

Artículo Especial

La utilidad de las Guías de Práctica Clínica depende de su calidad.

- El instrumento AGREE II para la evaluación de Guías y el sistema GRADE para calificar la evidencia y la fortaleza de las recomendaciones

The usefulness of the Clinical Practice Guidelines depends on their quality.

- The AGREE II instrument for the evaluation of Guidelines and the GRADE system to rate the evidence and the strength of the recommendations.

Hugo R. Ramos

Instituto Modelo de Cardiología Privado SRL, Facultad de Ciencias Médicas Universidad Nacional de Córdoba. Córdoba, Argentina.

INFORMACIÓN DEL ARTÍCULO

Recibido el 17 de Enero de 2020

Aceptado después de revisión
el 31 de Enero de 2020

www.revistafac.org.ar

El autor declara no tener
conflicto de intereses

Palabras clave:

Medicina basada en la evidencia.
Guías de práctica clínica.
Evidencia de calidad.
Recomendaciones.

Keywords:

Evidence-based medicine.
Practice guidelines.
Quality evidence.
Recommendations.

RESUMEN

El desarrollo de guías de práctica clínica (GPC) requiere de una extensa revisión y análisis de la evidencia disponible, por lo que utilizar una metodología sistematizada es lo recomendable para redactarlas de la mejor calidad posible. Existen dos herramientas usadas por muchas organizaciones científicas en el mundo: el instrumento **AGREE** (*Appraisal of Guidelines for REsearch & Evaluation*) y la sistemática **GRADE** (*Grading of Recommendations Assessment, Development and Evaluation*). El primero, sirve para evaluar la calidad de una GPC y establecer si pueden ser utilizadas o no en la práctica; la segunda, evalúa la calidad de la evidencia y en qué medida puede ser utilizada para formular recomendaciones. En este artículo, se describen brevemente los principios básicos de cada una, los objetivos, la metodología y las fuentes a las que se puede acceder para contar con ellas.

The usefulness of Clinical Practice Guidelines depends on their quality.

- The AGREE II instrument for the evaluation of Guidelines and the GRADE system to rate evidence and the strength of the recommendations

ABSTRACT

The development of clinical practice guidelines (CPG) demands an extensive review and analysis of the available evidence. Using a systematic methodology is recommendable in order to write it up with the best quality possible. There are two tools that provide the framework used by many scientific organizations in the world: the **AGREE** instrument (*Appraisal of Guidelines for REsearch & Evaluation*) and the **GRADE** system (*Grading of Recommendations Assessment, Development and Evaluation*). The AGREE instrument works assessing key components of methodological quality of the guidelines, and intends to establish if the recommendations may be useful or not in clinical practice; the GRADE system categorises the quality of evidence and the strength of recommendations. In this paper the principles of each system are succinctly described, the objectives, methodology, and the sources for access.

Las guías de práctica clínica (GPC) son desarrolladas por consenso de especialistas en determinadas áreas de salud, con el fin de aportar un resumen de la evidencia disponible y establecer recomendaciones de diagnóstico, manejo, tratamiento, seguimiento y/o evaluación de tecnologías en salud. El Instituto de Medicina de los EEUU las define como “declaraciones que incluyen recomendaciones que tienen

por finalidad optimizar el cuidado de los pacientes, por una revisión sistemática de la evidencia y una evaluación de los beneficios y riesgos de opciones alternativas de cuidado”¹. Pueden ser útiles para ayudar a los médicos asistenciales a tomar decisiones y compartirlas con sus pacientes, así como para definir políticas de salud. Por ello, la calidad con la que se elaboren las GPC es clave. Existen varias herramien-

tas para este fin pero aquí describiremos sólo las más utilizadas: AGREE II y GRADE.

Actualmente, muchas de las GPC son evaluadas siguiendo el instrumento *Appraisal of Guidelines for REsearch & Evaluation (AGREE)*, publicado originalmente en 2003 por un grupo de desarrolladores e investigadores de la Universidad McMaster, Ontario, Canadá²; fue actualizada en 2009 bajo la denominación **AGREE II**³ con puestas al día en 2013 y 2017. La actualización de 2017 contiene revisiones en las secciones de la introducción “AGREE Website: Resources and References”, “10 years of AGREE” y “Scoring the AGREE II”, pero básicamente el instrumento no ha sido modificado, por lo que todas las versiones de AGREE II son válidas para su uso⁴.

El **instrumento AGREE**, se trata de una herramienta que fue diseñada para evaluar el rigor metodológico y la transparencia con la que se elabora una guía; las actualizaciones que originaron el AGREE II permiten evaluar la calidad de las guías, provee una metodología para desarrollarlas y establece qué información deberían contener y cómo deberían ser presentadas. Ha sido traducido a 20 idiomas y está en proceso de traducción a otros siete. Dado que la calidad es el punto central, el Consorcio AGREE definió la calidad de las guías como “la confianza con la que los sesgos potenciales del desarrollo de guías han sido resueltos de forma adecuada y de qué manera las recomendaciones son válidas tanto interna como externamente y son aplicables a la práctica” (AGREE Next Steps Consortium, 2017). Para ello utiliza una lista de 23 ítems que están contenidos en 6 dominios de calidad (TABLA 1).

- **Dominio 1:** alcances y objetivos. Comprende los ítems 1 a 3 y está relacionado con el objetivo general de la guía, los problemas específicos de salud a tratar y la población blanco de la guía.
- **Dominio 2:** participación de los implicados. Incluye los ítems 4 a 6 y se concentra en describir si los implicados en el desarrollo de la guía son apropiados y si representa la visión de quienes serán los usuarios.
- **Dominio 3:** rigor en la elaboración. Comprende el método por el cual se recolecta y procesa la evidencia, los métodos mediante los que se hacen las recomendaciones y cómo será el proceso de actualización (ítems 7 a 14).
- **Dominio 4:** claridad de la presentación. Se refiere a la claridad del lenguaje utilizado, la estructura y formato de la guía (ítems 15 a 17).
- **Dominio 5:** aplicabilidad. En los ítems 18 a 21 trata de las posibles barreras o facilitadores para la implementación, estrategias para facilitar su uso, e implicancias de las recomendaciones sobre la disponibilidad de recursos.
- **Dominio 6:** independencia editorial. Se describe específicamente si existen conflictos de interés que podrían influir sobre las recomendaciones (ítems 22 y 23).

Este instrumento fue diseñado para el desarrollo de guías clínicas, pero no para evaluar documentos que describan problemas organizativos de la atención en salud y aun no se ha evaluado formalmente para examinar tecnologías en salud⁵.

TABLA 1.

Lista de verificación para guiar el informe de Guías de Práctica Clínica (AGREE II 2016, basado en ref 5 y 6 con modificaciones).

Dominio 1. Alcances y objetivos		
Ítems	Criterios	Pág n°
1. Objetivos.		
El/los objetivo/s general/es de la guía está/n específicamente descrito/s.	<ul style="list-style-type: none"> • Propósito/s para la salud (Ej: prevención, screening, diagnóstico, tratamiento, etc.) • Beneficio/s esperados o resultado/s • Objetivo/s (Ej: población de pacientes, sociedad) 	
2. Preguntas.		
El/los aspecto/s de salud cubierto/s por la guía está/n específicamente descrito/s.	<ul style="list-style-type: none"> • Población blanco • Intervención/es o exposición/es • Comparación/es (si corresponde) • Resultado/s • Escenario de cuidado de salud o contexto 	
3. Población.		
La población (pacientes, público, etc.) a la cual se pretende aplicar la guía está específicamente descrita.	<ul style="list-style-type: none"> • Población blanco, sexo, y edad • Condición clínica (si es relevante) • Severidad / estadio de la enfermedad (si es relevante) • Comorbilidades (si es relevante) • Poblaciones excluidas (si es relevante) 	

Dominio 2. Participación de los implicados.		
Ítems	Criterios	Pág n°
4. Miembros del grupo.		
El grupo que desarrolla la guía incluye individuos de todos los grupos profesionales relevantes. Informa todos los individuos que estuvieron comprometidos en el proceso de desarrollo. Esto puede incluir miembros del grupo directivo, el equipo de investigación comprometido en seleccionar y evaluar la evidencia e individuos comprometidos en la formulación de las recomendaciones finales.	<ul style="list-style-type: none"> • Nombre de los participantes • Disciplina/especialidad (Ej: neurocirujano, cirujano, metodólogo) • Institución (Ej: Hospital de Clínicas) • Localización geográfica (Ej: Buenos Aires, Argentina) • Descripción del rol de los miembros en el grupo de desarrollo de la guía 	
5. Preferencias y puntos de vista de la población blanco.		
Se han tenido en cuenta los puntos de vista de la población a la que dirigida (pacientes, público, etc.). Informa cómo los puntos de vista y preferencias de la población blanco fueron investigados/considerados y cuáles fueron los resultados.	<ul style="list-style-type: none"> • Declara el tipo de estrategia para capturar los puntos de vista/preferencias de los pacientes/público (Ej: participación en el grupo de desarrollo de la guía, revisión de la literatura de valores y preferencias) • Métodos por los que las preferencias y puntos de vista fueron investigados (Ej: evidencia de la literatura, encuestas, focus groups) • Resultados/información recolectada de pacientes/público • Cómo se usó la información recolectada para informar el proceso de desarrollo de la guía y/o la formación de recomendaciones 	
6. Usuarios blanco.		
Los usuarios blanco de la guía están claramente definidos.	<ul style="list-style-type: none"> • Audiencia a quien va dirigida (Ej: especialistas, médicos de familia, pacientes, clínicos o líderes institucionales/administradores) • Cómo puede ser usada la guía por su audiencia blanco (Ej: informar las decisiones clínicas, informar la política, informar los estándares de cuidado) 	
Dominio 3. Rigor en la elaboración.		
Ítems	Criterios	Pág n°
7. Métodos de investigación.		
Se han utilizado métodos sistemáticos para la búsqueda de la evidencia. Informa los detalles de la estrategia usada para buscar la evidencia.	<ul style="list-style-type: none"> • Base/s de datos electrónicas mencionadas o fuente/s desde donde se buscó la evidencia (Ej: MEDLINE, EMBASE, PsychINFO, CINAHL) • Períodos de tiempo de la búsqueda (Ej: 1 de enero de 2004 a 31 de marzo de 2008) • Términos que se utilizaron para la búsqueda (Ej: palabras del texto, términos indexados, subtítulos) • Estrategia completa de la búsqueda (Ej: posiblemente ubicado en el apéndice) 	
8. Criterios para seleccionar la evidencia.		
Los criterios para seleccionar la evidencia se describen con claridad. (Ej: inclusión y exclusión). Provee los fundamentos cuando es apropiado.	<ul style="list-style-type: none"> • Población blanco (características de los pacientes, público, etc.) • Diseño del estudio • Comparaciones (si es relevante) • Resultados • Idioma (si es relevante) • Contexto (si es relevante) 	

9. Fortalezas y limitaciones de la evidencia.	
Las fortalezas y limitaciones del conjunto de la evidencia están claramente descritas. Considera desde la perspectiva de los estudios individuales y del cuerpo de la evidencia completo a través de todos los estudios. Existen herramientas que pueden facilitar el reporte de este concepto.	<ul style="list-style-type: none"> • Diseño del/los estudio/s incluidos en el cuerpo de la evidencia • Limitaciones en la metodología del estudio (muestreo, ciego, ocultamiento de la asignación, métodos analíticos) • Pertinencia/ relevancia de los resultados primario y secundarios • Consistencia de los resultados a través de los estudios • Dirección de los resultados a través de los estudios • Magnitud de los beneficios versus magnitud de los daños • Aplicabilidad en el contexto de la práctica
10. Formulación de recomendaciones.	
Los métodos utilizados para formular las recomendaciones están claramente descritos. Describe cómo se alcanzaron las decisiones finales. Especifica las áreas de desacuerdo y los métodos usados para resolverlas.	<ul style="list-style-type: none"> • Proceso de desarrollo de recomendaciones (Ej: pasos usados para modificar la técnica Delphi, procedimientos que fueron considerados para votar) • Resultados del proceso de desarrollo de las recomendaciones (Ej: modo en que se alcanzó el consenso usando la técnica de Delphi modificada, resultados del procedimiento de votación) • Cómo el proceso influyó en las recomendaciones (Ej: influencia de la técnica de Delphi modificada en el resultado de las recomendaciones finales)
11. Consideración de los beneficios y riesgos.	
Al formular las recomendaciones han sido considerados los beneficios en salud, los efectos secundarios y los riesgos.	<ul style="list-style-type: none"> • Respalda los datos e informa los beneficios • Respalda los datos y reporta los daños/ efectos secundarios/ riesgos • Informa el balance entre beneficios y daños/ efectos secundarios/ riesgos • Las recomendaciones reflejan consideraciones de beneficios y daños/ efectos secundarios/ riesgos
12. Relación entre recomendaciones y evidencia.	
Hay una relación explícita entre cada una de las recomendaciones y las evidencias en las que se basan.	<ul style="list-style-type: none"> • Se describe cómo el grupo de desarrollo de la guía vinculó y usó la evidencia para informar las recomendaciones • Hay un vínculo entre cada recomendación y la evidencia clave (descripción en el texto y/ o lista de referencias) • Hay un vínculo entre las recomendaciones y la evidencia resumida, y/ o evidencia en tablas en la sección resultados de la guía
13. Revisión externa.	
La guía ha sido revisada por expertos externos antes de su publicación. Informa la metodología usada para conducir la revisión externa.	<ul style="list-style-type: none"> • Propósito e intención de la revisión externa (Ej: mejorar la calidad, reunir retroalimentación en el borrador de las recomendaciones, evaluar aplicabilidad y factibilidad, diseminar evidencia) • Métodos tenidos en cuenta para la revisión externa (Ej: escala de calificación, preguntas) • Descripción de los revisores externos, (Ej: número, tipo de revisores, afiliaciones) • Resultados/ información obtenida de la revisión externa (Ej: resumen de los hallazgos clave) • Cómo se usó la información recolectada para informar el proceso de desarrollo de la guía y/ o la formulación de las recomendaciones (Ej: el panel de la guía consideró los resultados de la revisión para formular las recomendaciones finales)
14. Procedimiento para actualización.	
Se incluye un procedimiento para actualizar la guía.	<ul style="list-style-type: none"> • Hay una declaración de que la guía será actualizada • Intervalo de tiempo explícito o criterio explícito para guiar decisiones acerca de cuándo se hará una actualización • Metodología del procedimiento de actualización

Dominio 4. Claridad de la presentación.		
Ítems	Criterios	Pág n°
15. Las recomendaciones son específicas y no son ambiguas.		
Describe cuáles opciones son apropiadas en cada situación y en qué grupos de población, de acuerdo a lo informado en el cuerpo de evidencia.	<ul style="list-style-type: none"> • Existe una declaración de la acción recomendada • Intención o propósito de la acción recomendada (Ej: mejorar la calidad de vida, reducir los efectos colaterales) • Población relevante (Ej: pacientes, público) • Advertencias o declaraciones especiales de calificación, si fuera relevante (Ej: pacientes o condiciones para quienes no aplican las recomendaciones) • Si hay incertidumbre acerca de las mejor/es opción/es de cuidado, debería estar establecido en la guía 	
16. Opciones de manejo.		
Las distintas opciones para el manejo de la enfermedad o condición de salud se presentan claramente.	<ul style="list-style-type: none"> • Descripción de las opciones de manejo • Población o situación clínica más apropiada para cada opción 	
17. Las recomendaciones clave son fácilmente identificables.		
Presenta las recomendaciones clave de modo que son fácilmente identificables.	<ul style="list-style-type: none"> • Las recomendaciones están resumidas en un box, tipeadas en negrita, subrayadas o presentadas como diagramas de flujo o algoritmos • Hay recomendaciones específicas agrupadas en una sección 	

Dominio 5. Aplicabilidad.		
Ítems	Criterios	Pág n°
18. La guía describe factores facilitadores y barreras para su aplicación.		
	<ul style="list-style-type: none"> • Tipos de facilitadores y barreras que se consideraron • Métodos por los cuales fue buscada la información respecto a los facilitadores y barreras para implementar las recomendaciones (Ej: retroalimentación con implicados clave, test piloto de las guías antes de su implementación amplia) • Información/descripción de los tipos de facilitadores y barreras que surgieron a partir de la investigación (Ej: los profesionales tienen las destrezas para proveer el cuidado que se recomienda, el equipamiento suficiente no está disponible para que todos los sujetos elegibles de la población reciban una mamografía) • Cómo influyó la información en el proceso de desarrollo de la guía y/o la formulación de las recomendaciones 	
19. Facilitadores y barreras para la aplicación.		
La guía proporciona consejos y/o herramientas sobre cómo las recomendaciones pueden ser llevadas a la práctica.	<ul style="list-style-type: none"> • Materiales adicionales para respaldar la implementación de la guía en la práctica. Por ejemplo: <ul style="list-style-type: none"> • Documentos de resumen de la guía • Vínculos con listas de verificación, algoritmos • Vínculos a manuales de instrucciones • Soluciones vinculadas con el análisis de barreras (ver ítem 18) • Herramientas para capitalizar facilitadores de la guía (ver ítem 18) 	

20. Se han considerado las posibles implicaciones de las recomendaciones sobre los recursos.	
Describe toda implicancia potencial sobre los recursos al aplicar las recomendaciones.	<ul style="list-style-type: none"> • Consideraciones que se tuvieron en cuenta acerca de tipos de costo (Ej: evaluaciones económicas, costos de adquisición de drogas) • Métodos por los que se buscó la información de costos (ej: un economista de salud fue parte del panel de desarrollo de la guía, uso de evaluaciones de tecnología en salud para drogas específicas, etc.) • Información / descripción de la información de costos que surgió de la investigación (Ej: costos de adquisición de drogas para tratamiento completo) • Cómo se utilizó la información recolectada para informar el proceso de desarrollo de la guía y/o la formulación de recomendaciones
21. Criterios de monitoreo y auditoría.	
La guía ofrece criterios para monitorización y/o auditoría	<ul style="list-style-type: none"> • Criterios para evaluar la implementación de la guía o adherencia a las recomendaciones • Criterios para evaluar el impacto de implementar las recomendaciones • Consejo sobre la frecuencia e intervalo de la medición • Definiciones operacionales sobre cómo se deberían medir los criterios

Dominio 6. Independencia editorial.

Ítems	Criterios	Pág n°
22. Financiamiento		
Los puntos de vista de la entidad financiadora no han influido en el contenido de la guía.	<ul style="list-style-type: none"> • El nombre del financiador o fuente del financiamiento (o declaración explícita de que no hubo financiamiento) • Declaración de que el financiamiento no influyó en el contenido de la guía 	
23. Conflictos de interés.		
Se han registrado y abordado los conflictos de intereses de los miembros del grupo elaborador de la guía.	<ul style="list-style-type: none"> • Tipos de conflictos de intereses que se consideraron • Métodos por los que se buscaron potenciales conflictos de intereses • Descripción de los conflictos de intereses • Cómo influyeron los conflictos de intereses en el proceso de la guía y desarrollo de las recomendaciones 	

Cómo valorar los scores. Cada uno de los ítems debe ser valorado con una escala de 1 a 7 puntos, donde un score de 1 punto significa en “fuerte discrepancia” debido a que la información disponible no es relevante o el concepto está pobremente descrito. Un score de 7 puntos significa “fuerte acuerdo” y se debería dar cuando la calidad de los datos es excelente y se corresponde con todos los criterios descritos en el Manual de Usuario. El puntaje intermedio de 2 a 6 dependerá de cuan completo y de calidad sea el reporte considerado en la revisión; mientras más sean los *criterios y consideraciones* que reúna el ítem, mas alto será el puntaje. Los *criterios* se refieren a elementos explícitos que definen al ítem, mientras que las *consideraciones* apuntan a

ayudar a completar la evaluación con la intervención de los revisores o evaluadores que utilizarán su juicio para evaluar la información. Es decir, que AGREE II no tiene una descripción específica para cada punto dado, sino que está basada en el juicio de los evaluadores, guiados por los criterios y consideraciones.

Algunas guías, debido a que podrían tener un objetivo limitado, pueden no incluir todos los ítems descritos en los dominios; sin embargo, AGREE II recomienda que “independientemente de la estrategia elegida, esto se debería decidir por adelantado y de manera explícita y, si algunos ítems son omitidos, se deberían hacer modificaciones apropiadas para calcular los scores de los dominios. Como

principio, la exclusión de ítems no se recomienda". (AGREE Next Steps Consortium, 2017).

Score de calidad. Se debe hacer una puntuación de la calidad de cada uno de los 6 dominios, puntuando en ellos cada ítem con la escala de 1 a 7 puntos. Cada revisor puntúa individualmente y se suma el puntaje dado a cada ítem del dominio. También se suma el total de los puntajes de los evaluadores y se aplica una fórmula para establecer la calidad de ese dominio. A continuación un "ejemplo" para el dominio 1 (tomado de: AGREE Next Steps Consortium, 2017).

Ejemplo: Evaluación del dominio 1 por 4 evaluadores de la guía.

	Ítem 1	Ítem 2	Ítem 3	Total
Evaluador 1	5	6	6	17
Evaluador 2	6	6	7	19
Evaluador 3	2	4	3	9
Evaluador 4	3	3	2	8
Total	16	19	18	53

Máximo score posible = 7 (fuerte acuerdo) x 3 (ítems) x 4 (evaluadores) = 84

Mínimo score posible = 1 (fuerte discrepancia) x 3 (ítems) x 4 (evaluadores) = 12

- Se aplica la siguiente fórmula:

$$\frac{\text{Score obtenido} - \text{Mínimo score posible}}{\text{Máximo score posible} - \text{Mínimo score posible}}$$

$$\frac{53 - 12}{84 - 12} \times 100 = \frac{41}{72} \times 100 = 0.5694 \times 100 = 57\%$$

No hay un patrón o valor de alta calidad o de baja calidad recomendado por el Consorcio AGREE, por lo que el usuario del instrumento debe tomar una decisión basado en el contexto en el que está siendo utilizado. Al completar los 23 ítems, dos evaluaciones globales deben considerar la calidad de la guía: 1) calidad global de la guía: en una escala de 7 puntos se califica como "la más alta calidad posible" o "la más baja calidad posible"; 2) recomendación para el uso: es decir se emite un juicio teniendo en cuenta los criterios considerados en el proceso de evaluación y si se recomienda su uso. Se pueden emitir las siguientes respuestas: "sí", "sí con modificaciones" o "no"⁶.

¿Cuánto tiempo lleva desarrollar una guía bajo los criterios AGREE II? Dado el detallado y complejo proceso, la duración media es de 12 a 18 meses.

¿Quiénes deberían evaluar la guía? Se recomienda que sean al menos dos revisores externos, pero preferentemente cuatro, para aumentar su confiabilidad.

Se han publicado muchos estudios utilizando el instrumento AGREE para evaluar la calidad de las guías y es la herramienta más utilizada en idioma inglés. Se ha visto que los dominios 3 y 5 son los que más influyen sobre las dos evaluaciones globales, aunque éstas son pobremente descritas por los mismos evaluadores de las guías⁷. Además,

en guías pediátricas se ha reportado que hasta 2/3 de ellas fueron consideradas de moderada a baja calidad utilizando el instrumento AGREE II, y el otro tercio fue considerado de buena calidad⁸; las calificaciones de baja calidad fueron debidas sobre todo al pobre desempeño en el dominio 3 ("Rigor en la elaboración") que mide la metodología de su desarrollo y en el dominio 6 ("Independencia editorial").

En esta publicación no se pretende hacer una revisión extensa del instrumento AGREE sino solamente una exposición sumaria de su historia, objetivos y metodología para evaluar las GPC, y estimular a los Comités de Especialidades, revisores y evaluadores a utilizar esta herramienta, que está en continua revisión porque no es perfecta, pero que provee una metodología trazable para mejorar la calidad de las GPC y la atención de los pacientes.

El sistema GRADE. Uno de los problemas que tienen quienes desarrollan una GPC es cómo calificar la evidencia y redactar adecuadamente las recomendaciones basadas en la calidad de aquella evidencia; a menudo se detectan inconsistencias entre la evidencia y las recomendaciones⁹. Por este motivo, en 2004 el Grupo de Trabajo GRADE (*Grading of Recommendations Assessment, Development and Evaluation*) publicó su propuesta para crear un sistema transparente, altamente estructurado, con el fin de ayudar a quienes confeccionan GPC para calificar la calidad de la evidencia y la fortaleza de sus recomendaciones^{10,11}. Aunque tiene críticas y observaciones para mejoría en sus definiciones y metodología¹², es una herramienta muy utilizada, incluso por la Organización Mundial de la Salud (OMS) para el desarrollo de sus guías éticas¹³. Para calificar la evidencia utiliza cuatro niveles y para calificar las recomendaciones utiliza dos¹⁴.

Definiciones de la calidad de la evidencia:

- 1) **Alta calidad:** se considera que es muy improbable que una investigación futura cambie la confianza en la estimación del efecto.
- 2) **Calidad moderada:** es probable que una futura investigación tenga un impacto importante en la confianza de la estimación del efecto y pueda cambiar la estimación.
- 3) **Baja calidad:** es muy probable que una futura investigación tenga un impacto importante en la estimación del efecto y es probable que cambie la estimación del efecto.
- 4) **Muy baja calidad:** es muy incierta cualquier estimación del efecto.

Los ensayos controlados aleatorizados se suelen considerar desde el principio como de alta calidad pero pueden bajar de categoría por varias razones: a) limitaciones del estudio; b) inconsistencia de los resultados; c) evidencia indirecta; d) imprecisión; e) reporte de sesgos. Del mismo modo, estudios observacionales que comienzan como de un nivel de baja calidad pueden elevar su categoría si la magnitud del efecto investigado es muy grande (por ejemplo: mejoría de la osteoartritis de cadera con el reemplazo de cadera), si hay una relación dosis respuesta, si todas las tendencias disminuyen la magnitud de un efecto del tratamiento⁹.

TABLA 2.

Factores que afectan la fortaleza de una recomendación (tomado de: AGREE Next Steps Consortium, 2017).

Factor	Recomendaciones fuertes. Ejemplos.	Recomendaciones débiles. Ejemplos.
Calidad de la evidencia	Muchos ensayos aleatorizados han mostrado el beneficio de los esteroides inhalados en el asma	Solamente series de casos han investigado la utilidad de la pleurodesis en el neumotórax
Incertidumbre en el balance entre los efectos deseables e indeseables	Aspirina en el infarto de miocardio reduce la mortalidad con mínimos efectos colaterales, inconvenientes y costo	Warfarina en pacientes de bajo riesgo con fibrilación auricular reduce en pequeña proporción el ACV pero aumenta el riesgo de sangrado y presenta inconvenientes sustanciales
Incertidumbre o variabilidad en los valores y preferencias	Pacientes jóvenes con linfoma invariablemente le dan un alto valor a los efectos de prolongar la vida que tiene la quimioterapia, por sobre los efectos tóxicos	Pacientes ancianos pueden no dar un alto valor a los efectos de prolongar la vida de la quimioterapia por sobre la toxicidad del tratamiento
Incertidumbre acerca de si la intervención representa un inteligente uso de los recursos	El bajo costo de la aspirina como profilaxis del ACV en pacientes con AIT	El alto costo de clopidogrel y de la combinación de dipiridamol y aspirina como profilaxis del ACV en pacientes con AIT

ACV: accidente cerebrovascular; AIT: ataque isquémico transitorio.

Graduación de las recomendaciones:

- 1) Fuerte:** cuando los efectos deseables de una intervención sobrepasan a los efectos indeseables, o cuando sucede lo contrario.
- 2) Débil:** cuando la solución es menos certera, debido a baja calidad de la evidencia o porque la evidencia sugiere que los efectos deseables y los efectos indeseables están estrechamente balanceados.

También hay otros factores que pueden influir para que una recomendación sea “fuerte” o “débil” como lo muestra la TABLA 2.

Tanto en el instrumento AGREE II como en GRADE, se debe usar el juicio de quienes intervienen, lo que deja una ventana abierta a la subjetividad de la evaluación, lo que ha sido criticado por algunos autores¹². Sin embargo, ambas herramientas son extremadamente útiles para elaborar GPC de una manera ordenada, transparente y sistematizada que mejora su calidad, aunque esto está en continua evolución^{15,16}. Series de artículos del grupo GRADE y autores independientes analizaron los distintos aspectos que influyen en la toma de decisiones para evaluar la evidencia; por ejemplo el riesgo de sesgo¹⁷, inconsistencia, imprecisión, sesgo de publicación, calificación y variabilidad¹⁸, la calificación de la evidencia^{19,20,21,22,23} etc.

Así, una GPC puede ser desarrollada bajo estándares internacionales, aunque no perfectos pero sí en condiciones trazables y de la mejor calidad disponible.

BIBLIOGRAFIA

- Graham RM, Mancher M, Miller-Wolman D, Greenfield S, Steinberg E, editors. Clinical practice guidelines we can trust. Washington: National Academies Press; 2011.

- AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Health Care* 2003; 12 (1): 18-23.
- Brouwers MC, Kho ME, Browman GP, et al, for the AGREE Next Steps Consortium. AGREE II: Advancing guideline development, reporting and evaluation in healthcare. *CMAJ* 2010; 182: E839-42.
- AGREE Next Steps Consortium (2017). The AGREE II Instrument [Electronic version]. Retrieved January, 15th, 2020, from <http://www.agreetrust.org>
- AGREE Next Steps Consortium (2009). El Instrumento AGREE II Versión electrónica. Disponible en: <http://www.agreetrust.org>; Versión en español: <http://www.guiasalud.es> / https://www.agreetrust.org/wp-content/uploads/2013/06/AGREE_II_Spanish.pdf
- Brouwers MC, Kerkvliet K, Spithoff K, et al on behalf of the AGREE Next Steps Consortium. The AGREE Reporting Checklist: a tool to improve reporting of clinical practice guidelines. *BMJ* 2016; 352: i1152.
- Hoffmann-Eßer W, Siering U, Neugebauer EAM, et al. Guideline appraisal with AGREE II: Systematic review of the current evidence on how users handle the 2 overall assessments. *PLoS ONE* 2017; 12 (3): e0174831.
- Bhatt M, Nahari A, Wang P-W, et al. The quality of clinical practice guidelines for management of pediatric type 2 diabetes mellitus: a systematic review using the AGREE II instrument. *Systematic Reviews* (2018) 7: 193.
- Guyatt GH, Oxman AD, Vist GE, et al. GRADE: what is “quality of evidence” and why is it important to clinicians? *Br Med J*. 2008; 336 (7651): 995-98.
- Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004; 328 (7454): 1490.
- Atkins D, Eccles M, Flottorp S, et al. Systems for grading the quality of evidence and the strength of recommendations I: Critical appraisal of existing approaches The GRADE Working Group. *BMC Health Serv Res* 2004; 4 (1): 38.
- Mercuri M, Baigrie B, Upshur REG. Going from evidence to recommendations: Can GRADE get us there? *J Eval Clin Pract*. 2018; 24 (5): 1232-39.
- World Health Organization. WHO Handbook for Guideline Development. 2nd ed. Geneva: WHO Press. at: <http://apps.who.int/medicinedocs/documents/s22083en/s22083en.pdf>; 2014
- Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; 336: 924-26.
- Alonso-Coello P, Schünemann HJ, Moberg J, 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.

Bibliografía completa disponible en www.revistafac.org.ar