

Anexo 7. Clasificación de la calidad de la evidencia y graduación de la fuerza de las recomendaciones

Para clasificar la calidad de la evidencia y graduar la fuerza de las recomendaciones de la presente guía de práctica clínica (GPC), y de acuerdo al manual metodológico, se han seguido las directrices del grupo de trabajo GRADE (*Grading of Recommendations of Assessment Development and Evaluation*)¹⁶³⁻¹⁶⁴.

Las principales etapas que conforman el sistema GRADE son:

1. Elaboración de las preguntas y definición de los desenlaces de interés

Las preguntas clínicas habitualmente se estructuran en forma de PICO, incorporando una población (**P**), una intervención a evaluar (**I**), un comparador (**C**), así como unos desenlaces (resultados, *outcomes* [**O**]) de interés (p.ej. progresión de la enfermedad, mortalidad o eventos adversos). El grupo elaborador define los desenlaces de interés a evaluar y los clasifica en no importantes, importantes pero no claves para la toma de decisiones o desenlaces clave para la toma de decisiones. GRADE propone una escala de nueve puntos para realizar esta clasificación siendo: de 1-3 desenlaces no importantes; de 4-6 desenlaces importantes pero no claves y, de 7-9 los desenlaces claves para la toma de decisiones. Son los desenlaces importantes y claves los que determinan la calidad global de la evidencia y el balance entre los beneficios y riesgos de las intervenciones evaluadas.

2. Clasificación de la calidad de la evidencia

La clasificación de la calidad de la evidencia se realiza para cada uno de los desenlaces de interés relevantes para la toma de decisiones y, por tanto, para una pregunta clínica es posible que diferentes desenlaces presenten diferentes calidades¹⁶⁵. La evaluación de la calidad de la evidencia y su clasificación nos permite conocer el grado de confianza que tenemos en los resultados de la literatura encontrados.

GRADE propone clasificar la calidad de la evidencia como alta, moderada, baja o muy baja¹⁶⁵. GRADE considera que los ECA tienen una “*calidad alta*” y que los estudios observacionales tienen una “*calidad baja*”. No obstante, GRADE sugiere una serie de aspectos que pueden disminuir o aumentar esta calidad. Los aspectos que pueden disminuir la calidad son: limitaciones de los estudios (limitaciones en el diseño y ejecución [riesgo de sesgo]), los resultados inconsistentes o imprecisos, la ausencia de evidencia directa y el sesgo de publicación. Los aspectos que pueden aumentar la calidad son: la fuerza de la asociación, el gradiente dosis-respuesta y si la consideración de otros potenciales factores de confusión que no se han tenido en cuenta en el estudio, en caso de presentarse, reforzarían sus resultados. La calidad de la evidencia para cada uno de los desenlaces incluidos en la presente GPC, se presenta en el margen derecho del texto de síntesis de la evidencia.

A continuación se describen cada uno los aspectos mencionados¹⁶⁵:

- **Limitaciones en el diseño o en la ejecución de los estudios (Riesgo de sesgo):** como pueden ser la ausencia de ocultamiento de la secuencia de aleatorización, el enmascaramiento inadecuado, las pérdidas importantes, la ausencia de análisis por intención de tratar o la finalización del estudio antes de lo previsto por razones de beneficio, entre otros.
- **Resultados inconsistentes:** estimaciones discrepantes del efecto de un tratamiento (heterogeneidad o variabilidad en los resultados) en los estudios disponibles sugieren diferencias reales en dichas estimaciones. Éstas pueden deberse a diferencias en la población, en la intervención, en los desenlaces o en la calidad de los estudios. Cuando existe heterogeneidad sin una potencial explicación razonable la calidad disminuye.
- **Ausencia de evidencia científica directa:** en el caso de ausencia de comparaciones directas entre dos tratamientos (comparación de cada tratamiento frente a placebo, pero no entre ellos) o de encontrarse estudios realizados en poblaciones diferentes a la población objeto de la guía, disminuye la calidad.
- **Imprecisión:** cuando los estudios disponibles incluyen relativamente pocos eventos y pocos pacientes y, por tanto, presentan intervalos de confianza amplios, la calidad disminuye.
- **Sesgo de notificación:** la calidad puede disminuir si se tiene la duda razonable sobre si los autores han incluido o no todos los estudios (por ejemplo, el sesgo de publicación en el contexto de una RS) o si los autores han incluido o no todas las variables relevantes de resultado (*outcome reporting bias*).

Por otro lado, en situaciones poco frecuentes, cuando se han evaluado estudios observacionales y existe un bajo riesgo de sesgo, existe una serie de factores que pueden aumentar su calidad:

- **Efecto importante:** cuando el efecto observado muestra una asociación fuerte ($RR > 2$ ó $< 0,5$) o muy fuerte ($RR > 5$ ó $< 0,2$) y consistente, basada en estudios sin factores de confusión. En estas ocasiones se puede considerar la calidad como moderada o incluso alta.
- La presencia de un **gradiente dosis-respuesta**.
- Situaciones en las cuales **todos los posibles factores de confusión podrían haber reducido la asociación observada**. En los casos en los que las pacientes que reciben la intervención de interés presentan un peor pronóstico y, aun así, presentan mejores resultados que el grupo control, aumenta nuestra confianza en que el potencial efecto sea real.

Posterior a la clasificación de la calidad de la evidencia para cada uno de los desenlaces de interés, GRADE propone realizar una evaluación de la calidad global. Para ello se sugiere tomar la calidad más baja de los desenlaces que se han considerado críticos e importantes para la toma de decisiones¹⁶⁷. La calidad global para cada una de las preguntas se presenta en el apartado ‘de la evidencia a la recomendación’.

B) Graduación de la fuerza de las recomendaciones

La fuerza de las recomendaciones nos informa de en qué medida el llevar a cabo una recomendación conlleva más beneficios que daños en los pacientes. GRADE propone varios factores que hay que tener en cuenta en la graduación de las recomendaciones^{164,166}. Estos son los principales:

- **Calidad de la evidencia:** Si la calidad de la evidencia no es alta, a pesar de que la magnitud sea importante, debería disminuir la confianza y por tanto la fuerza con la que se lleve a cabo una recomendación.
- **Valores y preferencias de los pacientes:** Se realizó una búsqueda de la literatura de estudios sobre valores y preferencias de los pacientes con enfermedad renal crónica. Para cada una de las preguntas incluidas en la presente GPC, los principales hallazgos fueron descritos narrativamente en el apartado del texto ‘valores y preferencias de los pacientes’ y considerados en el apartado correspondiente de la sección ‘de la evidencia a la recomendación’ de la pregunta.
- **Balance entre beneficios y riesgos:** Para realizar el balance entre los beneficios y los riesgos de las intervenciones evaluadas, deben considerarse las diferencias que hay entre los estimadores del efecto de los desenlaces considerados como claves e importantes. Si al realizar este balance, existe una gran diferencia entre los beneficios (ej. disminución de la progresión de la enfermedad) y los riesgos (ej. efectos adversos), es más probable de realizar una recomendación fuerte (a favor o en contra). Si la diferencia es pequeña, será más probable de hacer una recomendación débil.
- **Costes:** A diferencia de otras variables de resultado, los costes son mucho más variables en el tiempo, en distintas áreas geográficas y en función de diversas implicaciones. Así, aunque un coste elevado disminuye la probabilidad de realizar una recomendación fuerte, el contexto en el cual se realiza la recomendación será crítico en la valoración final de los mismos. En la presente GPC se realizó una búsqueda de la literatura de estudios de costes de manera global. De todos aquellos estudios económicos encontrados, se realizó una síntesis de los resultados que se incluyó en el apartado de ‘uso de recursos y costes’ del texto de cada una de las preguntas y fue considerada en el apartado de la sección ‘de la evidencia a la recomendación’.

En la graduación de la fuerza de las recomendaciones GRADE considera dos categorías: recomendaciones fuertes y recomendaciones débiles. En las recomendaciones **fuertes** el grupo elaborador confía en que los efectos beneficiosos superan a los perjudiciales o viceversa, que los daños superan a los beneficios. En el primer caso la recomendación es fuerte a favor. En el segundo es fuerte en contra. Este tipo de recomendaciones incluyen expresiones enfáticas, del tipo “se recomienda”.

Las recomendaciones **débiles** también pueden ser a favor o en contra. Una recomendación es débil a favor cuando el grupo elaborador concluye que los efectos beneficiosos de llevar a cabo la recomendación probablemente superan los perjudiciales, aunque no está completamente seguro. En cambio la recomendación es débil en contra, cuando los efectos adversos probablemente superan a los beneficiosos. La recomendación se redacta con la expresión “Se puede valorar” o “se sugiere”.

En el apartado de la evidencia a la recomendación de cada una de las preguntas incluidas en la presente GPC, se puede encontrar detallada la forma como el grupo elaborador ha tenido en cuenta cada uno de estos aspectos en la elaboración de las recomendaciones. Para las intervenciones de las que no se dispone de evidencia y el grupo elaborador ha realizado recomendaciones de investigación con el fin de resaltar la importancia de realizar proyectos en áreas específicas.

Información más detalla sobre la metodología GRADE se encuentra en la página de internet <http://www.gradeworkinggroup.org>. Así mismo existen otras publicaciones en castellano para usuarios de GPC^{168,169}.