

Anexo 2. Clasificación de la calidad de la evidencia y graduación de la fuerza de las recomendaciones mediante el sistema GRADE

Cuando se formulan recomendaciones, el grupo elaborador de la guía debe plantearse hasta qué punto se puede confiar en que llevar a cabo una recomendación conllevará más beneficios que daños. Esta decisión no es sencilla y está influida por múltiples factores que convierten esta etapa en una de las más complejas de la elaboración de una guía.

Los sistemas de formulación de recomendaciones se comenzaron a proponer hace más de dos décadas. Estos sistemas diferenciaban ya inicialmente el nivel de evidencia científica (la adecuación de los diferentes diseños de estudios para contestar a los diferentes tipos de preguntas) de la fuerza de las recomendaciones. Desde entonces los diferentes sistemas han ido evolucionando e incorporando otros aspectos más allá del diseño de los estudios que hay que tener en cuenta a la hora de formular recomendaciones.

Para clasificar la calidad de la evidencia y graduar la fuerza de las recomendaciones se han seguido las directrices del grupo de trabajo GRADE. Este grupo de trabajo intenta establecer un método de elaboración de recomendaciones explícito y transparente, de fácil manejo por los equipos elaboradores de GPC, con el objetivo de superar los inconvenientes de otros sistemas de elaboración de recomendaciones (Guyatt, 2008a, 2008b, 2011a).

Las principales etapas que conforman el sistema GRADE (*Grading of Recommendations of Assessment Development and Evaluation*) se describen a continuación:

A) Clasificación de la calidad de la evidencia científica

La clasificación de la calidad de la evidencia se realiza para cada una de los desenlaces de interés relevantes para la toma de decisiones y, por tanto, para una pregunta clínica es posible que diferentes desenlaces de interés reflejen una clasificación de la calidad diferente (Guyatt, 2011b). Inicialmente, se considera que los ECAs tienen una "*calidad alta*" y que los estudios observacionales tienen una "*calidad baja*". No obstante se sugieren una serie de aspectos que pueden disminuir la calidad de los ECAs o que la pueden aumentar en el caso de los estudios observacionales. La calidad de la evidencia científica se califica como alta, moderada, baja o muy baja (Balslem, 2011). La calidad de la evidencia se presenta en el margen derecho del texto de la guía y se modifica adecuadamente si la clasificación de la evidencia cambia para cualquier otro desenlace de interés que se discuta en el resumen de la evidencia. En la clasificación de la calidad global de la evidencia, el grupo elaborador ha considerado:

- Las variables relevantes para la toma de decisiones.
- Cuando éstas han mostrado un resultado inconsistente (hacia un beneficio y hacia un daño) se ha considerado la calidad global como la más baja entre las diferentes variables.
- Cuando éstas han mostrado un resultado consistente (hacia un beneficio o un daño) se ha considerado la calidad global como la más alta entre las diferentes variables.
- Cuando éstas han mostrado un resultado consistente pero el balance entre los beneficios y los perjuicios es incierto se ha considerado la calidad global como la más baja entre las diferentes variables.

Se han considerado los siguientes aspectos para disminuir la calidad de los ECAs (Balshem, 2011):

- **Limitaciones en el diseño o en la ejecución del ECA:** como pueden ser la ausencia de ocultamiento de la secuencia de aleatorización, enmascaramiento inadecuado, pérdidas importantes, ausencia de análisis por intención de tratar, finalización del estudio antes de lo previsto por razones de beneficio.
- **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 de interés o en la calidad de los estudios. Cuando exista heterogeneidad no explicada de manera 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 la extrapolación de los resultados de un estudio con un determinado fármaco al resto de fármacos de su misma familia en ausencia de un efecto de clase demostrado, también se considera evidencia científica indirecta. Es frecuente que existan grandes diferencias entre la población en la que se aplicarán las recomendaciones y la incluida en los estudios evaluados. Por último, deben ser también valorados los aspectos de la potencial aplicabilidad en nuestro entorno o la validez externa de la evidencia científica disponible.
- **Imprecisión:** cuando los estudios disponibles incluyen relativamente pocos eventos y pocas mujeres 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, cuando se han evaluado estudios observacionales se han tenido en cuenta una serie de factores que pueden aumentar su calidad:

- **Efecto importante:** cuando el efecto observado muestra una asociación fuerte ($RR > 2$ o $< 0,5$) o muy fuerte ($RR > 5$ o $< 0,2$) 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 mujeres que reciben la intervención de interés presentan un peor pronóstico y, aún así, presentan mejores resultados que el grupo control, es probable que el efecto observado real sea mayor.

Los miembros del grupo elaborador han trabajado las diferentes preguntas clínicas en el que se ha incluido el resumen de la evidencia disponible para cada una de las preguntas, y en el que se ha incluido una justificación razonada de los principales aspectos que han modificado la calidad de la evidencia. Además a modo de anexo a esta justificación se han incluido unas tablas GRADE de resumen de los hallazgos en los que se sintetiza la literatura disponible para cada desenlace de interés y se detallan los aspectos y las razones que han modificado la calidad de la evidencia (Guyatt, 2011c).

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 las mujeres. Existen una serie de factores que hay que tener en cuenta en la graduación de las recomendaciones (Guyatt, 2008b; Brożek, 2011):

- **Balance entre beneficios y riesgos:** Para realizar una adecuada valoración del balance entre los beneficios y los riesgos es necesario tener en cuenta el riesgo basal de la población a la que va dirigida la recomendación, y el efecto tanto en términos relativos como absolutos.
- **Calidad de la evidencia científica:** antes de llevar a cabo una recomendación es necesario conocer la certidumbre sobre la estimación del efecto observada. Si la calidad de la evidencia científica 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:** la incertidumbre sobre los valores y las preferencias de la población diana a la cual va dirigida la GPC, será otro de los factores que hay que tener en cuenta. El personal sanitario, el colectivo de mujeres o la sociedad en general deben ver reflejados sus valores y sus preferencias que deberían influir en la graduación de las recomendaciones.

- **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 graduar como fuerte una recomendación, el contexto será crítico en la valoración final.

La graduación de la fuerza de las recomendaciones 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. La recomendación se redacta con la expresión ‘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’.

Para las intervenciones de las que no se dispone de evidencia y el grupo elaborador quiere resaltar un determinado aspecto, se dispone de una serie de recomendaciones basadas en la experiencia clínica y el consenso del grupo elaborador, que se han identificado con el símbolo √.

Los miembros del grupo elaborador han formulado las recomendaciones a partir de los documentos de trabajo comentados anteriormente en los que para cada pregunta clínica se ha incluido un resumen de la evidencia y la justificación de la clasificación de la calidad de la evidencia. Este documento de trabajo también incluía un apartado en el que se discutían de aspectos identificados en la literatura relacionados con los factores que hay que tener en cuenta en la graduación de las recomendaciones. Además las recomendaciones se encabezaban con un apartado en el que se justificaban tanto los aspectos que habían determinado tanto la dirección como la fuerza de las recomendaciones. Todos estos documentos de trabajo se distribuyeron entre los componentes del grupo elaborador para su consideración y comentarios, y el enunciado final de las recomendaciones se aprobó en una serie de reuniones de trabajo celebradas en diferentes momentos del proceso.