

Metodología de investigación y lectura crítica de estudios

Metodología de calificación y resumen de las opiniones dentro de consensos formales

Ricardo Sánchez Pedraza¹
Luis Eduardo Jaramillo González²

Resumen

Actualmente el desarrollo y uso de guías de práctica clínica es un aspecto importante desde el punto de vista político y desde del cuidado de la salud. Los métodos de consenso se usan para desarrollar guías de práctica clínica, especialmente cuando la evidencia es insuficiente o contradictoria. Debido a sus características (ser un proceso sistemático y transparente), los consensos formales se prefieren sobre los no formales al desarrollar guías de práctica clínica. Este artículo describe algunos métodos utilizados en el proceso de calificación y resumen de los resultados de un consenso formal.

Palabras clave: guía de práctica clínica, consenso.

Title: Methodology of Qualification and Summary of the Opinions in Formal Consensus

Abstract

Currently, the development and use of clinical guidelines is an important political and healthcare issue. Consensus methods are used to develop clinical guidelines, mainly when there is insufficient or contradictory evidence. Formal rather than informal consensus methods are preferred in guideline development, because of its characteristics (being a systematic and transparent process). This paper describes some methods used to scoring and summarizing results of a formal consensus.

Key words: Practice guidelines, consensus.

¹ Psiquiatra, MSc en Epidemiología Clínica. Profesor titular de la Facultad de Medicina de la Universidad Nacional de Colombia. Grupo de Investigación Clínica del Instituto Nacional de Cancerología. Bogotá, Colombia.

² Psiquiatra, MSc en Farmacología. Profesor asociado de la Facultad de Medicina de la Universidad Nacional de Colombia. Bogotá, Colombia.

Introducción

El consenso es una metodología extensamente utilizada para la elaboración de guías de práctica clínica. Debe tenerse en cuenta que en dicho proceso pueden darse los siguientes escenarios (1-4):

1. No hay suficiente evidencia clínica para generar las recomendaciones de la guía.
2. Existe evidencia clínica contradictoria.
3. La evidencia clínica es robusta, pero la aplicación en ámbitos locales genera discrepancias entre los miembros del grupo que elabora la guía.
4. Existe evidencia no publicada que poseen algunos expertos, la cual no puede ser detectada por una revisión sistemática, y que puede ser necesaria para generar las recomendaciones de la guía.

Los métodos de consenso se pueden agrupar en dos grandes categorías: formales e informales. Los primeros están caracterizados por los siguientes elementos (5-8):

1. Existe como insumo una revisión sistemática sobre los tópicos de discusión.
2. El proceso de discusión, calificación de opiniones y síntesis de estas es sistemático; esto es, sigue un proceso estructurado que asegura su replicabilidad y transparencia.

Los anteriores elementos permiten que en un consenso formal se puedan integrar las opiniones de los expertos con la evidencia científica en el tema, utilizando técnicas grupales basadas en retroalimentación e iteración controladas, junto con métodos de calificación de las opiniones que permitan reducir los sesgos (9,10).

En el presente artículo se describirán algunos de estos métodos de calificación cuantitativa de las opiniones de los integrantes del grupo, así como algunas estrategias empleadas para la síntesis y presentación de los resultados. Dichas estrategias se aplican a cualquiera de los siguientes tres métodos recomendados en la elaboración de guías de práctica clínica (11,12):

1. Técnica Delphi (13-16): es una técnica que en su forma original plantea enviar cuestionarios por correo a un grupo de expertos, para que califiquen una serie de recomendaciones. Estas calificaciones se resumen y reenvían a los expertos, de manera iterativa, para evaluar la posibilidad de que haya cambios en la calificación que permitan llegar a un mayor grado de acuerdo. Existen modificaciones a este método (Delphi cara a cara, Delphi telefónico). El número recomendado de expertos oscila entre 6 y 30, aunque pueden manejarse grupos más grandes.

2. Grupo nominal (7): en esta técnica, bajo la coordinación de un facilitador, un grupo de 9 a 12 expertos se reúnen para analizar y calificar una serie de preguntas. Esta calificación se realiza de manera iterativa en dos rondas, buscando reducir la variabilidad de las opiniones, con el fin de alcanzar un acuerdo. Este tipo de técnica se ha utilizado ampliamente para generar recomendaciones en salud. La ventaja de esta técnica sobre la anterior es que permite interactuar a los expertos, pero anula, a la vez, el efecto de influencias de personajes con estilos dominantes sobre la opinión de otros.
3. Conferencia de consenso (11,17): es una técnica que permite una participación amplia de interesados. Un grupo de expertos efectúa una presentación sobre los tópicos de las preguntas ante un jurado y un público. El público puede participar haciendo preguntas y comentarios a los expertos. Tanto el contenido de las presentaciones como el resultado de la discusión con el público se utilizan a manera de insumo para que el jurado genere las recomendaciones.

Formatos para calificación

Los formatos utilizados para este fin están diseñados buscando obtener una calificación cuantitativa de una opinión. A partir de la propuesta del método de conveniencia RAND/UCLA, desarrollada en los años ochenta (18), se ha convenido en utilizar escalas ordinales con nueve niveles para calificar una recomendación. De acuerdo con estos niveles, un procedimiento en salud puede ser recomendado (apropiado), contraindicado (inapropiado) o estar en un nivel de incertidumbre, dependiendo del valor de la mediana (19). Adicionalmente, según el grado de dispersión de las calificaciones, puede calificarse sobre el nivel de acuerdo o consenso. Los formatos de calificación tienen la estructura que se presenta en la Figura 1 (18,19).

Los formatos para calificación presentan tres componentes:

1. Zonas de calificación: son espacios destinados para que se marque la calificación escogida. En la Figura 1, tales espacios corresponden a los círculos en la parte superior. No utilizar estas zonas puede generar con-

Figura 1. Escala para calificación en un consenso formal



fusiones, pues los calificadores pueden señalar opiniones en los puntos intermedios de la escala (cabe recordar que esta es una escala ordinal, en la que los valores son discretos y no admiten puntuaciones intermedias).

2. Zona de escala: muestra numéricamente los valores en los cuales se deberá efectuar la calificación (de los números 1 a 9). La zona de escala insinúa que hay tres regiones (de los números 1 a 3, de 4 a 6 y de 7 a 9), que corresponderán a la calificación del grado de recomendación.
3. Zona de anclaje: corresponde a los textos, debajo de los valores de la zona de escala, que relacionan algunos valores numéricos con un concepto verbal. Además del sistema usado en la Figura 1 (completamente en desacuerdo, en desacuerdo, etc.), y dependiendo de las características del enunciado que se califica, pueden usarse sistemas tales como (16,20):
 - a. Completamente inapropiado; parcialmente inapropiado; ni inapropiado ni apropiado; parcialmente apropiado; completamente apropiado.
 - b. Completamente contraindicado; parcialmente contraindicado; ni contraindicado ni indicado; parcialmente indicado; completamente indicado.

- c. Para nada; muy poco; algo; mucho; demasiado.
- d. Completamente en desacuerdo; en desacuerdo; algo en desacuerdo; algo de acuerdo; de acuerdo; completamente de acuerdo.
- e. Nunca está indicado; muy pocas veces está indicado; algunas veces está indicado; muchas veces está indicado; siempre está indicado.

Métodos de calificación

Dada la naturaleza ordinal del sistema de calificación, la medida de resumen utilizada en estos casos es la mediana (percentil 50) (18). Se pueden distinguir dos métodos de calificación:

1. Método descriptivo: tiene las siguientes fases:
 - a. Una vez se han recopilado todas las calificaciones de los expertos, se calcula la mediana de estas calificaciones y se ubican los puntos extremos del rango de respuestas (el puntaje más alto y el puntaje más bajo).
 - b. Si los puntos extremos del rango de respuestas se ubican dentro de alguna de las tres regiones de la zona de escala (1 a 3; 4 a 6; 7 a 9), se considera que hay acuerdo fuerte y se declara consenso. Si los puntos extremos

del rango se ubican dentro de dos regiones consecutivas (1 a 3 y 4 a 6, por ejemplo), se considera que hay acuerdo relativo. Si los puntos extremos del rango están dispersos entre dos regiones no consecutivas (1 a 3 y 7 a 9) se considera que no hubo consenso. En los casos en los que no hubo consenso o hubo acuerdo relativo se lanza una segunda ronda de calificación, durante la cual se utilizan preguntas con la estructura que se presenta en la Figura 2.

Según los resultados anteriores, puede requerirse una tercera ronda, en la cual se utilizará la misma metodología, y, adicionalmente, se enviará a los expertos el texto de los comentarios de

- quienes no estuvieron de acuerdo con la medida de resumen.
- c. Dependiendo del valor de la mediana, puede calificarse el grado de la recomendación, de la siguiente manera (5,18):
 - i. Si la mediana se ubica en la zona de la escala 7 a 9 se considera que la recomendación es adecuada, indicada o de primera línea.
 - ii. Cuando la mediana se ubica en la zona de la escala 1 a 3 se toma la recomendación como no indicada o no recomendada.
 - iii. Si la mediana se encuentra en la zona 4 a 6 no es posible plantear afirmaciones sobre

Figura 2. Ilustración de una segunda ronda en un consenso formal

¿Esta Ud. de acuerdo con el valor medio (mediana) obtenido por el conjunto de los expertos consultados?

Sí No

En caso de que no esté de acuerdo, ¿cuál es el nuevo valor que propone?

Completamente en desacuerdo En desacuerdo Ni de acuerdo ni en desacuerdo De acuerdo Completamente de acuerdo

En caso de que no esté de acuerdo y proponga un nuevo valor, ¿podría justificar su respuesta?

qué tan adecuada o indicada es la recomendación. En estos casos, así haya consenso, se recomienda lanzar una nueva ronda de consulta a los expertos, buscando un pronunciamiento sobre la indicación o contraindicación de la recomendación. En caso de no lograr un acuerdo en esta nueva ronda, se deben documentar los motivos presentados por los expertos que justificaron este punto de indecisión.

2. Método estadístico: se utiliza cuando el grupo de expertos tiene un tamaño de 20 o más personas. Se fundamenta en pruebas estadísticas que hacen comparaciones múltiples de medianas. Sus etapas son las siguientes (21):

a. Cuando se recolectan todas las calificaciones de los expertos, se calcula la mediana y se aplica un método estadístico de comparación. La prueba que se suele utilizar es el test de Kruskal-Wallis, con opción de comparaciones *post-hoc* cuando hay más de dos medianas; dependiendo del tamaño del grupo de expertos, puede considerarse la opción de usar niveles de significa-

ción menos exigentes; incluso, menores de 0,1. Por ejemplo, si se plantea un consenso sobre el manejo recomendado en pacientes con depresión mayor leve y se plantean como opciones *psicoterapia-inhibidores de recaptación de serotonina-antidepresivos tricíclicos*, cada una de estas tres opciones tendrá, luego de la calificación de los expertos, una mediana.

b. Aplicar criterios de decisión para definir consenso: en el caso de que se comparen más de dos opciones (y, por lo tanto, más de dos medianas) se tiene en cuenta como criterio para declarar consenso que exista una diferencia significativa, y que esta sea atribuible a alguna de las dos opciones con la mayor mediana. Tal situación puede darse en los siguientes casos:

i. La mediana mayor es diferente de todas las otras medianas.

ii. Todas las medianas son diferentes entre ellas.

iii. Las dos medianas más grandes son diferentes de la menor, pero entre las más grandes no hay diferencia. En este caso se declara consenso, pero cuando se redacte el documento final debe

aclararse que no hubo diferencia significativa entre las recomendaciones de primera y segunda línea.

En caso de que no haya diferencia significativa entre ningún par de medianas se considera que no hubo consenso.

Si la opción que se está calificando tiene sólo dos observaciones se declara consenso únicamente si hay diferencia significativa entre las medianas. Obviamente, en esta situación no son necesarias las pruebas *post-hoc*.

Adicionalmente, deben calcularse los intervalos de confianza para cada una de las medianas que se van a reportar.

Síntesis y presentación

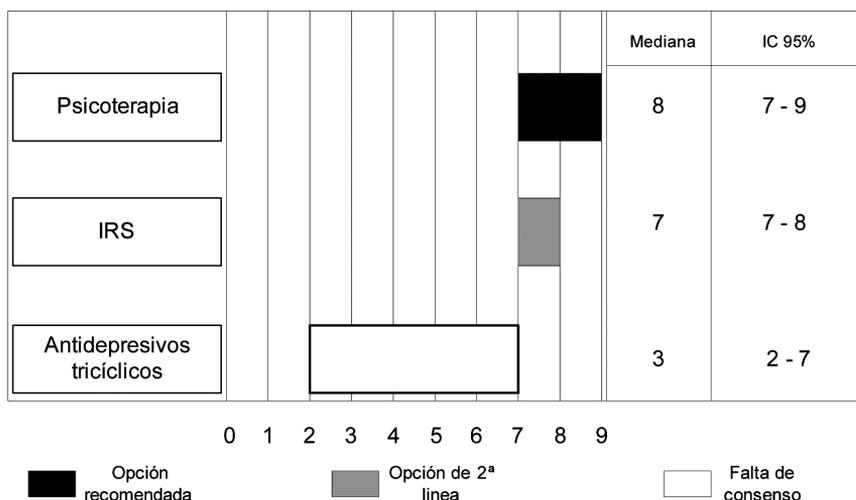
Además del texto que declara tanto los puntos de consenso como el grado de la recomendación (22), son útiles las herramientas gráficas que incorporan las mencionadas dimensiones. Estos gráficos deben contener la siguiente información (23):

1. Las diferentes opciones sometidas a consenso.
2. Las medianas junto con sus intervalos de confianza (o los límites del rango de respuestas, dependiendo del método utilizado).
3. Una convención que represente el grado de recomendación (usualmente se recurre a colores).

En la Figura 3 se presenta un ejemplo hipotético sobre la forma de

Figura 3. Presentación gráfica de los resultados del consenso

El manejo recomendado en pacientes con depresión mayor leve es:



representar gráficamente los resultados del consenso en relación con una recomendación terapéutica.

Referencias

1. Black N, Murphy M, Lamping D, McKee M, Sanderson C, Askham J, et al. Consensus development methods: a review of best practice in creating clinical guidelines. *J Health Serv Res Policy*. 1999;4(4):236-48.
2. Rycroft-Malone J. Formal consensus: the development of a national clinical guideline. *Qual Health Care*. 2001;10(4):238-44.
3. Ferlie EB, Shortell SM. Improving the quality of health care in the United Kingdom and the United States: a framework for change. *Milbank Q*. 2001;79(2):281-315.
4. Olmos ME, Sánchez R, Venegas MA. Los consensos de expertos: una metodología útil en la toma de decisiones en salud. *Revista Colombiana de Cancerología*. 2006;10(1):50-60.
5. Haute Autorité de Santé. Bases méthodologiques pour l'élaboration de recommandations professionnelles par consensus formalisé. Guide méthodologique. France: Haute Autorité de Santé; 2006.
6. Lomas J, Anderson G, Enkin M, Vayda E, Roberts R, MacKinnon B. The role of evidence in the consensus process. Results from a Canadian consensus exercise. *JAMA*, 1988;259(20):3001-5.
7. Dobbie A, Rhodes M, Tysinger JW, Freeman J. Using a modified nominal group technique as a curriculum evaluation tool. *Fam Med*. 2004;36(6):402-6.
8. Tong R. The epistemology and ethics of consensus: uses and misuses of 'ethical' expertise. *J Med Philos*, 1991;16(4):409-26.
9. Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. *Am J Public Health*. 1984;74(9):979-83.
10. Brook RH, McGlynn EA, Shekelle PG. Defining and measuring quality of care: a perspective from US researchers. *Int J Qual Health Care*. 2000;12(4):281-95.
11. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess*. 1998;2(3):i-iv, 1-88.
12. Murphy KJ, Kelekis DA, Rundback J, Gomolka B, Faciszewski T, Gangi A, et al. Development of a research agenda for skeletal intervention: proceedings from a multidisciplinary consensus panel. *J Vasc Interv Radiol*. 2008;19(5):631-8.
13. Astigarraga E. El método Delphi. San Sebastián: Universidad de Deusto; 2003.
14. Williams PL, Webb C. The Delphi technique: a methodological discussion. *J Adv Nurs*. 1994;19(1):180-6.
15. Powell C. The Delphi technique: myths and realities. *J Adv Nurs*. 2003;41(4):376-82.
16. Green B, Jones M, Hughes D, Williams A. Applying the Delphi technique in a study of GPs' information requirements. *Health Soc Care Community*. 1999;7(3):198-205.
17. Jones JH. Consensus methods for medical and health services research. En: Pope CM, Mays N, Editors. *Qualitative research in health care*. Massachusetts: BMJ books; 2000. 46-58.
18. Fitch KB, Bernstein SJ, Aguilar MD, Burnand B, La Calle JR, Lazaro P, et al. The RAND/UCLA appropriateness method. Santa Monica: RAND; 2001.
19. Coulter I, Adams A, Shekelle P. Impact of varying panel membership on ratings of appropriateness in consensus panels: a comparison of a multi- and single disciplinary panel. *Health Serv Res*. 1995;30(4):577-91.
20. Kravitz RL, Park RE, Kahan JP. Measuring the clinical consistency of panelists' appropriateness ratings: the case of coronary artery bypass surgery. *Health Policy*. 1997;42(2):135-43.
21. Colciencias, M.d.P.S.C. Manual metodológico para elaboración de Guías de Práctica Clínica (GPC) basadas en la evidencia, evaluaciones económicas

- de GPC y del impacto de la implementación de las GPC en el POS y la UPC del Sistema General de Seguridad Social en Salud Colombiano. Bogotá: Ministerio de Protección Social; 2009.
22. Oxman AD, Schünemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 14. Reporting guidelines. *Health Res Policy Syst.* 2006;4:26.
23. Sachs GS, Printz DJ, Kahn DA, Carpenter D, Docherty JP. The Expert Consensus Guideline Series: Medication Treatment of Bipolar Disorder 2000. *Postgrad Med.* 2000;Spec No:1-104.

Conflicto de interés: los autores manifiestan que no tienen ningún conflicto de interés en este artículo.

Recibido para evaluación: 30 de agosto del 2009

Aceptado para publicación: 18 de noviembre del 2009

Correspondencia
Ricardo Sánchez Pedraza
Departamento de Psiquiatría
Facultad de Medicina
Universidad Nacional de Colombia
Ciudad Universitaria, carrera 30, calle 45, edificio 471
Bogotá, Colombia
rsanchezpe@unal.edu.co