



Artículo especial

Calidad y fuerza: el sistema GRADE para la formulación de recomendaciones en las guías de práctica clínica

Quality and strength: The GRADE system for formulating recommendations in clinical practice guidelines

Pablo Alonso-Coello^{a,b,*}, David Rigau^{a,b}, Andrea Juliana Sanabria^a, Vicente Plaza^c, Marc Miravittles^d y Laura Martínez^{a,b}

^a Centro Cochrane Iberoamericano, Instituto de Investigación Biomédica (IIB-Sant Pau), Barcelona, España

^b CIBER de Epidemiología Clínica y Salud Pública (CIBERESP), Barcelona, España

^c Servicio de Neumología, Hospital de la Santa Creu i Sant Pau, Barcelona, España

^d Servicio de Neumología, Hospital Universitari Vall d'Hebron, Barcelona, España

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido el 3 de octubre de 2012

Aceptado el 7 de diciembre de 2012

On-line el 21 de febrero de 2013

Introducción

Durante la última década las guías de práctica clínica (GPC) han experimentado un auge muy importante, consolidándose como una herramienta para la toma de decisiones. Las GPC presentan de una manera sintética la mejor información disponible en forma de recomendaciones para la práctica clínica. Así mismo, en los últimos años ha habido avances relevantes por lo que se refiere a la metodología para su elaboración, actualización e implementación¹. Estos avances prestan una mayor atención a la composición multidisciplinar de los grupos elaboradores, incluyendo a los pacientes, la gestión de los conflictos de interés, así como la búsqueda exhaustiva de la literatura y la evaluación detallada de la calidad y la graduación de la fuerza de las recomendaciones, entre otros aspectos²⁻⁵.

El problema y una potencial solución

A pesar de los avances en el desarrollo de las GPC, su calidad es todavía mejorable⁶. Un aspecto que ha generado confusión es la presencia de diferentes sistemas para evaluar la calidad de la evidencia y la fuerza de las recomendaciones, los cuales presentan en mayor o menor medida limitaciones^{7,8}. Estos sistemas son claves a la hora de transmitir a los usuarios la confianza que pueden

depositar en la información dada en la GPC. Un ejemplo son las GPC sobre asma, pues sus principales guías utilizaban un sistema que no graduaba la fuerza de las recomendaciones⁹. Este aspecto es considerado actualmente crucial a la hora de formular recomendaciones, pues existen otros factores, además de la evidencia disponible y su calidad, que se deben tener en cuenta en la elaboración de las recomendaciones y la graduación de su fuerza (p. ej., el balance beneficio riesgo o los costes).

En este contexto, un grupo internacional de epidemiólogos, metodólogos y clínicos provenientes de las principales instituciones que elaboran GPC ha desarrollado una propuesta que tiene como objetivo consensuar un sistema común que supere las limitaciones de los sistemas previos^{10,11}. Este grupo de profesionales constituye el grupo de trabajo *Grading of Recommendations Assessment, Development and Evaluation* (GRADE). El sistema GRADE ha sido adoptado por más de 70 organizaciones en todo el mundo, algunas tan importantes como la Organización Mundial de la Salud (OMS), la Colaboración Cochrane, el *National Institute of Clinical Excellence* (NICE), el *Scottish Intercollegiate Guidelines Network* (SIGN) o publicaciones como *Clinical Evidence* o *Update* (<http://www.gradeworkinggroup.org/society/index.htm>). En nuestro entorno, el Programa Nacional de Elaboración de Guías de Práctica Clínica del Sistema Nacional de Salud (<http://www.guiasalud.es/web/guest/gpc-sns>), la Guía Española para el Manejo del Asma (GEMA), la Guía Española de la EPOC (GesEPOC) o la Sociedad Española de Medicina de Familia y Comunitaria (semFYC), entre otros, ya lo han adoptado o utilizado^{1,12-14}.

* Autor para correspondencia.

Correo electrónico: palonso@santpau.cat (P. Alonso-Coello).

¿Qué diferencia GRADE de otros sistemas?

Las principales diferencias del sistema GRADE con otros sistemas son:

- Valoración de la importancia relativa para los clínicos y pacientes de los desenlaces de interés.
- Delimitación clara entre calidad de la evidencia y graduación de fuerza de la recomendación.
- Disponer de criterios explícitos para aumentar o disminuir la calidad de la evidencia independientemente del diseño del estudio (ensayo clínico aleatorizado [ECA] u observacional).
- Consideración de los valores y preferencias de los pacientes en la formulación de recomendaciones.
- Proponer un proceso estructurado y explícito para la elaboración de recomendaciones.

Estas características, junto a un amplio consenso internacional, hacen del sistema GRADE un marco metodológico sistemático, explícito y transparente para la evaluación de la calidad de la evidencia y la graduación de la fuerza de las recomendaciones.

La importancia de delimitar la pregunta clínica y los desenlaces de interés

Una de las primeras etapas al desarrollar una GPC, independientemente del sistema utilizado para evaluar la calidad de la evidencia y la fuerza de las recomendaciones, consiste en la definición de la pregunta clínica por parte del grupo elaborador. Esta pregunta debe ser planteada en una forma estructurada, habitualmente utilizando un formato PICO de pregunta (Población, Intervención, Comparación y desenlaces [Outcome]).

El sistema GRADE da una especial importancia al planteamiento de los desenlaces de interés (p.ej., reagudización de la sintomatología o efectos adversos graves), pues estos serán los que delimiten el balance entre los beneficios y los riesgos de la intervención evaluada. El sistema GRADE señala que no todos los desenlaces de interés tienen la misma importancia y que, por tanto, solo los más importantes deberían influenciar nuestra evaluación de la calidad de la evidencia y la graduación de las recomendaciones¹⁵. En concreto, divide los desenlaces en claves, importantes pero no claves y no importantes, siendo los claves los que hay que tener más en cuenta. Esta propuesta lleva consigo que un grupo elaborador tenga que evaluar la importancia relativa de los desenlaces que incluya además la perspectiva de los pacientes.

En el caso de una GPC sobre el manejo de pacientes con rinitis alérgica y asma, el grupo elaborador consideró que, para el tratamiento preventivo del asma, la disminución de la sintomatología y la disminución de las exacerbaciones eran desenlaces clave para los pacientes (tabla 1)¹⁶. Como desenlaces importantes pero no claves se consideraron la calidad de vida y los efectos adversos. Los resultados espirométricos o gasométricos fueron considerados como desenlaces no importantes. Solo los desenlaces

claves e importantes fueron tenidos en cuenta a lo largo de todo el proceso.

La confianza en la evidencia disponible

Los usuarios de una guía necesitan saber qué confianza pueden depositar en los resultados disponibles de los estudios. Esta confianza es lo que constituye la denominada calidad de la evidencia. GRADE define la calidad como el grado de confianza que tenemos en que la estimación de un efecto es adecuada para apoyar una recomendación¹⁷. Por ejemplo, en pacientes con EPOC estable, el tratamiento combinado con un beta-2 de acción larga y corticoides reduce en un 28% el riesgo de exacerbaciones en comparación con placebo (RR, 0,72; IC 95%, 0,65-0,80)¹⁴. Esta disminución del 28% de riesgo de exacerbaciones con el tratamiento combinado es el efecto estimado de la intervención. La confianza en esta estimación depende de múltiples factores, como las limitaciones del diseño y la ejecución del estudio (riesgo de sesgo), la consistencia de los resultados o la precisión de los mismos, entre otros¹⁷. El sistema GRADE evalúa la calidad para cada uno de los desenlaces considerados como claves en una misma pregunta de interés. En el ejemplo del tratamiento de la EPOC avanzada, el grupo elaborador de la GPC podría considerar, además del riesgo de exacerbaciones, la mejora en la sintomatología nocturna. Así, sería posible realizar una evaluación de la calidad también para este desenlace y otros, si los hubiera.

El sistema GRADE categoriza la calidad de la evidencia en:

- **Calidad alta:** confianza alta en que la estimación del efecto a partir de la literatura disponible se encuentra muy cercana al efecto real.
- **Calidad moderada:** es probable que la estimación del efecto se encuentre cercana al efecto real, aunque podrían existir diferencias sustanciales.
- **Calidad baja:** la estimación del efecto puede ser sustancialmente diferente al efecto real.
- **Calidad muy baja:** es muy probable que la estimación del efecto sea sustancialmente diferente al efecto real.

¿Qué factores influyen en la calidad de la evidencia?

La tabla 2 resume los diferentes factores que pueden disminuir la confianza en la estimación del efecto observado, diferenciando los ECA (que inicialmente aportan una alta calidad para evaluar el efecto de las intervenciones) de los estudios observacionales (que inicialmente se considera que aportan una calidad baja). El cómputo de estos factores, que limitan nuestra confianza en los resultados, determinará que nuestra confianza aumente o disminuya. El sistema GRADE establece que la calidad global es la menor entre los desenlaces clave considerados¹⁸. Por último, reconoce que la opinión de los expertos influye en la evaluación de la evidencia disponible (independientemente del diseño) pero no la considera un tipo de evidencia en sí misma.

Tabla 1
Componentes PICO de la pregunta «¿Los antagonistas de los receptores de los leucotrienos deben ser usados para el tratamiento de asma en pacientes con rinitis alérgica y asma?»¹⁶

Pacientes	Intervención	Comparador	Desenlaces (outcomes) e importancia ^a
Pacientes con rinitis alérgica y asma	Corticoides inhalados + antagonistas de los receptores de los leucotrienos	Corticoides inhalados	- Disminución de la sintomatología diurna (7-9) - Disminución de la sintomatología nocturna (7-9) - Disminución las exacerbaciones (7-9) - Calidad de vida (4-6) - Eventos adversos (4-6) - Resultados espirométricos o gasométricos (1-3)

^a GRADE sugiere calificar la importancia de los desenlaces con una escala de nueve puntos: 1-3, desenlace no importante; 4-6, desenlace importante pero no clave para la toma de decisiones; 7-9, desenlace clave para la toma de decisiones.

Tabla 2
Evaluación de la calidad de la evidencia y factores modificadores

Diseño del estudio	Calidad de la evidencia	Disminuir si	Aumentar si
Ensayo controlado aleatorio→	Alta	<i>Limitaciones en el diseño:</i> 1 Importantes 2 Muy importantes	<i>Asociación fuerte:</i> 1 Fuerte 2 Muy fuerte
	Moderada	<i>Resultados inconsistentes:</i> 1 Inconsistentes 2 Muy inconsistentes	<i>Presencia de un gradiente dosis respuesta:</i> 1 Evidencia de un gradiente
Estudio observacional→	Baja	<i>Ausencia de evidencia directa:</i> 1 Indirecta 2 Muy indirecta	<i>Potenciales factores de confusión:</i> 1 Reduciría un efecto demostrado o 1 Sugeriría un efecto espurio cuando los resultados no muestran ningún efecto
	Muy baja	<i>Resultados imprecisos:</i> 1 Imprecisos 2 Muy imprecisos <i>Sesgo de publicación</i> 1 Probable 2 Muy probable	

Adaptado de: Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011;64:383-94.

A continuación se describen los factores que influyen en la calidad y, por tanto, la confianza que se puede depositar en la estimación de un efecto:

Limitaciones en el diseño o la ejecución

Las limitaciones en el diseño o la ejecución (riesgo de sesgo) difieren según se consideren ECA o estudios observacionales. En los ECA se consideran los siguientes factores: la ausencia de ocultamiento de la secuencia de aleatorización, un enmascaramiento inadecuado, unas pérdidas de seguimiento importantes y la ausencia de análisis por intención de tratar, la inclusión selectiva de los desenlaces de interés, además de otros factores menos frecuentes, como la finalización prematura de un estudio por razones de beneficio, el uso de medidas no validadas, el efecto arrastre (*carryover*) en los ensayos cruzados o la presencia de sesgo de reclutamiento en ensayos aleatorizados por grupos (*clusters*)¹⁹.

En los estudios observacionales se consideran: la presencia de criterios de selección de la población inapropiados, mediciones inapropiadas para la exposición o el desenlace de interés, el control inadecuado de los factores de confusión o el seguimiento incompleto¹⁹.

Resultados inconsistentes

La calidad de la evidencia disminuye si los resultados son inconsistentes o heterogéneos, es decir, si los distintos estudios muestran resultados muy diferentes. Se debe valorar además si tras explorar las razones que pudieran explicar la heterogeneidad observada (p.ej., diferencias en la población, la intervención, los desenlaces de

resultado o el riesgo de sesgo), esta inconsistencia persiste. En caso de no identificar las razones que podrían explicar esta variabilidad, la confianza disminuye, pues podrían existir diferencias reales entre las estimaciones del efecto proporcionadas por los estudios incluidos²⁰.

Por ejemplo, una revisión sistemática (RS) que evaluó la eficacia de la inmunoterapia alérgeno-específica frente a placebo en personas adultas con rinitis alérgica muestra que los resultados para los síntomas nasales son muy variables entre los diferentes estudios, sus intervalos de confianza no se solapan, la prueba de heterogeneidad es significativa y el estadístico I² presenta un valor elevado²¹. En situaciones como esta, hay una menor confianza en los resultados y se debería reducir la calidad por este motivo (fig. 1).

Ausencia de evidencia directa

En situaciones de ausencia de comparaciones directas entre las intervenciones consideradas, o de diferencias importantes entre los estudios disponibles y la población, las intervenciones o los desenlaces planteados en la pregunta de interés, podemos encontrarnos con que únicamente disponemos de información indirecta²².

Por ejemplo, en el caso de los desenlaces, no existe evidencia directa al evaluar la administración de 2 broncodilatadores de acción prolongada en la EPOC, frente a la administración de un broncodilatador beta-adrenérgico de acción prolongada junto con un corticoide inhalado. En este caso, un único ECA aporta resultados espirométricos pero no resultados de desenlaces clínicos relevantes para los pacientes (p.ej., mejora en la sintomatología). Nuestra confianza en que una mejoría en los resultados en las pruebas de

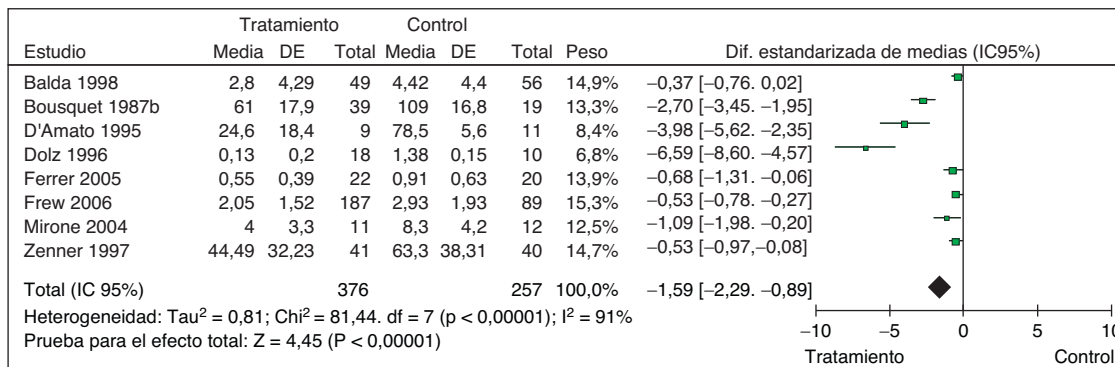


Figura 1. Eficacia de la inmunoterapia alérgeno-específica frente a placebo en personas adultas con rinitis alérgica. IC 95%: intervalo de confianza del 95%; DE: desviación estándar.
Adaptado de: Calderon MA, et al. Allergen injection immunotherapy for seasonal allergic rhinitis. Cochrane Database of Systematic Reviews 2007, Issue 1.

espirometría reflejen una mejoría en desenlaces más importantes para los pacientes es incierta y, por tanto, nuestra confianza es menor²³. Para la misma pregunta de interés, se ha publicado un metaanálisis que aporta resultados para la frecuencia de exacerbaciones basado en estudios que evaluaron la administración de 2 broncodilatadores o que, por otro lado, evaluaron la administración de un broncodilatador beta-adrenérgico de acción prolongada más un corticoide inhalado. No obstante, no se dispone de comparaciones directas entre estas estrategias de tratamiento, y a partir de este metaanálisis se dispone de una estimación indirecta y, por tanto, menos fiable o de menor calidad²⁴.

En el caso del tratamiento con antihistamínicos en pacientes con asma y rinitis alérgica, la evidencia de que se dispone es indirecta a causa de diferencias en la población: los ECA realizados incluyen hasta un 60% de pacientes sin asma al inicio del ensayo. En esta misma línea, otro ejemplo sería la evaluación de la eficacia de descongestionantes nasales como tratamiento de rescate en pacientes con rinitis alérgica. Los estudios encontrados analizan la eficacia de su uso regular (no de rescate), por lo cual dispondríamos también de evidencia indirecta. En ambos casos, la confianza que se tiene en los resultados de estos estudios para responder a las preguntas planteadas es, por tanto, menor¹⁶.

Resultados imprecisos

Para considerar que el efecto de una intervención es impreciso se debe valorar el estimador del efecto, preferiblemente en términos absolutos (en lugar de en términos relativos) y su intervalo de confianza correspondiente. Si al considerar un extremo u otro del intervalo de confianza para un desenlace, teniendo en cuenta los riesgos e inconvenientes de la intervención, nuestra recomendación cambiara, la confianza en el estimador del efecto, por impreciso, disminuiría. Así mismo, ante un intervalo de confianza preciso, si el número de eventos o el número de participantes evaluados en los diferentes estudios son escasos, se debe también considerar disminuir la confianza²⁵.

Por ejemplo, en una guía reciente se evaluó la eficacia de los antihistamínicos H₁ frente a placebo para reducir la aparición de asma en niños con diferentes tipos de alergia¹⁶. Los resultados de 3 ECA muestran que los antihistamínicos H₁ reducen el riesgo de desarrollar asma de forma no significativa. El beneficio absoluto muestra que un extremo del intervalo de confianza aporta un beneficio significativo (10 niños menos por cada 100 tratados van a desarrollar asma, en comparación con placebo), por lo que generaría una recomendación a favor, pero el otro extremo del mismo intervalo aporta un perjuicio para el tratamiento (31 niños más por cada 100 tratados van a desarrollar asma, en comparación con placebo), por lo que generaría una recomendación en contra. El panel de autores de la guía decidió reducir la confianza en este desenlace por impreciso¹⁶.

Sospecha de sesgo de publicación

Finalmente, hay situaciones en las que se sospecha que hay estudios, principalmente con resultados negativos, que no se han publicado y que por tanto existe una posible sobreestimación del efecto²⁶. Se debe explorar esta posibilidad si nos encontramos con un conjunto de ensayos de pequeño tamaño, positivos y financiados por la industria²⁷. Para detectar este posible sesgo existen pruebas estadísticas o gráficas (*funnel plot*). En estos casos se reduciría la confianza en la estimación de un efecto.

¿Qué factores aumentan la calidad de la evidencia?

Las situaciones que pueden justificar un aumento de nuestra confianza en los resultados de un conjunto de estudios son menos comunes y se aplican fundamentalmente en los estudios observacionales (cohortes y casos y controles), siempre que no coexistan otras limitaciones de diseño y ejecución (riesgo de sesgo)²⁸.

Asociación fuerte

Cuando los resultados de un estudio, sin otras limitaciones, muestran un efecto, protector o perjudicial, con una asociación fuerte (riesgo relativo u *odds ratio* >2 o <0,5) o muy fuerte (riesgo relativo u *odds ratio* >5 o <0,2), la confianza en estos resultados aumenta²⁸. Un ejemplo es la relación que se encuentra entre la mortalidad por cualquier causa y el uso de tabaco que es hasta 3 veces superior en los fumadores respecto a los no fumadores, en una cohorte prospectiva de médicos británicos²⁹. La confianza en esta asociación es, por tanto, al menos moderada.

Gradiente dosis-respuesta

Es también un motivo para aumentar la confianza en la estimación de un efecto la existencia de un claro gradiente dosis-respuesta, ya que nos aporta una mayor certidumbre sobre una potencial relación causa-efecto. Por ejemplo, se ha comprobado que el riesgo de desarrollar EPOC es proporcional al consumo acumulado de tabaco, siendo 2,6 veces superior en los fumadores de 15 a 30 paquetes año y 5,1 veces superior en los fumadores de más de 30 paquetes año³⁰. La existencia de este gradiente de asociación entre el factor estudiado y el efecto aumenta la confianza en la relación entre el tabaco y la EPOC.

Consideración de los potenciales factores de confusión y sesgo residuales

En ocasiones pueden existir situaciones en las que se observa un efecto asociado a una intervención y para la cual, después de analizar los potenciales factores que podrían reducir el efecto observado, se considera que estos factores, de existir, reforzarían las conclusiones obtenidas²⁸. Por ejemplo, una revisión sistemática de estudios observacionales mostró una tasa de mortalidad superior en hospitales de gestión privada con ánimo de lucro, en comparación con hospitales de gestión privada sin ánimo de lucro (*not for profit*), a pesar de que estos últimos posiblemente eran pacientes más graves³¹.

Para facilitar la evaluación de la calidad de la evidencia y sus factores modificadores, el sistema GRADE proporciona la posibilidad de sintetizar la evidencia en una tabla resumen de los hallazgos (*summary of findings* [SoF]), que refleja de manera estructurada el número de estudios por desenlace de interés, la calidad de la evidencia y los resultados observados en términos relativos y absolutos. Estas se generan con un programa informático de distribución libre denominado GRADEPro³².

Una GPC sobre rinitis alérgica y asma evaluó el uso de antagonistas de los leucotrienos por vía oral en monoterapia para el tratamiento de base del asma (uso de corticoides inhalados en pacientes con rinitis alérgica y asma)¹⁶, elaborándose una tabla resumen de la evidencia disponible para los desenlaces de interés (tabla 3). Respecto a las exacerbaciones que requerían el uso de corticoides sistémicos, se observó que, en términos absolutos, existía una clara disminución con el uso de corticoides inhalados respecto a los antagonistas de los leucotrienos. El grupo de pacientes en tratamiento con inhibidores de los receptores de los leucotrienos presentaba 30 exacerbaciones más por cada 1.000 pacientes, respecto al grupo en tratamiento con corticoides inhalados (calidad alta). Por otra parte, los inhibidores de los receptores de los leucotrienos presentaron menos efectos adversos (4 menos por cada 1.000) que los corticoides inhalados (calidad moderada).

¿Por qué se debe graduar la fuerza de las recomendaciones?

Los usuarios de las guías deben reconocer rápidamente hasta qué punto pueden confiar en que una recomendación va a aportar más consecuencias deseables que indeseables. La fuerza de la recomendación refleja un gradiente de confianza, con una mayor confianza en las recomendaciones fuertes y menor confianza en las

Tabla 3

Resumen de los hallazgos sobre antagonistas de los leucotrienos comparados con corticoides inhalados en el tratamiento del asma en pacientes con rinitis alérgica y asma

Desenlaces	Participantes (estudios) Seguimiento	Calidad de la evidencia (GRADE)	Efecto relativo (IC 95%)	Efecto absoluto
Exacerbación que requiere uso de esteroides sistémicos	1.018 (2) 6-40 semanas	⊕⊕⊕⊕ Alta	RR 1,56 (1,36 a 2,00)	30 más por 1.000 (desde 19 más a 53 más)
Ingreso hospitalario por exacerbación	3.189 (13) 6-40 semanas	⊕⊕⊕○ Moderada ^a	RR 1,62 (0,64 a 4,15)	2 más por 1.000 (desde 1 menos a 9 más)
Calidad de vida: cambio desde el nivel basal, medido con el cuestionario de calidad de vida de asma; mayor puntuación significa mejor	1.027 (2) 8-16 semanas	⊕⊕⊕○ Moderada ^b	-	DM -30 (-0,43 a -0,17)
Síntomas diurnos (menos significa mejor)	2.543 (6) 8-16 semanas	⊕⊕⊕⊕ Alta	-	DEM 0,29 (0,21 a 0,37)
Síntomas nocturnos (menos significa mejor)	1.995 (6) 8-16 semanas	⊕⊕⊕⊕ Alta	-	DEM 0,21 (0,13 a 0,30)
Días sin síntomas	1.328 (5) 8-16 semanas	⊕⊕⊕⊕ Alta	-	DM -11,47 (-15,72 a -7,23)
Efectos adversos	6.277 (16) 6-40 semanas	⊕⊕⊕○ Moderada ^c	0,99 (0,93 a 1,04)	4 menos por 1.000 (desde 29 menos a 17 más)

DEM: diferencia estandarizada de medias;DM: diferencia de medias; IC: intervalo de confianza; RR: riesgo relativo.

Adaptado de: Brozek JL, Bousquet J, Baena-Cagnani CE, et al. Allergic rhinitis and its impact on asthma (ARIA) guidelines: 2010 Revision. J Allergy Clin Immunol. 2010;126:466-76.

^a Imprecisión: el intervalo de confianza del 95% incluye la ausencia de efecto o 1% más de pacientes requieren hospitalización en términos absolutos.

^b Riesgo de sesgo, solo dos ensayos estudian este desenlace.

^c Imprecisión: el intervalo de confianza del 95% incluye la ausencia de efecto. Desde 3% menos a 2% más en términos absolutos.

Tabla 4

Implicaciones de la fuerza de las recomendaciones

	Recomendación fuerte	Recomendación débil
Para los pacientes	La mayoría de las personas estarían de acuerdo con la intervención recomendada y únicamente una pequeña parte no lo estarían	La mayoría de las personas estarían de acuerdo con la acción recomendada pero un número importante de ellas no
Para los profesionales sanitarios	La mayoría de los pacientes deberían recibir la intervención recomendada	Se reconoce que diferentes opciones son apropiadas para diferentes pacientes y que el médico tiene que ayudar a cada paciente a llegar a la decisión más consistente con sus valores y preferencias
Para los gestores	La recomendación puede ser adoptada como política sanitaria en la mayoría de las situaciones	Existe necesidad de un debate importante con la participación de los grupos de interés

Adaptado de: Guyatt et al.³³.

recomendaciones débiles. A su vez, la dirección de la recomendación puede ser a favor o en contra (fig. 2). Las recomendaciones, fuertes o débiles, tienen implicaciones diferentes en el caso de los pacientes, los profesionales sanitarios o los gestores (tabla 4).

Existen 4 factores fundamentales que influyen en la fuerza de las recomendaciones según GRADE: el balance entre los beneficios y riesgos, la calidad de la evidencia, los valores y preferencias de los pacientes y, finalmente, los costes o utilización de recursos³³.

Balance entre beneficios y riesgos

Es necesario conocer el balance entre el efecto para los desenlaces deseados y los indeseados. Para realizar este balance hay que asignar un peso o un valor a estos desenlaces. Esto se realiza implícitamente siempre que evaluamos los pros y los contras de una

decisión. No obstante, el grupo elaborador debe explicitar estos valores en la medida de lo posible. Cuando este balance presenta una diferencia importante entre estos dos tipos de desenlaces, es más probable realizar una recomendación fuerte. En el caso de que el balance esté más equilibrado, lo más apropiado es formular una recomendación débil. Por ejemplo, en el caso de los corticoides inhalados para el tratamiento de mantenimiento del asma persistente los beneficios son mayores que los riesgos e inconvenientes. La guía GEMA realizó en este contexto una recomendación fuerte a favor¹². No obstante, en el caso del asma grave mal controlada con corticoides inhalados y un agonista beta-2 de acción larga, formula una recomendación débil sugiriendo la utilización de corticoides orales debido a un balance más incierto entre beneficios y riesgos¹².

Calidad de la evidencia

Es crucial conocer hasta qué punto podemos confiar en la estimación del efecto para los desenlaces clave. Cuando la calidad es alta es más probable que se formule una recomendación fuerte y, al contrario, si la calidad es baja es más probable que se formule una débil. No obstante, existen situaciones que justifican una recomendación fuerte disponiendo únicamente de evidencia de calidad baja o muy baja. Por ejemplo, en mujeres embarazadas con asma, la guía GEMA 2009 formula una recomendación fuerte de no reti-

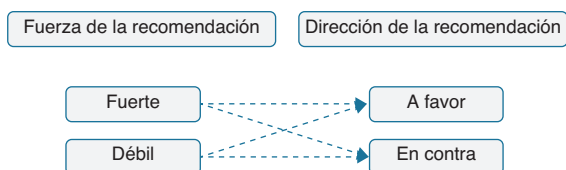


Figura 2. Fuerza y dirección de las recomendaciones.

rar el tratamiento de mantenimiento con corticoides más agonistas adrenérgicos beta-2 de acción prolongada por el riesgo bien conocido de exacerbación tras su supresión, a pesar de disponer de calidad de la evidencia baja sobre la toxicidad de esta combinación sobre el feto¹².

Valores y preferencias

GRADE incluye los valores y preferencias como otro de los factores a valorar a la hora de graduar la fuerza de las recomendaciones. A menudo los pacientes tienen diferentes puntos de vista acerca de lo que implica un desenlace (y por tanto un tratamiento) y, por otro lado, el punto de vista de los profesionales sanitarios a menudo difiere del de los pacientes³⁴. Por todo ello, se han de tener en cuenta los valores y preferencias de los pacientes a la hora de graduar la fuerza de una recomendación. Si la confianza sobre estos valores y preferencias disponibles es alta y la variabilidad es escasa es más probable formular una recomendación fuerte (y viceversa). Así mismo, los grupos elaboradores deberían explicitar de qué valores se trata y la fuente que han utilizado al formular sus recomendaciones (p. ej., los disponibles de la literatura o los estimados a través de su interacción con los pacientes en la toma de decisiones). En el caso de la guía mencionada anteriormente, para la pregunta sobre si tratar con antihistamínicos orales-H₁ en niños en edad preescolar, con otras afecciones alérgicas para prevenir sibilancias o asma, la guía explicitó que su recomendación asigna una mayor importancia a evitar los efectos secundarios de estos fármacos que en una muy incierta reducción del riesgo de desarrollar sibilancias o asma¹⁶.

Costes y uso de recursos

Los costes que se derivan de una decisión clínica son complejos de cuantificar porque a menudo se dispone de información desactualizada en el tiempo o de otros contextos sanitarios. El análisis económico debe realizarse tras la evaluación del balance riesgo-beneficio, y es importante explicitar la perspectiva del mismo (p. ej., la toma de decisiones del paciente o la perspectiva del sistema sanitario), ya que puede considerar los costes directos, indirectos o ambos o el uso de recursos a corto plazo o a más largo plazo. Un coste elevado disminuye la probabilidad de formular una recomendación fuerte a favor de una intervención; por el contrario, un coste bajo aumenta las probabilidades de formular una recomendación fuerte.

Integración de los factores

En el proceso de formulación de las recomendaciones es necesario integrar todos los factores que se han comentado para determinar la fuerza de las mismas. Este proceso requiere un balance ponderado y explícito de estos factores y, por tanto, es importante que las GPC reflejen este proceso de manera detallada.

En la **tabla 5** se presenta un ejemplo relacionado con la GPC de rinitis y asma, mencionada anteriormente, y el uso exclusivo de antagonistas de los receptores de los leucotrienos en el tratamiento del asma¹⁶. En relación con el balance beneficio-riesgo, los corticoides inhalados, respecto a los antagonistas de los leucotrienos, mostraron una disminución de las exacerbaciones, una mejora de los síntomas diurnos y nocturnos, así como un incremento en los días sin síntomas y en la calidad de vida. En relación a los efectos adversos, los antagonistas de los leucotrienos presentaron menos efectos que los corticoides inhalados, pero los resultados son imprecisos. Por otro lado, la calidad de la evidencia se evaluó como moderada debido a la imprecisión de este y otros los resultados en los desenlaces de interés evaluados (**tabla 3**). El coste de los dos tratamientos es diferente, ya que los corticoides

Tabla 5

Integración de los factores para graduar la fuerza de una recomendación

<i>Balance entre beneficios y riesgos</i>
Los antagonistas de los receptores de los leucotrienos en monoterapia oral para el tratamiento del asma son menos efectivos que los corticoides inhalados en la mejora de los síntomas de asma y en la disminución de las exacerbaciones que requieren uso de esteroides sistémicos (30 más por 1.000 pacientes tratados). Estos presentan un número de efectos adversos menor que los corticoides inhalados (4 menos por 1.000 pacientes tratados)
<i>Calidad de la evidencia</i>
La calidad de la evidencia es moderada debido a la imprecisión de los resultados en algunos estudios para los desenlaces clave considerados
<i>Valores y preferencias de los pacientes</i>
Los valores y las preferencias de los pacientes no son probablemente distintos para los desenlaces clave considerados. Es muy probable que la inmensa mayoría de los pacientes se muestren a favor de tomar corticoides
<i>Costes y uso de recursos</i>
Los corticoides tienen un menor coste que los leucotrienos
<i>Recomendación</i>
En pacientes con rinitis alérgica y asma se recomienda utilizar los corticoides inhalados en lugar de los antagonistas de los receptores de los leucotrienos como monoterapia para el tratamiento del asma (<i>recomendación fuerte, calidad de la evidencia moderada</i>)

inhalados son más baratos que los antagonistas de los receptores de los leucotrienos. La integración de estos factores condujo a la formulación de una recomendación fuerte a favor del uso de los corticoides inhalados frente a los antagonistas de los receptores de los leucotrienos orales como monoterapia para el control del asma.

Finalmente, otro aspecto de interés es la terminología utilizada en la redacción de las recomendaciones. El uso de términos específicos (palabras, números, letras, símbolos, etc.) debería describir de manera óptima la fuerza de las recomendaciones dadas. El uso de expresiones como «se recomienda/no se recomienda» cuando se habla de recomendaciones fuertes, o «se sugiere/no se sugiere» para recomendaciones débiles, son algunos ejemplos de redacción. No obstante, la información disponible actualmente sobre este tema es muy limitada³⁵. Estudios futuros, algunos de ellos impulsados por el propio grupo GRADE, abordarán este y otros temas sobre la óptima presentación y diseminación de las recomendaciones en salud³⁶.

Conclusión e implicaciones futuras

La formulación de recomendaciones es un proceso complejo que implica múltiples juicios y una inversión importante de recursos. GRADE ha puesto de manifiesto la complejidad inherente del proceso, pero a la vez proporciona una herramienta sistemática y estructurada que permite realizarlo de una manera explícita. Diferentes grupos pueden llegar a conclusiones diferentes con GRADE; no obstante, si se adhieren al proceso y lo hacen público, los usuarios podrán conocer si están de acuerdo con los juicios que han determinado las recomendaciones finales. GRADE cuenta con una aceptación muy importante en la comunidad internacional y está siendo adoptado por las principales instituciones elaboradoras de guías, internacionalmente y en nuestro entorno. En el caso de las guías en el campo de la neumología, guías de nuestro entorno como GEMA o GesEPOC ya lo han implantado. En el ámbito internacional, organizaciones como la *American Thoracic Society* o la *Global Initiative for Asthma* (GINA) ya lo utilizan o están comenzando a utilizarlo. Por todo ello, GRADE se está consolidando como la metodología que debería mejorar la calidad de las guías y, en definitiva, el cuidado de los pacientes.

Conflicto de intereses

PAC, DR, AJS y LM son miembros del grupo GRADE.

Bibliografía

- Grupo de trabajo sobre GPC. Elaboración de Guías de Práctica Clínica en el Sistema Nacional de Salud. Manual Metodológico. Madrid: Plan Nacional para el SNS del MSC. Instituto Aragonés de Ciencias de la Salud-I+CS; 2007. Guías de Práctica Clínica en el SNS: I+CS. N.º 2006/01.
- Qaseem A, Forland F, Macbeth F, Ollenschlager G, Phillips S, van derWees P. Guidelines International Network: toward international standards for clinical practice guidelines. *Ann Intern Med.* 2012;156:525–31.
- Laine C, Taichman DB, Mulrow C. Trustworthy clinical guidelines. *Ann Intern Med.* 2011;154:774–5.
- Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. *Can Med Assoc J.* 2010;182:E839–42.
- Guyatt G, Akl EA, Hirsh J, Kearon C, Crowther M, Gutterman D, et al. The vexing problem of guidelines and conflict of interest: a potential solution. *Ann Intern Med.* 2010;152:738–41.
- Alonso-Coello P, Irfan A, Sola I, Gich I, Delgado-Noguera M, Rigau D, et al. The quality of clinical practice guidelines over the last two decades: a systematic review of guideline appraisal studies. *Qual Saf Health Care.* 2010;19:e58.
- West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF, et al. Systems to Rate the Strength of Scientific Evidence: Summary en: Rockville MD. AHRQ Evidence Report Summaries. Agency for Healthcare Research and Quality (US). 2002.
- The GRADE Working Group. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches. *BMC Health Serv Res.* 2004;4:38.
- Plaza V, Bellido-Casado J, Alonso-Coello P, Rodrigo G. Asthma clinical practice guidelines: advantages and pitfalls. *Arch Bronconeumol.* 2009;45 Suppl 1:25–9.
- Guyatt GH, Oxman AD, Vist G, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al., for the GRADE Working Group. Rating quality of evidence and strength of recommendations GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ.* 2008;336:924–6.
- Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: A new series of articles in the Journal of Clinical Epidemiology. *J Clin Epidemiol.* 2010;64:380–2.
- GEMA 2009. Guía española para el manejo del asma. *Arch Bronconeumol.* 2009;45 Supl 7:1–35.
- Grupo de Trabajo de GesEPOC. Guía de Práctica Clínica para el diagnóstico y tratamiento de pacientes con enfermedad pulmonar obstructiva crónica (EPOC) – Guía Española de la EPOC (GesEPOC). *Arch Bronconeumol.* 2012;48 Supl 1: 2–58.
- Grupo de trabajo de la guía de práctica clínica sobre Atención Integral al paciente con Enfermedad Pulmonar Obstructiva Crónica (EPOC). Desde la Atención Primaria a la Especializada. Sociedad Española de Medicina de Familia (semFYC) y Sociedad Española de Neumología y Cirugía Torácica (SEPAR). 2010.
- Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol.* 2011;64:395–400.
- Brozek JL, Bousquet J, Baena-Cagnani CE, Bonini S, Canonica GW, Casale TB, et al. Allergic Rhinitis and its Impact on Asthma (ARIA) guidelines: 2010 revision. *J Allergy Clin Immunol.* 2010;124:66–76.
- Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines 3: rating the quality of evidence. *J Clin Epidemiol.* 2011;64: 401–6.
- Guyatt G, Oxman AD, Sultan S, Brozek J, Glasziou P, Alonso-Coello P, et al. GRADE guidelines 11-making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *J Clin Epidemiol.* 2013;66:151–7.
- Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines 4: rating the quality of evidence – risk of bias. *J Clin Epidemiol.* 2011;64:407–15.
- Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence – inconsistency. *J Clin Epidemiol.* 2011;64:1294–302.
- Calderon MA, Alves B, Jacobson M, Hurwitz B, Sheikh A, Durham S. Allergen injection immunotherapy for seasonal allergic rhinitis. *Cochrane Database of Systematic Reviews* 2007, Issue 1. Art. No.: CD001936. DOI: 10.1002/14651858.CD001936.pub2.
- Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence – indirectness. *J Clin Epidemiol.* 2011;64:1303–10.
- Rabe KF, Timmer W, Sagkriotis A, Viel K. Comparison of a combination of tiotropium plus formoterol to salmeterol plus fluticasone in moderate COPD. *Chest.* 2008;134:255–62.
- Mills EJ, Druyts E, Ghement I, Puhan MA. Pharmacotherapies for chronic obstructive pulmonary disease: a multiple treatment comparison meta-analysis. *Clin Epidemiol.* 2011;3:107–29.
- Guyatt G, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence – imprecision. *J Clin Epidemiol.* 2011;64:1283–93.
- Sterne JAC, Egger M, Moher D, editors. Chapter 10: Addressing reporting biases. En: Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Intervention.* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. [consultado 12 Jun 2012]. Disponible en: www.cochrane-handbook.org
- Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines 5: rating the quality of evidence – publication bias. *J Clin Epidemiol.* 2011;64:1277–82.
- Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol.* 2011;64:1311–6.
- Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years' observations on male British doctors. *BMJ.* 2004;328:1519.
- Miravittles M, Soriano JB, García-Río F, Muñoz L, Duran-Taulería E, Sánchez G, et al. Prevalence of COPD in Spain: Impact of undiagnosed COPD on quality of life and daily life activities. *Thorax.* 2009;64:863–8.
- Devereaux PJ, Choi PT, Lacchetti C, Weaver B, Schünemann H, Haines T, et al. A systematic review and meta-analysis of studies comparing mortality rates of private for-profit and private not-for-profit hospitals. *CMAJ.* 2002;166:1399–406.
- GRADEpro [programa informático]. Version 3.2 para Windows. Jan Brozek, Andrew Oxman, Holger Schünemann, 2008 [consultado 11 Feb 2013]. Disponible en: <http://ims.cochrane.org/gradepro>
- Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A, et al., GRADE Working Group. Going from evidence to recommendations. *BMJ.* 2008;336:1049–51.
- Devereaux PJ, Anderson DR, Gardner MJ, Putnam W, Flowerdew GJ, Brownell BF, et al. Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study. *BMJ.* 2001;323:1218–22.
- Akl EA, Guyatt GH, Irani J, Feldstein D, Wasi P, Shaw E, et al. «Might» or «suggest»? No wording approach was clearly superior in conveying the strength of recommendation. *J Clin Epidemiol.* 2012;65:268–75.
- Treweek S, Oxman AD, Alderson P, Bossuyt P, Brozek J, Davoli M, et al., the DECIDE Consortium. Developing and Evaluating Communication Strategies to Support Informed Decisions and Practice Based on Evidence (DECIDE): Protocol and preliminary results. *Implementation Sci.* 2013;8:6.