

Jerarquización del conocimiento científico. El sistema GRADE

Carlos Quereda Rodríguez-Navarro¹, Vicente Barrio Lucia², Fernando García López³

¹ Servicio de Nefrología. Hospital Ramón y Cajal. Madrid. ² Servicio de Nefrología. Hospital Infanta Sofía. San Sebastián de los Reyes. ³ Unidad de Epidemiología. Hospital Puerta de Hierro. Majadahonda (Madrid). Grupo de Nefrología Basada en la Evidencia de la Sociedad Española de Nefrología

Nefrología 2009;29(Sup. Ext. 6):7-14.

Constituye una aspiración irrenunciable de la medicina actual reclamada por todos sus protagonistas (enfermos, profesionales y administraciones), el que los actos médicos se sustenten en conocimientos científicos obtenidos de procesos de investigación clínica rigurosa.

Por otra parte, sobre todo desde el surgimiento de internet y su imparable desarrollo, las posibilidades de difusión del conocimiento clínico se han multiplicado de forma exponencial (revistas, sitios *web*, cursos *on line*, etc.). Paradójicamente, esta inflación de medios, y otros fenómenos (heterogeneidad en calidad y formato, complejidad metodológica, sesgos, *marketing*, etc.) que hemos analizado en un trabajo previo^{1,2}, explican la dificultad que tiene el médico para identificar los avances clave de investigación clínica que le permitan seleccionar con rigor los mejores instrumentos disponibles para el cuidado de sus enfermos.

Éste es el sentido de la metodología que ha ido desarrollándose en las últimas dos décadas (y que se ha dado en llamar Medicina Basada en la Evidencia) para proporcionar al profesional sanitario técnicas de evaluación de la calidad e impacto clínico del conocimiento contenido en la literatura médica, así como de documentos de síntesis que le permitan un acceso fácil y riguroso a éste. Entre éstos, tienen especial relevancia las revisiones sistemáticas (RS) y las guías de práctica clínica (GPC).

Las RS se realizan siempre como respuesta a problemas clínicos concretos, utilizando una metodología estructurada y sistemática para el análisis de la información, valoración crítica de los estudios y síntesis de éstos. Este rigor metodológico las diferencia positivamente de las revisiones clásicas de la literatura científica³⁻⁶.

CALIDAD DE LA EVIDENCIA Y GRADOS DE RECOMENDACIÓN

El estudio de una pregunta clínica mediante una RS puede concluir que no existen trabajos de investigación que avalen la res-

puesta o que éstos son escasos y de baja calidad o bien que diferentes estudios presentan resultados inconsistentes entre sí (sin causa justificada), ofreciendo escasa confianza en la validez de los mismos. Por el contrario, la intervención analizada puede estar refrendada por un metanálisis de alta calidad, que incluye el estudio de varios ensayos clínicos aleatorizados, multicéntricos, rigurosamente realizados, proporcionando su agregación un resultado muy preciso y, por tanto, de gran valor probatorio de su utilidad clínica.

De esta forma, las RS proporcionan resultados muy objetivos y reproducibles, que pueden ordenarse en niveles de calidad de la evidencia, que indican la confianza en que los valores inducidos por la intervención sobre las variables de resultado se aproximen al real en cada caso, y cuál será el efecto promedio esperable. Esta clasificación es enormemente útil para calificar la validez de cualquier intervención y para objetivar las «zonas grises» en las que el conocimiento clínico es muy escaso y que necesitan más investigación⁷⁻⁹.

La jerarquización de la evidencia, además de ofrecer una síntesis rápida y objetiva de la calidad metodológica del conjunto de trabajos de la literatura científica sobre una determinada intervención, constituye una de las bases para la definición de las recomendaciones utilizadas en las GPC¹⁰⁻¹³.

La mayoría de las sociedades científicas establecen, como uno de sus más importantes objetivos, la realización de GPC para reducir la variabilidad de la práctica clínica, fenómeno muy relevante en nuestra especialidad¹⁴⁻¹⁷. Con ello se pretende establecer patrones de actuación que mejoren los niveles asistenciales, aunque el tema no está exento de polémica^{8-10,17-20}. En nefrología, el trabajo DOPPS (*Dialysis Outcomes and Practice Patterns Study*) muestra cómo pueden mejorar las variables de resultado en pacientes en los que se sigue una serie de indicadores marcados por las GPC²¹. La Sociedad Española de Nefrología (S.E.N.) ha potenciado su propia línea de guías clínicas y recomendaciones²².

En las GPC, además del concepto de calidad de la evidencia, se introduce el de grado o fuerza de la recomendación que sustenta una propuesta de intervención¹⁰⁻¹³. Ambos conceptos, aunque relacionados y complementarios, se ocupan de aspectos distintos. La fuerza de una recomendación se apoya, deci-

Correspondencia: Carlos Quereda Rodríguez-Navarro
Servicio de Nefrología.
Hospital Ramón y Cajal. Madrid.
cquereda.hrc@salud.madrid.org

sivamente, en la calidad de la evidencia que la sustenta, ya que un alto nivel de recomendación debe basarse en la buena calidad metodológica de los trabajos analizados. Pero esto no es suficiente porque puede existir un alto nivel de evidencia del resultado beneficioso de una determinada intervención, pero la magnitud del efecto puede ser irrelevante clínicamente y no recomendable. Por tanto, es importante distinguir la significación estadística de la importancia clínica y mostrar objetivamente la magnitud del efecto sobre las variables primarias y su precisión (que se relaciona estrechamente con el tamaño muestral). Y, por último, el elemento más importante para decidir el grado de recomendación depende de la solución del binomio beneficio/perjuicio neto para la salud resultante de un análisis de varios factores (magnitud del efecto y daño), pero también de otros, como disponibilidad social y coste^{23,24}.

El objetivo fundamental de este trabajo consiste en analizar los principales sistemas de jerarquización de la evidencia y sus problemas, deteniéndonos fundamentalmente en el recientemente propuesto sistema GRADE²³⁻²⁸.

SISTEMAS DE JERARQUIZACIÓN DE LA EVIDENCIA Y GRADOS DE RECOMENDACIÓN

Desde el lanzamiento de las primeras GPC hace más de tres décadas^{19,29} se han utilizado numerosas escalas de gradación de la evidencia, que se caracterizan por su heterogeneidad y la presencia, en muchas de ellas, de serias deficiencias que impiden reflejar, con rigor, la calidad de los trabajos en los que se fundamentan las recomendaciones de práctica clínica. La mayoría adoptó un enfoque muy simple, fácil de entender, de aplicar y de presentar en los escritos de recomendaciones, lo que explica su éxito y difusión²⁹⁻³². Pero, en la mayor parte de los casos, el grado de evidencia se fija, casi exclusivamente, por el tipo de diseño de los estudios, considerando muchas de ellas que el ensayo clínico es la única garantía de acreditación de la intervención, sin tener en cuenta el análisis de calidad de los estudios o su consistencia y marginando completamente a los estudios observacionales²⁹⁻³⁴. Por otra parte, la mayoría adolece de un déficit de reproducibilidad: en el trabajo de Ferreira et al.³⁵ se muestra que la utilización de diferentes sistemas de jerarquización de la evidencia puede conducir a la formulación de distintas propuestas de actuación clínica en un mismo asunto.

En las primeras descripciones del grupo de medicina basada en la evidencia, publicadas en la revista JAMA^{36,37}, se establecen escalas de gradación de la evidencia que atienden, no sólo al tipo de diseño, sino también a cuestiones que se derivan de la calidad de éste y se acompañan de protocolos para graduar la fuerza de las recomendaciones. Una de las que ha adquirido mayor nivel de consenso y difusión ha sido la elaborada por el Centro de Medicina Basada en la Evidencia de la Universidad de Óxford (OCEBM)³⁸. En ésta se distinguen cinco niveles de evidencia y cuatro grados de recomendación. El pilar básico de la clasificación continúan siendo los tipos de diseño. Para los artículos que estudian intervenciones sobre tratamiento, el ni-

vel más alto de la escala lo ocupan las revisiones sistemáticas de ensayos clínicos aleatorizados (ECA), especialmente aquellos de alta calidad metodológica. Pero éstos pueden situarse en niveles de evidencia I o II, según la calidad de los trabajos. Los estudios observacionales se sitúan en un nivel inferior de evidencia (grados II a IV) sólo por encima de las opiniones de expertos o documentos de consenso (grado V), aunque se incluye el fenómeno «todo o nada» (*all & none*) en un nivel alto de evidencia (I c), para incluir a los estudios observacionales que muestran una importante magnitud del efecto que deja sin viabilidad, por razones éticas, la realización de ECA (p. ej., el tratamiento sustitutivo en la enfermedad renal crónica terminal).

El sistema contempla, además, cuatro grados de recomendación (de A a D), relacionados, fundamentalmente, con el nivel de calidad de la evidencia pero con la posibilidad de extrapolar a escalones más bajos, por razones clínicas. Por otra parte, la escala del OCEBM comprende, no sólo artículos de tratamiento, sino también de diagnóstico, pronóstico, estudios de prevalencia y análisis económico.

Esta escala ha sido ampliamente utilizada por instituciones de prestigio y, hasta ahora, por el Grupo de Nefrología Basada en la Evidencia de la S.E.N. Nosotros creemos que es una herramienta útil, que presenta como principal inconveniente la dificultad de jerarquizar cada uno de los diseños por la calidad del trabajo y la difusa consideración del contexto clínico.

Sin embargo, la situación dista de ser satisfactoria: la proliferación de escalas genera confusión y dudas en los usuarios, que se cuestionan la sensibilidad y reproducibilidad de los distintos modelos. Atkins et al.^{24,25} publicaron un trabajo en el que doce expertos evaluaron independientemente los seis sistemas de jerarquización de la evidencia más usados en el momento. Concluyeron que existían diferencias sustanciales en los criterios de gradación, con una baja sensibilidad y reproducibilidad de los mismos. La escala que encontraron más sensible y que respondía más adecuadamente a todo tipo de cuestiones es la del OCEBM.

La principal conclusión de ese trabajo fue constatar la inexistencia, hasta ese momento, de un modelo adecuado para la calificación de la evidencia y grados de recomendación que pudieran ser universalmente aceptados. Esto decidió a un grupo de expertos a constituirse en grupo para elaborar una propuesta que superara las limitaciones de los actuales sistemas de gradación de la evidencia (*Grading of Recommendations Assessment, Development and Evaluation Working Group* [GRADE])²³⁻²⁸, que se describe más adelante.

En nefrología la situación es similar, pero agravada por dificultades específicas de nuestra especialidad en relación con la disponibilidad de evidencia: dificultad y escasez de ECA^{39,40}, evolución silenciosa y prolongada de las enfermedades renales que obliga a utilizar variables subrogadas (p. ej., proteinuria), exclusión de enfermos con enfermedad renal en grandes ECA de afectación cardiovascular cuando ésta es su primera causa de mortalidad^{41,42}, y otras.

Diversas instituciones internacionales nefrológicas vienen realizando GPC desde hace unos 20 años⁴³⁻⁴⁶. Las primeras guías de *The National Kidney Foundation. Dialysis Outcomes and Quality Initiative* (DOQI) establecieron bases para el desarrollo de GPC en enfermedad renal crónica (ERC). En 1999 se habían desarrollado 114 guías DOQI, 56 basadas en el consenso de expertos, 31 mixtas y sólo 27 basadas en la evidencia⁴⁶⁻⁴⁹, lo que pone de manifiesto las dificultades existentes. En estas primeras GPC no se indicaban expresamente los criterios utilizados para evaluar la calidad de la evidencia que sustentaba las recomendaciones. Pero sí se especifica, de forma precisa y detallada, en las guías publicadas más recientemente en las que se ha adoptado, al igual que otras organizaciones internacionales, el modelo GRADE de elaboración de GPC y jerarquización de la evidencia, que se describen más adelante^{50,51}.

La EDTA-ERA viene realizando GPC desde hace más de 10 años^{44,45} en un contexto similar de problemas, por lo que se decidió pedir a un grupo de expertos que reconsiderara las condiciones en las que debían realizarse las GPC patrocinadas por esta institución y el sistema que adoptar para realizar la gradación de la evidencia. Las principales conclusiones fueron las siguientes:

- La falta de evidencia verificable en estudios de investigación clínica en numerosas áreas de nuestra especialidad^{39,40,44} explica que un número importante de las recomendaciones recogidas en las guías europeas son exclusivamente respaldadas por recomendaciones de expertos. Creían que, aun señalando expresamente este bajo nivel de evidencia, su presencia en un documento auspiciado por la EDTA-ERA podría generar equívocos, orientando la práctica clínica en un determinado sentido sin suficiente respaldo experimental. Por esta razón, el grupo acordó que la EDTA-ERA impulsara GPC sólo en aspectos sustentados por niveles de evidencia adecuados, el resto se recogerían como «Recomendaciones» y se publicarían en un formato completamente diferente^{44,45}. También se decidió el cambio de nombre del grupo responsable de su elaboración y publicación, de *European Renal Best Practice Guidelines* a *European Renal Best Practice*.
- Se reconocieron déficits de calidad metodológica en algunos aspectos de las guías⁴⁴, por lo que se decidió reconsiderar su sistemática de realización y clasificación en niveles de evidencia y grado de recomendación, adoptando como propias las decisiones de la iniciativa KDIGO^{44,45,51,52}. Ante las dudas acerca de la presencia de sesgos en las GPC patrocinadas por la industria⁵³, se decidió que las guías europeas de nefrología sólo admitieran patrocinios institucionales⁴⁴.

La última y más rigurosa iniciativa para propiciar el desarrollo de GPC en nefrología es la de KDIGO (*The Kidney Disease Global Outcomes improvement initiative*)⁵², una fundación independiente, declarada sin ánimo de lucro, y que tiene como principal objetivo: «mejorar el cuidado y los resultados del tratamiento de pacientes con enfermedad renal en todo el mundo, mediante la promoción, coordinación, colaboración e integración de GPC sobre temas nefrológicos».

Ante la importante variabilidad de los sistemas utilizados para el desarrollo de GPC en nefrología, KDIGO encargó en el año 2004 a una comisión de expertos la revisión de los conocimientos actuales sobre graduación de la evidencia, con el fin de llegar a un consenso para adoptar un sistema que conjugara, de forma razonable, la simplicidad con la complejidad y que cubriera los objetivos necesarios para un sistema de jerarquización eficaz y riguroso. Esta comisión, acordó adoptar el sistema GRADE para la sistemática de elaboración de las guías y para la clasificación de la calidad de la evidencia y la fuerza de las recomendaciones.

También realizaron adaptaciones particulares del GRADE para el caso de la nefrología, elaboradas con criterios estrictos y mencionando expresamente las mismas en las GPC. Las principales fueron: 1) la necesidad de extrapolar resultados obtenidos en población sin nefropatía a los pacientes con nefropatía e insuficiencia renal, y 2) la necesidad de agregación de datos cualitativos cuando no es posible utilizar variables cuantitativas.

EL MODELO GRADE PARA LA JERARQUIZACIÓN DE LA CALIDAD DE LA EVIDENCIA Y GRADOS DE RECOMENDACIÓN

El modelo GRADE^{23-28,52} presenta importantes fortalezas: es un sistema sencillo y fácil de implementar que desarrolla una metodología rigurosa de realización de revisiones sistemáticas y utiliza escalas estructuradas para calificar la calidad de la evidencia de los trabajos y la fuerza de la recomendación. El modelo ha sido evaluado en estudios piloto que afrontan distintos problemas clínicos, acreditando un adecuado nivel de sensibilidad y reproducibilidad⁵⁴. Por todo ello, ha alcanzando un alto nivel de consenso y aceptación internacional^{44,45,51,52} por parte de importantes instituciones nefrológicas (*Caring for Australians with renal impairment, United Kingdom Renal Association, European Best Practice Guidelines, Canadian Society of Nephrology. Kidney Disease Outcomes Quality Initiative*).

La tabla 1 muestra la sistemática que seguir en respuesta a preguntas clínicas concretas sobre el resultado de una intervención y, por tanto, para la realización de GPC establecida por GRADE^{23,52}.

Organización general y realización de revisiones sistemáticas

Los pasos 1 a 3 describen la estrategia inicial de búsqueda de la evidencia para contestar a preguntas clínicas sobre una intervención y realizar una GPC, incluyendo la organización general del trabajo (constitución del grupo y sistemática de realización, identificación de las intervenciones a estudiar, definición de las variables de resultado, incluyendo las de daño), la realización de RS para cada intervención de interés clínico (búsqueda, selección y tabulación de resultados para cada variable, determinando la magnitud del efecto y la calidad del estudio), y la pre-

Tabla 1. Sistemática básica para la realización de una guía clínica^a**ORGANIZACIÓN GENERAL y REALIZACIÓN DE REVISIONES SISTEMÁTICAS**

- Paso 1 **Organización general del trabajo.** Constitución del grupo y sistemática de trabajo, identificación de las intervenciones con relevancia clínica, definición de las variables de resultado, incluyendo las de daño, programar la realización de las revisiones sistemáticas necesarias
- Paso 2 **Revisión sistemática de la bibliografía para cada intervención de interés clínico.** Búsqueda, selección y tabulación de resultados para cada variable, determinando la magnitud del efecto y la calidad de cada trabajo
- Paso 3 **Preparación de perfiles de evidencia** para cada variable de resultado con importancia clínica y en cada subgrupo de enfermos, tabulando los principales resultados y la calidad metodológica de cada trabajo

ESTIMACIÓN DE LA CALIDAD DE LA EVIDENCIA

- Paso 4 **Análisis de calidad de la evidencia** para cada variable de resultado (criterios GRADE)
- Paso 5 **Valoración de la importancia relativa de cada variable** (incluir sólo aquellas con importancia clínica). Calificar la importancia como «crítica» o «importante pero no crítica»
- Paso 6 **Globalizar la calidad de la evidencia de una intervención**, asumiendo para el conjunto la calidad de la asistencia de las variables críticas con menor nivel de evidencia

ESTIMACIÓN DE LA FUERZA DE LA RECOMENDACIÓN

- Paso 7 **Análisis beneficio/daño** para las intervenciones en cada variable crítica (criterios GRADE)
- Paso 8 **Otras consideraciones:** magnitud del efecto, análisis de costes, disponibilidad, etc.
- Paso 9 **Fuerza de la Recomendación:** formular recomendaciones, calificando la fuerza de la recomendación, en cada punto de la GPC (Criterios GRADE)

^a Modificado de Guyatt GH, et al.²³ y de Uhlig K, et al.⁵².

paración de perfiles de evidencia para cada variable de resultado con importancia clínica y en cada subgrupo de enfermos, tabulando los principales resultados y la calidad metodológica de cada trabajo.

Estimación de la calidad de la evidencia

Del 4 a 6 se describen los pasos a seguir para jerarquizar la calidad de la evidencia. Para ello, GRADE desarrolla un análisis estructurado y explícito (cada revisor explica la naturaleza de sus decisiones) que considera, no sólo el diseño, sino también la calidad metodológica de los estudios, así como su consistencia, precisión y aplicabilidad. En primer lugar, se realiza un análisis de la calidad de la evidencia para cada variable de resultado (y subgrupo de pacientes), clasificándola en cuatro grados: calidad alta, moderada, baja y muy baja (grados del 1 al 4). Este índice expresa el grado de confianza que podemos tener en que la estimación del efecto de una intervención sobre la variable de interés se aproxime al real y será tanto mayor cuanto más alto sea el nivel de evidencia. La probabilidad de que una nueva investigación pueda cambiar los resultados será muy baja, en los casos calificados como de calidad de la evidencia muy alta e irá aumentando conforme ésta disminuya. Los puntos que se siguen en el GRADE para su valorar la calidad se esquematizan en la tabla 2.

- **Diseño:** Para la evaluación del nivel de evidencia se comienza analizando el tipo de diseño de los trabajos incluidos. Para estudios que deciden sobre la eficacia terapéutica, la producción de sesgos es mucho menos probable cuando se utiliza un ECA (si el planteamiento y el desarrollo son metodológicamente adecuados) que con cualquier otro tipo de diseño. Por esto, en la clasificación de GRADE si una intervención está respaldada inicialmente por ECA obtiene, como punto

de partida, la mayor calificación de la evidencia (grado 4 = calidad muy alta). Por el contrario, los estudios observacionales se califican inicialmente de calidad baja (grado 2).

Naturalmente, este planteamiento constituiría una simplificación inválida si no se realizan correcciones ya que, a veces, los estudios observacionales muestran resultados de gran valor probatorio mientras que los ECA pueden reducirlo por varias circunstancias: déficit de calidad metodológica, inconsistencia, aplicabilidad, presencia de sesgos.

- **Calidad metodológica:** Los ECA pueden disminuir su nivel de evidencia si existen limitaciones metodológicas importantes (en este caso, reducirán su nivel de evidencia en -1 grado) y si estas limitaciones son muy serias, disminuirían en -2 grados. Los elementos que el GRADE considera de forma sistemática para valorar la calidad metodológica son: el diseño del estudio, lo adecuado de su tamaño muestral, las técnicas de aleatorización y ocultación, variables de resultado estudiadas (incluyendo daño), el seguimiento, los perdidos a la evolución, el análisis estadístico y el análisis por intención de tratar. De esta forma, una recomendación formulada a partir de estudios con diseño de ECA puede ser calificada como de baja evidencia.
- **Consistencia:** El hallazgo de resultados discordantes entre trabajos semejantes con los que se estudia el efecto de una intervención cuestiona la credibilidad de aquéllos y, por tanto, si no se encuentra una explicación razonable a estas diferencias, reduce su nivel de evidencia. Si existe inconsistencia, el método GRADE disminuye en -1 el grado asignado hasta ese momento.
- **Aplicabilidad:** Se refiere a si la población o el diseño del estudio lo hacen asimilable a nuestra población de interés. Si existen diferencias sustanciales (raza, edad, sexo, comorbilidad, gravedad, subpoblación, etc.) de la población estudiada con la población diana, el grado de evidencia disminuye también entre -1 y -2 grados, según la importancia

Tabla 2. Criterios GRADE para jerarquizar el grado de calidad de la evidencia^a

PASO 1/ (grado inicial de evidencia según DISEÑO)	PASO 2/ Reducir nivel (0, -1, -2) si:	PASO 3/ Subir nivel (0, +1, +2) si:	NIVEL DE CALIDAD DE LA EVIDENCIA
ECA Calidad alta (Grado 4)	CALIDAD METODOLÓGICA Limitaciones metodológicas serias (-1) Limitaciones metodológicas muy serias (-2)	FUERZA DE LA ASOCIACIÓN (intervención vs. variable de resultado) - Magnitud efecto fuerte (+ 1) (RR >2 e IC <0,5) en 2 o más estudios observacionales	A: ALTO
ESTUDIOS CUASI-EXPERIMENTALES Calidad moderada (Grado 3)	CONSISTENCIA Trabajos con resultados inconsistentes (-1)	- Magnitud efecto muy fuerte (+ 2) (RR >5 e IC <0,2) - Todo o nada	B: MODERADO
E. OBSERVACIONAL Calidad baja (Grado 2)	APLICABILIDAD Diferencias en población, intervenciones o variables de resultado (-1 o -2)	- Gradiente de respuesta relacionado con dosis (+1)	C: BAJO
OTROS DISEÑOS = Calidad muy baja (Grados 1 o 0)	OTRAS Datos confusos o imprecisos (-1) Probabilidad de sesgos (-1)	FACTORES DE CONFUSIÓN +1 si perjudicaran el efecto	D: MUY BAJO

* Modificado de Guyatt GH, et al.²³ y de Uhlig K, et al.⁵².

de las diferencias. Lo mismo sucede si existen diferencias importantes entre las variables de resultado estudiadas o en la intervención (dosis o pauta administrada, etc.) o cuando la comparación entre dos intervenciones no se estudia directamente, sino utilizando otra o un grupo control o con el empleo de variables subrogadas de significación no claramente establecida. Finalmente, el nivel de la evidencia también disminuye (-1 grado) si en el trabajo se utilizan datos confusos o imprecisos o existe alta posibilidad de sesgos por alguna de las características del trabajo.

Por otra parte, como se ha descrito previamente, se considera que la probabilidad de sesgos es mayor, y por tanto el grado de evidencia menor, cuando se utilizan estudios observacionales. Por tanto, el GRADE les asigna una evidencia inicial baja (grado 2). Sin embargo, hay intervenciones con resultados muy convincentes que utilizan, exclusivamente, estudios observacionales. Anteriormente se ha descrito que el sistema de clasificación OCEBM concede un alto nivel de evidencia a los estudios observacionales que obtienen una gran magnitud del efecto (*all & none*). En el sistema GRADE se tiene también en cuenta esta circunstancia, valorando la fuerza de la asociación entre la intervención y la variable de resultado, de tal forma que para un estudio observacional con una magnitud del efecto fuerte, considerando éste cuando el RR (riesgo relativo) es mayor que 2 con un IC (intervalo de confianza) inferior a 0,5 se añade +1 grado a la calificación basal y si es muy fuerte (RR >5 e IC <0,2) se añaden +2 grados.

Por otra parte, también se considera que potencia el nivel de evidencia la presencia de factores de confusión que tienden a disminuir el efecto de la intervención o cuando existe un gradiente

de respuesta intervención-variable (+1 grado en ambos casos).

Siguiendo esta clasificación, una intervención cuyo efecto ha sido estudiado con ECA puede obtener un grado de evidencia bajo o muy bajo y, al contrario, un estudio observacional, una calificación de evidencia de calidad alta.

En una GPC el análisis del valor de una intervención (que hará recomendable o no su utilización en una determinada situación clínica) depende de la calidad de la evidencia y de la magnitud del efecto de ésta sobre diferentes variables de resultado (p. ej., supervivencia, síntomas cardiovasculares, ingresos, efectos secundarios, etc.) y cada una de éstas tiene una importancia diferente sobre las decisiones: es decir, hay variables de resultado que pueden estar sustentadas en un alto nivel de evidencia pero cuya importancia relativa en la toma de decisión es bajo. Por tanto, es importante considerar la importancia de cada variable en relación al conjunto de ellas, para establecer la recomendación final. El GRADE establece (paso 5) que, para este análisis, se incluyan sólo las variables con importancia clínica, ordenándolas de más a menos según el peso que estimemos que deben tener en la recomendación, calificando su importancia según corresponda en cada caso, como «crítica» o «importante pero no crítica». Por último, debemos globalizar la calidad de la evidencia de una intervención asumiendo para el conjunto la calidad de la asistencia de las variables críticas con menor nivel de evidencia.

Estimación de la fuerza de la recomendación

El sistema GRADE establece dos grados de recomendación, fuerte (nivel 1) y débil (nivel 2), aunque también se admiten re-

Tabla 3. Criterios GRADE. Estimación del grado de la recomendación^a

GRADO	PRERREQUISITOS	IMPLICACIONES
- Fuerte (nivel 1)	Calidad de la evidencia alta Balance favorable beneficio/daño	La mayoría de los especialistas bien informados elegirían esta opción La recomendación debe ser seguida
- Débil (nivel 2)	Calidad de la evidencia moderada o alta El balance beneficio/daño u otras aconsejan una recomendación débil	Muchos especialistas bien informados elegirían esta opción, pero una sustancial minoría no
- Basado en el consenso	Calidad de la evidencia baja, muy baja o ausente, pero con criterios firmes de que beneficio >> daño	Se recomienda seguir la recomendación pero prevalece el criterio individual

^a Modificado de Guyatt GH, et al.²³ y de Uhlig K, et al.⁵².

comendaciones fuera de gradación para casos con niveles bajos de evidencia, en los que el beneficio para la salud sea evidente a juicio de los expertos que desarrollan la GPC (basado en el consenso).

La calidad de la evidencia es el argumento inicial para determinar la fuerza de la recomendación y debe plantearse sólo cuando está respaldada por niveles adecuados de calidad metodológica. Pero no es el único: una determinada intervención puede estar sustentada en trabajos de alta calidad metodológica, pero no ser recomendable por otras razones y, al contrario, los niveles de evidencia pueden ser bajos o muy bajos, pero su recomendación, obligatoria por razones clínicas obvias (p. ej., anticoagulación en hemodiálisis).

El método GRADE (paso 7) establece la necesidad de realizar un balance del efecto neto para la salud de cada una de las intervenciones utilizadas en una GPC para matizar el énfasis o grado de recomendación. Las posibilidades de conclusión son: 1) hay un beneficio neto para la salud (beneficio>daño); 2) hay bastante equilibrio entre beneficio y daño (depende de circunstancias clínicas concretas); 3) es incierto si existe beneficio o daño, y 4) no hay un beneficio neto para la salud (daño>beneficio). Junto a esto, deben hacerse otras consideraciones (paso 8 del GRADE), como la valoración clínica de la magnitud del efecto, los costes, la disponibilidad, etc.

Con todo esto se realiza una estimación del grado de recomendación, teniendo en cuenta los criterios que se muestran en la tabla 3.

Las recomendaciones en una GPC deben basarse en niveles de calidad de la evidencia altos o moderados (A o B). El grado de recomendación fuerte (nivel 1) está respaldado por una alta calidad de la evidencia y un efecto neto para la salud favorable. Por esto, el texto especifica que «se recomienda» y se espera que la propuesta sea seguida por los médicos ya que la mayoría de los especialistas bien informados elegirían esta opción. En el grado de recomendación débil, la calidad de la evidencia es moderada o alta pero el balance beneficio/daño u otras razones aconsejan una recomendación débil. Muchos especialistas bien informados elegirían esta

opción, pero una sustancial minoría no lo haría, con argumentos razonables. La redacción de la propuesta en este caso, no la «recomienda», sino que la «sugiere». Se espera que los especialistas sigan la propuesta, pero su seguimiento está condicionado o es discrecional.

En la recomendación basada en el consenso, la calidad de la evidencia es baja, muy baja o ausente, por lo que la propuesta se realiza a partir de la opinión sustentada por expertos y la decisión, en casos en los que, aunque están sustentados en niveles muy bajos de evidencia metodológica, presentan, sin embargo, argumentos clínicos que muestran de forma clara y rotunda que el beneficio para la salud es claramente superior al daño (p. ej., el tratamiento con métodos de diálisis en la enfermedad renal crónica terminal) o, al menos, que los expertos consideran muy aconsejable (p. ej., una determinada frecuencia de revisiones). El seguimiento de estas recomendaciones es sugerido, pero prevalece el criterio individual del médico ante la situación de su paciente individual.

Como conclusión, creemos que los sistemas de jerarquización de la evidencia y grados de recomendación son importantes para objetivar de forma sencilla, pero con un importante trasfondo de rigor crítico, los distintos elementos en los que deben basarse las recomendaciones de práctica clínica. El modelo GRADE, aunque todavía presenta limitaciones, supone un avance importante en los sistemas de clasificación de calidad de la evidencia y grados de recomendación, en el que se tienen en cuenta de forma estructurada, explícita y sencilla, los distintos elementos que intervienen en ésta: diseño, calidad metodológica, consistencia, magnitud del efecto, precisión, aplicabilidad, presencia potencial de sesgos y balance beneficio/daño. Ha sido adoptado por las principales instituciones nefrológicas internacionales: KDIGO ha lanzado las primeras GPC siguiendo esta metodología, expuesta de una forma transparente y fácilmente comprensible, que facilita la comprensión de la recomendación y los matices (calidad metodológica y fuerza de la recomendación que modulan su aplicación)⁵⁵. De igual modo y por las mismas razones, nosotros propugnamos su adopción por la S.E.N. y por su grupo de Nefrología Basada en la Evidencia.

BIBLIOGRAFÍA

1. Quereda C, Pascual J, García López F, Alcázar R, Aljama P, Amenábar JJ, et al. Nefrología basada en investigación clínica: bases conceptuales para la constitución de un grupo de trabajo en Nefrología Basada en la Evidencia de la Sociedad Española de Nefrología. *Nefrología* 2006;26:163-72.
2. Quereda C, Pascual J, García López F, Alcázar R. ¿Por qué Nefrología Basada en la Evidencia? *Nefrología* 2006;26(Supl 5):3-5.
3. Oxman AD, Guyatt GH, et al. User's Guides to The Medical Literature VI. How to use an overview. *JAMA* 1994;272(17):1367-71.
4. Coronado BE, Bu D, Levey AS, Meyer KB. Methods of literature review and synthesis. *Am J Kidney Dis* 1998;32:S14-S55.
5. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF, for the QUOROM Group. Improving the quality of reports of meta-analyses of randomized controlled trials: the QUOROM statement. *Lancet* 1999;354:1896-900.
6. Oxman AD, editor. Section VI: Preparing and Maintaining Systematic Reviews: The Cochrane Collaboration Handbook. Oxford: Cochrane Collaboration, 1994.
7. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ; GRADE Working Group. Rating quality of evidence and strength of recommendations: What is "quality of evidence" and why is it important to clinicians? *BMJ* 2008;336:995-8.
8. De Vet HC, van Tulder MW, Bouter LM. Levels of evidence: intellectual aid or absolute judgment? *Clin Epidemiol* 2003;56:917-8.
9. Schünemann HJ, Best D, Vist G, et al. Letters, numbers, symbols, and words: how best to communicate grades of evidence and recommendations? *Can Med Assoc J* 2003;169:677-80.
10. Woolf SH, Grol R, Hutchinson A, et al. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. *BMJ* 1999;318:527-30.
11. Field MJ, Lohr KN. En: Field MJ, Lohr KN, editores). *Guidelines for Clinical Practice—From Development to Use*. National Academy Press: Washington, D.C., 1992.
12. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. *CMAJ* 1988;138:697-703.
13. Schunemann H, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: Integrating values and consumer involvement. *Health Res Policy Syst* 2006;5:4-22.
14. Nissenson AR, Collins AJ, Hurley J, et al. Opportunities for improving the care of patients with chronic renal insufficiency: current practice patterns. *J Am Soc Nephrol* 2001;12:1713-20.
15. Locatelli F, Zoccali C. Clinical policies on the management of chronic kidney disease patients in Italy. *Nephrol Dial Transplant* 2007;23:621-6.
16. Owen WF, Jr. Patterns of care for patients with chronic kidney disease in the United States: dying for improvement. *J Am Soc Nephrol* 2003;14:S76-S80.
17. Vanbelleghem H, Vanholder R, Levin NW, et al. The kidney disease: improving global outcomes website: comparison of guidelines as a tool for harmonization. *Kidney Int* 2007;71:1054-61.
18. Genuis SJ. The proliferation of clinical practice guidelines: professional development or medicine-by-numbers? *J Am Board Fam Pract* 2005;18:419-25.
19. Guidelines for the treatment of mild hypertension. Memorandum from a WHO/ISH meeting. *Bull World Health Organ* 1983;61:53-6.
20. Amerling R, Winchester JF, Ronco C. Guidelines have done more harm than good. *Blood Purif* 2008;26:73-6.
21. Port FK, Pisoni RL, Bommer J, et al. Improving outcomes for dialysis patients in the international Dialysis Outcomes and Practice Patterns Study. *Clin J Am Soc Nephrol* 2006;1:246-55.
22. Alcázar R, Quereda C. Nefrología Basada en la Evidencia de la SEN: Guías Clínicas. *Nefrología* 2006;26:53-8.
23. 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;336:1049-51.
24. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. GRADE Working Group Grading. Quality of evidence and strength of recommendations. *BMJ* 2004;328:1490-4.
25. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, et al., and The GRADE Working Group Systems for grading the quality of evidence and the strength of recommendations. Critical appraisal of existing approaches. GRADE Working Group. *BMC Health Serv Res* 2004;4:38.
26. Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al.; GRADE Working Group. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 2008;336:1106-10.
27. Guyatt GH, Oxman AD, Kunz R, Jaeschke R, Helfand M, Liberati A, Vist GE, Schünemann HJ; GRADE working group. Rating quality of evidence and strength of recommendations: Incorporating considerations of resources use into grading recommendations. *BMJ* 2008;336:1170-3.
28. Jaeschke R, Guyatt GH, Dellinger P, Schünemann H, Levy MM, Kunz R, et al.; GRADE working group. Use of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive. *BMJ* 2008;337:744-9.
29. Canadian Task Force on the Periodic Health Examination: The periodic health examination. *CMAJ* 1979;121:1193-254.
30. West S, King V, Carey TS, et al. Systems to rate the strength of scientific evidence. Evidence Report/Technology Assessment No. 47 (Prepared by the Research Triangle Institute-University of North Carolina Evidence-based Practice Center under Contract No. 290-97-0011). AHRQ Publication No. 02-E016. Agency for Healthcare Research and Quality: Rockville, MD, 2002.
31. US Preventive Services Task Force. Screening for adolescent idiopathic scoliosis: review article. *JAMA* 1993;269:2667-72.

32. Schunemann HJ, Jaeschke R, Cook DJ, Bria WF, El-Solh AA, Ernst A, et al. An official ATS statement: grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations. *Am J Respir Crit Care Med* 2006;174:605-14.
33. Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. *BMJ* 2004;328:39-41.
34. Victoria CG, Habicht JP, Bryce J. Evidence-based public health: moving beyond randomized trials. *Am J Public Health* 2004;94:400-5.
35. Ferreira PH, Ferreira ML, Maher CG, et al. Effect of applying different levels of evidence criteria on conclusion of Cochrane review of interventions for low back pain. *J Clin Epidemiol* 2002;55:1126-9.
36. Wilson MC, Hayward RS, Tunis SR, Bass EB, Guyatt GH, for the Evidence-Based Medicine Working Group: Users' guides to the medical literature. VIII. How to use clinical practice guidelines, B: What are the recommendations and will they help you in caring for your patients? *JAMA* 1995;274:1630-2.
37. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ, for the Evidence-Based Medicine Working Group: Users' guides to the medical literature. IX. A method for grading health care recommendations. *JAMA* 1995;274:1800-4.
38. Levels of Evidence OCEBM. Disponible en: http://www.cebm.net/levels_of_evidence.asp
39. Himmelfarb J. Chronic kidney disease and the public health: gaps in evidence from interventional trials. *JAMA* 2007;297:2630-3.
40. Strippoli GMF, Craig JC, Schena FP. The Number, Quality, and Coverage of Randomized Controlled Trials in Nephrology. *J Am Soc Nephrol* 2004;15:411-9.
41. Coca SG, Krumholz HM, Garg AX, et al. Under representation of renal disease in randomized controlled trials of cardiovascular disease. *JAMA* 2006;296:1377-84.
42. Sarnak MJ, Levey AS, Schoolwerth AC, et al. Kidney disease as a risk factor for development of cardiovascular disease: a statement from the American Heart Association Councils on kidney in cardiovascular disease, high blood pressure research, clinical cardiology, and epidemiology and prevention. *Hypertension* 2003;42:1050-65.
43. Kliger AS, Haley WE. Clinical Practice Guidelines in End-Stage Renal Disease: A Strategy for Implementation. *J Am Soc Nephrol* 1999;10:872-7.
44. Zoccali C, Abramowicz D, Cannata-Andia JB, Cochat P, Covic A, Eckardt KW. European best practice, quo vadis? From European best practice guidelines (EBPG) to European renal best practice (ERBP). *Nephrol Dial Transplant* 2008;23:2162-6.
45. Vanholder R, Abramowicz D, Cannata-Andia JB, Cocchi V, Cochat P, Covic A, et al. The future of European Nephrology Guidelines: a declaration of intent by European Renal Best Practice (ERBP). *NDT Plus* 2009;2:213-21.
46. Woolf SH. Manual for Conducting Systematic Reviews, U.S. Department of Health and Human Services, Public Health Service, Agency of Health Care Policy and Research, August 1996. Clinical practice guidelines: Hemodialysis adequacy and peritoneal dialysis adequacy. *Am J Kidney Dis* 1997;30(Suppl 2):S1-S136.
47. National Kidney Foundation Dialysis Outcomes Quality Initiative: Clinical practice guidelines: Vascular access and anemia of chronic renal failure. *Am J Kidney Dis* 1997;30(Suppl 3):S137-S240.
48. K/DOQI. Clinical practice guidelines for cardiovascular disease in dialysis patients. *Am J Kidney Dis* 2005;45:S1-S153.
49. National Kidney Foundation. K/DOQI clinical practice guidelines and clinical practice recommendations for anemia in chronic kidney disease. *Am J Kidney Dis* 2006;47:S1-S145.
50. National Kidney Foundation. K/DOQI Clinical Practice Guidelines and Clinical Practice Recommendations for diabetes and chronic kidney disease. Methods for evaluation of evidence. *Am J Kidney Dis* 2007;49(Suppl 2):148-54.
51. Guyatt GH, Oxman AD, Visit GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924-6.
52. Uhlig K, Macleod A, Craig J, Lau J, Levey AS, Levin A, et al. Grading evidence and recommendations for clinical practice guidelines in nephrology. A position statement from Kidney Disease: Improving Global Outcomes (KDIGO). *Kidney Int* 2006;70:2058-65.
53. Coyne DW. Influence of industry on renal guideline development. *Clin J Am Soc Nephrol* 2007;2:3-7.
54. GRADE Working Group. Systems for grading the quality of evidence and the strength of recommendations II: a pilot study of a new system for grading the quality of evidence and the strength of recommendations. *BMC Health Serv Res* 2005;5:25.
55. KDIGO Clinical Practice Guideline for the diagnosis, evaluation, prevention and treatment of Chronic Kidney Disease-Mineral and Bone Disorder. Methodological approach. *Kidney Int* 2009;76(Suppl 113):S9-S21.