

4.5 Diferenciar evidencia de calidad alta de evidencia de calidad baja

No toda evidencia es de calidad alta y confiable para tomar decisiones. Existen herramientas para muchas (pero no todas) las formas de evidencia para ayudar a hacer juicios sobre la confiabilidad de la evidencia (de un solo estudio o de un conjunto de estudios). Como describimos aquí, estas herramientas usan puntajes o calificaciones para ayudar a los usuarios a entender qué tanta confianza pueden depositar en la evidencia. Muchas revistas científicas actualmente exigen que los autores sigan estándares de reporte, como CONSORT para ensayos controlados aleatorizados y el PRISMA para síntesis de evidencia. La mayoría de las revistas científicas no exigen a los revisores el uso de herramientas específicas para evaluar la calidad de los estudios o la fuerza de las recomendaciones; como resultado, el hecho de estar publicado en una revista científica que incluya revisión por pares no es un buen indicador de calidad.

Cuestión	Respuesta
Estudios (y guías) varían en calidad (o confiabilidad)	<ul style="list-style-type: none"> Herramientas de evaluación de calidad (o de evaluación crítica) han sido desarrolladas para diseños específicos de estudios (p. ej. Ensayos controlados aleatorizados), para categorías generales de diseños de estudios (p. ej. Estudio observacional, investigación cualitativa, y síntesis de evidencia), y para guías – ver el anexo al final de este capítulo (sección 4.16) para ver ejemplos (RoB2, ROBINS-I, checklist JBI, AMSTAR, y AGREE II) Las herramientas pueden ofrecer un resumen de la evaluación (p. ej. Bajo riesgo de sesgo usando RoB2 o ROBINS-I), un puntaje que algunas agrupan en rangos (p. ej. Calidad alta usando AMSTAR), una serie de puntajes (p. ej. Seis dominios usando AGREE II), o una serie de consideraciones que pueden informar un resumen de la valoración (p. ej. Lista de verificación JBI)
Conjuntos de evidencia pueden variar en su certeza (o en la confianza que se puede depositar en ellos)	<ul style="list-style-type: none"> Se han desarrollado herramientas de evaluación de certeza para conjuntos de estudios que abordan la misma pregunta (p. ej. El efecto de una intervención en un desenlace específico o el significado que los ciudadanos le dan a un fenómeno particular) – ver en la sección 4.16 dos ejemplos (GRADE y GRADE CERQual) Las herramientas pueden ofrecer un resumen de la valoración sobre la confianza en que el efecto verdadero es similar al efecto estimado (p. ej. Alta certeza con GRADE) o que el fenómeno de interés está bien representado por los hallazgos de un estudio cualitativo (con GRADE CERQual) Un resumen de la valoración sobre la certeza en la estimación de un efecto es más útil que un test de significancia estadística que demuestre que una intervención ‘funciona’ o ‘no funciona’ (que sucederá por azar una cada 20 veces si la significancia estadística es establecida al nivel de 0.05)
Las recomendaciones varían en su fuerza	<ul style="list-style-type: none"> Se han desarrollado herramientas de evaluación de la fuerza de las recomendaciones de guías (p. ej. GRADE, además de jerarquizar la certeza de un conjunto de evidencia, como se describió previamente) – ver en la sección 4.16 un ejemplo Las herramientas pueden ofrecer un resumen de la valoración donde la mayoría de los tomadores de decisiones elegiría proceder con una intervención (p. ej. Recomendación fuerte con GRADE) o donde la mayoría necesitaría valorar cuidadosamente las ventajas y las desventajas de una intervención
Puede ser difícil juzgar algunas fuentes de evidencia (o enfoques usados para generarla)	<ul style="list-style-type: none"> No existen herramientas ampliamente aceptadas para evaluar qué tanta confianza puede ser depositada en: <ul style="list-style-type: none"> Un experto, aunque sí existen ejemplos para pronosticar, como El Proyecto del Buen Juicio (volveremos sobre la opinión de expertos posteriormente en este capítulo y, en el caso de opinión de expertos sobre parámetros de modelos, en la sección 4.16) Los modelos usados en la generación de algunas formas de evidencia (que abordamos en la sección 4.7 cuando hablamos sobre modelos de cambio climático y en la sección 4.16) Un algoritmo de inteligencia artificial usado en la generación de algunos tipos de evidencia, aunque algunos ejemplos como TRIPOD están empezando a emerger (3)

Distinguir entre evidencia de calidad alta y evidencia de calidad baja es particularmente desafiante cuando la evidencia está integrada en tableros, modelos y otros formatos, y cuando hay conflictos de interés en juego. Mencionaremos esto último en las [secciones 4.12, 4.14, y 4.16](#). Aunque no es el centro de este informe, distinguir los ‘datos crudos’ de calidad alta de los de calidad baja también puede ser un desafío, y organizaciones como UNICEF han desarrollado estructuras de calidad de datos para ayudar con esto (bit.ly/3DQQRrv).

Algunos ‘repositorios de evidencia’, como Social Systems Evidence y el inventario de COVID-19 Evidence Network to support Decision-making (COVID-END) (descrito en la [sección 4.6](#)), usan algunas de estas herramientas para que los tomadores de decisiones y el personal que los apoya se puedan enfocar en síntesis de evidencia de calidad alta o puedan entender que están usando la mejor evidencia disponible en cuanto a síntesis de evidencia (si no es de calidad alta).

La pandemia de COVID-19 exigió a los tomadores de decisiones tomar decisiones difíciles en cortos períodos de tiempo, inicialmente con poca evidencia (y con frecuencia indirecta), y luego, con el tiempo, con estudios, conjuntos de evidencia, y recomendaciones desarrolladas usando un proceso robusto. Para apoyar la toma de decisiones sobre COVID-19 basada en conjuntos de evidencia (en lugar de estudios individuales), COVID-END destaca en su inventario de ‘mejores’ síntesis de evidencia las que están actualizadas (basados en las fechas de la búsqueda de la evidencia), que sean de calidad alta (basados en la herramienta AMSTAR), y que provean una evaluación de la certeza de la evidencia (basados en la herramienta GRADE).

Así como no toda la evidencia es de calidad alta, no toda la evidencia global será aplicable en un contexto determinado. Por ejemplo, una síntesis de evidencia que contenga estudios realizados exclusivamente en países de ingresos altos tendrá limitada aplicabilidad en algunos países de ingresos bajos. Pueden haber diferencias importantes en las condiciones de base, en las realidades y limitaciones particulares de un territorio, y en características estructurales del sistema local (p. ej. Sistema de salud nacional o sistema subnacional de educación). Una herramienta SUPPORT puede ayudar a las personas a pensar sobre la aplicabilidad local de los hallazgos de una síntesis de evidencia y a considerar de qué manera las ideas de la síntesis todavía pueden ser aprovechadas, incluso si los hallazgos no son aplicables.⁽⁴⁾

El razonamiento Bayesiano ha cosechado creciente atención como manera de re-crear deliberadamente nuestros ‘mapas mentales’ sobre desafíos y maneras para abordarlos, no reemplazando todo lo que creíamos saber con nueva información, sino modificando nuestra comprensión en un grado apropiado. El grado depende de qué tanta confianza se tiene en su conocimiento preexistente (la probabilidad ‘previa’ de que algo fuera cierto) y cuánta confianza puede depositar en el nuevo conocimiento. Más confianza puede ser depositada en el nuevo conocimiento si este proviene de síntesis de evidencia de calidad alta que incluyen estudios realizados en contextos similares al suyo.



Productora e intermediaria de evidencia, Gillian Leng

Ejecutiva experimentada que lidera una agencia de evaluación y guía de tecnología que apoya la toma de decisiones sanitarias y sociales realizada por gobiernos, prestadores de servicios y pacientes

El Reino Unido ha dirigido trabajos durante muchos años para motivar las síntesis y el uso de la evidencia – desde el primer ensayo controlado aleatorizado para prevenir el escorbuto en marineros, hasta los más recientes e innovadores What Works Centres para promover el uso de evidencia en una gama de áreas políticas. Como parte de este movimiento basado en la evidencia, en los últimos 20 años el Instituto Nacional para la Calidad de la Sanidad y de la Asistencia (NICE, por sus siglas en inglés) ha transformado el uso de la evidencia en la práctica sanitaria, así como en las iniciativas de salud pública y asistencia social más amplias.

La pandemia de COVID-19 ha reforzado dramáticamente la necesidad de evidencia de calidad alta para informar la política y la práctica, y también ha resaltado las consecuencias negativas de las redes sociales y la información falsa asociada. En este contexto, el trabajo de la Comisión Global de Evidencia para Abordar Desafíos Sociales es de crucial importancia, y debería ser percibido como una lectura esencial para todos los formuladores de políticas gubernamentales en el mundo.

