Guía clínica de La evaluación de la calidad de la evidencia y la graduación de la fuerza de las recomendaciones: el sistema GRADE

La necesidad de graduar las recomendaciones en las guías de práctica clínica (GPC)

Las guías de práctica clínica (GPC) han experimentado una enorme transformación durante las últimas dos décadas. Los dos aspectos en los que se han sentido de manera más importante estos cambios han sido en la búsqueda exhaustiva de la literatura, y la formulación de las recomendaciones (Grupo de trabajo sobre GPC, 2007).

En el caso concreto de la formulación de recomendaciones el grupo elaborador de una guía (GEG) se debe plantear hasta qué punto puede afirmarse que el adherirse a una recomendación conllevará más beneficios que daños. Esta decisión no es sencilla y está influida por múltiples factores que convierten a esta etapa en una de las más complejas dentro del proceso de elaboración de una guía. De acuerdo al instrumento AGREE (Appraisal of Guidelines Research and Evaluation), la etapa de formulación de recomendaciones, se incluye dentro del área denominada "Rigor en la elaboración" (Brouwers M, 2010). En esta área, según el instrumento AGREE, deben contemplarse los siguientes aspectos: la metodología explícita, la relación entre beneficios y riesgos y la relación entre las recomendaciones y la evidencia científica.

Preguntas incluidas en el instrumento AGREE en el apartado de “Rigor en la elaboración” relacionados con la formulación de recomendaciones:

- ¿Los métodos para formular las recomendaciones están claramente descritos?
- ¿Al formular las recomendaciones, los beneficios de salud, efectos secundarios y los riesgos, han sido considerados?
- ¿Hay una relación clara entre cada una de las recomendaciones y las evidencias en las que se basan?

Hasta la fecha se han utilizado múltiples sistemas para la evaluación de la calidad de la evidencia y graduación de la fuerza de las recomendaciones. Desde hace tiempo ya los sistemas diferenciaban el nivel o la calidad de la evidencia de la fuerza de la recomendación (Canadian Task Force on Preventive Health Care, 2007). La calidad se equiparaba, por lo general, a la adecuación de los diseños de los estudios para contestar a los diferentes tipos de preguntas y a la ausencia o no de sesgos inherentes al tipo de estudio. Progresivamente las diferentes propuestas han ido evolucionando e incorporando otros aspectos más allá del diseño de los estudios y sus sesgos.

Al ser las GPC una herramienta en un campo en constante evolución es comprensible que existan múltiples sistemas que utilizan diferentes propuestas, a menudo con limitaciones importantes (Agency for Healthcare Research and Quality, 2002). Generalmente las propuestas sitúan a las revisiones sistemáticas de ensayos clínicos aleatorizados (ECA) en el nivel más alto de evidencia, seguidas, en orden descendente, por los ECA individuales, estudios de cohortes, estudios de casos y controles y series de casos. Finalmente, la opinión de los expertos es considerada en algunos sistemas como la categoría más baja de evidencia.

Otras de las limitaciones presentes, a menudo, en los sistemas actuales es la falta de transparencia en el paso de la evidencia a la recomendación y en la valoración del balance entre beneficios y riesgos, así como la indefinición en los criterios explícitos para evaluar la calidad de la evidencia y graduar la fuerza de las recomendaciones. Todo ello puede contribuir al desconcierto de los usuarios finales de las GPC y a generar problemas de comunicación y desconfianza en su aplicación. La evaluación de la preferencias de los usuarios finales o de la adherencia hacia las recomendaciones GPC en función del sistema utilizado es un tema todavía abierto en las agendas de investigación de muchas instituciones.

Un sistema para clasificar la calidad de la evidencia y graduar la fuerza de las recomendaciones debe combinar la sencillez con la explicitud. Recientemente un grupo internacional de elaboradores de GPC, epidemiólogos y clínicos pertenecientes a las principales instituciones que desarrollan GPC, ha elaborado una nueva propuesta que tiene como objetivo consensuar un sistema común que supere las limitaciones detectadas hasta el momento en los sistemas previos. Este grupo de profesionales constituye el grupo de trabajo GRADE (Grading of Recommendations Assessment, Development and Evaluation) (Guyatt G, 2008). Este sistema, denominado GRADE como el propio grupo de trabajo, está comenzando a ser utilizado por diferentes instituciones elaboradoras de guías o que formulan recomendaciones en salud. El grupo dispone de una web propia (http://www.gradeworkinggroup.org/) en la cual se proporciona acceso a materiales y publicaciones, algunas de ellas en castellano.

El sistema GRADE

El sistema GRADE tiene algunos aspectos diferenciales que se pueden resumir en tres. Inicialmente categoriza los desenlaces de interés (ej. muerte) y su importancia relativa. Posteriormente evalúa la calidad de la evidencia en cuatro categorías para cada uno de los desenlaces para después obtener una calidad global de evidencia. Finalmente gradúa la fuerza de las recomendaciones en dos únicas categorías (recomendaciones fuertes o débiles).

Es importante partir de una definición común de los dos conceptos clave en la formulación de las recomendaciones: la *calidad de la evidencia* (conocida también como nivel de evidencia) y la *fuerza de la recomendación*. El sistema GRADE propone las siguientes definiciones:

- Calidad de la evidencia: indica hasta qué punto nuestra confianza en la estimación de un efecto es

adecuada para apoyar una recomendación.
- Fuerza de la recomendación: indica hasta qué punto podemos confiar si poner en práctica la recomendación conllevará más beneficios que riesgos.

GRADE al tener dos etapas claramente diferenciadas (Calidad y Fuerza) puede ser utilizado para evaluar únicamente la calidad de la información disponible sobre un tema (p.ej. al evaluar la calidad en una revisión sistemática) o además para formular recomendaciones en el marco de la elaboración de una guía. En el caso concreto de las revisiones sistemáticas el sistema GRADE se ha adoptado por parte de la Colaboración Cochrane para evaluar la calidad de la evidencia de sus revisiones sistemáticas.

Clasificación de la importancia relativa de los desenlaces de interés

En esta etapa inicial es necesario que el grupo elaborador defina los desenlaces de interés para las preguntas que abordará su guía y califique la importancia relativa de las mismas. Estos desenlaces pueden tener una importancia muy diferente a la hora de tomar decisiones y, por otra parte, a menudo los pacientes y cuidadores pueden valorar su importancia de manera diferente a los clínicos.

Se sugiere clasificar su importancia mediante la siguiente escala de nueve puntos:

- 1-3: desenlace no importante. Estos desenlaces no juegan un papel importante en la formulación de las recomendaciones y no se incluyen en la tabla de evaluación de la calidad o de los resultados.
- 4-6: desenlace importante pero no clave para la toma de decisiones.
- 7-9: desenlace clave para la toma de decisiones.

Por ejemplo, en una GPC sobre la prevención de la osteoporosis, un grupo elaborador puede seleccionar los siguientes desenlaces para evaluar el efecto del tratamiento con bifosfonatos en mujeres osteoporóticas (Figura 1). Probablemente el grupo valorará los desenlaces de fractura de cadera, fractura clínica vertebral y la calidad de vida como desenlaces clave para la toma de decisiones (7-9), mientras que considerará la fractura no vertebral y la fractura vertebral radiológica como importantes pero no claves (4-6). Finalmente puede considerar como no importantes la DMO y los marcadores óseos (1-3).

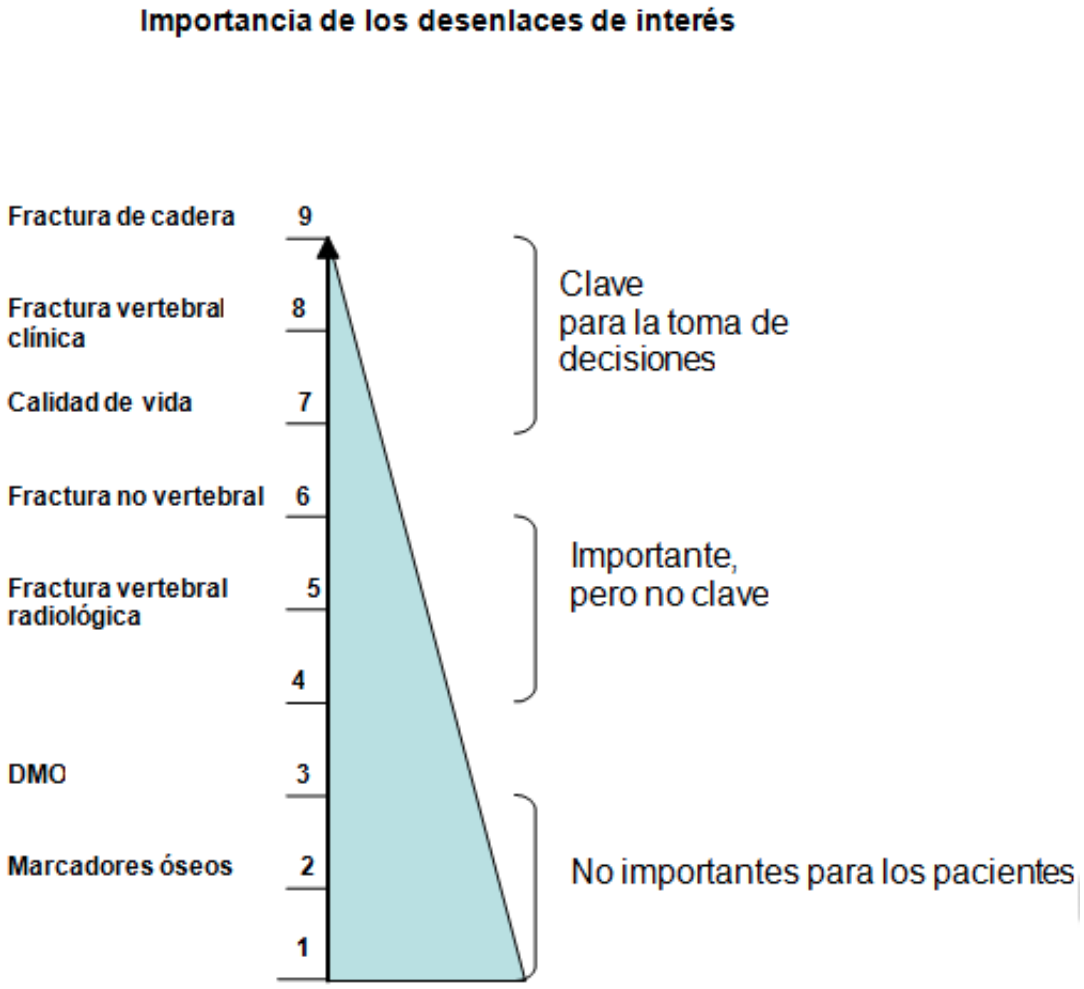

Figura 1.
Jerarquía de los desenlaces de interés de acuerdo a la importancia para los pacientes en la evaluación de los bifosfonatos en mujeres con un riesgo alto de fractura.

Evaluación de la calidad de la evidencia

Según el sistema GRADE la calidad de la evidencia en las GPC refleja el grado de confianza que tenemos en que la estimación de un efecto es adecuada para apoyar una recomendación (Balshem H, 2011). Esta evaluación de la calidad de la evidencia se realiza para cada uno de los desenlaces de interés (p.ej. reagudización en un paciente EPOC) considerados como claves o importantes, de forma que los distintos desenlaces evaluados para una intervención pueden presentar distinta calidad dependiendo del número y características de los estudios que los hayan incluido.

La calidad de la evidencia según GRADE tiene un significado mas amplio que la evaluación del riesgo de sesgo. Es por tanto posible disponer, para un determinado desenlace, de un conjunto de estudios sin limitaciones en el diseño (con un bajo riesgo de sesgo) pero que la confianza en que el estimador del efecto sirva para apoyar una recomendación se encuentre limitada por otros factores (Guyatt G, 2011).

Aunque la calidad de la evidencia es un continuo, se propone la clasificación en cuatro categorías por

razones prácticas. El significado de calidad alta, moderada, baja o muy baja se presenta en la tabla 1. En la propuesta GRADE los estudios se clasifican de forma que los ECAs corresponden de entrada a "calidad alta" y los observacionales a "calidad baja". Existen además otros cinco factores que pueden justificar disminuir la calidad y otros tres que pueden justificar aumentar la calidad de la evidencia (tabla 2).

Tabla 1. Significado de los niveles de calidad de la evidencia

Calidad de la evidencia	Definición
Alta	Hay una confianza alta en que el estimador del efecto se encuentra muy cercano al efecto real.
Moderada	Hay una confianza moderada en el estimador del efecto: es probable que el estimador del efecto se encuentre cercano al efecto real pero hay la posibilidad que existan diferencias substanciales.
Baja	La confianza en el estimador del efecto es baja: el estimador del efecto puede ser substancialmente diferente al efecto real.
Muy baja	Hay una confianza muy baja en el estimador del efecto: es muy probable que el estimador del efecto sea substancialmente diferente al efecto real.

Tabla 2. Evaluación de la calidad de la evidencia según el tipo de diseño de los estudios

Diseño de estudio	Calidad de la evidencia inicial	Disminuir si	Aumentar si	Calidad del conjunto de la evidencia
Ensayos clínicos aleatorizados	Alta	Limitaciones en el diseño o la ejecución Importantes (-1) Muy importantes (-2) Inconsistencia	Fuerza de asociación Fuerte (+1) Muy fuerte (+2)	Alta ⊖⊖⊖⊖

		Inconsistencia Importante (-1) Muy importante (-2)	Gradiente dosis-respuesta Presente (+1)	Moderada ⊕⊕⊕⊖
		Incertidumbre en que la evidencia sea directa Importante (-1) Muy importante (-2)	Consideración de los posibles factores de confusión que habrían reducido el efecto (+1) Sugerirían un efecto espurio si no hay efecto (+1)	Baja ⊕⊕⊖⊖
Estudios observacionales	Baja	Imprecisión Importante (-1) Muy importante (-2) Sesgo de publicación Importante (-1) Muy importante (-2)		Muy baja ⊕⊖⊖⊖

Aspectos que pueden disminuir la calidad

Tanto para los ECA, que de antemano se consideran como calidad alta para evaluar el efecto de diferentes intervenciones, como para los estudios observacionales, que de antemano se consideran como calidad baja, existen factores que pueden disminuir la calidad o más concretamente, la confianza en la estimación del efecto observado para los distintos desenlaces evaluados (Balshem H, 2011).

- Limitaciones en el diseño o la ejecución: en los ECA se consideran; 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, notificación selectiva de las variables de resultado de interés, además de otros factores menos frecuentes como la finalización prematura del 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 (Guyatt GH, 2011).

En los estudios observacionales se consideran; la presencia de unos criterios de selección de la población inapropiados, las mediciones inapropiadas para la exposición o el desenlace de interés, el control inadecuado de los factores de confusión o el seguimiento incompleto (Guyatt GH, 2011).

- Resultados inconsistentes: se debe considerar disminuir la calidad de la evidencia si persisten diferencias importantes en los resultados de los distintos estudios disponibles (heterogeneidad), tras haber explorado las fuentes de esta variabilidad, principalmente las diferencias en la población, la intervención, las variables de resultado o al riesgo de sesgo de los estudios. En estas circunstancias pueden existir diferencias reales entre las estimaciones del efecto y, por este motivo, disminuye nuestra confianza en el resultado global (Guyatt GH, 2011).
- Ausencia de evidencia directa: la confianza en un estimador común se reduce si se detectan existen diferencias importantes entre la población donde se aplicará la GPC y la incluida en los estudios o, por ejemplo, entre la intervención que se pretende evaluar y la que se practicó en los estudios disponibles. También se debe considerar disminuir la calidad de la evidencia ante la ausencia de comparaciones directas entre las intervenciones de interés (Guyatt GH, 2011).
- Imprecisión: para evaluar la imprecisión se debe considerar el intervalo de confianza del estimador del efecto, preferiblemente del efecto absoluto -más que el relativo- de una intervención. Si una decisión clínica o recomendación pudieran ser diferentes dependiendo de si se interpreta un extremo u otro del intervalo de confianza como el efecto real de una intervención, la confianza en el estimador del efecto debería disminuir. Además, incluso en los estimadores del efecto con un intervalo de confianza considerado como preciso, si el número de eventos o el número de sujetos evaluados en los diferentes estudios son escasos, se debe considerar disminuir la calidad (Guyatt GH, 2011).
- Sesgo de publicación: la calidad, y por tanto nuestra confianza, puede disminuir si se tiene la duda razonable que, en el contexto de una revisión sistemática, no se hayan incluido todos los estudios realizados, situación probable con los estudios que han mostrado resultados “negativos”. Existen pruebas estadísticas que pueden orientar a la existencia de un sesgo de publicación como la revisión de los gráficos de embudo (*funnel plots*). Esta situación puede llegar a ser más frecuente y difícil de detectar en los estudios observacionales (Guyatt GH, 2011).

Aspectos que pueden aumentar la calidad

Las situaciones que pueden llevar a incrementar la confianza con los resultados de los estudios son poco comunes e implican sobre todo a estudios observacionales. Estos aspectos se deberían evaluar después de constatar que no se encuentra presente ninguna condición para disminuir la calidad de la evidencia

debido a un potencial riesgo de sesgo (Guyatt GH, 2011).

- Fuerza de la asociación: cuando el efecto observado muestra una asociación fuerte (riesgo relativo u *odds ratio* >2 ó <0,5) o muy fuerte (riesgo relativo u *odds ratio* >5 ó <0,2), es improbable que este efecto sea debido a una sobreestimación y que el efecto observado se encuentre muy cercano al efecto real (Guyatt GH, 2011).
- Presencia de un gradiente dosis respuesta: la existencia de un gradiente dosis-respuesta es un factor reconocido para establecer con mayor certidumbre una relación causa-efecto. En estas circunstancias este gradiente puede aumentar nuestra confianza en los resultados (Guyatt GH, 2011).
- Consideración de los potenciales factores de confusión: pueden existir situaciones en las que se observa un efecto asociado a una intervención y que, después de considerar los potenciales factores de confusión (que podrían explicar el efecto detectado), se llegue a la conclusión de que éstos no alteran el resultado sino que incluso refuercen la asociación (Guyatt GH, 2011).

Una vez evaluada la calidad de la evidencia científica para cada variable de resultado, el sistema GRADE propone que las variables de resultado clave son las únicas que se tienen que tomar en cuenta para evaluar la calidad de forma global. Por ejemplo, en el caso del tratamiento de las mujeres con osteoporosis y fractura vertebral previa, el GEG puede considerar como variables de resultado clave la fractura de cadera y la fractura vertebral pero no la densidad mineral ósea o los marcadores óseos en orina. Por tanto a la hora de evaluar la calidad global tendrán en cuenta la calidad de las variables de resultado clave y en el caso de que la calidad sea diferente, se tendrá en cuenta la calidad más baja de todas las variables clave.

Graduación de la fuerza de las recomendaciones

La fuerza de una recomendación refleja el grado hasta el que podemos confiar en que los efectos deseados de una intervención sean superiores a los adversos (Guyatt GH, 2008). En relación a dos estrategias terapéuticas refleja hasta qué punto podemos confiar en que existe un beneficio neto a favor de una de ellas.

El sistema GRADE considera dos categorías en relación a la fuerza de las recomendaciones:

- Recomendación fuerte: se formula una recomendación fuerte a favor de una intervención cuando se confía en que los efectos deseables de la intervención superan a los indesables; o fuerte en contra en la situación inversa, en la que los efectos indeseables de la intervención superan los deseables.
- Recomendación débil: las recomendaciones débiles, tanto a favor como en contra de una intervención, se formulan cuando no se disponen de pruebas concluyentes sobre los efectos de la intervención. En la Tabla 3 se resumen las implicaciones de la fuerza de las recomendaciones desde diferentes perspectivas (pacientes, clínicos y gestores).

Tabla 3. Implicaciones de las recomendaciones

Implicaciones de una recomendación fuerte	
Para los pacientes	La mayoría de las personas estarían de acuerdo con la intervención recomendada y únicamente una pequeña parte no lo estarían.
Para los clínicos	La mayoría de los pacientes deberían recibir la intervención recomendada.
Para los gestores	La recomendación puede ser adoptada como política sanitaria en la mayoría de las situaciones.
Implicaciones de una recomendación débil	
Para los pacientes	La mayoría de las personas estarían de acuerdo con la acción recomendada pero un número importante de ellos no.
Para los clínicos	Se reconoce que diferentes opciones serán 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	Existe necesidad de un debate importante con la participación de los grupos de interés.

Factores a tener en cuenta en la graduación de las recomendaciones

- Balance entre beneficios y riesgos: al elaborar una recomendación se debe considerar el balance entre beneficios y riesgos de la intervención. Por ejemplo, en el asma moderado-grave los corticoides inhalados presentan importantes beneficios en términos de una reducción del número y gravedad de las exacerbaciones junto con bajo riesgo de eventos adversos sistémicos. Sin embargo, en niños pequeños con sibilancias recurrentes el balance es más incierto y su uso debería sopesarse individualmente. Por tanto en el primer caso sería más adecuada una recomendación fuerte a favor y en el segundo una débil a favor (en función de una valoración individualizada del balance riesgo beneficio).
- Calidad de la evidencia: antes de formular una recomendación es necesario conocer la certidumbre sobre la estimación del efecto observado. Si la calidad de la evidencia no es alta, a pesar de que la magnitud del efecto sea importante, debería disminuir la confianza, y por tanto la fuerza con la que se lleve a cabo una recomendación. Por ejemplo, las medias de compresión graduada producen un amplio efecto aparente en la reducción de la trombosis venosa profunda en los individuos que hacen viajes largos en avión. Sin embargo, los ensayos aleatorizados en los que se basa el efecto presentan problemas metodológicos (Geerts W, 2005). Así que, a pesar del amplio beneficio aparente, el uso de

medias de compresión sólo merece una recomendación débil a favor.

- Valores y preferencias: los valores y preferencias de la población diana a la cual va dirigida la guía es otro de los factores a tener en cuenta. Los valores y preferencias del personal sanitario, el colectivo de pacientes o la sociedad en general deberían influir en el proceso de formulación y graduación de las recomendaciones. Por ejemplo, la warfarina reduce el riesgo de ictus en aproximadamente el 65% pero aumenta el de hemorragia gastrointestinal grave en los pacientes con fibrilación auricular. La mayoría de pacientes otorgan una elevada importancia a evitar un ictus y estan dispuestos a aceptar un riesgo de hemorragia del 22% para reducir su probabilidad de experimentar un ictus en un 8%. Sin embargo, existe una gran diversidad de valores y preferencias entre los pacientes (Devereaux PJ, 2001). Estos datos indican que una recomendación fuerte de la administración de warfarina sólo estaría justificada en los pacientes con riesgo elevado de ictus.
- Costes: los costes, a diferencia de otras variables de resultado, son mucho más variables en el tiempo, en distintas áreas geográficas y en función de las implicaciones de los recursos. Así, aunque un coste elevado disminuye la probabilidad de graduar una recomendación como fuerte el contexto será crítico en la valoración final.

Así, la fuerza de las recomendaciones está determinada por el equilibrio entre las consecuencias deseables e indeseables de las estrategias de tratamiento, la calidad de la evidencia, la variabilidad en los valores y preferencias y el uso de recursos. A continuación dos ejemplos de dos recomendaciones, una fuerte y una débil.

Por ejemplo, el tratamiento con corticoides para el manejo de las exacerbaciones en la enfermedad pulmonar obstructiva crónica (EPOC) presenta un balance beneficio-riesgo favorable (reduce el riesgo de fracaso terapéutico, la necesidad de tratamiento adicional, la estancia hospitalaria y aumenta la capacidad funcional, aunque presenta eventos adversos leves). La calidad de la evidencia es moderada. No hay razones para pensar que la mayoría de los pacientes, una vez informados, no aceptarían el tratamiento. Y finalmente, el coste de los corticoides es muy bajo. En conjunto, por tanto, tenemos evidencia de calidad moderada (disminuye algo nuestra confianza en la estimación del efecto) y no hay otros factores adicionales que nos hagan disminuir nuestra confianza en que el llevar a cabo esta recomendación conllevaría más beneficios que daños. Por tanto, el grupo elaborador de una guía podría probablemente llevar a cabo una recomendación fuerte a favor de la intervención con corticoides en las exacerbaciones de la EPOC (Management of exacerbations of COPD Thorax, 2004).

Otro ejemplo sería la prevención del ictus en pacientes con fibrilación auricular (Riesgo intermedio; $CHADS_2$=1) el tratamiento con anticoagulantes orales (dicumarínicos), comparado con aspirina. En este caso los dicumarínicos (warfarina), muestran un balance beneficio riesgo favorable (9 ictus menos y 3 sangrados extracraneales no fatales más cada 1000 pacientes tratados), siendo la calidad de la evidencia moderada (You JJ, 2012). En cuanto a los valores y preferencias los pacientes dan, en términos generales, tres veces más importancia a evitar un ictus que a sufrir un sangrado extracraneal grave. No obstante estas estimaciones de los valores y preferencias son muy variables. En cuanto al coste ambas

medicaciones son poco costosas. En conjunto, por tanto, tenemos evidencia de calidad moderada (disminuye algo nuestra confianza en la estimación del efecto) y el coste es bajo. No obstante el balance riesgo beneficio es poco marcado y los valores y preferencias de los pacientes son muy variables. Por tanto hay algunos factores que disminuyen nuestra confianza en que el llevar a cabo esta recomendación conllevaría más beneficios que daños (hay un porcentaje razonable de pacientes que podrían preferir tomar aspirina en lugar de anticoagulantes orales). Por tanto, el grupo elaborador de una guía podría probablemente llevar a cabo una recomendación débil a favor de los anticoagulantes, comparados con la aspirina, en esta población (You JJ, 2012).

Tabla 4. Dicumarínicos (warfarina) comparados con aspirina en pacientes con fibrilació auricular (Riesgo intermedio de ictus; $CHADS_2$= 1)

Variables de resultado	Participantes (estudios) Seguimiento	Calidad de la evidencia (GRADE)	Efecto Relativo (IC95%)	Efectos absolutos esperados a un año	
				Riesgo con aspirina	Diferencia de riesgo con dicumarínico (warfarina) (IC95%)
				Riesgo intermedio de ictus ($CHADS_2$=1)	
Muerte	5.855 (10) Media 1.8 años	Moderada ⊕⊕⊕○ debido a imprecisión	RR 0.97 (0.85 - 1.12)	47 muertes por 1.000 pacientes	1 muerte menos cada 1.000 (de 7 menos a 6 más)
				Riesgo intermedio de ictus ($CHADS_2$=1)	
Ictus Incluye los ictus isquémicos no fatales isquémicos y hemorrágicos	6.526 (11) Media 1.8 años	Alta ⊕⊕⊕⊕	RR 0.48 (0.33 - 0.70)	17 ictus cada 1.000 pacientes	9 ictus menos por 1.000 (desde 5 menos hasta 11 menos)
				Riesgo intermedio de ictus ($CHADS_2$=1)	
Sangrado grave		Moderada			3 sangrados

Incluye los sangrados mayores no fatales extracraneales	6.526 (11) Media 1.8 años	⊕⊕⊖⊖ debido a imprecisión	RR 1.42 (0.89 - 2.29)	8 sangrados graves cada 1.000 pacientes	más cada 1.000 pacientes (desde 1 menos hasta 10 más)
				Riesgo intermedio de ictus ($CHADS_2$=1)	
Embolismo sistémico	6.526 (11) Media 1.8 años	Moderada ⊕⊕⊕⊖ debido a imprecisión	RR 0.81 (0.40 - 1.64)	3 embolismos cada 1.000 pacientes	1 embolismo menos cada 1.000 pacientes (desde 2 menos hasta 2 más)
Inconvenientes del tratamiento	N/A	Alta ⊕⊕⊕⊕	Dicumarínicos >aspirina	Dicumarínicos: medicación diaria, modificaciones de los estilos de vida, controles analíticos frecuentes y visita médicas. Aspirina: medicación diaria.	

Bibliografía:
You JJ, Singer DE, Howard PA, et al. Antithrombotic therapy for atrial fibrillation: antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest 2012; 141(suppl 2):e531S–e575S17.

Implicaciones para el futuro

El sistema GRADE es un sistema exhaustivo, transparente y explícito para la evaluación de la calidad y la formulación de recomendaciones. Estas características conllevan que, a pesar de su atractivo enfoque, sus usuarios tengan que realizar juicios a la hora de decidir el grado de confianza que tienen en los resultados de la investigación o en que adherirse a una recomendación conlleve más beneficios que riesgos. GRADE pone de manifiesto la importante incertidumbre alrededor de la mayoría de las decisiones clínicas de cada día. El sistema ya ha sido adoptado por numerosas instituciones tan importantes como la OMS, la Colaboración Cochrane, el Instituto NICE (National Institute of Clinical Excellence), SIGN

(Scottish Intercollegiate Guidelines Network) o publicaciones como Clinical Evidence o Uptodate, entre otras.

Es importante señalar que el uso incompleto o de forma parcial del sistema GRADE por algunas iniciativas puede llevar a confusión por parte de los usuarios de guías o incluso de los elaboradores. Por tanto, es importante conocer el sistema con cierta profundidad tanto para aplicarlo como para reconocer su correcta aplicación. Actualmente durante el año 2012-2013 terminará de publicarse la serie de artículos en la revista Journal of Clinical Epidemilogy, dirigida fundamentalmente a elaboradores de GPC pero que, facilitará su utilización a todos los usuarios que quieran profundizar en su aplicación. Existe una serie dirigida más a los profesionales de la salud publicada en el British Medical Journal.

El Programa Nacional de Elaboración de Guías de Práctica Clínica ha incluido al sistema GRADE dentro de su manual metodológico y está prevista próximamente su actualización (Grupo de trabajo sobre GPC, 2007). Durante algún tiempo en el Programa cohabitarán dos sistemas, el sistema SIGN y el sistema GRADE, para en un futuro realizar las guías exclusivamente con GRADE.

Bibliografía

- Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, Vist GE, Falck-Ytter Y, Meerpohl J, Norris S, Guyatt GH. GRADE guidelines 3: rating the quality of evidence-introduction. J Clin Epidemiol. 2011 Jan 6.
- Brouwers M, Kho ME, Browman GP, Cluzeau F, Feder G, Fervers B, Hanna S, Makarski J on behalf of the AGREE Next Steps Consortium. AGREE II: Advancing guideline development, reporting and evaluation in healthcare. Can Med Assoc J. Dec 2010, 182: E839-842.
- Canadian Task Force on Preventive Health Care. History and methods. Disponible en: http://www.canadiantaskforce.ca/index.html Acesso: 5 de septiembre de 2012.
- Chronic Obstructive Pulmonary Disease. National clinical guideline on management of chronic obstructive pulmonary disease in adults in primary and secondary care. Management of exacerbations of COPD Thorax. 2004; 59 (Suppl 1): i131-i156. Disponible en:http://thorax.bmj.com/cgi/reprint/59/suppl_1/i131 Acesso: 22 de diciembre de 2007.
- 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.
- Geerts W, Ray JG, Colwell CW, Bergqvist D, Pineo GF, Lassen MR, et al. Prevention of venous thromboembolism. Chest. 2005;128(5):3775-6.
- 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. Nº 2006/0I.
- 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 Aug 12; 64(12):1283-93.

- Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A, Schünemann HJ; GRADE Working Group. Rating quality of evidence and strength of recommendations: Going from evidence to recommendations. BMJ. 2008 May 10;336(7652):1049-51.
- 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 Aug 2; 64(12):1294-302.
- 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 Aug 1; 64(12):1303-10.
- 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 Aug 1; 64(12):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 Aug 1; 64(12):1311-6.
- 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 Jan 20; 64(4):407-15.
- Guyatt G, Oxman AD, Vist G, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ. GRADE: An Emerging Consensus on Rating Quality of Evidence and Strength of Recommendations. BMJ (Serie de 5 artículos. Fecha prevista de publicación en 2008).
- Systems to Rate the Strength of Scientific Evidence. Summary, Evidence Report/Technology Assessment: Number 47. AHRQ Publication No. 02-E015, March 2002. Agency for Healthcare Research and Quality, Rockville, MD. Disponible en:http://www.ahrq.gov/clinic/epcsums/strengthsum.htm. Acceso: 15 de mayo de 2007.
- You JJ, Singer DE, Howard PA, et al. Antithrombotic therapy for atrial fibrillation: antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest 2012; 141(suppl 2):e531S–e575S.