BMJ Open

Do guidelines offer implementation advice to target users?: A meta-review of guideline applicability.

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-007047
Article Type:	Research
Date Submitted by the Author:	28-Oct-2014
Complete List of Authors:	Gagliardi, Anna; University Health Network, Toronto General Research Institute Brouwers, Melissa; McMaster University,
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Evidence based practice
Keywords:	Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Change management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

Do guidelines offer implementation advice to target users?: A meta-review of guideline applicability.

Anna R Gagliardi, PhD, University Health Network, Toronto, Canada

Melissa C Brouwers, PhD, McMaster University, Hamilton, Canada

University Health Network, Toronto, Canada, M5G2C4, Scientist; McMaster University, Hamilton, Canada, L8V1C3, Associate Professor

Corresponding author:
Anna R Gagliardi
University Health Network
Toronto General Research Institute
200 Elizabeth Street
Toronto, Ontario, Canada, M5G2C4
TEL 416.340.4800
anna.gagliardi@uhnresearch.ca

Keywords: guidelines, implementation, quality appraisal, systematic review

Word count: 2,076

ABSTRACT

Objective: Providers and patients are most likely to use and benefit from guidelines accompanied by implementation support. Guidelines published in 2007 and earlier assessed with AGREE scored poorly for applicability, which reflects the inclusion of implementation instructions or tools. The purpose of this study was to examine the applicability of guidelines published in 2008 or later, and identify factors that may have influenced applicability.

Design: Meta-review of studies that used AGREE to assess guidelines published in 2008 or later.

Data Sources: MEDLINE and EMBASE were searched from 2008 to July 2014, and the reference lists of eligible items. Two individuals independently screened results for English language studies that reviewed guidelines of which at least half were published in 2008 or later using AGREE and reported all domain scores, and extracted data. Descriptive statistics were calculated across all domains. Data were observed for trends between applicability scores, other domain scores, and guideline or developer features.

Results: Of 245 search results, 53 were retrieved as potentially relevant, and 12 studies were eligible for review. The mean and median domain scores for applicability across all 254 guidelines in all studies were 40.9% and 39.0% (interquartile range 30.7% to 49.1%), respectively. Applicability scored lower than all other domains. While applicability was higher among guidelines in which other AGREE domains also scored well, factors such as version of AGREE, timing of guideline publication, nature or complexity of guideline topic, country of guideline publication and type of guideline developer did not appear to influence applicability.

Conclusions: Despite increasing recognition of the need for implementation tools, guidelines continue to lack such resources. To improve health care delivery and associated outcomes,

further research is needed to establish the type of tools both needed and desired by healthcare providers and consumers, and methods for developing high quality tools.

Strengths and limitations of this study

- This study found that, among 254 guidelines published from 2001 to 2013 (70.0% published in 2008 or later) described in 12 systematic reviews published from 2014 to 2014, the mean and median domain scores for applicability were 40.9% and 39.0%, respectively, and applicability scored lower than the other five AGREE domains
- Applicability of guidelines, which refers to the inclusion of implementation instructions and tools, did not improve subsequent to the publication of two similar meta-reviews in 2010 and 2012, respectively, which examined a total of 654 guidelines published from 1980 to 2007
- Timing of guideline publication, guideline topic, country of guideline publication, and type of guideline developer did not appear to be associated with applicability
- Our literature search may not have identified all relevant studies, few studies were eligible, and the AGREE instrument may not objectively appraise guidelines or high AGREE scores may not be a determinant of guideline use therefore further research is needed to identify strategies that promote and support the development of guideline implementation tools

BACKGROUND

Guidelines play a fundamental role in healthcare planning, delivery, evaluation and quality improvement. However, they are not consistently translated into policy or practice.[1-3] Interviews with users found they were frustrated with the vast number of guidelines and uncertain about how to implement them given numerous interacting contextual challenges.[4-6] Greenhalgh et al. described this as an evidence based medicine "crisis" and called for guideline-based tools that could be used by providers and patients to clarify the goals of care, quality and completeness of evidence, and relevance of potential benefits and harms.[7] Pronovost also advocated for the development of implementation tools such as instructions for assessing barriers and choosing corresponding implementation strategies, and point-of-care checklists that integrate recommendations for patients with comorbid conditions.[8]

Considerable evidence supports the assertion that guidelines featuring implementation instructions or tools such as those recommended by Greenhalgh et al.[7] and Pronovost[8] are more likely to be used.[9-11] For example, a systematic review of 68 studies of provider adherence to asthma guidelines found that decision support tools (electronic or paper-based guideline summaries, algorithms, history-taking template, asthma status reminders) increased prescribing and provision of patient self-education or action plans, and was the only intervention studied that reduced emergency department visits.[9] A Cochrane systematic review of eight studies found that mailing of print summaries improved compliance with care delivery recommendations.[10] A systematic review of 100 randomized/non-randomized studies involving 3,826 practitioners/practices caring for more than 92,895 patients found that nearly two thirds of

studies resulted in improved guideline adherence for diagnosis, prevention, disease management and prescribing.[11]

The Appraisal of Guidelines, Research and Evaluation (AGREE) instrument and its revised version, AGREE II, can be used to develop or appraise guidelines and related material in separate documents that may be published or publicly available on web sites.[12-13] AGREE II consists of 23 items grouped in six domains: scope and purpose, stakeholder involvement, rigor of development, clarity and presentation, applicability, and editorial independence.[13] The domain of applicability includes four items related to planning, undertaking and evaluating implementation – facilitators and barriers of guideline implementation, resource considerations, monitoring or audit criteria, and implementation instructions or tools similar to those recommended by Greenhalgh et al.[7] and Pronovost[8], and for which there is evidence of association with guideline use.[9-11] A meta-review by Alonso-Coello et al. of 42 studies in which 626 guidelines on a range of topics published in various countries from 1980 to 2007 were assessed with AGREE found that most guidelines scored low for applicability (mean 22%, 95% CI 20.4 to 23.9) relative to all other domains.[14] Another meta-review by Knai et al. of 28 European guidelines on a range of topics published from 2000 to 2007 similarly found that most guidelines scored low for applicability (mean 44%, range 0% to 100%) relative to all domains but editorial independence.[15] Although scoring reflects all domain items, not only the presence of implementation tools, the finding that applicability consistently scored lower than other domains across multiple years and types of guidelines are striking.

Limited use of guidelines contributes to omission of beneficial therapies, preventable harm, suboptimal patient outcomes or experiences, or waste of resources.[7,8] Alonso-Coello et al. and Knai et al.[14,15] showed that few guidelines featured implementation tools, which improve guideline use, but both studies were based on guidelines published in 2007 or earlier. This study reviewed the applicability of guidelines published in 2008 or later given emerging views and evidence regarding the need for implementation tools. A secondary purpose was to identify factors that may have influenced applicability. The findings may reveal whether additional guidance is needed to promote the development of guideline implementation tools, thereby enabling guideline use, and improved care delivery and associated outcomes.

METHODS

Approach

We conducted a meta-review of studies that used the original AGREE instrument or AGREE II (henceforth referred to collectively as AGREE) to evaluate the quality of guidelines. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) criteria guided reporting of the methods and findings (eTable 1).[16] A protocol was not registered, and ethics review was not required.

Searching and screening

MEDLINE and EMBASE were searched from 2008 to July 2014 for English language studies that assessed guidelines using AGREE. The search strategy (Table 1) was based on terms used to index previous meta-reviews.[14,15] The references of all eligible studies were also screened. Titles and abstracts of search results were reviewed independently by the principal investigator and a

trained research assistant. All items selected by at least one reviewer were retrieved for further assessment. Studies were eligible if they were systematic reviews, more than half of the guidelines they evaluated were published in 2008 or later, scores for all AGREE domains were reported, and either domain score was reported, or scores for individual items such that domain score could be calculated. Eligible studies reviewed all guidelines in a particular country, or all guidelines on a particular topic, clinical condition or type of patient management. Studies were not eligible if they primarily compared underlying evidence, development methods or recommendations across guidelines and did not report domain scores; evaluations served as a baseline needs assessment in a country new to guideline development; or guidelines were sampled from one organization only. Studies in the form of abstracts, letters, commentaries, or editorials were not eligible.

Table 1. Literature search strategy

Database: Ovid MEDLINE(R) without Revisions <1996 to June Week 2 2014> Search Strategy:

1	Practice Guidelines as Topic/st [Standards] (5129)
2	quality control/ (27721)
3	AGREE.mp. (14714)
4	2 or 3 (42328)
5	1 and 4 (204)
6	limit 5 to (english language and yr="2008-Current") (117)
7	limit 6 to (comment or editorial or interview or lectures or letter or news) (15)
8	6 not 7 (102)

Data collection and analysis

There are frameworks that identify multiple, often interacting factors that influence guideline use,[4-6] but there are no frameworks that identify why some guidelines and not others feature implementation tools. We postulated that type of developer, nature or complexity of guideline

topic, or year produced may have influenced decisions about whether to develop implementation tools. Data were collected on year published, clinical topic, range of years during which guidelines were published, number of guidelines appraised, number of appraisers, version of AGREE (original or II), and domain scores. Data were extracted and tabulated by the principal investigator, then independently reviewed by the research assistant. Descriptive statistics were calculated for all domains (mean, median, range, interquartile range). Correlation across domains was not undertaken due to the small number of eligible studies. Instead data were observed for trends between applicability and other domain scores, AGREE version, guideline publication date, country in which guidelines were produced, type of guideline developers, and guideline topic. The methodological quality of eligible studies was not scored since AMSTAR is appropriate for assessing systematic reviews of randomized controlled trials.[17] However, most of its 11 items (items 1,2,5-9) were screening criteria and therefore present across all studies.

RESULTS

Study characteristics

The initial search resulted in 245 articles, 53 were retrieved as potentially relevant, and 12 were eligible for review (Figure 1). Studies were published from 2010 to 2014 and reviewed 254 guidelines (range 7 to 101) published from 2001 to 2013 on numerous topics (Table 2).[18-29] Of these, 178 (70.0%) were published in 2008 or later. AGREE was applied by two reviewers in 7 studies and 4 reviewers in 5 studies. Ten of 12 studies used AGREE II.

Table 2. Data extracted from eligible studies

Study	Guideline topic	AGREE version (n appraisers)	Guidelines appraised (n), years published (2008+)	AGREE domain scores (%, range)					
				scope & purpose	stakeholder involvement	rigour of development	clarity of presentation	applicability	editorial independence
Tan 2010 (17)	psoriasis vulgaris	original (4)	8 2007-2009 (7)	97.0 94.0-100.0	64.5 58.0-81.0	98.0 89.0-100.0	97.3 92.0-100.0	74.1 53.0-89.0	87.0 50.0-100.0
Greuter 2012 (18)	diabetes in pregnancy	original (2)	8 2003-2010 (5)	55.6 28.0-94.0	26.7 4.0-100.0	40.0 12.0-93.0	72.5 63.0-100.0	58.4 33.0-94.0	18.8 0.0-50.0
Arevalo-Rodriguez 2012 (19)	dementia in Alzheimer's disease	II (4)	15 2005-2011 (9)	75.4 26.0-100.0	55.3 17.0-88.0	55.3 16.0-92.0	80.0 49.0-93.0	32.4 3.0-93.0	47.5 4.0-94.0
Al-Ansary 2013 (20)	hypertension	II (4)	11 2006-2011 (8)	47.6 25.0-83.0	39.4 12.5-75.0	30.3 8.3-86.4	65.7 44.4-88.9	38.1 16.6-72.0	47.6 4.1-88.0
Holmer 2013 (21)	glycemic control type 2 diabetes	II (2)	24 2007-2012 (19)	64.0 6.0-94.0	52.0 6.0-94.0	48.0 0.0-81.0	81.0 61.0-94.0	43.0 21.0-83.0	26.0 0.0-75.0
Lopez-Vargas 2013 (22)	chronic kidney disease	II (2)	11 2002-2011 (9)	75.0 25.0-100.0	63.0 14.0-97.0	67.0 20.0-96.0	81.0 64.0-94.0	46.0 10.0-90.0	67.0 0.0-100.0
Sabharwal 2013 (23)	cardiology, cardiac surgery	II (2)	101 2001-2011 (65)	85.1 61.0-100.0	58.5 39.9-80.6	46.0 16.7-83.3	81.8 27.8-100.0	22.4 4.2-54.2	28.8 0.0-62.5
Rohde 2013 (24)	aphasia in stroke management	II (2)	19 2001-2010 (13)	60.2 16.0-78.0	43.5 8.3-75.0	42.0 3.0-81.0	65.8 30.5-97.0	39.9 8.0-75.0	36.3 1.0-87.5
Lee 2014 (25)	acute procedural pain, pediatrics	II (2)	18 2001-2013 (12)	80.9 36.0-100.0	50.5 27.8-100.0	47.3 8.3-96.9	85.7 36.1-100.0	25.5 0.0-100.0	28.7 0.0-100.0
Nuckols 2014 (26)	opioid use for chronic pain	II (4 to 6)	13 2009-2012 (13)	69.0 39.0-89.0	52.0 23.0-77.0	48.0 20.0-84.0	71.0 37.0-93.0	37.0 13.0-56.0	44.0 0.0-88.0
Sabharwal 2014 (27)	orthopedic thrombo-prophylaxis	II (2)	7 2009-2013 (7)	91.8 6.0-100.0	83.0 73.8-90.5	88.1 81.3-96.4	91.8 88.1-97.6	63.4 42.8-80.3	70.4 60.7-75.0
Larmer 2014 (28)	management of osteoarthritis	II (4)	19 2001-2013 (11)	86.8 72.0-100.0	51.4 6.0-100.0	67.6 50.0-96.0	72.5 39.0-100.0	10.0 0.0-42.0	31.6 0.0-100.0
Mean				74.0	53.3	56.5	78.8	40.9	44.5
Median				75.2	52.0	48.0	80.5	39.0	40.2
Range				47.6-97.0	26.7-83.0	30.3-98.0	65.7-97.3	10.0-74.1	18.8-87.0
Interquartile range (difference)				63.1-85.5 (22.4)	48.8-59.6 (10.9)	45.0-67.2 (22.2)	72.1-82.8 (10.7)	30.7-49.1 (18.4)	28.8-52.5 (23.7)

Applicability scores

The mean and median domain scores for applicability across all guidelines in all studies were 40.9% and 39.0%, respectively. The interquartile range (30.7% to 49.1%, difference 18.4%) indicates that the applicability domain score was relatively consistent across studies. These results are higher than those reported by Alonso-Coello et al. (mean 22%, 95% CI 20.4 to 23.9%)[14] but somewhat lower than the findings of Knai et al. (mean 44%, range 0% to 100%).[15]

Factors influencing applicability

The mean and median applicability scores across all guidelines in all studies were lower than those of the other five AGREE domains (Table 2). The mean applicability score was highest in studies by Tan (74.1%)[18] and Sabharwal (63.4%)[28] which also featured high scores in most AGREE domains, unlike other eligible studies. These two studies varied in factors that may have influenced applicability and therefore could not be distinguished from others based on AGREE version (Tan original AGREE; Sabharwal AGREE II), timing of guideline publication (Tan 2007 to 2009; Sabharwal 2009 to 2013), nature or complexity of guideline topic (Tan management of psoriasis vulgaris; Sabharwal prevention of thromboprophylaxis during orthopedic surgery), country (both included guidelines from multiple, developed countries) or type of developer (Tan medical specialty societies; Sabharwal government, international consortium, and medical specialty societies).

DISCUSSION

Providers and patients are most likely to benefit from guidelines featuring implementation tools.[5,6,9-11] The applicability of guidelines published in 2008 or later has not markedly improved

compared with guidelines published in 2007 or earlier, and remains low compared with other AGREE domains.[14,15] While applicability was higher among guidelines in which other AGREE domains also scored well, associated factors were not identified. These findings are concerning given the intensity and cost of efforts to generate an ever-increasing body of guidelines which are not used. Although multiple factors other than implementation tools influence guideline use, including patient, provider, institutional and system-level issues, implementation tools are meant to overcome many of these barriers.[4-6] Furthermore, guideline developers, implementers and researchers said that, in comparison with other approaches for implementing guidelines, developing implementation tools was more feasible, could be widely applied, and was therefore more likely to impact guideline use.[30]

Several issues may limit the interpretation and use of these findings. The literature search may not have identified all relevant studies, however, we searched the two most relevant medical databases, and screening and data extraction were undertaken independently by two individuals to improve reliability. Few studies were eligible and, because we relied on published meta-reviews, the sample of guidelines was non-random. However eligible studies included 254 guidelines with a variety of characteristics so the findings may be generalizable across other guidelines. Eligible studies included guidelines that were published prior to 2008, however the majority (70.0%) were published in 2008 or later, none of the eligible studies were included in the previously published Alonso-Coello[14] and Knai[15] studies, and if we had imposed such strict eligibility criteria there would have remained only two eligible studies including 20 guidelines.[26,27] Others have noted several limitations of the AGREE instrument which was used to score guidelines in eligible studies.[17] For example, scoring of domain items can be subjective,

and domain or overall score has not been definitively associated with guideline use. However, AGREE remains the internationally-accepted gold standard for appraising guidelines. While this study did not find that guideline or developer characteristics were linked with applicability, it was exploratory in nature and examined preliminary hypotheses so ongoing research is needed to investigate the influence of these, and other factors. For example, acquisition and analysis of the content of individual guidelines included in systematic reviews that were eligible for this study may identify guideline characteristics that were unique to those in the Tan[18] and Sabharwal[28] studies which resulted in higher applicability scores. Alternatively, further investigation of other factors, for example, the characteristics and workflow of the intended users of these guidelines may provide some insight on why implementation tools were created for these guidelines. Despite these potential limitations, this study underscores the urgent need to create impetus and guidance that would support the development of guideline implementation tools.

AGREE and other initiatives such as GRADE and GLIA have improved the description of guideline methods, evidence and recommendations.[31,32] However, there has been far less scrutiny of accompanying implementation tools. We interviewed international guideline developers who said there was a demand for such resources among their constituents but they required guidance for developing implementation tools.[30] We and others analyzed guideline development and implementation instructional manuals and found that they lacked guidance for developing implementation tools.[33,34] Therefore we consulted with members of the international guideline community to generate a 12-item framework that can serve as the basis for evaluating, and endorsing or adapting existing guideline implementation tools, or developing new tools.[35]

Additional research is needed to examine the type of tools that are most needed and preferred by different types of guideline users, the types of implementation tools best suited for different guidelines, and the features of implementation tools that are associated with guideline use. Pronovost noted that developers may lack relevant expertise to develop implementation tools and encouraged them to partner with others such as implementation or social scientists.[8] Coordinating complex, protracted partnerships involving numerous stakeholders with differing interests can be challenging.[36,37] However, the Choosing Wisely initiative, in which numerous specialty societies and consumer groups partnered to develop shared decision making tools demonstrates that partnership is indeed possible when there is a widely recognized need for improvement.[38] Still, further research is needed to identify the capacity, including skills and resources needed to develop implementation tools.

AUTHOR CONTRIBUTIONS

Both ARG and MCB conceived the study, interpreted data, drafted the manuscript and approved the final version of this manuscript. ARG was responsible for collecting and analyzing data. Both authors agree to be accountable for the work, and provide data on which the manuscript was based. This research was conducted without peer-reviewed funding. all authors, external and internal, had full access to all of the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. ARG affirms that affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

FUNDING

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

COMPETING INTERESTS

All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare no relationships with companies that might have an interest in the submitted work in the previous 3 years; their spouses, partners, or children have no financial relationships that may be relevant to the submitted work; and have no non-financial interests that may be relevant to the submitted work.

DATA SHARING

All relevant data is available in the manuscript and supplemental files.

REFERENCES

1. McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. *NEJM*. 2003;348(26):2635-2645.
2. Sheldon TA, Cullum N, Dawson D, et al. What's the evidence that NICE guidance has been implemented? Results from a national evaluation using time series analysis, audit of patients' notes, and interviews. *BMJ*. 2004; 329(7473):999.
3. Runciman WB, Hunt TD, Hannaford NA, Hibbert PD, Westbrook JI, Coiera EW et al. CareTrack: assessing the appropriateness of health care delivery in Australia. *Med J Aust*. 2012;197(2):100-105.
4. Francke AL, Smit MC, de Veer AJE, Mistiaen P. Factors influencing the implementation of clinical guidelines for health care professionals. *BMC Med Inform Dec Mak*. 2008;8:38.
5. Mickan S, Burls A, Glasziou P. Patterns of "leakage" in the utilization of clinical guidelines: a systematic review. *Postgrad Med J*. 2011;87(1032):670-679.
6. McKillop A, Crisp J, Walsh K. Practice guidelines need to address the "how" and the "what" of implementation. *Primary Health Care Res Develop*. 2012;13(1):48-59.
7. Greenhalgh T, Howick J, Maskrey N for the Evidence Based Renaissance Group. Evidence based medicine: a movement in crisis? *BMJ*. 2014;348:g37225.
8. Pronovost PJ. Enhancing physicians' use of clinical guidelines. *JAMA*. 2013;310(23):2501-2502.
9. Okelo SO, Butz AM, Sharma R, et al. Interventions to modify health care provider adherence to asthma guidelines: a systematic review. *Pediatrics*. 2013;132(3):517-534.

10. Murthy L, Shepperd S, Clarke MJ, et al. Interventions to improve the use of systematic reviews in decision-making by health system managers, policy makers and clinicians. *Cochrane Database Syst Rev*. 2012;9:CD009401.
11. Garg AX, Adhikari NK, McDonald H et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes. *JAMA*. 2005;293:1223-1238.
12. Cluzeau F, Burgers J, Brouwers M, et al. 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(1):18-23.
13. Brouwers M, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting and evaluation in healthcare. *CMAJ*. 2010;182:e839-842.
14. Alonso-Coello P, Irfan A, Sola I, et al. The quality of clinical practice guidelines over the last two decades: a systematic review of guideline appraisal studies. *Qual Saf Health Care*. 2010;19(6):e58.
15. Knai C, Brusamento S, Legido-Quigley H, et al. Systematic review of the methodological quality of clinical guideline development for the management of chronic disease in Europe. *Health Policy*. 2012;107:157-167.
16. Moher D, Liberati A, Tetzlaff J, Altman DG and The PRISMA Group: Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *BMJ*. 2009;339:b2535.
17. Shea B J, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol*. 2007;7:10.

18. Tan JK, Wolfe BJ, Bulatovic R, Jones EB, Lo AY. Critical appraisal of quality of clinical practice guidelines for treatment of psoriasis vulgaris, 2006-2009. *J Invest Dermatol*. 2010; 130(10):2389-2395.

19. Greuter MJ, van Emmerik NM, Wouters MG, van Tulder MW. Quality of guidelines on the management of diabetes in pregnancy: a systematic review. *BMC Pregnancy Childbirth*. 2012;12:58.

20. Arevalo-Rodriguez I, Pedraza OL, Rodriguez A, et al. Alzheimer's disease dementia guidelines for diagnostic testing: A systematic review. *Am J Alzheimer's Dis & Other Dementias*. 2012;28(2): 111-119.

21. Al-Ansary LA, Tricco AC, Adi Y, et al. A systematic review of recent clinical practice guidelines on the diagnosis, assessment and management of hypertension. *PLOS ONE*. 2013;8(1):e53744.

22. Holmer HK, Ogden LA, Burda BU, Norris SL. Quality of clinical practice guidelines for glycemic control in type 2 diabetes mellitus. *PLOS ONE*. 2013;8(4):e58625.

23. Lopez-Vargas PA, Tong A, Sureshkumar P, Johnson DW, Craig JC. Prevention, detection and management of early chronic kidney disease: a systematic review of clinical practice guidelines. *Nephrology*. 2013;18(9):592-604.

24. Sabharwal S, Patel V, Nijjer SS, et al. Guidelines in cardiac clinical practice: evaluation of their methodological quality using the AGREE II instrument. *J R Soc Med*. 2013; 106(8):315-322.

25. Rohde A, Worrall L, Le Dorze G. Systematic review of the quality of clinical guidelines for aphasia in stroke management. *J Eval Clin Pract*. 2013;19:994-1003.

26. Lee GY, Yamada J, Kyololo O, Shorkey A, Stevens B. Pediatric clinical practice guidelines for acute procedural pain: a systematic review. *Pediatrics*. 2014;133(3):500-515.

27. Nuckols TK, Anderson L, Popescu I, et al. Opioid prescribing: a systematic review and critical appraisal of guidelines for chronic pain. *Ann Intern Med*. 2014;160(1):38-47.

28. Sabharwal S, Gauher S, Kyriacou S, Patel V, Holloway I, Athanasiou T. Quality assessment of guidelines on thromboprophylaxis in orthopaedic surgery. *Bone Joint J*. 2014;96(1):19-23.

29. Larmer PJ, Reay ND, Aubert ER, Kersten P. Systematic review of guidelines for the physical management of osteoarthritis. *Arch Phys Med Rehab*. 2014;95:375-89.

30. Gagliardi AR. "More bang for the buck": exploring optimal approaches for guideline implementation through interviews with international developers. *BMC Health Serv Res*. 2012;12:404.

31. Schunemann HJ, Oxman AD, Brozek J, et al. GRADE: assessing the quality of evidence for diagnostic recommendations. *Evid Based Med*. 2008;13(6):162-163.

32. Shiffman RN, Dixon J, Brandt C, et al. The GuideLine Implementability Appraisal (GLIA): development of an instrument to identify obstacles to guideline implementation. *BMC Med Inform Decis Mak*. 2005;5:23.

33. Gagliardi AR, Brouwers MC. Integrating guideline development and implementation: Analysis of guideline development manual instructions for generating implementation advice. *Implement Sci*. 2012;7:67.

34. Schunemann HJ, Wiercioch W, Etxeandia I, Falavigna M, Santesso N, Mustafa R et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. *CMAJ*. 2014;186(3):E123-142.

35. Gagliardi AR, Brouwers MC, Bhattacharyya O. A framework of the desirable features of guideline implementation tools (GItools): Delphi survey and assessment of GItools. *Implement Sci*. 2014;9:98.

36. King G, Currie M, Smith L, Servais M, McDougall J. A framework of operating models for interdisciplinary research programs in clinical service organizations. *Eval Program Plan*. 2008;31:160-173.

37. Hall JG, Buchan A, Cribb A, et al. A meeting of the minds: interdisciplinary research in the health sciences in Canada. *CMAJ*. 2006;175(7):763-771.

38. Choosing Wisely: an initiative of the ABIM Foundation. 2013 (http://www.choosingwisely.org).

FIGURE LEGEND

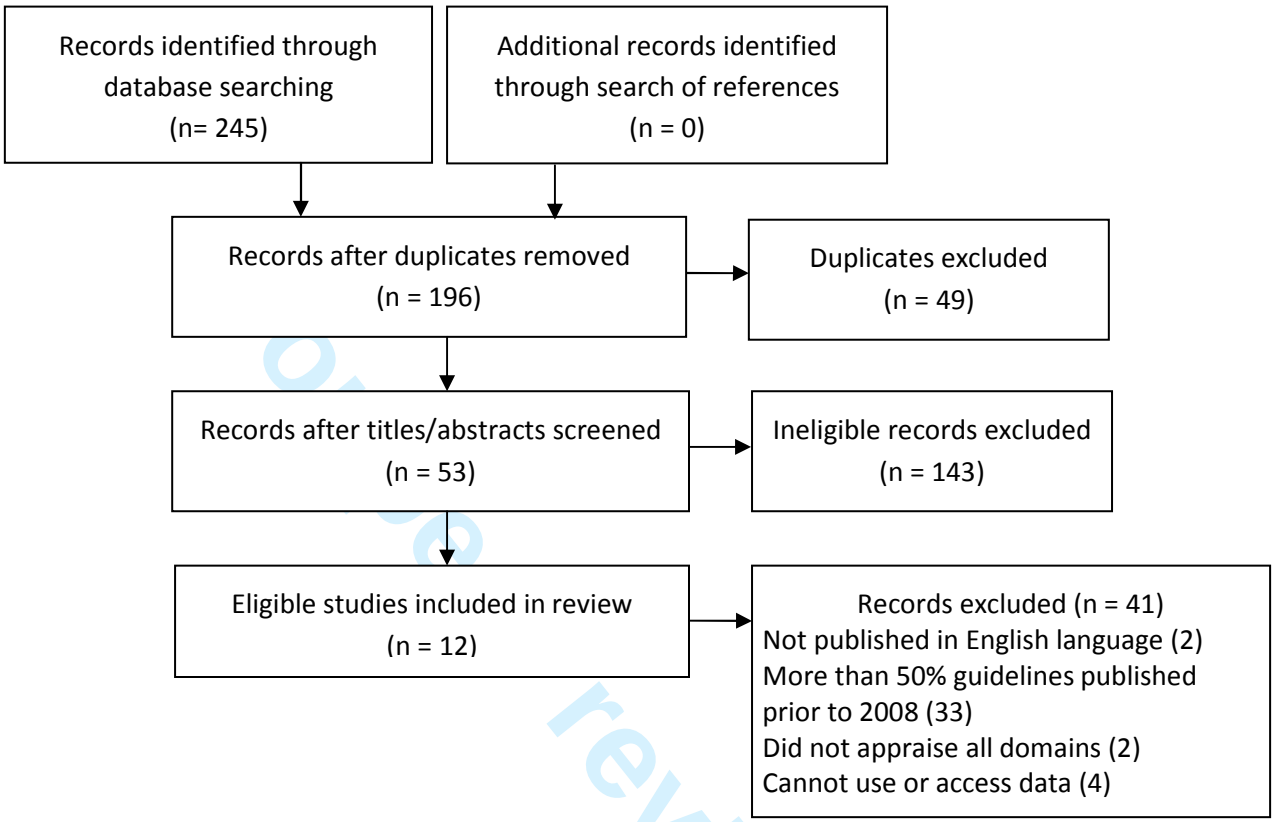
Figure 1. PRISMA flow diagram

eTable 1. PRISMA Checklist

Section/topic	#	Checklist item	Reported on page #
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
Abstract	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
Introduction			
Rationale	3	Describe the rationale for the review in the context of what is already known.	4-5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	6
Methods			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	6
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	8
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	8
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Table 1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	8-9
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	8-9
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	8-9
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	8-9
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	8-9
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	8-9

eTable 1. PRISMA Checklist, contd.

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	8-9
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	8-9
Results			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	9, Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	11, Table 2
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	n/a
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Table 2
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	11
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	n/a
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	11
Discussion			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	11-12
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	12-13
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	13-14
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	15



BMJ Open

Do guidelines offer implementation advice to target users?: A systematic review of guideline applicability.

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-007047.R1
Article Type:	Research
Date Submitted by the Author:	12-Jan-2015
Complete List of Authors:	Gagliardi, Anna; University Health Network, Toronto General Research Institute Brouwers, Melissa; McMaster University,
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Evidence based practice
Keywords:	Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Change management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

Do guidelines offer implementation advice to target users?: A systematic review of guideline applicability

Anna R Gagliardi, PhD, University Health Network, Toronto, Canada

Melissa C Brouwers, PhD, McMaster University, Hamilton, Canada

University Health Network, Toronto, Canada, M5G2C4, Scientist; McMaster University, Hamilton, Canada, L8V1C3, Associate Professor

Corresponding author:
Anna R Gagliardi
University Health Network
Toronto General Research Institute
200 Elizabeth Street
Toronto, Ontario, Canada, M5G2C4
TEL 416.340.4800
anna.gagliardi@uhnresearch.ca

Keywords: guidelines, implementation, quality appraisal, systematic review

Word count: 2,439

ABSTRACT

Objective: Providers and patients are most likely to use and benefit from guidelines accompanied by implementation support. Guidelines published in 2007 and earlier assessed with AGREE scored poorly for applicability, which reflects the inclusion of implementation instructions or tools. The purpose of this study was to examine the applicability of guidelines published in 2008 or later, and identify factors associated with applicability.

Design: Systematic review of studies that used AGREE to assess guidelines published in 2008 or later.

Data Sources: MEDLINE and EMBASE were searched from 2008 to July 2014, and the reference lists of eligible items. Two individuals independently screened results for English language studies that reviewed guidelines using AGREE and reported all domain scores, and extracted data. Descriptive statistics were calculated across all domains. Multi-level regression analysis with a mixed effect models identified factors associated with applicability.

Results: Of 245 search results, 53 were retrieved as potentially relevant, and 20 studies were eligible for review. The mean and median domain scores for applicability across 137 guidelines published in 2008 or later were 43.6% and 42.0% (interquartile range 21.8% to 63.0%), respectively. Applicability scored lower than all other domains, and did not markedly improve compared with guidelines published in 2007 or earlier. Country (United Kingdom) and type of developer (disease-specific foundation, non-profit health care system) appeared to be associated with applicability when assessed with AGREE II (not original AGREE).

Conclusions: Despite increasing recognition of the need for implementation tools, guidelines continue to lack such resources. To improve health care delivery and associated outcomes,

further research is needed to establish the type of implementation tools needed and desired by healthcare providers and consumers, and methods for developing high quality tools.

Strengths and limitations of this study

- This study found that, among 137 guidelines published from 2008 to 2013 described in systematic reviews published from 2010 to 2014, the mean and median domain scores for applicability were 43.6% and 42.0%, respectively, and applicability scored lower than the other five AGREE domains
- Applicability of guidelines, which refers to the inclusion of implementation instructions and tools, did not improve subsequent to the publication of two similar meta-reviews in 2010 and 2012, respectively, which examined a total of 654 guidelines published from 1980 to 2007
- Country (United Kingdom) and type of developer (disease-specific foundation, non-profit health care system) appeared to be associated with applicability when assessed using AGREE II (not original AGREE) though these findings should be interpreted with caution
- Our literature search may not have identified all relevant studies, the AGREE instrument may not objectively appraise guidelines, or high AGREE scores may not be a determinant of guideline use therefore further research is needed to identify strategies that promote and support the development of guideline implementation tools

BACKGROUND

Guidelines play a fundamental role in healthcare planning, delivery, evaluation and quality improvement. However, they are not consistently translated into policy or practice.[1-3] Interviews with users found they were frustrated with the vast number of guidelines and uncertain about how to implement them given numerous interacting contextual challenges.[4-6] Greenhalgh et al. described this as an evidence based medicine "crisis" and called for guideline-based tools that could be used by providers and patients to clarify the goals of care, quality and completeness of evidence, and relevance of potential benefits and harms.[7] Pronovost also advocated for the development of implementation tools such as instructions for assessing barriers and choosing corresponding implementation strategies, and point-of-care checklists that integrate recommendations for patients with comorbid conditions.[8]

Considerable evidence supports the assertion that guidelines featuring implementation instructions or tools such as those recommended by Greenhalgh et al.[7] and Pronovost[8] are more likely to be used.[9-11] For example, a systematic review of 68 studies of provider adherence to asthma guidelines found that decision support tools (electronic or paper-based guideline summaries, algorithms, history-taking template, asthma status reminders) increased prescribing and provision of patient self-education or action plans, and was the only intervention studied that reduced emergency department visits.[9] A Cochrane systematic review of eight studies found that mailing of print summaries improved compliance with care delivery recommendations.[10] A systematic review of 100 randomized/non-randomized studies involving 3,826 practitioners/practices caring for more than 92,895 patients found that nearly two thirds of

studies resulted in improved guideline adherence for diagnosis, prevention, disease management and prescribing.[11]

The Appraisal of Guidelines, Research and Evaluation (AGREE) instrument and its revised version, AGREE II, can be used to develop or appraise guidelines and related material in separate documents that may be published or publicly available on web sites.[12-13] AGREE II consists of 23 items grouped in six domains: scope and purpose, stakeholder involvement, rigor of development, clarity and presentation, applicability, and editorial independence.[13] The domain of applicability includes four items related to planning, undertaking and evaluating implementation – facilitators and barriers of guideline implementation, resource considerations, monitoring or audit criteria, and implementation instructions or tools similar to those recommended by Greenhalgh et al.[7] and Pronovost[8], and for which there is evidence of association with guideline use.[9-11] A meta-review by Alonso-Coello et al. of 42 studies in which 626 guidelines on a range of topics published in various countries from 1980 to 2007 were assessed with AGREE found that most guidelines scored low for applicability (mean 22%, 95% CI 20.4 to 23.9) relative to all other domains.[14] Another meta-review by Knai et al. of 28 European guidelines on a range of topics published from 2000 to 2007 similarly found that most guidelines scored low for applicability (mean 44%, range 0% to 100%) relative to all domains but editorial independence.[15] Although scoring reflects all domain items, not only the presence of implementation tools, the finding that applicability consistently scored lower than other domains across multiple years and types of guidelines is striking.

Limited use of guidelines contributes to omission of beneficial therapies, preventable harm, suboptimal patient outcomes or experiences, or waste of resources.[7,8] Alonso-Coello et al. and Knai et al.[14,15] showed that few guidelines featured implementation tools, which improve guideline use, but both studies were based on guidelines published in 2007 or earlier. This study reviewed the applicability of guidelines published in 2008 or later given emerging views and evidence regarding the need for implementation tools. A secondary purpose was to identify factors associated with applicability. The findings may reveal whether additional guidance is needed to promote the development of guideline implementation tools, thereby enabling guideline use, and improved care delivery and associated outcomes.

METHODS

Approach

We conducted a meta-review of studies that used the original AGREE instrument or AGREE II (henceforth referred to collectively as AGREE) to evaluate the quality of guidelines. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) criteria guided reporting of the methods and findings (eTable 1).[16] A protocol was not registered, and ethics review was not required.

Searching and screening

MEDLINE and EMBASE were searched from 2008 to July 2014 for English language studies that assessed guidelines using AGREE. The search strategy (Table 1) was based on terms used to index previous meta-reviews.[14,15] The references of all eligible studies were also screened. Titles and abstracts of search results were reviewed independently by the principal investigator and a

trained research assistant. All items selected by at least one reviewer were retrieved for further assessment. Studies were eligible if they were systematic reviews, one or more of the guidelines they evaluated were published in 2008 or later, guidelines were assessed by at last two reviewers, scores for all AGREE domains were reported, and either domain score, or scores for individual items such that domain score could be calculated were reported for each guideline. Eligible studies reviewed all guidelines in a particular country, or all guidelines on a particular topic, clinical condition or type of patient management. Studies were not eligible if they compared guideline content only, for example, underlying evidence, development methods or recommendations across guidelines, and did not report domain scores; evaluations served as a baseline needs assessment in a country new to guideline development since they had not yet developed capacity for generating guidelines; or guidelines were sampled from and assessed by the same organization since this would not reflect a range of factors of interest that might influence applicability, and potentially bias the assessment. Studies in the form of abstracts, letters, commentaries, or editorials were not eligible.

Table 1. Literature search strategy

Database: Ovid MEDLINE(R) without Revisions <1996 to June Week 2 2014> Search Strategy:

---

1 Practice Guidelines as Topic/st [Standards] (5129)
2 quality control/ (27721)
3 AGREE.mp. (14714)
4 2 or 3 (42328)
5 1 and 4 (204)
6 limit 5 to (english language and yr="2008-Current") (117)
7 limit 6 to (comment or editorial or interview or lectures or letter or news) (15)
8 6 not 7 (102)

Data collection and analysis

There are frameworks that identify multiple, often interacting factors that influence guideline use,[4-6] but there are no frameworks that identify why some guidelines and not others feature implementation tools. We postulated that type of developer, nature or complexity of guideline topic, year produced or AGREE version may have influenced decisions about whether to develop implementation tools. For eligible reviews, data were collected on year published, clinical topic, version of AGREE, range of years during which guidelines were published, number of guidelines appraised, and number of guidelines appraised that were published in 2008 or later. For individual guidelines published in 2008 or later included in eligible reviews, data were collected on date published, country, type of developer (professional organization, disease-specific organization, government agency, non-profit agency, health care system, academic organization), AGREE version and domain scores. Data were extracted and tabulated by the principal investigator, then independently reviewed by the research assistant. Descriptive statistics were calculated for all domains (mean, median, range, interquartile range). We tested the association between applicability score and guideline publication date, country, and type of developer using mixed effect models accounting for the review source as a nested variable. A secondary analysis was conducted testing the association between applicability and the covariates using the same statistical procedure stratified by AGREE. We used SAS 9.4 (SAS Institute, Cary NC) to conduct the analysis. All p values were two-sided, and reported as being statistically significant on the basis of a significance level of 0.05. The methodological quality of eligible studies was not scored since AMSTAR is appropriate for assessing systematic reviews of randomized controlled trials.[17] However, most of its 11 items (items 1,2,5-9) were screening criteria and therefore present across all studies.

RESULTS

Study characteristics

The search resulted in 245 articles, 53 were retrieved, and 20 were eligible for review (Figure 1). Studies were published from 2010 to 2014 and reviewed 254 guidelines (range 5 to 24) published from 1992 to 2013 on numerous topics (Table 2).[18-37] Guidelines were appraised with the original AGREE instrument in 9 studies and AGREE II in 11 studies. Of the guidelines included in eligible studies, 137 were published in 2008 or later.

Guideline characteristics

Of 137 guidelines, 33 (24.1%) were published in 2008, 37 (27.0%) in 2009, 28 (20.4%) in 2010, 22 (16.1%) in 2011, 14 (10.2%) in 2012, and 3 (2.2%) in 2013. Almost half were published by professional associations or societies (67, 48.9%). The remaining guidelines were published by government agencies (36, 26.3%), disease-specific organizations (16, 11.7%), non-profit health delivery systems (10, 7.3%), academic organizations (7, 5.1%) and one by the World Health Organization. Most guidelines were developed in the United States (46, 33.6%), United Kingdom (25, 18.2%) and Canada (20, 14.6%) and 13 (9.5%) by international groups. Several countries produced one or more guidelines included in the sample including Argentina, Australia, Brazil, Finland, France, Italy, Japan, Malaysia, Mexico, the Netherlands, New Zealand, Saudi Arabia, Singapore, South Africa, Spain, Sweden, and Turkey. Most guidelines were appraised using AGREE II (103, 75.2%).

Table 2. Characteristics of eligible studies

Study	Clinical Topic	AGREE version	Number of guidelines	Publication date of guidelines	Number of guidelines published 2008+
Lee 2014[18]	acute procedural pain, pediatrics	2	18	2001 to 2013	12
Nuckols 2014[19]	opioid use for chronic pain	2	13	2009 to 2012	12
Sabharwal 2014[20]	orthopedic thrombo-prophylaxis	2	7	2009 to 2013	7
Larmer 2014[21]	management of osteoarthritis	2	19	2001 to 2013	11
Lytras 2014[22]	occupational asthma	2	7	1992 to 2012	3
Al-Ansary 2013[23]	hypertension	2	11	2006 to 2011	8
Holmer 2013[24]	glycemic control type 2 diabetes	2	24	2007 to 2012	19
Lopez-Vargas 2013[25]	chronic kidney disease	2	11	2002 to 2011	6
Rohde 2013[26]	aphasia in stroke management	2	19	2001 to 2010	13
Greuter 2012[27]	diabetes in pregnancy	1	8	2003 to 2010	5
Arevalo-Rodriguez 2012[28]	dementia in Alzheimer's disease	2	15	2005 to 2011	9
Seixas 2012[29]	management of attention-deficit hyperactivity disorder	1	13	2001 to 2011	3
Pillastrini 2012[30]	management of low back pain in primary care	1	13	2002 to 2009	3
Tong 2011[31]	screening and follow-up of living kidney donors	2	10	1996 to 2010	3
De Hert 2011[32]	screening and monitoring of cardiometabolic risk in schizophrenia	1	18	2004 to 2010	7
Hurkmans 2011[33]	physiotherapy use in rheumatoid arthritis	1	8	2002 to 2009	2
Fortin 2011[34]	chronic conditions relevant to primary care	1	16	2003 to 2009	4
Tan 2010[35]	psoriasis vulgaris	1	8	2007 to 2009	7
McNair 2010[36]	primary care of lesbian, gay and bisexual people	1	11	1997 to 2010	2
Mahmud 2010[37]	preconception care of women with diabetes	1	5	2001 to 2009	2

Applicability scores

Table 3 summarizes scores for all AGREE domains. The mean and median domain scores for applicability across all guidelines were 43.6% and 42.0%, respectively. These were lower than the mean and median of all other domains for guidelines in the sample. These results are higher than those reported by Alonso-Coello et al. (mean 22%, 95% CI 20.4 to 23.9%)[14] and similar to the findings of Knai et al. (mean 44%, range 0% to 100%).[15] The spread across range and interquartile range for each domain shows that scores for all domains were inconsistent across guidelines but more so for editorial independence, rigour of development, then applicability, following by remaining domains.

Table 3. Domain scores for guidelines published in 2008 or later

Data	AGREE domain scores (%)					
	scope and purpose	stakeholder involvement	rigour of development	clarity of presentation	applicability	editorial independence
Mean	73.9	55.00	57.3	76.3	43.6	48.8
Median	78.0	53.0	57.3	83.0	42.0	50.0
Range	6.0 to 100.0	8.0 to 100.0	0.0 to 100.0	14.0 to 100.0	0.0 to 100.0	0.0 to 100.0
Interquartile range (difference)	61.1 to 94.0 (32.9)	39.0 to 73.0 (34.0)	31.0 to 81.0 (50.0)	64.0 to 91.7 (27.7)	21.8 to 63.0 (41.2)	12.5 to 79.0 (66.5)

Factors influencing applicability

An analysis of factors associated with applicability appears in Table 4. The estimated intra-class correlation was 0.47. Applicability mean score differed by year of guideline publication. Guidelines published in 2010 and 2012 were associated with higher applicability score than those published in 2013. The differences in mean applicability score for 2010 and 2012 were 26.5 ($p< 0.03$) and 28.3 ($p<0.02$), respectively. With respect to country, the highest mean applicability score was for guidelines developed in the United Kingdom. Guidelines developed by international groups, in Canada, or the United States had significantly lower applicability

scores compared with the United Kingdom. As for type of developer, disease-specific foundations and non-profit health care systems were associated with higher applicability scores than professional guideline developers. Mean applicability score differences were 16.2 (p=0.01) and 14.9 (p<0.04), respectively. When stratified by version of the AGREE instrument, 34 studies were included in the analysis for AGREE and 103 for AGREE II. The association between applicability score, year, country, and type of guideline developer remained significant for AGREE II only and not for AGREE (data not shown).

Table 4. Observed, adjusted and mean applicability score difference using mixed effect model controlling for publication year, country, and type of guideline developer nested within the review source, 2008-2013

Factor	Observed mean applicability score	Adjusted mean applicability score (95% Confidence Interval)	Mean applicability score difference	P value
Guideline publication year				
2008	47.0	47 (33.2 to 60.7)	24.0	0.48
2009	47.7	45.8 (32.1 to 59.5)	22.8	0.06
2010	47.7	49.5 (36.0 to 63.0)	26.5	*0.03
2011	31.1	41.3 (26.6 to 55.9)	18.3	0.14
2012	41.9	51.3 (35.2 to 67.4)	28.3	*0.02
2013	16.3	23 (-3.1 to 49.1)	Ref	
Country				
International group	35.9	35.7(21.9 to 49.4)	-26.1	*< 0.001
Canada	47.9	42.0 (26.7 to 57.3)	-19.8	*< 0.001
United States	38.6	32.3 (18.2 to 46.4)	- 29.5	*< 0.001
United Kingdom	64.4	61.8 (46.3 to 77.3)	Ref	
Type of guideline developer				
Disease-specific foundation	50.0	56.5 (42.7 to 70.3)	16.2	*0.01
Non-profit health care system	49.7	55.2 (38.5 to 71.9)	14.9	*0.04
Government agency	48.1	44.4 (31.4 to 57.4)	4.1	0.4
Academic organization	40.7	53.6 (34.1 to 73.1)	13.3	0.14
Professional organization	39.7	40.3 (28.7 to 51.9)	Ref	

* All p values two-sided, significance level 0.05

DISCUSSION

Providers and patients are most likely to benefit from guidelines featuring implementation tools.[5,6,9-11] The applicability of guidelines published in 2008 or later has not markedly improved compared with guidelines published in 2007 or earlier, did not increase over time from 2008 to 2013, and remains low compared with other AGREE domains.[14,15] Guidelines published in the United Kingdom, or by disease-specific foundations or non-profit health care systems appeared to be associated with higher applicability scores, though only when assessed using AGREE II. These findings are concerning given the intensity and cost of efforts to generate an ever-increasing body of guidelines which are not used. Although multiple factors other than implementation tools influence guideline use, including patient, provider, institutional and system-level issues, implementation tools are meant to overcome many of these barriers.[4-6] Furthermore, guideline developers, implementers and researchers said that, in comparison with other approaches for implementing guidelines, developing implementation tools was more feasible, could be widely applied, and was therefore more likely to impact guideline use.[38]

Several issues may limit the interpretation and use of these findings. The literature search may not have identified all relevant studies, however, we searched the two most relevant medical databases, and screening and data extraction were undertaken independently by two individuals to improve reliability. We relied on published meta-reviews so the sample of guidelines was non-random. However eligible studies included 137 guidelines published in 2008 or later with a variety of characteristics so the findings may be generalizable across other guidelines. Others have noted several limitations of the AGREE instrument which was used to score guidelines in eligible studies.[17] For example, scoring of domain items can be subjective, and domain or overall

score has not been definitively associated with guideline use. With respect to applicability, this information may be more likely found outside the guideline document compared with content reflecting other domains, rendering an assessment of applicability more challenging. However, AGREE remains the internationally-accepted gold standard for appraising guidelines. It is notable that associations between applicability scores and other factors were revealed by AGREE II[13] in which the definition and instructions for scoring of applicability were elaborated compared with the original AGREE instrument.[12] While this study found that country (United Kingdom) and type of developer (disease-specific foundation, non-profit health care system) were associated with applicability score, the finding may not be meaningful, in part because all non-profit health care systems were located in the United States and not the United Kingdom, and because fewer guidelines were produced by non-professional organizations. However, this study was exploratory in nature and examined preliminary hypotheses so ongoing research is needed to investigate the influence of these, and other factors, perhaps by repeating the same analysis once more meta-reviews were published. Alternatively, further investigation of other factors, for example, the characteristics and workflow of the intended users of these guidelines may provide some insight on why implementation tools were created for these guidelines. Despite these potential limitations, this study underscores the urgent need to create impetus and guidance that would support the development of guideline implementation tools.

AGREE and other initiatives such as GRADE and GLIA have improved the description of guideline methods, evidence and recommendations.[39,40] However, there has been far less scrutiny of accompanying implementation tools. We interviewed international guideline developers who said there was a demand for such resources among their constituents but they required guidance

for developing implementation tools.[38] We and others analyzed guideline development and implementation instructional manuals and found that they lacked guidance for developing implementation tools.[41,42] Therefore we consulted with members of the international guideline community to generate a 12-item framework that can serve as the basis for evaluating, and endorsing or adapting existing guideline implementation tools, or developing new tools.[43]

Additional research is needed to examine the type of tools that are most needed and preferred by different types of guideline users, the types of implementation tools best suited for different guidelines, and the features of implementation tools that are associated with guideline use. Pronovost noted that developers may lack relevant expertise to develop implementation tools and encouraged them to partner with others such as implementation or social scientists.[8] Coordinating complex, protracted partnerships involving numerous stakeholders with differing interests can be challenging.[44,45] However, the Choosing Wisely initiative, in which numerous specialty societies and consumer groups partnered to develop shared decision making tools demonstrates that partnership is indeed possible when there is a widely recognized need for improvement.[46] Still, further research is needed to identify the capacity, including skills and resources needed to develop implementation tools.

AUTHOR CONTRIBUTIONS

Both ARG and MCB conceived the study, interpreted data, drafted the manuscript and approved the final version of this manuscript. ARG was responsible for collecting and analyzing data. Both authors agree to be accountable for the work, and provide data on which the manuscript was based. This research was conducted without peer-reviewed funding. all authors, external and

internal, had full access to all of the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. ARG affirms that affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

FUNDING

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

COMPETING INTERESTS

All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare no relationships with companies that might have an interest in the submitted work in the previous 3 years; their spouses, partners, or children have no financial relationships that may be relevant to the submitted work; and have no non-financial interests that may be relevant to the submitted work.

DATA SHARING

All relevant data is available in the manuscript and supplemental files. No additional data available.

REFERENCES

1. McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. *NEJM*. 2003;348(26):2635-2645.
2. Sheldon TA, Cullum N, Dawson D, et al. What's the evidence that NICE guidance has been implemented? Results from a national evaluation using time series analysis, audit of patients' notes, and interviews. *BMJ*. 2004; 329(7473):999.
3. Runciman WB, Hunt TD, Hannaford NA, Hibbert PD, Westbrook JI, Coiera EW et al. CareTrack: assessing the appropriateness of health care delivery in Australia. *Med J Aust*. 2012;197(2):100-105.
4. Francke AL, Smit MC, de Veer AJE, Mistiaen P. Factors influencing the implementation of clinical guidelines for health care professionals. *BMC Med Inform Dec Mak*. 2008;8:38.
5. Mickan S, Burls A, Glasziou P. Patterns of "leakage" in the utilization of clinical guidelines: a systematic review. *Postgrad Med J*. 2011;87(1032):670-679.
6. McKillop A, Crisp J, Walsh K. Practice guidelines need to address the "how" and the "what" of implementation. *Primary Health Care Res Develop*. 2012;13(1):48-59.
7. Greenhalgh T, Howick J, Maskrey N for the Evidence Based Renaissance Group. Evidence based medicine: a movement in crisis? *BMJ*. 2014;348:g37225.
8. Pronovost PJ. Enhancing physicians' use of clinical guidelines. *JAMA*. 2013;310(23):2501-2502.
9. Okelo SO, Butz AM, Sharma R, et al. Interventions to modify health care provider adherence to asthma guidelines: a systematic review. *Pediatrics*. 2013;132(3):517-534.

10. Murthy L, Shepperd S, Clarke MJ, et al. Interventions to improve the use of systematic reviews in decision-making by health system managers, policy makers and clinicians. *Cochrane Database Syst Rev*. 2012;9:CD009401.
11. Garg AX, Adhikari NK, McDonald H et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes. *JAMA*. 2005;293:1223-1238.
12. Cluzeau F, Burgers J, Brouwers M, et al. 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(1):18-23.
13. Brouwers M, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting and evaluation in healthcare. *CMAJ*. 2010;182:e839-842.
14. Alonso-Coello P, Irfan A, Sola I, et al. The quality of clinical practice guidelines over the last two decades: a systematic review of guideline appraisal studies. *Qual Saf Health Care*. 2010;19(6):e58.
15. Knai C, Brusamento S, Legido-Quigley H, et al. Systematic review of the methodological quality of clinical guideline development for the management of chronic disease in Europe. *Health Policy*. 2012;107:157-167.
16. Moher D, Liberati A, Tetzlaff J, Altman DG and The PRISMA Group: Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *BMJ*. 2009;339:b2535.
17. Shea B J, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol*. 2007;7:10.

18. Lee GY, Yamada J, Kyololo O, Shorkey A, Stevens B. Pediatric clinical practice guidelines for acute procedural pain: a systematic review. *Pediatrics*. 2014;133(3):500-515.

19. Nuckols TK, Anderson L, Popescu I, et al. Opioid prescribing: a systematic review and critical appraisal of guidelines for chronic pain. *Ann Intern Med*. 2014;160(1):38-47.

20. Sabharwal S, Gauher S, Kyriacou S, Patel V, Holloway I, Athanasiou T. Quality assessment of guidelines on thromboprophylaxis in orthopaedic surgery. *Bone Joint J*. 2014;96(1):19-23.

21. Larmer PJ, Reay ND, Aubert ER, Kersten P. Systematic review of guidelines for the physical management of osteoarthritis. *Arch Phys Med Rehab*. 2014;95:375-89.

22. Lytras T, Bonovas S, Chronis C, et al. Occupational asthma guidelines: a systematic quality appraisal using the AGREE II instrument. *Occup Environ Med*. 2014;71: 81-86.

23. Al-Ansary LA, Tricco AC, Adi Y, et al. A systematic review of recent clinical practice guidelines on the diagnosis, assessment and management of hypertension. *PLOS ONE*. 2013;8(1):e53744.

24. Holmer HK, Ogden LA, Burda BU, Norris SL. Quality of clinical practice guidelines for glycemic control in type 2 diabetes mellitus. *PLOS ONE*. 2013;8(4):e58625.

25. Lopez-Vargas PA, Tong A, Sureshkumar P, Johnson DW, Craig JC. Prevention, detection and management of early chronic kidney disease: a systematic review of clinical practice guidelines. *Nephrology*. 2013;18(9):592-604.

26. Rohde A, Worrall L, Le Dorze G. Systematic review of the quality of clinical guidelines for aphasia in stroke management. *J Eval Clin Pract*. 2013;19:994-1003.

27. Greuter MJ, van Emmerik NM, Wouters MG, van Tulder MW. Quality of guidelines on the management of diabetes in pregnancy: a systematic review. *BMC Pregnancy Childbirth*. 2012;12:58.

28. Arevalo-Rodriguez I, Pedraza OL, Rodriguez A, et al. Alzheimer's disease dementia guidelines for diagnostic testing: A systematic review. *Am J Alzheimer's Dis & Other Dementias*. 2012;28(2): 111-119.

29. Seixas M, Weiss M, Muller U. Systematic review of national and international guidelines on attention-deficit hyperactivity disorder. *J Psychopharm*. 2012;26(6): 753-765.

30. Pillastrini P, Gardenghi I, Bonetti F, et al. An updated overview of clinical guidelines for chronic low back pain management in primary care. *Joint Bone Spine*. 2012;79: 176-185.

31. Tong A, Chapman JR, Wong G, et al. Screening and follow-up of living kidney donors: A systematic review of clinical practice guidelines. *Transplant*. 2011;92(9): 962-972.

32. De Hert M, Vancampfort D, Correll CU, et al. Guidelines for screening and monitoring of cardiometabolic risk in schizophrenia: systematic evaluation. *Br J Psych*. 2011;199: 99-105.

33. Hurkmans EJ, Jones A, Li LC, et al. Quality appraisal of clinical practice guidelines on the use of physiotherapy in rheumatoid arthritis: a systematic review. *Rheumatol*. 2011;50: 1879-1888.

34. Fortin M, Contant E, Savard C, et al. Canadian guidelines for clinical practice: an analysis of their quality and relevance to the care of adults with comorbidity. *BMC Fam Pract*. 2011;12: 74.

35. Tan JK, Wolfe BJ, Bulatovic R, Jones EB, Lo AY. Critical appraisal of quality of clinical practice guidelines for treatment of psoriasis vulgaris, 2006-2009. *J Invest Dermatol*. 2010; 130(10):2389-2395.

36. McNair RP, Hegarty K. Guidelines for the primary care of lesbian, gay and bisexual people: A systematic review. *Ann Fam Med*. 2010;8: 533-541.

37. Mahmud M, Mazza D. Preconception care of women with diabetes: a review of current guideline recommendations. *BMC Women's Health*. 2010;10: 5.

38. Gagliardi AR. "More bang for the buck": exploring optimal approaches for guideline implementation through interviews with international developers. *BMC Health Serv Res*. 2012;12:404.

39. Schunemann HJ, Oxman AD, Brozek J, et al. GRADE: assessing the quality of evidence for diagnostic recommendations. *Evid Based Med*. 2008;13(6):162-163.

40. Shiffman RN, Dixon J, Brandt C, et al. The GuideLine Implementability Appraisal (GLIA): development of an instrument to identify obstacles to guideline implementation. *BMC Med Inform Decis Mak*. 2005;5:23.

41. Gagliardi AR, Brouwers MC. Integrating guideline development and implementation: Analysis of guideline development manual instructions for generating implementation advice. *Implement Sci*. 2012;7:67.

42. Schunemann HJ, Wiercioch W, Etxeandia I, Falavigna M, Santesso N, Mustafa R et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. *CMAJ*. 2014;186(3):E123-142.

43. Gagliardi AR, Brouwers MC, Bhattacharyya O. A framework of the desirable features of guideline implementation tools (GItools): Delphi survey and assessment of GItools. *Implement Sci*. 2014;9:98.

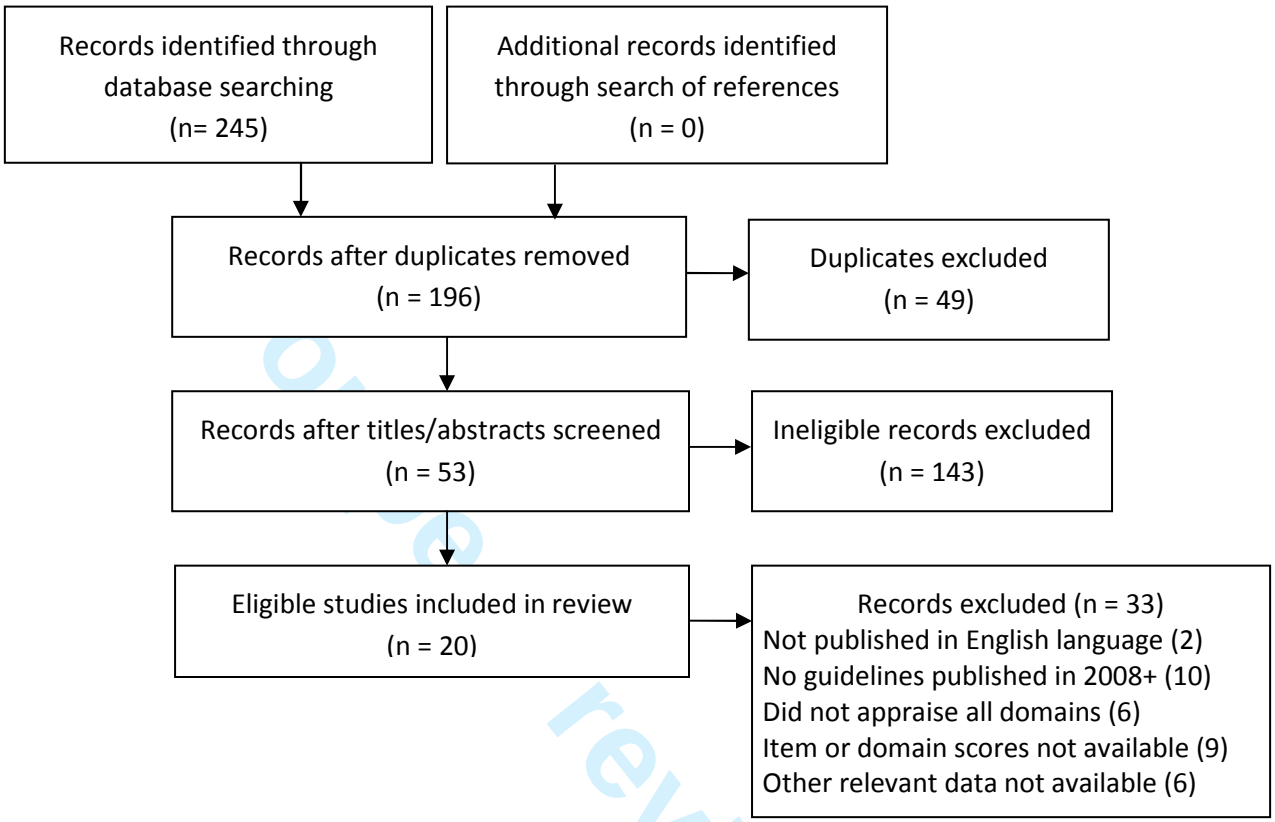
44. King G, Currie M, Smith L, Servais M, McDougall J. A framework of operating models for interdisciplinary research programs in clinical service organizations. *Eval Program Plan*. 2008;31:160-173.

45. Hall JG, Buchan A, Cribb A, et al. A meeting of the minds: interdisciplinary research in the health sciences in Canada. *CMAJ*. 2006;175(7):763-771.

46. Choosing Wisely: an initiative of the ABIM Foundation. 2013 (http://www.choosingwisely.org).

FIGURE LEGEND

Figure 1. PRISMA flow diagram



eTable 1. PRISMA Checklist

Section/topic	#	Checklist item	Reported on page #
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
Abstract	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
Introduction			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5-7
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	7
Methods			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	7
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	7-8
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	7-8
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	8, Table 1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	7-8
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	9
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	9
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	9
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	9
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	9

eTable 1. PRISMA Checklist, contd.

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	9
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	9
Results			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	10, Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	11, Table 2
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	n/a
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	12, Table 3
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	12, Table 3
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	n/a
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	12-13, Table 4
Discussion			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	14
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	14-15
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	14-16
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	17