

¿Las guías que nos guían son fiables? Evaluación de las guías de práctica clínica españolas

M. A. Navarro Puerto^a, F. Ruiz Romero^a, A. Reyes Domínguez^a, I. Gutiérrez Ibarlucea^b, T. Hermosilla Gago^c, C. Alonso Ortiz del Río^a, R. Rico Iturrioz^b y J. Asúa Batarrita^b

^a Servicio de Medicina Interna. Hospital Universitario Valme. Sevilla. ^b Servicio de Evaluación de Tecnologías Sanitarias. OSTEBA. Gobierno Vasco. Vitoria. ^c Agencia de Evaluación de Tecnologías Sanitarias de Andalucía. AETSA. Sevilla. Consejería de Salud de la Junta de Andalucía.

Fundamento y objetivo. Hay gran proliferación de guías de práctica clínica (GPC), existiendo dudas sobre inconsistencias entre las recomendaciones que realizan e incongruencias internas que comprometen el papel de las guías como instrumentos de ayuda a los clínicos. El estudio pretende analizar la calidad de un conjunto de GPC españolas, así como valorar la factibilidad de uso y consistencia del Instrumento AGREE en su versión en español, diseñado para la evaluación de GPC. **Método.** Se identificaron 278 GPC producidas en España entre enero de 1990 y diciembre del 2002. Se recuperaron, mediante solicitud a sus autores, originales de 61 guías a las que se pasó el Instrumento AGREE de forma independiente por cuatro evaluadores. Se analizó la calidad de las guías, la factibilidad y reproducibilidad del AGREE. **Resultados.** Para todos los atributos del AGREE, excepto en la independencia editorial, más del 60% de las guías evaluadas obtienen una puntuación de mala calidad, no habiendo ningún dominio en el que más del 10% de las guías sea excelente. Las áreas de peor calificación son las de aplicabilidad, participación y rigor en el desarrollo de las guías. De las 61 guías evaluadas, los revisores a 25 las califican de «muy poco recomendables», a 26 de «no recomendables» y tan sólo a 6 y 4 guías como «recomendables» o «muy recomendables», respectivamente. La consistencia entre evaluadores fue alta. **Conclusiones.** La calidad media de las guías españolas evaluadas es muy baja. El Instrumento AGREE, en su versión en español es consistente para las guías españolas y de uso fácil.

PALABRAS CLAVE: guías de práctica clínica, evaluación, AGREE.

Navarro Puerto MA, Ruiz Romero F, Reyes Domínguez A, Gutiérrez Ibarlucea I, Hermosilla Gago T, Alonso Ortiz del Río C, Rico Iturrioz R, Asúa Batarrita J. ¿Las guías que nos guían son fiables? Evaluación de las guías de práctica clínica españolas. *Rev Clin Esp.* 2005;205(11):533-40.

Are the guidelines guiding us reliable? Evaluation of the Spanish clinical practice guidelines **Background and objective.** There is a large proliferation of clinical practice guidelines (CPG), there being doubts on inconsistencies between the recommendations made and internal incongruencies, that affect the role of the guidelines as instruments helping the clinicians. This study aims to analyze the quality of a group of Spanish CPG and to assess the feasibility of use and consistence of the AGREE Instrument in its Spanish version, designed to evaluate the CPG. **Method.** A total of 278 CPG produced in Spain between January 1990 and December 2002 were identified. The originals of 61 guidelines were recovered from their authors and passed to the AGREE Instrument independently by four raters. Quality of the guidelines feasibility and reproducibility of AGREE were analyzed. **Results.** For all the attributes of AGREE except in the editorial independence, more than 60% of the guidelines evaluated obtained a poor quality score, there being no domain in which more than 10% of the guidelines was excellent. The worst grading areas are those of applicability, participation and rigour of development of the guidelines. Of the 61 guidelines evaluated, the reviewers graded 25 as «very poor quality», 26 as «non recommendable» and 6 and 4 as «recommendable» or «highly recommendable», respectively. Interrater consistency was high. **Conclusions.** The mean quality of the Spanish guidelines evaluated is very low. The AGREE Instrument in its Spanish version is consistent for the Spanish guidelines and easy to use.

KEY WORDS: clinical practice guidelines, evaluation, AGREE.

Introducción

Las guías de práctica clínica (GPC) ocupan actualmente un papel central en las estrategias de mejora en la toma de decisiones de los sistemas de salud, tanto como herramientas aisladas o incorporadas a programas de gestión clínica de procesos¹. Además, exis-

Correspondencia: M. A. Navarro Puerto.
Servicio de Medicina Interna.
Hospital Universitario Valme.
41014 Sevilla.
Correo electrónico: masunnp@telefonica.net

Aceptado para su publicación el 14 de julio de 2004.

te un gran entusiasmo de los profesionales para incorporarse a la producción de GPC, a juzgar por la gran cantidad de documentos denominados «guías» que se encuentran disponibles².

Esta proliferación de GPC ha supuesto que muchas de ellas versen sobre los mismos temas, apreciándose inconsistencias y notables discrepancias en las recomendaciones que se dan sobre las mismas patologías³. La variabilidad en las guías identifica un problema de calidad y rigor del fundamento científico de las GPC españolas. A veces ante el panorama de guías disponibles, da la impresión que la prioridad de sus productores es publicar guías, olvidándose en parte de los requisitos mínimos que deben cumplir para gozar de una aceptable calidad⁴.

La validez y fiabilidad de las guías es especialmente relevante, ya que precisamente las GPC se proponen como un nexo de transferencia rápida entre la investigación y la asistencia, siendo por tanto requisito indispensable de las GPC aparecer como herramientas en las que los clínicos puedan confiar.

En nuestro país hay múltiples iniciativas y escasa sistemática en la producción de GPC, con repeticiones e incongruencias entre ellas. El estudio REDEGUIAS ha identificado 1.084 documentos denominados «guías» en España desde 1990 hasta 2002⁵. Es habitual una escasa fundamentación de la evidencia científica de las recomendaciones, algunas sin aportar bibliografía, sin estimación de costes de su aplicación o sin programar futuras actualizaciones⁶. Además existe un problema semántico denominándose «guías» a documentos que son protocolos de actuación interna en unidades asistenciales.

Junto a la proliferación hay un déficit en la evaluación de la calidad de las guías y de su idoneidad para ser implantadas, así como en la comprobación de los presuntos efectos beneficiosos para los pacientes y las organizaciones asistenciales que se les supone².

En algunos países como Alemania⁷, Australia⁸, Canadá⁹, EE.UU.¹⁰, Nueva Zelanda¹¹ y Reino Unido¹² se ha observado la necesidad de disponer de programas que gobiernen la iniciativa en la producción de las GPC y el análisis de la calidad de las mismas. Estos países han comenzado a definir criterios para juzgar la calidad de las GPC según el rigor en su producción, fundamentación científica y actualización. Igualmente, mediante un proceso de selección, pretenden evitar redundancias e inconsistencias entre las guías que se ofrecen a los profesionales.

Dos iniciativas han surgido en nuestro medio para ordenar dicha actividad, el proyecto GUIASALUD¹³ y el proyecto REDEGUIAS⁵. El proyecto GUIASALUD consiste en la creación de un registro de GPC que mejore la disponibilidad de las guías producidas en España aprovechando las facilidades que ofrece la informática. Por su parte, el proyecto REDEGUIAS pretende identificar las guías producidas en España y analizar la metodología de su desarrollo y su calidad. Así, una vez conocidas las principales deficiencias en la calidad de las GPC españolas, desarrollar una serie de recursos y herramientas para mejorar su elaboración, facilitar su implantación y evaluar los resultados del uso de las guías.

En otro artículo se expone cómo se realizó una recuperación y análisis de criterios y sistemas de evaluación disponibles con el fin de establecer un estándar para la evaluación de GPC¹⁴. De dicho análisis se determinó que el Instrumento AGREE con algunas adaptaciones puede resultar un buen marco teórico para analizar las GPC españolas¹⁵.

El objetivo de este artículo es la evaluación y análisis de la calidad de un conjunto de GPC españolas, así como valorar la factibilidad de uso y consistencia del Instrumento AGREE.

Material y métodos

En este estudio se considera GPC aquel documento de recomendaciones elaborado según un criterio científico explícito y reproducible para facilitar la toma de decisiones en la asistencia sanitaria. No se consideran guías los protocolos u otros documentos de obligado cumplimiento para la organización de la asistencia¹⁶.

De entre las 1.084 identificadas como producidas en España en el período 1990-2002, cumplían este criterio 278 documentos. Mediante solicitud escrita a sus autores, con cuatro recuerdos sucesivos, se recuperaron los originales impresos de 61 GPC, que constituyen la muestra analizada (anexo 1).

A las 61 guías se les aplicó el Instrumento AGREE de forma individual e independiente por cuatro evaluadores. Este consta de 23 criterios distribuidos en 6 áreas que analizan, respectivamente, los siguientes campos: 1) objetivo de la guía; 2) tipo de participantes implicados en su elaboración; 3) rigor en el desarrollo; 4) claridad en la presentación; 5) aplicabilidad, y 6) independencia editorial de sus autores. Aporta información diferencial sobre cada área y de conjunto sobre la calidad de las guías evaluadas.

Previo al pase del AGREE se realizó un taller de entrenamiento consistente en la lectura conjunta de los criterios que lo componen y de las instrucciones para su uso. También se pasó el Instrumento de forma conjunta por los cuatro evaluadores a dos guías, previamente conocidas por su buena y mala calidad, y se discutieron las discrepancias interevaluador en la calificación que otorgaban. Los cuatro evaluadores desconocían las guías que iban a ser evaluadas. De los cuatro, dos son clínicos con experiencia en el desarrollo de GPC y dos con limitada experiencia en su desarrollo. Para facilitar la puntuación y documentación de la evaluación los datos se recogían en una plantilla común para todos los evaluadores, que contenía una identificación de las GPC, y una sinopsis de los 23 criterios agrupados en las 6 áreas. Además, cada evaluador aportó su opinión global sobre la guía mediante la pregunta ¿recomendaría esta GPC? Asimismo, los evaluadores debían hacer constar el tiempo utilizado en la revisión de cada guía y su impresión sobre la utilidad de la herramienta de evaluación. Esta plantilla era independiente para cada guía y se archivaba por cada evaluador sin conocimiento de los otros. El análisis conjunto de los resultados se hizo a partir de las plantillas completadas por los cuatro evaluadores.

La calidad de la guía en cada uno de sus 6 áreas se determinó mediante el porcentaje del máximo de puntuación posible, tal como recomiendan los autores del AGREE¹⁵. Para hacer más operativo el análisis, en este estudio categorizamos este porcentaje en cuartiles: 1) menos del 25% de puntuación o muy baja puntuación; 2) entre 25%-50% de puntuación o baja puntuación; 3) entre 50%-75% de puntuación o alta puntuación, y 4) más del 75% de puntuación o muy alta puntuación.

ANEXO 1

Guías recuperadas para evaluación

Guía para el manejo del paciente con enfermedad vascular cerebral aguda. Sociedad Española de Neurología	Guía clínica sobre hernias de la región inguinal. Hospital de Laredo
Guía práctica del tratamiento del dolor. Hospital Universitario de La Princesa	Guía clínica de la colecistectomía. Hospital de Laredo
Guía para la prevención y el control de la infección nosocomial. Hospital Universitario de La Princesa	Vía clínica de la lumbalgia. Servicios de Traumatología y Reumatología. Hospital de Laredo
Guía práctica de profilaxis y tratamiento de la enfermedad tromboembólica. Hospital Universitario de La Princesa	Vía clínica del AVCA. Servicios de Medicina Interna y Rehabilitación. Hospital de Laredo
Guía de actuación en urgencias ante el dolor sugestivo de cólico renal. Grupo de trabajo multidisciplinario. Hospital Universitario de La Princesa.	Tratamiento antibiótico de las infecciones comunitarias respiratorias en la infancia. Grupo Investigador del Proyecto FIS: 97/0053. Hospital Virgen de la Concha
Guía de práctica clínica para el manejo de la artritis reumatoide en España. Sociedad Española de Reumatología	Carcinoma broncopulmonar no microcítico. Protocolo-guía de diagnóstico y tratamiento. Hospital de Txagorritxu
Guía de práctica clínica sobre hipertensión arterial. Asistencia Sanitaria de Osakidetza	Protocolo-guía del carcinoma de mama. Hospital de Txagorritxu
Guía de práctica clínica del asma. Comarca Guipúzcoa Este. Osakidetza	Neoplasia colorrectal. Protocolo-guía de diagnóstico y tratamiento. Hospital de Txagorritxu
Protocolo de ayuda al fumador. Comarca Guipúzcoa Este. Osakidetza	Protocolo-guía de tumores de laringe. Servicios de Otorrinolaringología. Hospitales Santiago y Txagorritxu
Guía práctica para el cuidado y prevención de las úlceras por presión. Consejería de Sanidad de la Generalitat Valenciana. Atención Primaria Área 12	Enfermedad tromboembólica venosa. Orientaciones dirigidas al paciente y sus cuidadores. Sociedad Española de Medicina Interna
Protocolo de quemaduras. Consejería de Sanidad de la Generalitat Valenciana. Atención Primaria Área 12.	Insuficiencia cardíaca. Recomendaciones diagnósticas y terapéuticas. Grupo de trabajo de ICC de Medicina Interna. Sociedad Española de Medicina Interna
Guía de práctica clínica para la hemorragia digestiva alta. Servicio de Digestivo. Hospital Universitario de Canarias	Preguntas claves en el asma. Almirall Prodesfarma
Guía de práctica clínica en la enfermedad inflamatoria intestinal. Servicio de Digestivo. Hospital Universitario de Canarias	Dislipemias. Guía práctica. Grup de Lípids del Bages i del Berguedá (GLIBB). Sociedad Española de Arteriosclerosis
Guía clínica de la pancreatitis aguda. Servicio de Digestivo. Hospital Universitario de Canarias	Directrices para la práctica clínica en patología cardiocirculatoria. HTA. Hipercolesterolemia. Hospital de la Santa Creu i Sant Pau
Alimentación enteral del recién nacido a término. Hospital Universitario de Canarias	El paciente con síntomas digestivos. Guía práctica de actuación diagnóstico-terapéutica. Hospital Clínico Universitario de Santiago de Compostela
Cuidados del recién nacido a término sano. Hospital Universitario de Canarias	Guía de práctica clínica de HTA para Atención Primaria. Gerencia de Atención Primaria de Lugo
Convulsiones neonatales. Hospital Universitario de Canarias	Guía de atención a adolescencia. Gerencia de Atención Primaria de Lugo
Guía clínica para el tratamiento inmunosupresor en el trasplante renal. Hospital Universitario de Canarias	Documento de consenso 2002 sobre detección, prevención y tratamiento de la nefropatía diabética en España. Grupo Español de Estudio de la Nefropatía Diabética. Sociedad Española de Nefrología
Guía clínica para el tratamiento del síndrome nefrótico idiopático en el niño y de la glomerulonefritis. Hospital Universitario de Canarias	Guía de prevención del cáncer en Atención Primaria. Sociedad Española de Medicina Familiar y Comunitaria
Guía práctica para el manejo de la nefropatía lúpica. Hospital Universitario de Canarias	Guía de salud mental en Atención Primaria. Sociedad Española de Medicina Familiar y Comunitaria
Guía clínica para el manejo de la glomerulonefritis membranosa idiopática. Hospital Universitario de Canarias	Guía rápida del tratamiento de las urgencias médicas en la consulta dental. Colegio de Odontólogos y Estomatólogos de Cataluña
Guía clínica para el manejo de la hepatitis vírica B y C en diálisis y tras el trasplante renal. Hospital Universitario de Canarias	Conductas de actuación en la enfermedad inflamatoria crónica intestinal. Sociedad Valenciana de Patología Digestiva
Guía clínica para el manejo de las alteraciones del metabolismo fosfocálcico en la insuficiencia renal. Hospital Universitario de Canarias	Situaciones clínicas en asma. GlaxoSmithKline
Guía clínica para el manejo de las complicaciones agudas de la hemodiálisis. Hospital Universitario de Canarias	Guía para el control de la infección en Atención Primaria. Sociedad Española de Medicina General. Sociedad Española de Quimioterapia. Asociación Española de Pediatría. Sociedad Española de Medicina Rural y Generalista
Manejo del paciente con enfermedad por reflujo gastroesofágico. Asociación Española de Gastroenterología y Sociedad Española de Medicina Familiar y Comunitaria	Temblor. Guía clínica. GlaxoSmithKline
Manejo del paciente con rectorragia. Asociación Española de Gastroenterología y Sociedad Española de Medicina Familiar y Comunitaria	Guía de práctica clínica sobre enfermedades autoinmunes. Sociedad Española de Medicina Interna
Protocolo de actuación en el dolor crónico oncológico. Proyecto Algos. Laboratorios Janssen-Cilag	Guía clínica del paciente asmático. Grupo Multidisciplinar. Hospital de Laredo
Programa de embarazo. Hospital Reina Sofía de Córdoba	Guía de práctica clínica sobre diabetes mellitus tipo 2, obesidad y riesgo vascular. Sociedad Española de Medicina Interna
Parto pretérmino. Hospital Reina Sofía de Córdoba.	Guía de práctica clínica para el uso racional de las heparinas de bajo peso molecular. Área Hospitalaria Juan Ramón Jiménez
Profilaxis y prevención del estreptococo grupo B. Hospital Reina Sofía de Córdoba	
Asistencia al parto de nalgas. Hospital Reina Sofía de Córdoba	
Angina inestable. Hospital de Laredo	

El análisis de factibilidad de la evaluación se realizó analizando el tiempo de duración medio de cada pase del instrumento, el número de criterios del Instrumento que se quedan sin completar y la opinión subjetiva de los evaluadores sobre la utilidad de la herramienta para el juicio sobre la guía.

La consistencia del Instrumento AGREE se midió mediante el acuerdo simple entre evaluadores. Para ello, se determinó el rango de dispersión de las puntuaciones entre los cuatro evaluadores, en cada una de las seis áreas para las 61 guías analizadas, respecto a la mediana de puntuación, excluida la del evaluador más disperso. Se acordó eliminar el valor más alejado de la mediana de todas las puntuaciones porque al ser más de dos evaluadores en múltiples calificaciones la dispersión en las puntuaciones es por azar muy amplia, pretendiéndose así controlar dicho fenómeno¹⁷. Para las áreas que tienen tres criterios (alcance y aplicabilidad) se admitió una variabilidad máxima tolerada de dos niveles con respecto a la mediana; para las que tienen cuatro criterios (participación de los implicados y claridad en la presentación), tres niveles, y para las áreas con siete criterios (rigor en la elaboración), cuatro niveles.

Resultados

Los resultados del análisis de las GPC españolas mediante el Instrumento AGREE, el rango y la puntuación media de valoración obtenido por cada una de las 61 guía en sus 6 áreas y el porcentaje de la puntuación global obtenida en cada área, así como si la GPC merece ser recomendada a juicio de los evaluadores, se recogen en el anexo 2.

La tabla 1 muestra la distribución de las guías evaluadas por cuartiles según el porcentaje del máximo de puntuación obtenido para las 6 áreas.

Se aprecia que para todos los atributos que juzga el AGREE, excepto en la independencia editorial, más del 60% de las guías evaluadas obtienen una puntuación de mala calidad, no habiendo ningún dominio en el que más del 10% de las guías sea excelente. Las áreas de peor calificación son las de aplicabilidad, participación y rigor en el desarrollo de las guías.

La puntuación por áreas coincide con la valoración global de las 61 guías evaluadas. A la pregunta ¿recomendaría usted la guía?, los revisores califican 25 de ellas como «muy poco recomendables», 26 como «no recomendables» y tan sólo 6 y 4 guías como «recomendables» o «muy recomendables», respectivamente. Los resultados del estudio de reproducibilidad del Instrumento AGREE, medido por la consistencia entre observadores, para tres observadores fue excelente y se recogen para cada área en la tabla 2.

Respecto a la factibilidad de uso del AGREE, el tiempo de duración medio empleado en la evaluación de la guía, que incluía el primer contacto con cada guía y la lectura de su metodología de desarrollo y una lectura superficial de su contenido, fue de 12 minutos, con un rango de 5 a 47 minutos.

No ha habido ningún elemento del Instrumento que se quedara sin completar en ninguna guía, lo cual ilustra la flexibilidad del AGREE para adaptarse a cualquier tipo de guía.

La opinión subjetiva de los evaluadores sobre la utilidad de la herramienta para hacer un juicio sobre cada guía ha sido en todos los casos de 4, en una esca-

la 1-5 (muy baja-muy alta utilidad), coincidiendo los cuatro evaluadores con la recomendación realizada por los productores del Instrumento, en el sentido de que debe emitirse una valoración global «cualitativa» de cada guía, sobre si se recomienda o no la guía en cuestión.

Discusión

Los principales hallazgos de nuestro estudio son: primero, que la calidad media de las guías elaboradas en España es muy baja, y segundo, que la utilización de la lista de criterios del Instrumento AGREE en su versión en español es factible y consistente para el tipo de guías habitualmente disponibles en España.

Respecto a la calidad de las GPC, junto a una minoría de guías buenas o excelentes (16% de las evaluadas) se aprecia un gran conjunto de guías (25%) de muy mala calidad. Resalta que en los atributos principales para la fiabilidad de las guías (rigor científico y claridad con la que se elaboran) y en las características que mejor relacionan su utilización en la asistencia clínica (alcance y aplicabilidad), la puntuación obtenida por un 90% y un 80% de las guías evaluadas es de mala o muy mala calidad, respectivamente, para cada atributo.

De tal hallazgo se sacaría una triple consecuencia que invalidaría la estrategia de mejora de la calidad apoyado en las GPC, o bien las ayudas asistenciales que utilizan nuestros profesionales son poco fiables, o éstos no utilizan las guías precisamente porque no las aprecian como de buena calidad, y en cualquier caso hay un desperdicio notable de recursos por el esfuerzo que supone la realización y aplicación de guías con dudosa fiabilidad.

Otros autores han señalado previamente resultados comparables. Así, en un estudio en el que encuentran una gran variabilidad de recomendaciones de protocolos sobre hipertensión arterial en una región española³, se describen como principales defectos de estos protocolos la indefinición de las recomendaciones, la dificultad para identificar las recomendaciones separadas de las opiniones que se expresan y la escasa consistencia con documentos que debieran ser de referencia. Esta última explicación, que tiene que ver con la revisión crítica de la bibliografía, también se ha reseñado por la baja citación en las guías de revisiones sistemáticas disponibles públicamente y de calidad reconocida (metaanálisis y revisiones Cochrane)⁶.

Evaluaciones internacionales refieren hallazgos similares, que precisamente motivaron la implantación de programas para mejorar la calidad de las guías que se producen¹⁸. En España una mayoría de las guías están elaboradas por sociedades científicas. Ello también ha podido contribuir a la baja calidad detectada, ya que tanto en estudios que comparan guías de un país¹⁹, como de varios países²⁰ parece que las producidas por grupos profesionales o sociedades científicas son de menor calidad que las guías producidas por instituciones relacionadas con la mejora de la calidad de la asistencia. En nuestro caso las 10 guías calificadas como de mejor calidad eran promovidas por

ANEXO 2
Calificación obtenida por las guías de práctica clínica mediante el instrumento AGREE

Guía	Alcance RMI: 3-12 Rango*/media** global***	Participación RMI: 4-16 Rango*/media** global***	Rigor RMI: 7-28 Rango*/media** global***	Claridad RMI: 4-16 Rango*/media** global***	Aplicabilidad RMI: 3-12 Rango*/media** global***	Independencia editorial RMI: 2-8 Rango*/media** global***	Recomendaciones por evaluador (muy poco: 1; muy alta: 4)
1	3-4/3,25 3	4-6/5,25 10	7/7 0	4-6/5 8	3/3 0	2-5/3,5 25	2
2	3-5/4 11	5-6/5,5 13	7/7 0	4,6/5,25 10	3/3 0	2-5/3,5 25	2
3	3-7/5,5 28	4-7/5,5 13	7/7 0	4-6/5,5 13	3-4/3,25 3	2-5/3,5 25	2
4	5-9/7,25 47	4-10/6,25 19	7/7 0	4-12/8 33	3/3 0	2-5/3,5 25	2
5	9-11/10 78	8-9/8,25 35	12-14/12,75 27	13-14/13,5 79	3/3 0	7-8/7,5 92	3
6	3-7/5 22	4-8/6,5 21	7-10/8 5	4-11/8,25 35	3-4/3,5 6	2-5/2,75 13	3
7	9-12/11,25 92	8-14/11,25 60	25-28/27,25 96	15-16/15,75 98	3-12/7 44	5-8/7,25 88	4
8	8-10/9 67	6-9/7,25 27	16-18/20,5 64	15-16/15,5 96	5-8/6,5 39	2-5/4,25 38	4
9	40-10/7,5 50	6-10/8 33	7-9/7,5 2	4-8/6 17	3/3 0	2-5/4,25 38	2
10	5-10/7,5 50	4-10/6,5 21	7-8/7,25 1	6-10/7,5 29	3-6/4 11	2-5/3,5 25	2
11	3-7/5,25 25	4-5/4,25 2	7/7 0	4-9/5,5 13	3-4/3,25 3	2-5/2,75 13	2
12	3-6/4,5 17	5-7/6 17	7/7 0	4-10/6 17	3/3 0	2/2 0	1
13	3-4/3,75 8	4/4 0	7/7 0	5-8/6 17	3/3 0	5-8/5,75 63	1
14	3-4/3,75 8	4/4 0	7/7 0	5-8/6,5 21	3/3 0	5/5 50	1
15	3-4/3,75 8	4/4 0	7/7 0	5-10/7,75 31	3/3 0	5/5 50	1
16	3-4/3,75 8	4/4 0	7/7 0	4-5/4,75 6	3/3 0	2-5/4,25 38	1
17	3-5/4 11	4/4 0	7/7 0	4/4 0	3/3 0	2-5/4,25 38	1
18	3-4/3,25 3	4/4 0	7/7 0	4-7/5 8	3/3 0	2-5/4,25 38	1
19	3-5/3,75 8	4/4 0	7/7 0	5-6/5,75 15	3/3 0	2,5/4,25 38	1
20	3-5/4,25 14	4/4 0	7/7 0	5-8/6,5 21	3/3 0	2-5/4,25 38	1
21	3-5/4,25 14	4/4 0	7/7 0	5-6/5,25 10	3/3 0	2-5/4,25 38	1
22	3-5/4 11	4/4 0	7/7 0	5-6/5,5 13	3/3 0	5/5 50	1
23	3-5/4 11	4/4 0	7/7 0	4-5/4,75 6	3/3 0	5/5 50	1
24	3-5/4 11	4/4 0	7/7 0	4-6 8	3/3 0	5/5 50	1
25	3-4/3,5 6	4/4 0	7/7 0	4-5/4,75 6	3/3 0	5/5 50	1
26	9-12/8,75 64	10-12/11 58	22-28/25,25 87	16/16 100	3-7/4 11	5-8/7,25 88	4
27	12/12 100	6-12/9,5 46	22-28/20,25 63	16/16 100	3-10/4,75 19	5-8/7,25 88	4
28	5-7/6 33	4-5/4,25 2	7-8/7,25 1	6-10/8 33	3/3 0	2/2 0	1
29	5-12/8,5 61	4-7/5,5 13	7-18/10,75 18	4-10/7,25 27	3-6/4,75 19	2-5/2,25 38	3
30	3-5/4 11	4/4 0	7-10/8,5 7	4-9/6 17	3/3 0	2-5/4,25 38	2
31	3-5/4,25 14	3-4/3,75 0	7-9/7,75 4	4-7/5,75 15	3/3 0	2-5/4,25 38	2
32	3-5/4,5 17	4/4 0	7-8/7,5 2	4-7/5,25 10	3/3 0	2-5/4,25 38	2
33	3-5/4 11	4/4 0	7/7 0	4-7/4,75 6	3/3 0	2-5/3,5 25	1
34	3-5/3,75 8	4/4 0	7/7 0	4-5/4,25 0	3/3 0	2-5/4,25 38	1
35	3-6/4,25 14	4/4 0	7/7 0	4-5/4,25 2	3/3 0	2,5/4,25 38	1

ANEXO 2

Calificación obtenida por las guías de práctica clínica mediante el instrumento AGREE (continuación)

Guía	Alcance RMI: 3-12 Rango*/media** global***	Participación RMI: 4-16 Rango*/media** global***	Rigor RMI: 7-28 Rango*/media** global***	Claridad RMI: 4-16 Rango*/media** global***	Aplicabilidad RMI: 3-12 Rango*/media** global***	Independencia editorial RMI: 2-8 Rango*/media** global***	Recomendaciones por evaluador (muy poco: 1; muy alta: 4)
36	5-9/6,25 36	4-5/4,5 4	7/7 0	4/4 0	3/3 0	2-5/4 38	1
37	3-4/3,25 3	4-5/4,25 2	7/7 0	4-6/4,5 4	3/3 0	2-5/3,5 25	1
38	3-5/3,5 6	4/4 0	7/7 0	4-6/4,75 6	3/3 0	2-5/4,25 38	1
39	10-11/10-5 83	4-6/5 8	10-20/14,75 37	9-14/10,75 56	3-7/4,5 17	2-5/3,75 25	3
40	3-6/4,5 17	4-5/4,75 6	7/7 0	4-7/5,75 15	3-4/3,25 3	5/5 50	2
41	3-9/5,5 28	4-8/6,25 19	8-12/9,5 12	4-14/8,25 35	3-4/3,25 3	5/5 50	3
42	3-8/5,25 25	4-5/4,5 4	7/7 0	6-15/9 42	3-4/3,25 3	5/5 50	2
43	4-7/5,75 31	4-7/5,75 15	8-10/8,5 7	5-6/5,75 15	3/3 0	5/5 50	2
44	3-5/4,25 14	5-8/6,75 2	7/7 0	4-5/4,25 2	3/3 0	2/2 0	1
45	4-6/4,5 17	7-8/7,25 27	7-9/7,5 2	4-7/5,5 13	3-4/3,25 3	2-5/2,75 13	2
46	7/7 44	7-8/7,5 29	14-15/14,5 36	9-10/9,5 46	3-4/3,5 6	5/5 50	3
47	3-9/5,25 25	4/4 0	7-8/7,25 1	4-16/8 33	3/3 0	2-5/2,75 13	2
48	5-12/8 56	6-9/6,75 23	7-10/8 5	7-15/9 42	3-4/3,25 3	2-5/2,75 13	2
49	5-6/5,25 25	4-7/5,25 10	7-13/9 10	4-11/7 25	3-6/3,75 0	2/2 0	2
50	3-5/3,5 6	4-5/5,25 2	7/7 0	4-11/6,75 23	3/3 0	2/2 0	1
51	6-10/7,75 53	4-8/6 17	7/7 0	5-10/6,25 19	3-4/3,5 6	2-5/2,75 13	2
52	11-12/11,25 92	5-9/7 25	7-9/7,5 2	4-9/6,75 23	3-6/4,75 19	2-5/4,25 38	2
53	5-12/8,5 61	5-8/7 25	7-9/7,75 4	6-14/8,75 40	3-6/4 11	2/2 0	2
54	5-6/5,5 28	4-5/4,25 2	7-8/7,25 1	4-8/5 8	3-5/3,5 6	2-5/2,75 13	2
55	5-12/8,25 58	6-8/6,75 23	7-8/7,25 1	6-9/7,25 27	3-4/3,25 3	2/2 0	2
56	3-7/4,75 19	6-9/7,25 27	7-9/7,5 2	4-7/6,25 19	3-4/3,25 3	2/2 0	2
57	3-7/4,25 14	4-6/5,25 10	7/7 0	4-10/6,25 19	3/3 0	2/2 0	1
58	3-4/3,25 3	4-6/5 8	7/7 0	4-7/5,5 13	3/3 0	2/2 0	1
59	3-6/4 11	4-7/5,25 10	7/7 1	4-11/6,5 21	3/3 3	2/2 0	1
60	3-5/4,25 14	6-8/7,5 9	7-8/7,25 1	4-9/5,75 15	3-4/3,25 3	2/2 0	2
61	3-4/3,25 3	4/4 0	7/7 0	4-7/5,25 10	3/3 0	2/2 0	1

*Rango de valoración obtenido por cada guía en sus 6 áreas. **Puntuación media obtenido por cada guía en sus 6 áreas. ***Porcentaje de la puntuación global obtenida en cada área. RMI: rango máximo individual para cada área del AGREE. El orden seguido en la relación de este anexo no se corresponde con el orden de la lista del anexo 1.

este tipo de instituciones con participación colaborativa de sociedades profesionales. Este dato concuerda con estudios que confirman que las GPC de mejor calidad se elaboran por equipos pluridisciplinarios, con apoyo metodológico y dentro de programas de elaboración de guías; por el contrario, las de peor calidad surgen como iniciativas espontáneas, elaboradas por un solo grupo profesional y si están dentro de programas se realizan en mayor cantidad (y probablemente

menos cuidadas) que las institucionales²⁰. Se ha descrito que guías producidas por una sola sociedad científica o grupo profesional a veces están más preocupados por acotar espacios de poder o justificar prácticas establecidas que por definir el estado objetivo del conocimiento en un campo^{5,17,21}. El volumen de producción también puede explicar la baja calidad que hemos encontrado. Mientras una sociedad científica española en dos años elaboró

TABLA 1
Distribución de las guías en función de la puntuación alcanzada en cada una de las áreas evaluadas por el instrumento AGREE

Porcentaje de guías	Categorización en cuartiles del porcentaje del máximo de puntuación posible			
	< 25% (muy baja)	25%-50% (baja)	50%-75% (alta)	> 75% (muy alta)
Áreas				
Alcance	57,3	19,7	14,8	8,2
Participación	80,3	16,4	3,3	0,0
Rigor	88,5	4,9	3,3	3,3
Claridad	67,2	23,0	1,6	8,2
Aplicabilidad	96,7	3,3	0,0	0,0
Independencia	33,3	38,3	21,7	6,7

36 guías, instituciones de gran prestigio y recursos en la elaboración de guías como Duodecim en Finlandia, SIGN en Escocia y AHRQ en EE.UU. elaboraron 20, 50 y 17 GPC en 6, 8 y 7 años, respectivamente^{6,20}; además la media del coste en la revisión bibliográfica de éstas fue de 250.000 dólares, cifra lejana a los recursos que se utilizan en nuestro medio²².

Otro aspecto que contribuye a la disminución de calidad de las guías evaluadas es la obsolescencia de sus evidencias, que ha sido cifrada en unos 3 años²². En el área «rigor en la elaboración» que evalúa dicho atributo en el AGREE, sólo un 7% de nuestras guías han obtenido puntuación buena o muy buena. Por tanto, parece claro que nuestros profesionales en gran medida presentan un déficit de habilidad para elaborar este tipo de documentos³, bien por insuficiencia en la evaluación crítica de las evidencias científicas o por falta de destreza al elaborar y ponderar las recomendaciones. A ello sin duda contribuye la confusión existente por la diversidad de aproximaciones disponibles sobre gradación de recomendaciones, que se está intentando clarificar mediante una propuesta de síntesis entre las más aceptadas^{6,23}.

Respecto a la consistencia y precisión de la herramienta de evaluación, todos los 23 criterios agrupados en 6 atributos han podido ser cotejados en todas las guías, los valores de puntuación han sido muy similares en los cuatro evaluadores (concordancia simple entre 88%-98%) y hay una gran consistencia entre la puntuación obtenida por cada guía y la valoración global que recibía sobre si la guía merecía ser promocionada y difundida. Además, el observador extremo no era siempre el mismo.

TABLA 2
Reproducibilidad interevaluador del instrumento AGREE en cada una de sus seis áreas

	% acuerdo simple
Alcance	90,17
Participación	96,73
Rigor	96,73
Claridad	93,45
Aplicabilidad	98,36
Independencia	88,53

Además de la facilidad de uso del Instrumento hay que comentar el escaso tiempo medio que se ha precisado en su cotejo, aparte de la lectura detallada de la guía. En nuestro estudio la media de tiempo ha sido notablemente menor que en otros estudios¹⁸. Creemos que ello se debe a la baja calidad de la mayoría de las GPC evaluadas que no aportaban información sobre la metodología de su desarrollo. Este dato de baja calidad simplificó notablemente la evaluación, pues el tiempo utilizado para evaluar las 10 guías que obtuvieron mejor calificación fue siempre superior a 30 minutos. El AGREE también aporta un elemento de valoración objetivo junto al juicio subjetivo sobre la guía.

Entre las deficiencias del AGREE, resaltar la necesidad de un entrenamiento previo a su uso, así como de homogeneizar criterios entre los evaluadores, lo cual le resta utilidad en la vida real, donde no siempre será posible encontrar cuatro evaluadores, ni proceder con un período de aprendizaje. Desconocemos cómo afectaría la subjetividad del juicio de los evaluadores en la reproducibilidad de la evaluación en caso de que no se haga entrenamiento conjunto previo, siendo ésta una de las limitaciones del estudio.

Otra limitación del estudio es la derivada del tamaño de la muestra de guías evaluadas, que apenas representa un 21% de las identificadas en España, a pesar del insistente contacto con los autores (cuatro recuerdos) para tratar de recuperar los originales de las guías. Pensamos, no obstante, que la generalizabilidad de los resultados no se ve comprometida por ello, ya que las guías incluidas no fueron seleccionadas *ex profeso* y además la calificación puede estar incluso sobreestimada, ya que dichas guías se seleccionaron porque fueron enviadas por sus autores, y es de suponer que éstos enviaron las que consideraron mejores.

Las implicaciones prácticas del estudio son: 1) para mejorar la calidad media de las GPC disponibles en España habría que hacer un esfuerzo de formación metodológica de sus productores; 2) debe buscarse la cooperación de sociedades y grupos profesionales con instituciones centradas en la calidad de la asistencia, como las agencias de evaluación de tecnologías o la colaboración Cochrane, y por último, 3) debe introducirse una cierta programación que evite redundancias y facilite el apoyo a su desarrollo.

Este tipo de propuesta se está desarrollando por iniciativas como la Red de Investigación en Medicina Basada en la Evidencia, recientemente reconocida por el Instituto de Salud Carlos III.

En futuros estudios debe profundizarse en los criterios del contenido de las guías que disminuyen su calidad, tales como la falta de correlación entre evidencias y recomendaciones, precisión con la que éstas están redactadas o elementos más proclives a la obsolescencia de las guías, entre otros²⁴.

En resumen, nosotros encontramos que las GPC pueden ser evaluadas en la calidad de su desarrollo de modo objetivo y reproducible mediante la versión española del Instrumento AGREE y que la evaluación muestra una escasa fiabilidad de las guías, ya que sólo un 15 % de las guías elaboradas en España se pue-

den considerar de calidad buena o muy buena. Ante tal situación hemos confirmado que el usuario de guías clínicas puede confiar en el Instrumento AGREE que le permite discernir *a priori* si la guía que se le presenta merece su consideración.

Agradecimientos

Este trabajo se ha hecho dentro del proyecto REDEGUIAS financiado por el FIS 01/1057. También es parte de las actividades financiadas por la RED_MBE FIS: G03/090, y Consejería de Salud de Junta de Andalucía: PI 48/02. Agradecemos a Eduardo Briones Pérez de la Blanca e Ignacio Marín León por su colaboración en el diseño del trabajo. No ha existido conflicto de intereses en el desarrollo del estudio.

BIBLIOGRAFÍA

1. Calderón Sandubete E, Cotos Cancas R, Marín León I, Briones Pérez de la Blanca E, Pérez Cano B, Grilo Reina A y Grupo CAMBIE. Mapa de evidencias para el tratamiento de la angina inestable en la cabecera del paciente. Etapa preliminar para el desarrollo de una guía de práctica clínica. *Med Clin (Barc)*. 2000;114 Supl 2:19-23.
2. Wolf S, Grol R, Hutchinson A, Eccles M, Grinshaw J. Potential benefits, limitations, and harms of clinical guidelines. *Br Med J*. 1999;318:527-30.
3. Saturno Hernández PJ, Antón Botella JJ, Murcia Payá JF. Variabilidad en la protocolización de la captación, diagnóstico y valoración inicial del paciente hipertenso en 40 centros de salud. *Med Clin (Barc)*. 2000;114 Supl 2:14-8.
4. Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are Guidelines Following Guidelines? The methodological quality of clinical practice guidelines in the Peer Reviewed Medical Literature. *JAMA*. 1999;281:1900-5.
5. REDEGUIAS. Informe de Evaluación de Tecnologías FIS 01/1057. ISCIII. 2003. Disponible en: <http://www.redeguias.edu.es>
6. Marzó M, Alonso P, Bonfill X. *Med Clin (Barc)*. 2002;118 Supl 3:30-5.
7. Ollenschlaeger G. Clinical practice guidelines in Germany. The German Guidelines Clearinghouse. 20 febrero 1999. Disponible en: bmj.com/cgi/eletters/318/7182/527/DC1/1#2178
8. National Health and Medical Research Council. A guide to the development, implementation and evaluation of clinical practice guidelines. Canberra: National Health and Medical Research Council; 1998.
9. Canadian Medical Association. Guidelines for Canadian Practice Guidelines. 1994. Ottawa.
10. National Guideline Clearinghouse. AHRQ. Septiembre 2002. Disponible en: www.ahcpr.gov/clinic/cpgsix.htm
11. National Health and Medical Research Council. Camberra. New Zealand Guidelines Group NZCFE-BNNZHTA. New Zealand Evidence-Based Health Care Bulletin. 2000;1.
12. SIGN guidelines: an introduction to SIGN methodology for the development of evidence-based clinical guidelines. SIGN publication n.º 39, 1999. Edimburgh.
13. GUIASALUD. Disponible en: <http://www.guiasalud.org>
14. Rico Iturrioz R, Gutiérrez-Ibarluzea I, Asua Batarrita J, Navarro Puerto MA, Reyes Domínguez A, Marín León I, et al. *Rev Esp Salud Pública*. 2004;78:447-57.
15. The AGREE Collaboration. Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument. Septiembre 2002 Disponible en: www.agreecollaboration.org en español: <http://www.agreecollaboration.org/pdf/Spain.pdf>.
16. Field MJ, Lohr KN. Guidelines for clinical practice. From development to use. Washington DC: National Academy Press; 1992.
17. Fleiss JL. The measurement of interrater agreement. En: Statistical methods for rates and proportions, 2.ª ed. Nueva York: Wiley and Sons; 1981. p. 211-36.
18. Burguers JS, Grol R, Klazinga NS, Makela M, Zaat J for the AGREE Collaboration. Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs. *Int J Qual Health Care*. 2003;15:31-45.
19. Grilli R, Magrini N, Penna A, Mura G, Liberati A. Practice guidelines developed by specialty societies. The need for a critical appraisal. *Lancet*. 2000;355:103-6.
20. Burgers JS, Cluzeau FA, Hanna SE, Hunt C, Grol R. Characteristics of high-quality guidelines: evaluation of 86 clinical guidelines developed in ten European countries and Canada. *Int J Technol Assess Health Care*. 2003; 19:148-56.
21. Vader JP, Burnand B, Froehlich F, Dupriez K, Larequi-Lauber T, Pache I, et al. Appropriateness of upper gastrointestinal endoscopy: Comparison of American and Swiss criteria. *Int J Qual Health Care*. 1997;9:87-92.
22. Shekelle P, Ortiz E, Rhodes S, Morton S, Eccles M, Grimshaw J, et al. Validity of the Agency for Healthcare Research and Quality Clinical Practice Guidelines. How quickly do guidelines become updated? *JAMA*. 2001;286:1461-7.
23. Atkins D, Best D, Briss PA, Eccles M, Falck Ytter Y, Flottorp S, for the Grade Working Group. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328:1490-8.
24. Grol R, Eccles M, Maisonneuve H, Woolf S. Developing clinical practice guidelines: the european experience. *Dis Manage Health Outcomes*. 1998;4:255-66.