten que esto se haga de manera sistemática y transparente, como SHELF – see bit.ly/30nteC4)</p>
---------------------	--



Enfoques usados con ciertos tipos de evidencia para los cuales no existen aún herramientas de evaluación de calidad	
Inteligencia artificial	No existen aún herramientas ampliamente aceptadas