

# Metodología GRADE, parte 1: cómo evaluar la certeza de la evidencia

## The GRADE approach, Part 1: how to assess the certainty of the evidence

Matías Kirmayr<sup>a</sup> , Carlos Quilodrán<sup>a</sup> , Bárbara Valente<sup>b</sup> , Cristóbal Loezar<sup>a,c</sup> , Luis Garegnani<sup>d</sup> , Juan Víctor Ariel Franco<sup>d</sup> 

<sup>a</sup> Escuela de Medicina, Universidad de Valparaíso, Valparaíso, Chile

<sup>b</sup> Carrera de Medicina, Universidad Nacional de La Matanza, Buenos Aires, Argentina

<sup>c</sup> Centro Interdisciplinario de Estudios en Salud (CIESAL), Universidad de Valparaíso, Valparaíso, Chile

<sup>d</sup> Centro Cochrane Asociado, Departamento de Investigación, Instituto Universitario Hospital Italiano, Buenos Aires, Argentina

\*Autor de correspondencia [jvaf85@gmail.com](mailto:jvaf85@gmail.com)

**Citación** Kirmayr M, Quilodrán C, Valente B, Loezar C, Garegnani L, Franco JVA. The GRADE approach, Part 1: how to assess the certainty of the evidence. *Medwave* 2021;21(02):e8109

Doi [10.5867/medwave.2021.02.8109](https://doi.org/10.5867/medwave.2021.02.8109)

**Fecha de envío** 22/9/2020

**Fecha de aceptación** 28/12/2020

**Fecha de publicación** 17/03/2021

**Origen** No solicitado.

**Tipo de revisión** Revisión por pares externa, por tres árbitros a doble ciego.

**Palabras clave** Evidence-Based Medicine, GRADE approach, Systematic reviews as topic

## Resumen

La certeza de la evidencia para las intervenciones es la seguridad o confianza de que el verdadero efecto se encuentra dentro de un rango particular o en relación con un umbral. En la nueva pirámide de la evidencia, las revisiones sistemáticas representan la lupa a través de la cual se evalúa dicha certeza. La metodología GRADE (Grading of Recommendations Assessment, Development and Evaluation) surge como respuesta a la existencia de múltiples sistemas de clasificación de la evidencia y ofrece un proceso transparente y estructurado para desarrollar y presentar resúmenes de evidencia considerando la certeza de esta y, en un segundo paso, la fuerza de las recomendaciones que informan. El proceso GRADE comienza con una pregunta explícita, que incluye las especificaciones de todos los desenlaces importantes y críticos. Los principales dominios que se utilizan para valorar la certeza de la evidencia son: 1. El riesgo de sesgo, 2. La inconsistencia, 3. Evidencia indirecta, 4. Imprecisión, y 5. El sesgo de publicación. Los factores que pueden aumentar la certeza de la evidencia son: 1. Gradiente dosis respuesta, 2. Gran magnitud del efecto, 3. Efecto de los potenciales factores de confusión residual. Finalmente, las tablas de resumen de hallazgos (Summary of Findings) sintetizan el proceso de manera simplificada y con un lenguaje controlado. El

objetivo general de esta revisión narrativa es abordar los principales conceptos básicos teóricos y prácticos de la metodología GRADE de manera simplificada y con ejemplos prácticos.

## Abstract

The certainty of the evidence for interventions is the certainty or confidence that the true effect is within a particular range or relative to a threshold. In the new pyramid of evidence, systematic reviews represent the magnifying glass through which this certainty is evaluated. The GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach arises in response to the existence of multiple evidence classification systems, and it offers a transparent and structured process to develop and present summaries of evidence considering its certainty and, in a second step, the strength of the recommendations that they inform. The GRADE process begins with an explicit question that includes all important and critical outcomes explicitly. The main domains used to assess the certainty of the evidence are risk of bias, inconsistency, indirectness of evidence, imprecision, and publication bias. The factors that can increase the certainty of the evidence are dose-response gradient, large magnitude of an effect,

and effect of plausible residual confounding. Finally, the Summary of Findings (Summary of Findings) tables summarize the process in a simplified way and with controlled language. This narrative review's purpose is to address the GRADE approach's theoretical and practical underlying concepts in a simplified way and with practical examples.

## Ideas clave

- La metodología GRADE establece criterios unificados y transparentes para graduar la certeza de la evidencia y la fuerza de las recomendaciones.
- El alcance de la metodología GRADE incluye la valoración de la certeza de la evidencia y la fortaleza de las recomendaciones sobre intervenciones, pruebas diagnósticas y modelos pronósticos.
- Algunas de las especificidades de GRADE para distintos tipos de evidencia están en continuo desarrollo.
- Este trabajo aborda los principales conceptos básicos teóricos y prácticos de la metodología GRADE para la valoración de la certeza de la evidencia de intervenciones en un lenguaje amigable y con la exposición de ejemplos, dirigido a la formación de estudiantes de pre y posgrado.

## Introducción

### Antecedentes de la metodología GRADE

El concepto de incorporar “evidencia” en la práctica clínica fue acuñado en Canadá por un grupo de médicos internistas y epidemiólogos clínicos de la Facultad de Medicina de la Universidad de McMaster, quienes consolidaron el concepto de medicina basada en la evidencia, dado que el incremento progresivo y abrumador de las publicaciones científicas no se asociaban con la certeza de estas. La medicina basada en la evidencia se sustenta en pilares (preferencias de los pacientes, pericia clínica y mejor evidencia disponible) los cuales contribuyeron al cambio de paradigma para estudiar y ejercer las ciencias médicas. Para simplificar el proceso de valorar cuál era la mejor evidencia disponible se necesitaba objetivar la “certeza” de esta.

La primera jerarquización de la evidencia fue formulada por la *Canadian Task Force on the Periodic Health Examination* en 1979, y fue adaptada cinco años más tarde por la *US Preventive Services Task Force* (USPSTF). Otra jerarquización ampliamente aceptada fue establecida en 1995 por el epidemiólogo David L. Sackett. Por lo tanto, en la actualidad coexisten diversas clasificaciones de la evidencia que se basan principalmente en el diseño de los estudios y habitualmente jerarquizan “niveles” de evidencia (I, IIa, IIb, III y demás). A su vez, existen varias clasificaciones de las recomendaciones según criterios predefinidos (grado A, B, C, etc.). El principal problema de estos sistemas es la falta de interoperabilidad, de manera tal que no siempre un “nivel I” o una recomendación “grado A” significa lo mismo para cada sistema<sup>1</sup>. Es por eso que surge GRADE, *Grading of Recommendations Assessment, Development and Evaluation*, como una definición unificada de la

certeza de la evidencia que puede servir de insumo para la formulación de recomendaciones<sup>2</sup>.

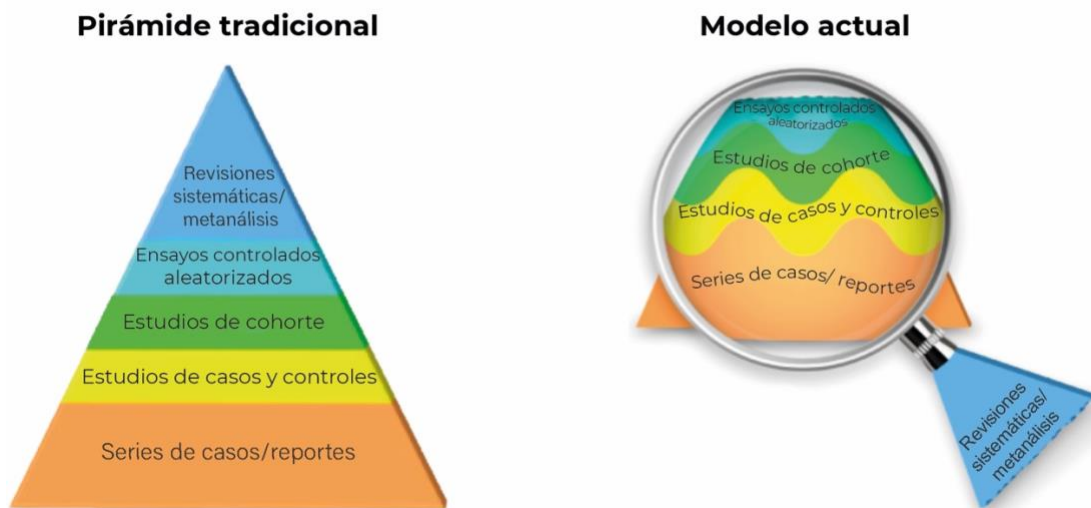
La certeza de la evidencia para las intervenciones es la seguridad o confianza de que el verdadero efecto se encuentra dentro de un rango particular o en relación con un umbral<sup>3,4</sup>. Esta definición tiene dos implicaciones contextuales importantes:

- 1) Las revisiones sistemáticas no deberían hacer recomendaciones, pues en estas la certeza de la evidencia refleja el grado de confianza en que una estimación del efecto es correcta.
- 2) Los paneles que desarrollan recomendaciones para guías de práctica clínica deben emitir juicios sobre la certeza de la evidencia relacionada al contexto específico en el que se está utilizando la evidencia<sup>3</sup>.

### GRADE en la nueva pirámide de la evidencia

El nivel de evidencia representado por la antigua o tradicional pirámide de la evidencia otorgaba un nivel estático para los distintos diseños, posicionando a las revisiones sistemáticas y metanálisis en el nivel más alto de la pirámide. El modelo actual entiende a las revisiones sistemáticas como una lupa a través del cual se ve y aplica un análisis de la evidencia, siendo la certeza de la evidencia variable y dinámica para distintos estudios con un mismo diseño, tal como lo muestra la Figura 1<sup>5</sup>. A modo de ejemplo, para responder una pregunta terapéutica, un ensayo clínico puede no ser siempre el mejor diseño. Si este incorpora pocos pacientes, con sesgos relevantes, puede generar estimaciones menos confiables que un estudio de cohortes con gran cantidad de pacientes y un adecuado análisis de sus variables. En este ejemplo, la lupa que nos permite diferenciar los estudios en su estructura interna sería la revisión sistemática.

Figura 1. Nueva pirámide de la evidencia.



Fuente: adaptada y traducida desde el artículo: *Proposed new evidence-based medicine pyramid*<sup>8</sup>.

En concordancia, el método GRADE surge como un marco para evaluar certeza de evidencia y para formular recomendaciones. Es así como para una determinada pregunta clínica terapéutica, un estudio observacional podría eventualmente tener mayor certeza de evidencia que un ensayo clínico aleatorizado o viceversa. Se debe tener en consideración que los estudios observacionales habitualmente se ubican por debajo de los ensayos clínicos, por debilidades intrínsecas de su diseño (por ejemplo, generalmente no establecen causalidad). Es importante también aclarar que este abordaje no elimina la necesidad inevitable de juicios con cierto grado de subjetividad, pero provee un marco para la sistematización y transparencia de los mismos<sup>6</sup>.

Este artículo corresponde al tercero de una serie metodológica de revisiones narrativas acerca de tópicos generales en bioestadística y epidemiología clínica, las que exploran y resumen en un lenguaje amigable, artículos publicados disponibles en las principales bases de datos y textos de consulta especializados. La serie está orientada a la formación de estudiantes de pre y posgrado. Es realizada por la Cátedra de Medicina Basada en Evidencia de la Escuela de Medicina de la Universidad de Valparaíso, Chile en colaboración con el Departamento de Investigación del Instituto Universitario del Hospital Italiano de Buenos Aires, Argentina y el Centro Evidencia UC, de la Universidad Católica, Chile. El objetivo de este manuscrito es abordar los principales conceptos básicos teóricos y prácticos de

la metodología GRADE para la valoración de la certeza de la evidencia de intervenciones.

## La metodología GRADE

### ¿Qué es la metodología GRADE y cuál es su alcance?

Esta iniciativa surgió como respuesta a la existencia de múltiples sistemas de clasificación de la evidencia (por ejemplo, CTFPHC<sup>7</sup>, USPSTF<sup>8</sup>, CEBM, AHA/ACC<sup>9</sup>, entre otros). A partir de un análisis exhaustivo de todos los sistemas de recomendaciones vigentes, la metodología GRADE ofrece un proceso transparente y estructurado para desarrollar y presentar síntesis de evidencia, incluida su certeza, para la confección de tablas resumen de hallazgos en revisiones sistemáticas y formulación de recomendaciones en la atención médica<sup>6</sup>. El alcance de la metodología GRADE incluye la valoración de la certeza de la evidencia y la fortaleza de las recomendaciones sobre intervenciones, pruebas diagnósticas y modelos pronósticos. Algunas de las especificidades de GRADE para distintos tipos de evidencia están en continuo desarrollo.

### ¿Qué significan las categorías otorgadas por GRADE?

Antes de evaluar los distintos pasos del proceso propuesto por GRADE, es conveniente saber qué significa cada categoría de certeza de evidencia<sup>10</sup>.

Tabla 1. Significado de los cuatro niveles de evidencia GRADE\*.

Nivel de certeza	Definición
<b>ALTO</b>	Estamos muy seguros de que el verdadero efecto se acerca al estimado.
<b>MODERADO</b>	Tenemos una confianza moderada en la estimación del efecto: el efecto verdadero es probable que esté cerca de la estimación del efecto, pero hay posibilidad de que sea sustancialmente diferente.
<b>BAJO</b>	Nuestra confianza en la estimación del efecto es limitada: el verdadero efecto puede ser sustancialmente diferente del estimado.
<b>MUY BAJO</b>	Tenemos muy poca confianza en la estimación del efecto: el efecto verdadero es probable que sea sustancialmente diferente al estimado.

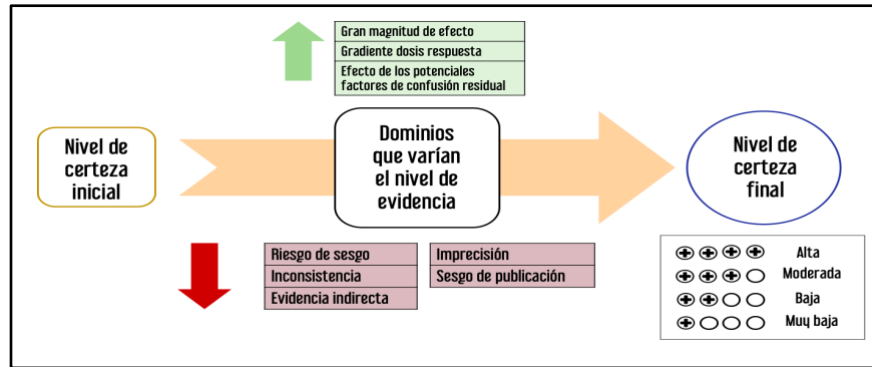
Fuente: adaptación del artículo de la serie GRADE: *GRADE guidelines: 3. Rating the quality of evidence*<sup>10</sup>

## Pasos propuestos por GRADE para evaluar la certeza de la evidencia

El proceso GRADE comienza con una pregunta explícita, que incluye las especificaciones de todos los desenlaces importantes y

críticos (tema abordado en una próxima entrega de esta serie). A partir de esta pregunta y según el diseño metodológico utilizado, se define una certeza inicial, que luego se analiza de acuerdo con distintos dominios, modificando la certeza de la evidencia, para llegar finalmente a una “certeza final”. La Figura 2 muestra un resumen de los distintos pasos de la metodología GRADE.

Figura 2. Resumen de la metodología GRADE.



Fuente: figura confeccionada por los autores.

Como se ha anticipado, el nivel de evidencia inicial otorgada por el diseño del estudio varía según distintos factores o dominios que lo influyen. Estos pueden disminuir la certeza de la evidencia o aumentarla<sup>10</sup>. Generalmente, los resultados derivados de ensayos clínicos comienzan con un alto nivel de certeza de evidencia inicial y aquellos derivados de los estudios observacionales con un bajo nivel de certeza, aunque esto excepcionalmente podría variar.

Dentro de los dominios que bajan la certeza, se encuentran:

1. El riesgo de sesgo.
2. La inconsistencia.
3. Evidencia indirecta.
4. Imprecisión.
5. El sesgo de publicación.

Estos se registran con valores (-1) o (-2) dependiendo si el resultado del análisis del dominio se considera “serio” o “muy serio”, respectivamente.

Los factores que aumentan la certeza de la evidencia son:

1. Gradiente dosis respuesta.
2. Gran magnitud del efecto.
3. Efecto de los potenciales factores de confusión residual.

### Ejemplos 1a y 1b. Niveles de riesgo de sesgo.

1a. Bajo riesgo de sesgo	1b. Alto riesgo de sesgo
<p>Una revisión sistemática sobre el uso de medias de compresión graduada para la prevención de la trombosis venosa profunda incluyó 20 ensayos controlados aleatorizados con 1681 participantes. En general, los estudios incluidos presentaron un bajo riesgo de sesgo al aplicar herramientas que lo evalúan. Por lo tanto, se mantuvo la evidencia de certeza alta sobre la afirmación de que las medias de compresión graduadas son efectivas para reducir el riesgo de trombosis venosa profunda en los pacientes hospitalizados sometidos a cirugía general y ortopédica, con o sin otros métodos de trombotoprofilaxis de base, cuando resulte clínicamente apropiado<sup>13</sup>.</p>	<p>Una revisión sistemática incluyó tres ensayos clínicos aleatorizados con el objetivo de comparar la cirugía versus el tratamiento conservador en el manejo de la hernia de disco lumbar. El beneficio de la cirugía en la reducción de los síntomas después de un año o más es incierto, puesto que se consideró que los ensayos tenían un inadecuado ocultamiento de la asignación y los desenlaces fueron evaluados por cirujanos utilizando una herramienta no validada de calificación, sin cegamiento. Estas características constituyen un alto riesgo de sesgo que justifica una disminución de dos niveles de evidencia: de alta a baja certeza de la evidencia<sup>14,15</sup>.</p>

2) Inconsistencia (heterogeneidad): la inconsistencia se refiere a la variación de las estimaciones del efecto (incluyendo el intervalo de confianza) habiendo excluido las principales causas de dicha variación<sup>16</sup>. A modo de ejemplo, el efecto de un medicamento puede variar de un estudio a otro debido a la variabilidad de los elementos de la pregunta PICO (pacientes, intervención, comparación, *outcome* o desenlace) en cada uno de los estudios. Sin embargo, si dicha variabilidad no justifica las diferencias en la estimación del efecto, la certeza de la evidencia disminuye.

Para evaluar la inconsistencia en la estimación del efecto, inicialmente se debe evaluar el grado de superposición entre los resultados de los estudios. Si existe solapamiento entre los intervalos de confianza de las estimaciones de cada estudio primario, se

establece que hay consistencia entre los estudios. Por el contrario, si no hay solapamiento en los intervalos de confianza, estos estudios son inconsistentes entre sí. Esto resulta más fácil de entender a través del análisis de un *Forest plot* (Ejemplo 2).

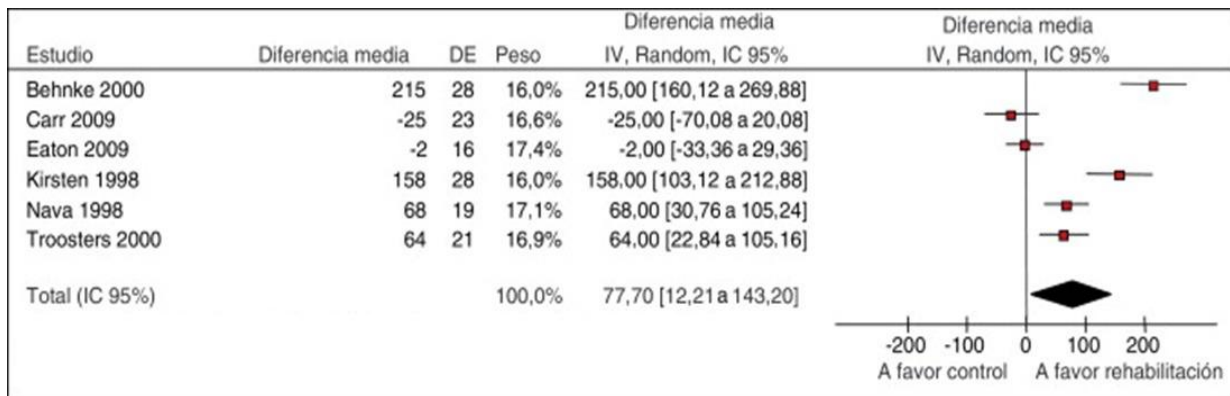
Además, es posible acudir a pruebas estadísticas como la prueba de heterogeneidad, en que un valor *p* bajo indica inconsistencia y el estadístico  $I^2$ , que cuantifica la proporción de la variación en las estimaciones puntuales debido a diferencias entre estudios: a mayor  $I^2$  mayor es la inconsistencia (< 40% es bajo, 30 a 60% moderado, 50 a 90% sustancial y 75 a 100% considerable)<sup>16</sup>.

Es importante aclarar que este dominio no tiene relación con la línea de no efecto.

**Ejemplo 2.** Evaluación de la inconsistencia.

Una guía de práctica clínica para el tratamiento de pacientes con enfermedad pulmonar obstructiva crónica, evaluó la eficacia de la rehabilitación respiratoria en el contexto de una exacerbación en términos de la tolerancia al ejercicio evaluada con la prueba caminata de 6 minutos. Una revisión sistemática que analizó seis estudios para esta prueba mostró unos resultados muy variables entre los estudios, con intervalos de confianza que no se solapan, y se objetivó heterogeneidad estadística ( $I^2 = 93\%$ ), como se puede ver en la Figura 3. En situaciones como esta, la confianza en los resultados disminuye al menos un nivel. Por lo tanto, la certeza de la evidencia catalogada como alta para una revisión sistemática de ensayos clínicos aleatorizados pasaría a ser moderada.

**Figura 3.** Ejemplo de Inconsistencia.



Metanálisis de la rehabilitación versus cuidado habitual en las exacerbaciones de la enfermedad pulmonar obstructiva crónica. Prueba de tolerancia al ejercicio a los 6 minutos.

Heterogeneidad:

TAU<sup>2</sup>: 6179,55

Chi-cuadrado: 71,60

df: 5 ( $p < 0,00001$ )

$I^2$ : 93%

Prueba de efecto global Z: 2,33 ( $p = 0,02$ )

IC: intervalo de confianza.

DE: desviación estándar.

Fuente: figura extraída y traducida desde el artículo *Pulmonary rehabilitation following exacerbations of chronic obstructive pulmonary disease*<sup>17</sup>.

3) Evidencia indirecta: la evidencia indirecta se refiere a los problemas emergentes en la concordancia entre los elementos de la pregunta pacientes, intervención, comparación, *outcome* o desenlace de la evidencia y la misma pregunta de la revisión sistemática o guía que usa dicha evidencia. La evidencia puede ser indirecta con respecto a la pregunta de interés cuando hay diferencias en los pacientes, las intervenciones o en los resultados<sup>18</sup>. Esto significa que

el resultado planteado en la pregunta no es respondido por el estudio, sino que en su reemplazo utiliza un resultado distinto, pero que indirectamente podría brindar información de lo establecido en nuestra pregunta.



### Ejemplo 3. Evaluación de evidencia indirecta - efecto del alendronato en fracturas patológicas.

#### Pregunta clínica

Una revisión sistemática evalúa el efecto del alendronato en prevención de fracturas patológicas a 10 años en mujeres con osteoporosis. Si un estudio incluido toma como desenlace principal la mejoría en la densitometría ósea en lugar de alteraciones radiográficas características de las fracturas, no está midiendo directamente el efecto en la presencia o ausencia de fracturas, sino que utiliza un resultado intermedio (subrogado) que predice la incidencia de fracturas (desenlace indirecto). A su vez, si un profesional discute las opciones del tratamiento de la osteoporosis con una mujer de 88 años considerando la evidencia de ensayos clínicos en los cuales la media de edad era de 61 años, puede considerar que está extrapolando los resultados (población indirecta). En ambos casos, asumiendo que son ensayos clínicos aleatorizados, pasarían de un alto nivel de certeza a un moderado nivel de esta.

4) Imprecisión<sup>19</sup>: se establece que los resultados son imprecisos cuando los estudios incluyen relativamente pocos pacientes y pocos eventos; consecuentemente tienen intervalos de confianza amplios alrededor del efecto, lo cual no permite estimar claramente la magnitud de este.

La precisión se puede evaluar de dos maneras:

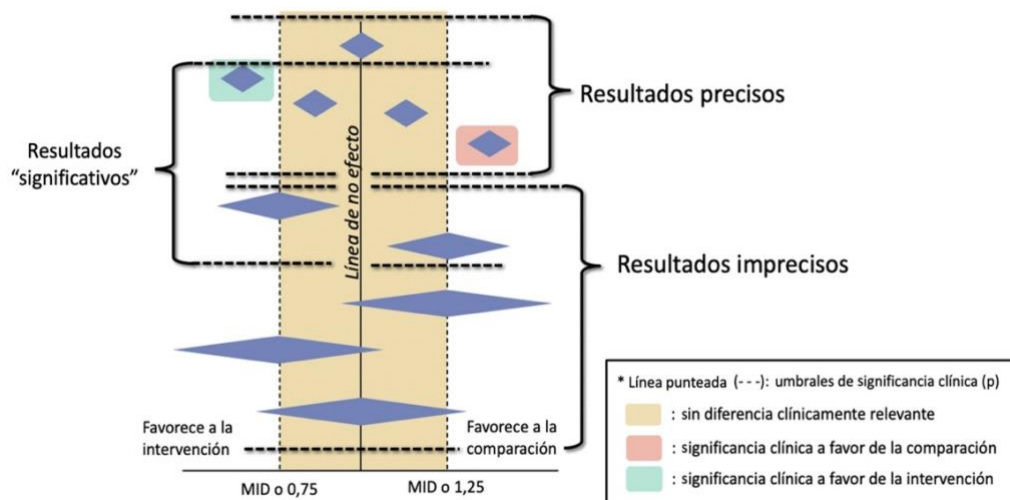
- Examinando los extremos del intervalo de confianza de la estimación del efecto en relación con un umbral.
- Evaluando la cantidad de datos disponibles con relación a un tamaño óptimo de información (*optimal information size*).

Para el primer abordaje es necesario definir qué umbral se utiliza para estimar que un efecto es clínicamente relevante y su magnitud (efecto pequeño o sustancial). Habitualmente se puede tomar como

referencia la diferencia mínima importante, por ejemplo, para los desenlaces continuos (Figura 4). Para el segundo abordaje es necesario realizar un cálculo de tamaño muestral de un ensayo clínico con poder adecuado para el desenlace que se evalúa. Si el intervalo de confianza cruza el umbral establecido o la cantidad de datos no alcanza el tamaño óptimo de información, se considerará bajar un nivel la certeza.

Gráficamente, lo que debe analizarse es el intervalo de confianza del metanálisis (no de los estudios individuales) representado por las figuras de diamantes en el *Forest plot* que se muestra como ejemplo en la Figura 4.

Figura 4. Figura explicativa del concepto de imprecisión en relación con la significancia clínica.



MID: diferencia mínima importante.  
Fuente: figura confeccionada por los autores.

La diferencia mínima importante (MID por su sigla en inglés, *Minimal Important Difference*) corresponde a la menor diferencia de puntuación en el desenlace de interés que los pacientes consideran importante, y que podría indicar un cambio en el tratamiento<sup>11</sup>. Este concepto, que será profundizado en un futuro artículo de la serie, relaciona la magnitud del cambio con las decisiones de tratamiento en la práctica clínica, enfatizando que prima la percepción del paciente por sobre la significancia estadística. Este concepto es

aplicable a desenlaces con valores continuos como aquellos que utilizan escalas (por ejemplo, la escala visual analógica para determinar la intensidad del dolor), mientras que en un desenlace dicotómico (como la mortalidad) la estimación de efecto es mediante los riesgos relativos, *Odds ratio* o *Hazard ratio*. Para los desenlaces dicotómicos no hay umbrales preestablecidos como la diferencia mínima importante, sin embargo, se toman como valor de referencia orientativos los riesgos relativos de 0,75 y 1,25 para evaluar la

precisión de los resultados. Estos valores umbrales relativos se traducen en diferencia de riesgos absolutas, que son las más importantes para la toma de decisiones. Por ejemplo, un riesgo relativo de 0,75 podría implicar una reducción del riesgo absoluto de mortalidad de 40 a un 30% (10% de diferencia de mortalidad). Si el

evento es más raro, un riesgo relativo de 0,75 puede ser menos relevante. Por ejemplo, podría implicar una reducción en la mortalidad específica por cáncer de 4 por 1000 a 3 por 1000 (1 por mil de diferencia de mortalidad).

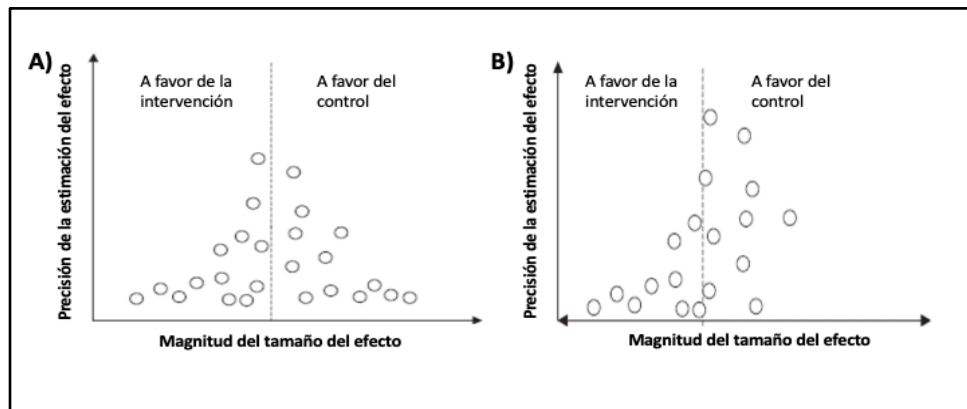
**Ejemplos 4a y 4b.** Precisiones de valor continuo y desenlace dicotómico.

<p><b>4a. Valor continuo impreciso.</b></p> <p>En el caso del hombro doloroso, la escala más utilizada es la <i>Constant Murley Functional Assessment</i> graduada del 0 al 100. En esta escala, una diferencia mínima importante corresponde a una variación de 10,4 puntos<sup>20</sup>. Por ejemplo, si en un estudio se determina que luego de una cirugía determinada, el dolor de los pacientes disminuye 11 puntos (intervalo de confianza del 95%; de -9 a -12) en esta escala, se podría considerar como una intervención clínicamente significativa. Sin embargo, al examinar los extremos del intervalo de confianza, se observa imprecisión al incluir un escenario en el cual esta diferencia puede ser clínicamente insignificante (porcentaje de pacientes que disminuirían su dolor entre 9 y 10,3).</p>	<p><b>4b. Desenlace dicotómico preciso.</b></p> <p>En el síndrome de parto prematuro, el riesgo absoluto de mortalidad intrahospitalaria en pacientes recién nacidos entre 24 y 31 semanas es alrededor del 20%. La administración de corticoides antenatales demostró un riesgo relativo de 0,5 (intervalo de confianza del 95%; de 0,4 a 0,6)<sup>21</sup> en mortalidad intrahospitalaria respecto a la no administración. Por su intervalo de confianza, se sabe que la intervención se encuentra por debajo del umbral (0,75 usado como valor referencial). Por lo tanto, es una intervención que muestra un efecto sustancial en todo el rango de efectos posibles. Esto se traduce también como una reducción del riesgo absoluto de mortalidad neonatal entre 8 y 12%.</p>
--	--

5) Sesgo de publicación<sup>22</sup>: es posible que existan estudios cuyos resultados se opongan a los resultados del metanálisis y que estos no hayan sido incluidos, ya sea porque no fueron publicados o porque se encuentran en revistas no indexadas. Típicamente, estos estudios suelen tener resultados “negativos” (es decir, que no muestran un efecto significativo a favor de la intervención). En ese caso, se habla de sesgo de publicación. No obstante, este sesgo es difícil de detectar, en consecuencia, resulta difícil bajar el nivel de evidencia por esta causa y si se hace será un peldaño (-1).

La forma de analizarlo es con un gráfico llamado *Funnel Plot* (Figura 5), que representa cómo se distribuyen los estudios en relación al efecto real del estudio; cuando tiene forma de embudo no habría sesgo. Si existe este sesgo, veremos un espacio entre los estudios dando esta imagen de pirámide truncada, como ocurre en la Figura 5, con estudios que contradicen la intervención.

**Figura 5.** Ejemplo de *Funnel plot*.



A) Distribución piramidal uniforme, sin sesgo de publicación.

B) Distribución de pirámide truncada, con sesgo de publicación

Fuente: figura extraída y traducida desde *GRADE guidelines: 5. Rating the quality of evidence-publication bias*<sup>22</sup>.

## B. Dominios que aumentan el nivel de evidencia

Si bien el mayor protagonismo se lo llevan los factores que disminuyen el nivel de certeza, es importante saber que existen factores que pueden aumentarlo. Para ello se requieren dos condiciones: que el estudio sea observacional y que no presente ninguna de las limitaciones (sesgos) previamente comentadas. Existen tres motivos para aumentar en un punto su nivel de certeza:

**1) Gran magnitud de efecto:** si el cuerpo de la evidencia arroja estimados grandes o muy grandes de la magnitud del efecto de una intervención, es posible tener más confianza acerca de los resultados. En tales ocasiones, a pesar de que los estudios observacionales tienen mayor probabilidad de sobreestimar el efecto verdadero, el diseño del estudio que es más propenso al sesgo es poco probable que explique la totalidad del beneficio o riesgo aparente. Por ejemplo, un metanálisis de estudios observacionales mostró que los

cascos de bicicleta reducen el riesgo de lesiones en la cabeza en ciclistas por un amplio margen (razón de probabilidades *Odds ratio* 0,31; intervalo de confianza del 95%: 0,26 a 0,37). Este gran efecto, en ausencia de un sesgo obvio que pudiera crear la asociación, sugiere una calificación de evidencia de calidad moderada en vez de baja<sup>23</sup>.

**2) Gradiente dosis respuesta:** la existencia de un efecto dosis-respuesta también aumenta el nivel de certeza. Por ejemplo, en usuarios de anticoagulantes hay un gradiente dosis respuesta entre niveles más altos de INR (*International Normalized Ratio*, indicador del grado de anticoagulación) y el riesgo aumentado de sangrado. Esto aumentaría nuestra confianza en cuanto a qué niveles supra terapéuticos de anticoagulación podrían aumentar el riesgo de sangrado.

**3) Efecto de los potenciales factores de confusión residual:** Al momento de evaluar la evidencia, se considera la posibilidad de que existan factores de confusión que reduzcan o incrementen un determinado efecto. Cuando el resultado obtenido no se ve afectado por estos factores, es posible aumentar la certeza en un nivel. Por

ejemplo, si en un análisis se considera que las personas que usan condón es probable que tengan más parejas sexuales, sería esperable que en esta población el riesgo de infección por virus de la inmunodeficiencia humana (VIH) sea mayor. Sin embargo, este factor no redujo el efecto demostrado del preservativo, por lo que constituye una medida tan efectiva para la prevención de la transmisión del virus de la inmunodeficiencia humana que supera a este factor de confusión<sup>24</sup>.

## Tablas resumen de hallazgos o *Summary of Findings*

Todo este proceso planteado por GRADE se resume de manera bastante sencilla en las tablas resumen de hallazgos (*Summary of Findings Tables*) o tablas SoF por su sigla en inglés. Estas tablas muestran todo el proceso por cada desenlace, con efecto relativo, absoluto, certeza de evidencia (con su justificación), y el mensaje clave en lenguaje controlado.

**Figura 6.** Tabla resumen de hallazgos acerca de tratamiento con sialogogos y tratamiento no farmacológico para xerostomía.

TRATAMIENTO FARMACOLÓGICO CON SIALOGOGO MÁS TRATAMIENTO NO FARMACOLÓGICO COMPARADO CON SÓLO TRATAMIENTO NO FARMACOLÓGICO PARA XEROSTOMÍA						
Componentes de la pregunta PICO		Pacientes	Personas con xerostomía.			
Intervención		Tratamiento farmacológico con sialogogo más tratamiento no farmacológico.				
Comparación		Solo tratamiento no farmacológico.				
Desenlaces	Efecto relativo (IC 95%)	Efecto absoluto estimado*			Certeza de la evidencia (GRADE)	Mensajes clave en términos sencillos
	Estudios/pacientes	SIN sialogogo	CON sialogogo	Diferencia (IC 95%)		
Reducción de sensación de boca seca (mayor o igual a 25 mm en escala VAS)**	RR 1,95 (1,62 a 2,36) -- 4 ensayos/ 616 pacientes [4, 9, 10, 12]	315 por 1000	614 por 1000	Diferencia: 299 pacientes más (199 a 460 más)	⊕⊕⊕⊕ Alta	El uso de sialogogos en pacientes con xerostomía reduce la sensación de boca seca.
Flujo salival no estimulado (ml/min)	-- 2 ensayos/ 298 pacientes [9, 12]	Aumento de 0,06 ml/min	Aumento de 0,2 ml/min	DM: 0,14 ml/min más (0,01 a 0,27 más)	⊕⊕⊕○ <sup>1</sup> Moderada	El uso de sialogogos en pacientes con xerostomía probablemente aumenta el flujo salival.
Efectos adversos	La tasa de abandono por efectos adversos fue de un 30% comparado con un 16% en el grupo placebo [2]. Los efectos adversos más frecuentes fueron sudoración (60% en dosis mayores y 10% en dosis menores, versus 7% con placebo) y alteración de frecuencia miccional (10-15% vs 1-6%) [9, 12]***				⊕⊕⊕○ <sup>1</sup> Moderada	El uso de sialogogos en pacientes con xerostomía probablemente aumenta los efectos adversos.

PICO: pacientes, intervención, comparación, *outcome* o desenlace.

IC 95%: intervalo de confianza al 95%.

RR: riesgo relativo

DM: diferencia de media.

GRADE: *Grading of Recommendations Assessment, Development and Evaluation*.

\*Riesgo SIN tratamiento farmacológico está basado en el riesgo del grupo control en los estudios. El riesgo CON tratamiento farmacológico (y su intervalo de confianza) está calculado a partir del efecto relativo (y su intervalo de confianza).

\*\*El cambio general promedio de los síntomas subjetivos de boca seca(xerostomía) fueron evaluados mediante un cambio en una escala analógica visual (VAS) de 0 a 100 milímetros, o se valoró subjetivamente utilizando un resultado categórico de mejora en comparación con el valor inicial.

\*\*\*Efectos adversos observados al utilizar pilocarpina 5 miligramos 4 veces al día por 12 semanas.

<sup>1</sup>Se disminuyó un nivel de certeza por riesgo de sesgo, ya que la mayoría de los estudios presentan alto riesgo de sesgo de selección.

Fecha de elaboración de la Tabla: noviembre de 2018.

Fuente: adaptación esquemática, extraída del Ministerio de Salud Chileno en su idioma original<sup>25</sup>.



**Ejemplo 5.** Explicación tabla resumen de hallazgos para tratamiento con sialogogos más tratamiento no farmacológico comparado con solo tratamiento no farmacológico para xerostomía.

En este ejemplo se menciona el tratamiento con sialogogos asociado a manejo no farmacológico. En la parte superior de la tabla, se encuentran los tres primeros componentes de la pregunta pacientes, intervención, comparación, *outcome* o desenlace, y los desenlaces se encuentran a continuación desglosados en la tabla, el primero de estos es:

Reducción de la sensación de boca seca: en este caso, el riesgo relativo es de 1,95 (1,62 a 2,36). Esto significa que hay “más riesgo” de que se reduzca la sensación de boca seca (lo que es bueno). También se muestra la cantidad de ensayos clínicos y el número de pacientes; y por separado el grupo control e intervención y la diferencia entre estos con su respectivo intervalo de confianza del 95% (en este caso 299 pacientes más por cada 1000 al manejo con sialogogos, con una dispersión desde 199 hasta 460 pacientes por cada 1000).

#### Aclaraciones

En cuanto a la certeza de evidencia, esta es alta, según lo expresado en la tabla por medio de figuras de cruces (+ muy baja; ++ baja; +++ moderada; y ++++ alta).

En el caso del segundo y tercer desenlace (flujo salival no estimulado y efectos adversos), la certeza es moderada. Mediante un superíndice (1) se explicitan los motivos de esta certeza referenciados al pie de la tabla. En este caso se argumentó el riesgo de sesgo, dado que la mayoría de los estudios presentaba sesgo de selección, disminuyendo en un nivel la certeza.

## Lenguaje controlado

Una vez que se obtuvo la certeza de evidencia, esta debe comunicarse mediante un lenguaje controlado, que consiste en transmitir en palabras simples todo el proceso anteriormente realizado. En la tabla resumen de hallazgo se explicita en la casilla de la derecha titulada “mensaje clave en términos sencillos”, en donde, frases como “probablemente aumenta el flujo salival/efectos adversos” reflejan certeza moderada; y afirmaciones como “reduce la sensación de boca seca” reflejan certeza alta. Cabe señalar que no existe una traducción oficial al español, por lo que se trata de una propuesta propia de adaptación.

## Conclusiones

Actualmente la metodología GRADE es uno de los principales métodos utilizados internacionalmente en el desarrollo de guías y recomendaciones clínicas, ya que provee un marco claro y transparente para valorar y presentar de la certeza de la evidencia. Si bien el grupo GRADE aborda de manera acabada cada uno de los conceptos en su extensa serie sobre esta metodología, al inicio su entendimiento y aplicación puede resultar abrumadora para los principiantes.

Por medio de descripciones sencillas y acompañadas de ejemplos sobre los elementos fundamentales de la metodología GRADE, este artículo busca entregar un acercamiento simplificado y resumido para su entendimiento y aplicación por el público general, estudiantes y los nóveles metodólogos.

## Notas

### Roles de contribución

MK: conceptualización, metodología, investigación, redacción borrador original, revisión y edición, visualización, administración de proyecto. CQ, BV, JVAF: conceptualización, metodología, investigación, redacción borrador original, revisión y edición, visualización. CL, LG: conceptualización, metodología, revisión y edición, visualización.

### Conflictos de intereses

Los autores completaron la declaración de conflictos de interés de ICMJE y declararon que no recibieron fondos por la realización de este artículo; no tienen relaciones financieras con organizaciones que puedan tener interés en el artículo publicado en los últimos tres años y no tienen otras relaciones o actividades que puedan influenciar en la publicación del artículo. Los

formularios se pueden solicitar contactando al autor responsable o al Comité Editorial de la Revista.

### Financiamiento

Los autores declaran que no poseen fuentes de financiamiento externas asociadas a la realización de este artículo.

### Agradecimientos

Agradecemos a la Cátedra de Medicina Basada en Evidencia de la Escuela de Medicina de la Universidad de Valparaíso, Chile por impulsar esta serie y la colaboración del Departamento de Investigación del Instituto Universitario del Hospital Italiano de Buenos Aires, Argentina.

## Referencias

1. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. *Plast Reconstr Surg.* 2011 Jul;128(1):305-310. | CrossRef | PubMed |
2. Franco JVA, Arancibia M, Meza N, Madrid E, Kopitowski K. Clinical practice guidelines: Concepts, limitations and challenges. *Medwave.* 2020 Apr 30;20(3):e7887. | CrossRef | PubMed |
3. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ; GRADE Working Group. What is "quality of evidence" and why is it important to clinicians? *BMJ.* 2008 May 3;336(7651):995-8. | CrossRef | PubMed |
4. Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, et al. The GRADE Working Group clarifies the construct of certainty of evidence. *J Clin Epidemiol.* 2017 Jul;87:4-13. | CrossRef | PubMed |
5. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. *Evid Based Med.* 2016 Aug;21(4):125-7. | CrossRef | PubMed |
6. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol.* 2011 Apr;64(4):383-94. | CrossRef | PubMed |
7. Birtwhistle R, Pottie K, Shaw E, Dickinson JA, Brauer P, Fortin M, et al. Canadian Task Force on Preventive Health Care: we're back! *Can Fam Physician.* 2012 Jan;58(1):13-5. | PubMed |
8. Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, et al. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med.* 2001 Apr;20(3 Suppl):21-35. | CrossRef | PubMed |
9. Correction to: 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation.* 2019 Sep 10;140(11):e649-e650. | CrossRef | PubMed |

10. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol*. 2011 Apr;64(4):401-6. | CrossRef | PubMed |
11. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence-- study limitations (risk of bias). *J Clin Epidemiol*. 2011 Apr;64(4):407-15. | CrossRef | PubMed |
12. Higgins JPT, Savović J, Page MJ, Elbers RG, Sterne JAC. Assessing risk of bias in a randomized trial. *Cochrane Handbook for Systematic Reviews of Interventions*. 2019. 205–28. [Internet] | CrossRef |
13. Sachdeva A, Dalton M, Lees T. Graduated compression stockings for prevention of deep vein thrombosis. *Cochrane Database Syst Rev*. 2018 Nov 3;11(11):CD001484. | CrossRef | PubMed |
14. Gibson JN, Waddell G. Surgical interventions for lumbar disc prolapse. *Cochrane Database Syst Rev*. 2007 Jan 24;(1):CD001350. | CrossRef | PubMed |
15. Higgins JPT. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1*. The Cochrane Collaboration. 2008 [cited 2020 Sep 11]. [Internet]. | Link |
16. 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 Dec;64(12):1294-302. | CrossRef | PubMed |
17. Puhan MA, Gimeno-Santos E, Scharplatz M, Troosters T, Walters EH, Steurer J. Pulmonary rehabilitation following exacerbations of chronic obstructive pulmonary disease. *Cochrane Database Syst Rev*. 2011 Oct 5;(10):CD005305. | CrossRef | PubMed |
18. 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 Dec;64(12):1303-10. | CrossRef | PubMed |
19. Guyatt GH, 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 Dec;64(12):1283-93. | CrossRef | PubMed |
20. Donado-Gómez JH, Jiménez-Cotes EA, Cardona LM. Diferencia mínima clínicamente importante en algunas escalas de medición utilizadas en enfermedades musculoesqueléticas. *Revista Colombiana de Reumatología*. 2014 Sep 1;21(3):125–32.
21. Norman M, Piedvache A, Børch K, Huusom LD, Bonamy AE, Howell EA, et al. Association of Short Antenatal Corticosteroid Administration-to-Birth Intervals With Survival and Morbidity Among Very Preterm Infants: Results From the EPICE Cohort. *JAMA Pediatr*. 2017 Jul 1;171(7):678-686. | CrossRef | PubMed |
22. 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 Dec;64(12):1277-82. | CrossRef | PubMed |
23. Schünemann H. *The GRADE handbook*. Cochrane Collaboration. 2013.
24. Cáceres CF. Intervenciones para la prevención del VIH e ITS en América Latina y Caribe: una revisión de la experiencia regional. *Cadernos de Saúde Pública*. 2004;20(6):1468–85.
25. Ministerio de Salud de Chile. Informe de búsqueda y síntesis de evidencia de efectos deseables e indeseables. *Guía de Práctica Clínica Salud Oral Integral: Tratamiento y Rehabilitación Oral en Personas Adultas y Personas Mayores con Edentulismo Parcial o Total*. 2019. [Internet]. | Link |

**Correspondencia a**  
Potosí 4265, Buenos Aires,  
Argentina



Esta obra de *Medwave* está bajo una licencia Creative Commons Atribución-No Comercial 3.0 Unported. Esta licencia permite el uso, distribución y reproducción del artículo en cualquier medio, siempre y cuando se otorgue el crédito correspondiente al autor del artículo y al medio en que se publica, en este caso, *Medwave*.