

Análisis crítico de un artículo: Análisis de una revisión sistemática sobre calidad de guías de práctica clínica

BLANCA PEÑALOZA^{1,2}

Critically appraised article

Pablo Alonso-Coello, Affan Irfan, Ivan Sola, Ignasi Gich, Mario Delgado-Noguera, David Rigau, Sera Tort, Xavier Bonfill, Jako Burgers, Holger Schunemann. The quality of clinical practice guidelines over the last two decades: a systematic review of guidelines appraisal studies. Qual Saf Health Care 2010; 19: e58.

Background: Despite the increasing number of manuals on how to develop clinical practice guidelines (CPGs) there remain concerns about their quality. The aim of this study was to review the quality of CPGs across a wide range of health-care topics published since 1980. **Methods:** The authors conducted a literature search in MEDLINE to identify publications assessing the quality of CPGs with the Appraisal of Guidelines, Research and Evaluation (AGREE) instrument. For the included guidelines in each study, the authors gathered data about the year of publication, institution, country, healthcare topic, AGREE score per domain and overall assessment. **Results:** In total, 42 studies were selected, including a total of 626 guidelines, published between 1980 and 2007, with a median of 25 CPGs. The mean scores were acceptable for the domain 'Scope and purpose' (64%; 95% CI 61.9 to 66.4) and 'Clarity and presentation' (60%; 95% CI 57.9 to 61.9), moderate for domain 'Rigour of development' (43%; 95% CI 41.0 to 45.2), and low for the other domains ('Stakeholder involvement' 35%; 95% CI 33.9 to 37.5, 'Editorial independence' 30%; 95% CI 27.9 to 32.3, and 'Applicability' 22%; 95% CI 20.4 to 23.9). From those guidelines that included an overall assessment, 62% (168/270) were recommended or recommended with provisos. There was a significant improvement over time for all domains, except for 'Editorial independence'. **Conclusions:** This review shows that despite some increase in quality of CPGs over time, the quality scores as measured with the AGREE Instrument have remained moderate to low over the last two decades. This finding urges guideline developers to continue improving the quality of their products. International collaboration could help increasing the efficiency of the process.

¹Departamento de Medicina Familiar.

²Unidad de Medicina Basada en Evidencia, Escuela de Medicina. Pontificia Universidad Católica de Chile. Santiago de Chile.

Correspondencia a:
Dra. Blanca Peñaloza
bpenalo@med.puc.cl

Conclusión de la revisora

Estudios metodológicos que evaluaron la calidad de guías de práctica clínica utilizando el instrumento AGREE muestran que un porcentaje importante de las guías evaluadas presentan limitaciones metodológicas importantes.

Pregunta clínica

¿Las guías de práctica clínica elaboradas para informar la toma de decisiones en la atención sanitaria, cumplen con los criterios de calidad actualmente aceptados?

Introducción

Las guías de práctica clínica (GPC) han sido definidas por el *US Institute of Medicine* como “enunciados elaborados sistemáticamente para asistir las decisiones de los profesionales sanitarios y sus pacientes acerca del cuidado de salud apropiado en circunstancias clínicas específicas”¹. Recientemente, esta misma organización precisa que las GPC son “enunciados que incluyen recomendaciones dirigidas a optimizar el cuidado de los pacientes y que están informadas por una revisión sistemática de la evidencia y una evaluación de los beneficios y daños de opciones de cuidado alternativas”².

Las GPC han ido tomando un rol protagónico en los últimos años, como una metodología para apoyar el proceso de toma de decisiones en salud y constituyen una herramienta útil para llevar a cabo una práctica clínica basada en evidencia.

Para que las guías cumplan su objetivo es fundamental que todos los pasos necesarios para su elaboración sean llevados a cabo de manera rigurosa. La iniciativa del Grupo Colaborativo AGREE (*Appraisal of Guidelines, Research, and Evaluation in Europe*) para evaluar la calidad del proceso de elaboración de una GPC ha permitido disponer de un instrumento que evalúa la calidad de dicho proceso. El grupo AGREE ha desarrollado 2 herramientas de evaluación: AGREE³ y una versión más reciente, AGREE II⁴.

Ambos instrumentos evalúan el proceso de elaboración de una guía en 6 dominios: 1) ámbito y propósito de la guía; 2) participación de los grupos de interés; 3) rigurosidad en el desarrollo (búsqueda y selección de la evidencia que sostiene las recomendaciones y la metodología de estas últimas); 4) claridad y presentación; 5) aplicabilidad e 6) independencia editorial del grupo que elabora la guía. Existe un último ítem de evaluación global, que corresponde a una apreciación subjetiva de si la guía es recomendada para su uso en la práctica clínica, de acuerdo al criterio del evaluador, considerando el resultado de la evaluación por cada dominio.

AGREE II tiene algunas diferencias en cada dominio con el instrumento original, pero la diferencia más importante es la incorporación en el dominio 3 del ítem que considera la inclusión de las fortalezas y debilidades de la evidencia utilizada para construir la guía,

El siguiente artículo corresponde al análisis crítico de una revisión sistemática de estudios metodológicos que han evaluado la calidad de GPC publicadas durante los últimos 20 años utilizando el instrumento AGREE. Esta revisión fue realizada antes de la aparición de AGREE II.

Características del estudio

Tipo de estudio: Revisión sistemática de estudios metodológicos que evaluaron guías de práctica clínica (GPC) utilizando el instrumento AGREE. La búsqueda se realizó en MEDLINE desde el año 2003, año de la publicación del instrumento AGREE, hasta julio de 2008. No se limitó la búsqueda por idioma. No se menciona si se limitó la inclusión de los estudios por su estatus de publicación.

Los autores incluyeron estudios que evaluaran GPC en cualquier temática, mediante la utilización del instrumento AGREE aplicado por al menos dos revisores. Los resultados de interés fueron los puntajes promedio (y su intervalo de confianza de 95%) obtenidos por las guías evaluadas, para cada uno de los dominios del instrumento AGREE y para el puntaje global.

Principales hallazgos: La revisión identificó 42 estudios que evaluaron un total de 625 GPC publicadas entre 1980 y 2007, La mayoría de las guías habían sido publicadas en los últimos 10 años; 41% elaboradas en Europa y 41% en Estados Unidos de Norteamérica. Sólo 2% de las guías evaluadas habían sido elaboradas en Sudamérica. Asimismo, 60% de las GPC evaluadas habían sido elaboradas por sociedades científicas y sólo 20% por organizaciones gubernamentales. Los tópicos de las guías más frecuentes fueron los de las áreas de medicina interna/intensiva (28%), pediatría (22%) y oncología (20,5%).

Validez (riesgo de sesgo) del estudio

1. Pregunta específica y focalizada	Sí
2. Búsqueda amplia y completa	No*
3. Criterios de inclusión explícitos y adecuados	Sí
4. Evaluación de calidad de los estudios incluidos	No
5. Dos revisores independientes	Parcialmente**

* Sólo se buscó en MEDLINE. **Sólo “screening” e inclusión.

Resultados

Tabla 1. Evaluación de las GPC, en los 6 dominios del instrumento AGREE (n = 608)

Dominio	Puntaje por dominio (0-100) Promedio (IC 95%)
Ámbito y propósito	64 (61,9-66,4)
Participación de los grupos de interés	35 (33,9-37,5)
Rigurosidad en el desarrollo	43 (41-45,2)
Claridad y presentación	60 (57,9-61,9)
Aplicabilidad	22 (20,4-23,9)
Independencia editorial	30 (27,9-32,3)
Evaluación global* N° guías recomendadas o recomendadas con condiciones	168 (62%)

*Total submuestra de 207 guías.

Comentarios

Comentarios acerca de la validez (riesgo de sesgo)

La principal limitación metodológica de esta revisión es la búsqueda reducida sólo a MEDLINE, sin estrategias adicionales para localizar estudios no publicados. Esta limitación probablemente podría explicar por qué la mayoría de las guías evaluadas provienen de países anglosajones. Tal vez si los autores hubieran incluido métodos para localizar estudios no publicados, el porcentaje de guías de otros países habría tenido mayor representatividad. Asimismo, la búsqueda puede también haber determinado que los autores encontrarán más guías elaboradas por sociedades científicas (62%) que por agencias gubernamentales (20%), las cuales podrían publicar menos en la literatura científica tradicional.

Otro aspecto del riesgo de sesgo de esta revisión es que por el tipo de estudios que decidieron incluir, no se plantean métodos de evaluación del riesgo de sesgo de dichos estudios, a los que sólo se les puso como condición para ser incluidos usar el instrumento AGREE y el que éste haya sido aplicado por más de un evaluador. Adicionalmente los autores de la revisión reportaron el porcentaje de estudios que midieron el acuerdo interevaluador, que sólo corresponde a 39% de los estudios selec-

cionados. No hay información sobre los valores de acuerdo encontrados en estos estudios. No existen otros criterios estandarizados de evaluación de riesgo de sesgo que se puedan aplicar a este tipo de estudios que evalúan la calidad de una guía.

Comentarios acerca de los resultados

Esta revisión incluye múltiples estudios que evalúan la calidad de guías, lo cual permite obtener una visión mucho más amplia que la que se puede obtener con los estudios individuales. Además, a pesar de las limitaciones de la búsqueda, permite tener una perspectiva de cuántas guías han sido evaluadas en el período de tiempo que los autores analizaron.

Llama la atención que a pesar de las limitaciones en la estrategia de búsqueda, los autores hayan encontrado un número tan alto de estudios de evaluación de guías en áreas tan diversas, lo que puede representar el interés y preocupación creciente por la calidad de la gran cantidad de guías que se producen anualmente.

En términos generales, las guías encontradas en esta revisión evaluadas con el instrumento AGREE obtienen un puntaje moderado a bajo. Esto resulta aún más preocupante al ver los resultados en el dominio “rigor en la elaboración”, lo que pone mayores dudas acerca de su calidad. Particularmente, en este punto se podría especular que si se hubiera utilizado AGREE II, la calificación en este dominio hubiera sido todavía más baja, considerando las exigencias mayores que hace el nuevo instrumento con respecto a la evaluación de la evidencia utilizada en la construcción de la guía.

Otros aspectos que llaman la atención por su baja evaluación son los dominios de independencia editorial y de aplicabilidad. El primero es considerado hoy en día un punto de alta relevancia en la calidad de una guía y en su credibilidad². El segundo punto, se relaciona con la poca incorporación de aspectos relacionados con la implementación de la guía, fundamental para la obtención de los resultados deseados en los pacientes.

Los aspectos mejor evaluados en las guías son los dominios correspondientes a ámbito y propósito, y claridad y presentación, que son aquellos dominios más relacionados con aspectos formales de las guías.

Los puntajes de evaluación global son signi-

ficativamente mejores en las guías de los últimos 5 años, comparadas con las guías realizadas antes del año 2003. En los dominios específicos, las guías publicadas después del 2003 son superiores a las previas en “ámbito y propósito”, “participación de los grupos de interés” y “claridad y presentación”, pero no en los otros dominios.

Sin embargo, de acuerdo a lo referido recientemente por el *Institute of Medicine*², a pesar de que el instrumento utilizado por esta revisión es el único validado y globalmente aceptado para evaluar la calidad de una GPC, una limitación importante es que AGREE falla en evaluar la calidad de la evidencia que subyace a las recomendaciones de la guía, tema fundamental para el objetivo de la guía de contribuir a la toma de decisiones informada por la mejor evidencia disponible. Este último punto condiciona además la necesidad de actualizar las recomendaciones en concordancia con los cambios que puedan producirse con la nueva evidencia disponible.

Comentarios acerca de la aplicabilidad

Las guías encontradas que fueron evaluadas corresponden en su mayoría a GPC elaboradas en Europa y Estados Unidos de Norteamérica (82% del total), por lo que no pueden considerarse representativas del universo de guías disponibles. Sólo un pequeño porcentaje corresponde a guías realizadas en nuestra Región. En efecto, uno de los estudios incluidos corresponde a un estudio chileno⁵ que evaluó la calidad de guías relevantes para la atención primaria realizadas en nuestro país. Los resultados de este estudio y de otro publicado recientemente por el mismo grupo⁶ son semejantes, en cuanto ponen de manifiesto las importantes limitaciones metodológicas que tienen las guías evaluadas, especialmente en la selección de la evidencia y en la independencia editorial de las recomendaciones.

Finalmente, a pesar de los bajos resultados de las evaluaciones encontradas, un factor que podría considerarse como esperanzador es la tendencia a la mejoría en el tiempo de la calidad de las guías, siendo las mejor evaluadas las realizadas en el último año. Sin duda, no puede considerarse garantía de calidad el que una guía haya sido realizada recientemente, pero es una señal de mejoría en el complejo proceso de elaboración de buenas GPC que contribuyan a mejorar la toma de decisiones en salud.

Referencias

1. Committee on Quality of Health Care in America IoM. *Crossing the quality chasm: a new health system for the 21st century*. Washington: National Academy Press; 2001.
2. Institute of Medicine. *Clinical Practice Guidelines We Can Trust*. Washington DC: The National Academies Press; 2011.
3. The Appraisal of Guidelines Research and Evaluation in Europe (AGREE) Collaborative Group. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Health Care* 2003; 12: 18-23.
4. Brouwers M, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, Fervers B, Graham ID, Grimshaw J, Hanna S, Littlejohns P, Makarski J, Zitzelsberger L for the AGREE Next Steps Consortium. AGREE II: Advancing guideline development, reporting and evaluation in healthcare *CMAJ* 2010; 182 (18): E839-E842. doi: 10.1503/cmaj.090449
5. Pantoja T, Strain H, Valenzuela L. [Clinical practice guidelines in primary health care: A critical appraisal.]. *Rev Med Chile* 2007 Oct; 135 (10): 1282-90.
6. Pantoja T, Valenzuela L, Leniz J, Castañon C. Guías de Práctica Clínica en el Régimen de Garantías en Salud: una evaluación crítica de su calidad. *Rev Med Chile* 2012; 140: 1391-400.