The Development of Cancer Screening Guidelines

Robert A. Smith, PhD

KEYWORDS

• Cancer screening • Guidelines • Evidence-based • Guideline development

KEY POINTS

- Cancer screening guidelines have evolved over the years from being principally expert-driven advice to evidence-driven advice to clinicians, the public, and policy makers.
- Guideline differences are best explained by differences in guideline development methodology, the culture of the guideline developing organization, and the timing of the most recent update.
- Online tools exist to compare guidelines for trustworthiness, and can be used to judge how well a guideline meets Institute of Medicine standards.

INTRODUCTION

In the 1970s, a new generation of clinicians began to challenge conventional medical practice by simply asking, "how do we know this intervention works?" What came to be eventually known as "evidence-based medicine" (EBM)[1] grew from a movement where clinicians were expected to abandon *expert-based medicine*, that is, the practice of medicine based on enduring conventional clinical wisdom and what experts recommended and instead, manage patients based on what the evidence showed was effective. In a published summary of an oral history of EBM organized by the editors of the *Journal of the American Medical Association* and the *British Medical Journal,*[2] Smith and Rennie report this period was defined by the application of increasingly critical appraisal of expert-based medicine, where supporting evidence was sought, and often little was found.[3] During this period, leaders in this movement promoted rigorous research designs, in particular randomized controlled trials (RCTs)[4]; the new field of clinical epidemiology evolved to study clinical practice[4–10]; and organizations dedicated to the systematic evaluation and synthesis of research evidence, such as the Canadian Task Force on Preventive Health Care,[11] the United States Preventive Services Task Force,[12] and The Cochrane Collaboration,[13] were established.

Cancer Screening, Cancer Prevention and Early Detection Department, Center for Cancer Screening, American Cancer Society, 250 Williams Street, Northwest, Suite 600, Atlanta, GA 30303, USA
E-mail address: robert.smith@cancer.org

Med Clin N Am 104 (2020) 955–970
https://doi.org/10.1016/j.mcna.2020.08.009
medical.theclinics.com

In the 1991 editorial where he coined the expression "evidence-based medicine," Guyatt described a scenario where an internist questions the extent of her knowledge in a clinical scenario, conducts a literature search, appraises the citations she receives, and after reviewing the relevant article departs from the clinical course she would have followed had she not sought supporting evidence.[1] The clinician, practicing EBM, used her skills at "literature retrieval, critical appraisal, and information synthesis," which resulted in improved patient care.

As a practical matter, individual clinicians cannot routinely evaluate the literature to ensure the delivery of EBM to the full spectrum of medical care. However, increasingly today's learning pathways are evidence based, and much of basic and on-going modern medical education now is based on applying the principles of EBM to the synthesis of medical information. Clinical practice guidelines (CPGs) serve that same purpose, and with regular updates based on the accumulation of new evidence, existing CPGs can be affirmed or updated. Today, expert groups under the auspices of national health systems and professional societies fulfill that role, ideally following a rigorous methodology to synthesize, assess, and regularly update the clinical and scientific evidence to provide clinicians and the public with recommendations and guidance that is accurate and based on the latest scientific evidence. When patients may wish to make preference-sensitive decisions about undertaking an intervention, decision aids based on these assessments of the evidence of the benefits, limitations, and harms of the intervention can be useful for decision-making.

THE EVOLUTION OF CANCER SCREENING GUIDELINES/RECOMMENDATIONS

The earliest guidance about cancer screening was expert based. Dr. George Papanicolaou devoted many years attempting to persuade clinicians that cytology could be used to diagnose cervical cancer at an early treatable stage. After more than a decade of accumulating evidence, the publication of results in the *American Journal of obstetrics and Gynecology* from a RCT in 1941,[14] and a monograph 2 years later,[15] his ideas, enhanced by the work of others,[16] began to gain widespread acceptance.[17] The American Cancer Society (ACS) had supported Dr. Papanicolaou's research and in 1948 held an interdisciplinary conference to review and promote implementation of cervical cancer screening at a time when pathologists were not persuaded that a malignancy could be identified through the exfoliation of cancer cells.[18–20] In 1957, in what may have been the first cancer screening guideline, the ACS promoted annual screening with the Pap test,[19] and other organizations over time, such as the American College of Obstetricians and Gynecologists, issued similar guidance, with the annual interval very likely chosen for convenience.[17]

According to Winawer, nearly a century ago the concept of the adenoma-carcinoma sequence in the natural history of colorectal cancer was advanced by Lockhart-Mummery and Dukes in 1927,[21] and on-going work at St. Marks Hospital in London in the 1930s demonstrated that patients diagnosed at an earlier stage had better survival. These observations led to efforts to identify colorectal cancer early; although there was speculation that occult bleeding had to be present for some significant duration of time before symptomatic bleeding was apparent, but before the late 1960s, there was no reliable method to detect occult blood. Instead, early efforts to detect colorectal cancer in symptomatic adults focused on rigid sigmoidoscopy. However, in 1948, Gilbertsen and Nelms[22] at the University of Minnesota launched the first screening study based on the concept of detecting occult colorectal cancer in asymptomatic adults using a rigid sigmoidoscope. Although the study methodology had significant limitations, the investigators reported both lower than expected incidence of

colorectal cancer and better survival among individuals who had undergone screening. Over time additional studies of sigmoidoscopy were conducted, and in 1961, Day advocated for sigmoidoscopy as part of a cancer detection examination.[23] In 1967 Gregor reported that occult, early stage colorectal cancer could be detected at home using a new guaiac card test (Hemoccult) to detect occult blood[24] and in 1969 endorsed routine testing for the presence of fecal occult blood with the guaiac impregnated cards.[25] By 1974, the ACS was promoting annual stool testing for occult blood in patients older than 40 years, with periodic proctoscopy after age 40 or 50 years (the panel had varying opinions) for adults based on risk and air-contrast barium enema for high-risk patients.[26]

In 1973, encouraging, early results from the Health Insurance Plan (HIP) of Greater New York RCT of breast cancer screening[27] led the ACS and the National Cancer Institute (NCI) to launch the Breast Cancer Detection Demonstration Project (BCDDP). As of 1974, the ACS advised monthly breast self-examination (BSE) beginning at age 21 years, periodic clinical breast examination (CBE), with the interval (6 months–5 years) based on risk (not specified), and periodic mammography for women at high risk.[26] Because early HIP study results for women younger than 50 years were not encouraging, in 1977 a decision was made by the ACS and NCI to restrict BCDDP participation to average-risk women aged 50 years and older and to only offer screening to women younger than 50 years if they were at higher than average risk.[28] This joint statement issued by the ACS and NCI was the first formal breast cancer screening guideline.[29] Annual mammography was recommended for women aged 50 years and older; annual mammography was recommended for women aged 40 to 49 years, only if they had a personal history of breast cancer or a family history of breast cancer (mother or sister); annual screening was recommended for women aged 35 to 39 years if they had a personal history of breast cancer.[29] In addition, the ACS endorsed the importance of periodic CBE and monthly BSE.[29]

In 1980, the ACS adopted a formal evidence-based approach to guideline development that was led by David Eddy, MD, an early leader in the EBM movement, and colleagues at Stanford University.[30] Although prior guidelines had been based on evidence and expert opinion, the lack of methodologic rigor in study designs, the evaluation of the medical literature, and criticism of health screening associated with the growth of EBM led the ACS to subject its current recommendations[26] to the scrutiny of an outside expert who would apply modern principles of EBM to refine the recommendations.[20] The resulting guidelines for the cancer-related checkup were perhaps the very first application of EBM to cancer screening.

A NEW ERA IN GUIDELINE DEVELOPMENT

Since this early period, numerous North American groups have issued cancer screening guidelines for average risk adults, including the ACS, US Preventive Services Task Force (USPSTF), Canadian Task Force on Preventive Health Care, the American College of Physicians (ACP), American Academy of Family Physicians, American College of Radiology, and others. Over time, these groups have tended to adhere to different guideline development methodologies, ranging from little evidence of any systematic methodology to a well-documented, formal process; they have examined different evidence, with different rules for study inclusion and exclusion; and they have brought different values and judgments to assessing the balance of benefits and harms. Different schedules and frequency of updating recommendations also results in guideline differences, mainly due to differences in organizational perspective, methodology, and evidence reviews and guideline updates having taken

place in different time periods, with the most recent guideline differing simply because important new evidence had become available. These differences have contributed to a long history of variance in recommendations for cancer screening that has fueled controversies and frustrated policy makers, clinicians, and the target populations.

In 2011, 2 reports from the Institute of Medicine (IOM) established guidelines for systematic evidence reviews and guideline development.[31,32] The reports were motivated by concerns that the proliferation of CPGs had been accompanied by uneven quality, including incomplete inclusion of evidence, inclusion of poor-quality evidence, variable quality in guideline development, lack of transparency, concerns about conflicts of interest among guideline development group members and sponsoring organizations, and the difficulty reconciling conflicting guidelines. At the time of the of IOM publications, there were nearly 2700 CPGs in the Agency for Health Care Research's (AHRQ) National Guidelines Clearinghouse.[32] These 2 reports represented not only guidance for guideline development but also new benchmarks for evaluating the trustworthiness, transparency, and rigor of an evidence-based guideline. In 2011, the ACS also revised its guideline development process to be adherent with the IOM recommendations.[33]

The IOM outlined 8 principles and procedures for developing trustworthy CPGs, which are show in **Table 1** and outlined below in greater detail.[32]

Transparency

The importance of transparency is to ensure that the guideline development process, including the systematic review methodology, protocols, rules for decision-making,

Table 1
Institute of Medicine standards for developing trustworthy clinical practice guidelines

Standards	IOM Recommendations
Transparency	The process and funding of guideline development should be completely specified.
Conflicts of interest	Conflicts of interest include commercial, institutional, professional, and intellectual conflicts, all of which must be openly declared. Members should divest conflicting financial relationships.
Group composition	The guideline group should include multidisciplinary methodological experts, clinicians, and patient advocates.
Systematic review of evidence	The guidelines should be based on a systematic literature review that meets the standards set by the IOM.
Grading strength of recommendations	For each recommendation, the text should explain the evidence and the reasoning, the balance of benefits and harms, and should indicate the level of confidence in the recommendation.
Articulation of recommendations	Recommendations should be clearly stated and actionable.
External review	The draft guidelines should be posted for public comment, and the final guidelines should be revised as appropriate before peer review.
Updating	Guidelines should be updated when new evidence could result in modifying the recommendations.

Data from Institute of Medicine. Clinical Practice Guidelines We Can Trust. . Washington, DC: National Academies Press; 2011.

disclosures and real or potential conflicts of interest, sources of funding support, etc. are clearly and completely disclosed; in other words, who developed the guideline and the process by which it was derived. The IOM also stressed that it was important to explain the rules for inclusion and exclusion of evidence, how the data were interpreted, the basis for assessing the magnitude of the benefits and harms, and the basis for judgments about the balance of benefits and harms.[32]

Conflicts of Interest and Group Composition

In 2009, the IOM defined a conflict of interest (COI) as "A set of circumstances that creates a risk that professional judgment or actions regarding a primary interest will be unduly influenced by a secondary interest."[34] Note that the IOM did not state that the circumstances *would* create a COI, only that they *could* create a COI. In this respect, the common label of any interest as a COI creates a defensive situation where the participant must regard an interest and a COI as equivalent, as both an actual COI or the perception of a COI are treated as the same for purposes of avoiding suspicion that the guideline was not trustworthy.

Potential COIs may be financial, professional (the latter may be intertwined), institutional, or ideological. Financial COI occur when income is directly tied to guideline issues, that is, clinical services, industry-sponsored research, investments, consulting, etc. These COI may exist with individuals or the organization sponsoring the development of the guideline. With professional COIs, decisions about utilization of a screening tests may be perceived as going beyond evidence to promote greater use of a screening technology, which may also be perceived as a financial COI. This judgment does not presume that COI is inherent, only that it may be, and even if there is only the perception of a COI, confidence in the trustworthiness of a guideline can be diminished. Institutional COI may occur when a guideline panel member is associated with an organization with an interest in the guideline topic, or an institutional COI may exist if the organization developing the guideline has a financial relationship with commercial entities with an interest in the guideline outcome. A shortcoming of the IOM statement on COI in guideline development is the neglect of professional specialization or ideological bias where there is no direct or indirect potential for financial gain. A professional specialization may also be associated with a bias for or against screening, or an individual may have an ideological bias associated with a career-long orientation that has been unwavering in support or lack of support for screening.

Management of COI was a cornerstone of the IOM report on the development of trustworthy guidelines,[32] although the recommendations do not entirely overcome the tradeoffs between avoiding real or potential COI and the need for the clinical expertise of specialists. The IOM report cited strategies taken by some organizations to address COI, including omission from guideline development panels for any COI, a financial threshold, balancing membership in a guideline panel to minimize the number of members with COI, and allowing participation, but requiring recusal from specific deliberations and/or decision-making. The IOM report recommended that before selection of a guideline development panel, potential members should disclose "all current and planned commercial (including services from which a clinician derives a substantial proportion of income), noncommercial, intellectual, institutional, and patient–public activities pertinent to the potential scope of the CPG."[32] The IOM report concluded that when possible, guideline development panel members should not have any COIs. However, the IOM recognized that exclusion of experts because of COI could leave a panel without needed expertise and recommended that experts with COI should be a minority of members and that

chairs and co-chairs of the panel should not have any COIs.[32] Although this recommendation seems straightforward enough, it is not entirely feasible for a specialty organization that wishes to develop a clinical practice guideline and avoid *all* potential COI. It is not realistic to expect that a specialty organization would recruit a nonspecialist panel from outside the organization to develop their guideline, so a reasonable approach to avoiding COI can be based on recruiting some nonconflicted methodologists to be on the guideline development panel, with remaining members being specialists with minimum COI (ie, excluding members with investments, significant consulting relationships, etc.).

As a real-world example, the ACS approach to COI includes full disclosure of all potential financial, professional, institutional, and ideological COI, for which the latter includes a history of academic writing and presentations that are pertinent to a guideline under development. These disclosure statements are reviewed to determine if any interests are determined to represent a concerning level of real or potential or perceived COI. If so, the guideline development group member will be asked to recuse themselves from the development of the guideline. All disclosures are included as an appendix to the final guideline article. The ACS also separates the process of receiving expert input from the process determining the guideline and writing the guideline. Members of the ACS guideline development group include one patient advocate, and the remaining 11 members are generalist health care professionals and primary care physicians with expertise in the interpretation of evidence regarding benefits, limitations, and harms of clinical interventions, with some members having experience in the evaluation of screening. For each new guideline, the ACS establishes an expert advisory committee who are asked to consult with the guideline development panel on a regular basis and review draft protocols and systematic review methodology and early and final drafts of the guideline. This approach provides the guideline writing group with appropriate specialty expertise while also protecting it from the appearance of specialty COI.

Systematic Evidence Review

A systematic review of relevant evidence is an essential component of a credible, trustworthy guideline development process, and the companion report to *Clinical Practice Guidelines We Can Trust*[32] was *Finding What Works in Health Care: Standards for Systematic Reviews*.[31] The IOM defines a systematic review as "a scientific investigation that focuses on a specific question and uses explicit, preplanned scientific methods to identify, select, assess, and summarize the findings of individual, relevant studies."[31] The same principles described earlier for ensuring transparency and trustworthiness by avoiding bias and COI also apply to systematic evidence reviews. First, it is important to avoid bias and COI in the choice of the review team, and second, the systematic review must be guided by a detailed methodology for identification of evidence, criteria for inclusion and exclusion of evidence, and how the evidence will be evaluated. The IOM report summarizes (1) standards for initiating a systematic review, which mostly pertain to defining the scope of the topic and developing the protocol; (2) standards for literature searches and critical appraisal of studies; (3) standards for synthesizing the body of evidence; (4) standards for reporting the results of systematic reviews; and (5) issues related to the relationship between the systematic review team and the guideline writing panel. With respect to the relationship between the review team and the writing panel, the IOM describes various degrees of interaction, ranging from complete isolation of the systematic review team from the guideline development panel to the guideline development panel conducting the

systematic review and writing the guideline. Although the IOM report tends to favor more versus less isolation between the 2 groups, there is clear value to some interaction to ensure that the final systematic review meets the needs of the guideline development panel.

Perhaps the best-known example of this process are the systematic evidence reviews conducted for the USPSTF by the AHRQ Evidence-Based Practice Centers. These systematic reviews are the basis for the USPSTF's assessment of the scientific evidence for clinical preventive services. A condensed version of the review usually accompanies the publication of the recommendation statement. Systematic reviews are archived in the National Library of Medicine and can be accessed at https://www.ncbi.nlm.nih.gov/books/NBK43437/. These reviews generally accompany the recommendations. The USPSTF updated their methods for evidence reviews and recommendation development in 2007.[35]

Grading the Strength of the Evidence and Recommendations

It has become increasingly accepted that a key methodological element of a high-quality clinical practice guideline is an assessment of the quality of the evidence, which is tied to the strength of the recommendation. Ultimately, the strength of the recommendation reflects the possibility that new evidence might result in a different recommendation, the degree of certainty that desirable outcomes outweigh undesirable outcomes, and the degree of confidence that all patients would accept the intervention as worth undertaking.

The assessment of the quality of the evidence essentially is a measure of the confidence in the conclusions derived from the appraisal of the evidence. This degree of confidence is linked to research designs, which means that the highest quality evidence derives from RCTs, followed by controlled trials without randomization, cohort studies or case-control studies, and uncontrolled case series. Each of these methodologies must also be assessed for the quality of their design, sample size, etc., to arrive at an overall quality rating. In a well-designed systematic review, if a study is accepted for initial inclusion, at least 2 individuals independently will rate the quality of the study, and if there is disagreement, a final determination will be reached by consensus or by another reviewer. Studies will then receive a score (1–4, with subdivisions), a rating (good, fair, poor), or a letter grade (A, B, C). Other factors that may be considered in the overall assessment of the evidence are the generalizability of the studies, number of good-quality studies, and consistency of the findings in the literature. Rating evidence is intended to ensure that studies receive systematic scrutiny, study strengths and weaknesses are identified, and subjectivity is minimized. Still, as would be expected, there is considerable subjectivity and variation in judgment about the strength of evidence, even when using the same system, and quite often a key step in a systematic evidence review that is intended to convey transparency, that is, how judgments about study quality were reached, is not transparent at all.

The most common system for grading evidence and recommendations is the Grades of Recommendation, Assessment, Development, and Evaluation (GRADE), which is used by more than 100 organizations in 19 countries.[36,37] GRADE shares a common feature with most grading systems in that the system of evidence grading, and the strength of the recommendation highly depends on the ranking of the study methodology. Most of these systems were designed to evaluate therapeutic interventions in individuals who are being treated for a condition, and thus ideally there should be a sufficient number of RCTs from which to assess the efficacy of the intervention. In contrast, there are very few RCTs of screening, they vary considerably in their quality, and may reflect older technology and protocols. Because new RCTs are unlikely to be

funded, and would face ethical challenges anyway, modern evaluations of screening will be carried out with observational studies. Further, initial trials will be conducted in average risk populations, with demonstrations of efficacy applied to higher risk populations, for which RCTs are especially difficult. This means that under these grading systems, most of the study designs are inherently judged to be of moderate to low quality (typically "low"), and the strength of the recommendations rarely qualify as "strong;" rather, the next recommendation rating in the scale is "weak," although it is acceptable to substitute "qualified." A weak recommendation means "trade-offs [between benefits and harms] are less certain, either because of low-quality evidence or because evidence suggests that desirable and undesirable effects are closely balanced."[32] Essentially, when using the GRADE system, even against the backdrop of evidence from RCTs, a very well-designed observational study that has favorable findings usually will be judged as low- or moderate-quality evidence, resulting in a weak recommendation. GRADE does allow for a strong recommendation if the study is well designed and there is a clear dose-response relationship or a large observed effect, but it seems clear that in practice, a strong evidence grade is limited for RCTs. Observational studies commonly are judged to be second-class citizens. Although guideline developers understand that their recommendation carries the full confidence of the issuing organization that the intervention is recommended fully and not with hesitation, referring physicians and patients may interpret the recommendation language as conveying low confidence. This is a situation that must be the focus of further attention in guideline development methodology.

Articulation of Benefits and Harms

Guideline developers are expected to assess the evidence for harms associated with screening, and there is an expectation that there should be an assessment of whether benefits outweigh harms. This assessment may be associated with considerable subjectivity. Challenges include the comparison of different data sources, that is, intention-to-treat effects of benefit from a meta-analysis versus observational data from adults all of whom were exposed to the intervention. It is also clear that benefits associated with screening, such as avoiding a diagnosis of an advanced breast cancer or death from breast cancer, are very different metrics compared with being recalled for further imaging or undergoing a biopsy. Studies of harms may also have subjective elements in their methodology not easily discerned by the systematic review team, who may place greater scrutiny on study methodology influences on estimates of benefit than they do on studies of harms. It should be understood that subjectivity is unavoidable in the assessment of the balance of benefits and harms, and thus what is important is clear articulation of the basis for subjective judgments. For example, the USPSTF places strong emphasis on the recall rate and estimates of overdiagnosis as important harms in breast cancer screening.[38] In contrast, the ACS stated that it did not regard being recalled for further evaluation as an important harm, and although overdiagnosis was judged to be an important harm, the data were insufficient to estimate the magnitude of overdiagnosis as a harm with any measurable confidence.[39]

External Review

Once a guideline is developed, there is value in subjecting it to external review from subject-matter experts (including those who may have been advisors during the guideline development process), likely guideline advocates as well as likely detractors, and key stakeholder organizations that represent a broad spectrum of positions. The ACS and the USPSTF each subjects their draft guidelines to external review before

finalization, and these reviews not only have resulted in changes in narrative to improve clarity but in a few instances external reviews have resulted in significant changes in the recommendation statements.

In addition to feedback on the guideline or recommendation, external reviewers may identify small details that require correction, logic that is unclear and poorly explained, or gaps in logic and flaws in methodology. Reviewers may identify flaws in the underlying evidence for a recommendation that may appropriately weaken the strength of the recommendation. External reviewers may identify implications and consequences of a new guideline or guideline change that may not have been anticipated or fully appreciated. The review period should be regarded as a key opportunity to correct errors, improve the narrative, and even rethink a recommendation. It provides an opportunity for the guideline development panel to reflect on the entirety of their effort and to be sure that the guideline development process and the recommendations stand up to scrutiny. To be sure, some feedback may be inflammatory, baseless and ideological, and thus entirely useless other than providing a preview to how an organization will respond publicly to the new guideline once it is released. However, it is important to remember that guideline development is a rather insular endeavor; guideline implementation takes a village, and thus feedback from end-users is a valuable step in the process.

Guideline Updates

At the most basic level, a clinical practice guideline should reflect the current state of the evidence. Clinicians, policy makers, and the public expect that a guideline reflects the most up-to-date evidence and reasonably expect that when it does not, it will be updated. Shekelle and colleagues[40] outlined 6 situations that should lead to updating a guideline: a change in evidence related to benefits and harms; a change in important outcomes; a change in available interventions; a change in evidence that current practice is optimal; a change in values placed on outcomes (benefits or harms); and a change in available resources.

Given that guideline development is a major investment in time and resources, and more now than ever given the IOM guidance, a guideline should be updated periodically, and in the interim, there should be periodic reassurance that the current guideline still reflects best practice. If not, there should be public notice that an update is underway. The IOM emphasized the following best practices to reflect the currency of an existing guideline and considerations for periodic updates. First, a guideline must clearly identify the period from which the existing evidence is drawn. Second, the scientific literature must be monitored to identify relevant new evidence that could alter existing recommendations or reaffirm the current recommendation. Third, when new evidence may lead to a modification of the current guideline (new technology, new evidence related to the intervention protocol, new evidence on harms, or modification of the target population), a guideline update process should be initiated.[32]

GUIDELINE EVALUATION

Despite an extensive and growing literature on systematic reviews and guideline development, CPGs still vary in quality, and it is difficult for users of guidelines to scrutinize the lengthy checklist of esoteric criteria that determines where a guideline lands on a scale of trustworthiness. The ECRI Guidelines Trust evaluates registered guidelines for trustworthiness, applying a TRUST (**T**ransparency and **R**igor **U**sing **S**tandards of **T**rustworthiness) scorecard based on IOM standards.[41,42] These scores are available for most guidelines on their Website.[42] An additional tool for addressing the

Table 2
AGREE reporting checklist, 2016

AGREE Reporting Checklist
2016

This checklist is intended to guide the reporting of clinical practice guidelines.

CHECKLIST ITEM AND DESCRIPTION	REPORTING CRITERIA	Page #
DOMAIN 1: SCOPE AND PURPOSE		
1. OBJECTIVES *Report the overall objective(s) of the guideline. The expected health benefits from the guideline are to be specific to the clinical problem or health topic.*	☐ Health intent(s) (i.e., prevention, screening, diagnosis, treatment, etc.) ☐ Expected benefit(s) or outcome(s) ☐ Target(s) (e.g., patient population, society)	
2. QUESTIONS *Report the health question(s) covered by the guideline, particularly for the key recommendations.*	☐ Target population ☐ Intervention(s) or exposure(s) ☐ Comparisons (if appropriate) ☐ Outcome(s) ☐ Health care setting or context	
3. POPULATION *Describe the population (i.e., patients, public, etc.) to whom the guideline is meant to apply.*	☐ Target population, sex and age ☐ Clinical condition (if relevant) ☐ Severity/stage of disease (if relevant) ☐ Comorbidities (if relevant) ☐ Excluded populations (if relevant)	
DOMAIN 2: STAKEHOLDER INVOLVEMENT		
4. GROUP MEMBERSHIP *Report all individuals who were involved in the development process. This may include members of the steering group, the research team involved in selecting and reviewing/rating the evidence and individuals involved in formulating the final recommendations.*	☐ Name of participant ☐ Discipline/content expertise (e.g., neurosurgeon, methodologist) ☐ Institution (e.g., St. Peter's hospital) ☐ Geographical location (e.g., Seattle, WA) ☐ A description of the member's role in the guideline development group	
5. TARGET POPULATION PREFERENCES AND VIEWS *Report how the views and preferences of the target population were sought/considered and what the resulting outcomes were.*	☐ Statement of type of strategy used to capture patients'/publics' views and preferences (e.g., participation in the guideline development group, literature review of values and preferences) ☐ Methods by which preferences and views were sought (e.g., evidence from literature, surveys, focus groups) ☐ Outcomes/information gathered on patient/public information ☐ How the information gathered was used to inform the guideline development process and/or formation of the recommendations	
6. TARGET USERS *Report the target (or intended) users of the guideline.*	☐ The intended guideline audience (e.g. specialists, family physicians, patients, clinical or institutional leaders/administrators) ☐ How the guideline may be used by its target audience (e.g., to inform clinical decisions, to inform policy, to inform standards of care)	

(*continued on next page*)

Table 2
(*continued*)

DOMAIN 3: RIGOUR OF DEVELOPMENT		
7. SEARCH METHODS *Report details of the strategy used to search for evidence.*	☐ Named electronic database(s) or evidence source(s) where the search was performed (e.g., MEDLINE, EMBASE, PsychINFO, CINAHL) ☐ Time periods searched (e.g., January 1, 2004 to March 31, 2008) ☐ Search terms used (e.g., text words, indexing terms, subheadings) ☐ Full search strategy included (e.g., possibly located in appendix)	
8. EVIDENCE SELECTION CRITERIA *Report the criteria used to select (i.e., include and exclude) the evidence. Provide rationale, where appropriate.*	☐ Target population (patient, public, etc.) characteristics ☐ Study design ☐ Comparisons (if relevant) ☐ Outcomes ☐ Language (if relevant) ☐ Context (if relevant)	
9. STRENGTHS & LIMITATIONS OF THE EVIDENCE *Describe the strengths and limitations of the evidence. Consider from the perspective of the individual studies and the body of evidence aggregated across all the studies. Tools exist that can facilitate the reporting of this concept.*	☐ Study design(s) included in body of evidence ☐ Study methodology limitations (sampling, blinding, allocation concealment, analytical methods) ☐ Appropriateness/relevance of primary and secondary outcomes considered ☐ Consistency of results across studies ☐ Direction of results across studies ☐ Magnitude of benefit versus magnitude of harm ☐ Applicability to practice context	
10. FORMULATION OF RECOMMENDATIONS *Describe the methods used to formulate the recommendations and how final decisions were reached. Specify any areas of disagreement and the methods used to resolve them.*	☐ Recommendation development process (e.g., steps used in modified Delphi technique, voting procedures that were considered) ☐ Outcomes of the recommendation development process (e.g., extent to which consensus was reached using modified Delphi technique, outcome of voting procedures) ☐ How the process influenced the recommendations (e.g., results of Delphi technique influence final recommendation, alignment with recommendations and the final vote)	
11. CONSIDERATION OF BENEFITS AND HARMS *Report the health benefits, side effects, and risks that were considered when formulating the recommendations.*	☐ Supporting data and report of benefits ☐ Supporting data and report of harms/side effects/risks ☐ Reporting of the balance/trade-off between benefits and harms/side effects/risks ☐ Recommendations reflect considerations of both benefits and harms/side effects/risks	
12. LINK BETWEEN RECOMMENDATIONS AND EVIDENCE *Describe the explicit link between the recommendations and the evidence on which they are based.*	☐ How the guideline development group linked and used the evidence to inform recommendations ☐ Link between each recommendation and key evidence (text description and/or reference list) ☐ Link between recommendations and evidence summaries and/or evidence tables in the results section of the guideline	

(*continued on next page*)

Table 2
(continued)

13. EXTERNAL REVIEW *Report the methodology used to conduct the external review.*	☐ Purpose and intent of the external review (e.g., to improve quality, gather feedback on draft recommendations, assess applicability and feasibility, disseminate evidence) ☐ Methods taken to undertake the external review (e.g., rating scale, open-ended questions) ☐ Description of the external reviewers (e.g., number, type of reviewers, affiliations) ☐ Outcomes/information gathered from the external review (e.g., summary of key findings) ☐ How the information gathered was used to inform the guideline development process and/or formation of the recommendations (e.g., guideline panel considered results of review in forming final recommendations)	
14. UPDATING PROCEDURE *Describe the procedure for updating the guideline.*	☐ A statement that the guideline will be updated ☐ Explicit time interval or explicit criteria to guide decisions about when an update will occur ☐ Methodology for the updating procedure	
DOMAIN 4: CLARITY OF PRESENTATION		
15. SPECIFIC AND UNAMBIGUOUS RECOMMENDATIONS *Describe which options are appropriate in which situations and in which population groups, as informed by the body of evidence.*	☐ A statement of the recommended action ☐ Intent or purpose of the recommended action (e.g., to improve quality of life, to decrease side effects) ☐ Relevant population (e.g., patients, public) ☐ Caveats or qualifying statements, if relevant (e.g., patients or conditions for whom the recommendations would not apply) ☐ If there is uncertainty about the best care option(s), the uncertainty should be stated in the guideline	
16. MANAGEMENT OPTIONS *Describe the different options for managing the condition or health issue.*	☐ Description of management options ☐ Population or clinical situation most appropriate to each option	
17. IDENTIFIABLE KEY RECOMMENDATIONS *Present the key recommendations so that they are easy to identify.*	☐ Recommendations in a summarized box, typed in bold, underlined, or presented as flow charts or algorithms ☐ Specific recommendations grouped together in one section	
DOMAIN 5: APPLICABILITY		
18. FACILITATORS AND BARRIERS TO APPLICATION *Describe the facilitators and barriers to the guideline's application.*	☐ Types of facilitators and barriers that were considered ☐ Methods by which information regarding the facilitators and barriers to implementing recommendations were sought (e.g., feedback from key stakeholders, pilot testing of guidelines before widespread implementation) ☐ Information/description of the types of facilitators and barriers that emerged from the inquiry (e.g., practitioners have the skills to deliver the recommended care, sufficient equipment is not available to ensure all eligible members of the	

(*continued on next page*)

Table 2
(*continued*)

	population receive mammography) ☐ How the information influenced the guideline development process and/or formation of the recommendations	
19. IMPLEMENTATION ADVICE/TOOLS *Provide advice and/or tools on how the recommendations can be applied in practice.*	☐ Additional materials to support the implementation of the guideline in practice. For example: ▫ Guideline summary documents ▫ Links to check lists, algorithms ▫ Links to how-to manuals ▫ Solutions linked to barrier analysis (see Item 18) ▫ Tools to capitalize on guideline facilitators (see Item 18) ▫ Outcome of pilot test and lessons learned	
20. RESOURCE IMPLICATIONS *Describe any potential resource implications of applying the recommendations.*	☐ Types of cost information that were considered (e.g., economic evaluations, drug acquisition costs) ☐ Methods by which the cost information was sought (e.g., a health economist was part of the guideline development panel, use of health technology assessments for specific drugs, etc.) ☐ Information/description of the cost information that emerged from the inquiry (e.g., specific drug acquisition costs per treatment course) ☐ How the information gathered was used to inform the guideline development process and/or formation of the recommendations	
21. MONITORING/ AUDITING CRITERIA *Provide monitoring and/or auditing criteria to measure the application of guideline recommendations.*	☐ Criteria to assess guideline implementation or adherence to recommendations ☐ Criteria for assessing impact of implementing the recommendations ☐ Advice on the frequency and interval of measurement ☐ Operational definitions of how the criteria should be measured	
DOMAIN 6: EDITORIAL INDEPENDENCE		
22. FUNDING BODY *Report the funding body's influence on the content of the guideline.*	☐ The name of the funding body or source of funding (or explicit statement of no funding) ☐ A statement that the funding body did not influence the content of the guideline	
23. COMPETING INTERESTS *Provide an explicit statement that all group members have declared whether they have any competing interests.*	☐ Types of competing interests considered ☐ Methods by which potential competing interests were sought ☐ A description of the competing interests ☐ How the competing interests influenced the guideline process and development of recommendations	

From Brouwers MC, Kerkvliet K, Spithoff K, on behalf of the AGREE Next Steps Consortium. The AGREE Reporting Checklist: a tool to improve reporting of clinical practice guidelines. BMJ 2016;352:i1152. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5118873/. With permission.

variability in guideline quality and measuring the thoroughness and trustworthiness of a CPG is the Appraisal of Guidelines for Research and Evaluation (AGREE II) instrument, which measures 23 items in 6 quality domains, including (1) scope and purpose, (2) stakeholder involvement, (3) rigor of development, (4) clarity of presentation, (5) applicability, and (6) editorial independence (**Table 2**).[43] The ACP issues CPG updates by reviewing CPGs from other organizations and scrutinizing them with the AGREE II instrument.[44,45]

SUMMARY

The IOM reports that standards for systematic reviews and guideline development were prompted by a growing body of evidence revealing serious shortcomings in transparency and trustworthiness in the development of CPGs. Some organizations may have met these standards, but they were poorly documented; others had serious deficiencies ranging from glaring neglect of COI to weak scientific justification of recommendations. The IOM standards and the Appraisal of Guidelines Research and Evaluation (AGREE)[43] checklist each provides sound guidance for ensuring that a clinical practice guideline is credible, trustworthy, and can measure up to scrutiny. However, these recommendations should be regarded as a yardstick for both best practices and how the guideline may be assessed by outside groups. Ransohoff and colleagues[46] published a commentary on the new standards for trustworthiness that acknowledged the importance of the new standard, but rightfully pointed out that they mostly represented consensus judgments rather than practices based on evidence of their value, and that although well intentioned, they truly imposed an impractical and inflexible standard for trustworthiness. Ransohoff and colleagues[46] cited a recent study[47] that showed poor adherence to the new IOM standards among 114 clinical practice guidelines. Having failed to meet the new standards, were they all untrustworthy?

Guideline development methodology will continue to evolve. It is important to recognize that guidelines differ not only due to variations in guideline development methodology, including not only what evidence is included in the guideline review, but how it is interpreted, but also the judgment that a guideline developing group brings to interpretations about the balance of benefits and harms. What is essential in producing a trustworthy guideline is that both the process and the values and judgments that are the basis for the recommendations are clearly described.

DISCLOSURES

The work was supported by the American Cancer Society.

REFERENCES

1. Guyatt G. Evidence-based medicine. ACP J Club 1991;A-16.
2. Evidence-based medicine: an oral history. JAMA Network. 2018. Available at: https://ebm.jamanetwork.com/. Accessed July 19, 2020.
3. Smith R, Rennie D. Evidence-based medicine–an oral history. JAMA 2014;311:365–7.
4. Wilson JMG, Jungner G. Principles and practice of screening for disease. Geneva (Switzerland): World Health Organization; 1968.
5. Feinstein AR. Boolean algebra and clinical taxonomy. I. analytic synthesis of the general spectrum of a human disease. N Engl J Med 1963;269:929–38.
6. Feinstein AR. Scientific methodology in clinical medicine. i. introduction, principles, and concepts. Ann Intern Med 1964;61:564–79.
7. Cochrane AL. Effectiveness and efficiency: random reflections on health services. London (United Kingdom).: Nuffield Provincial Hospitals Trust; 1972.
8. Sackett DL. How to read clinical journals, I: why to read them and how to start reading them critically. Can Med Assoc J 1981;124:555–8.
9. Feinstein AR. Clinical epidemiology : the architecture of clinical research. Amsterdam (the Netherlands): Elsevier - Health Sciences Division; 1985.

10. Sackett DL, Haynes RB, Guyatt GH, et al. Clinical epidemiology: a basic science for clinical medicine. 2nd edition. Boston: Little Brown; 1991.
11. History of the Canadian task force on preventive health care. 2019. Available at: https://canadiantaskforce.ca/about/history/. Accessed July 19, 2020.
12. About the USPSTF. Agency for Healthcare Research and Quality. 2020. Available at: https://www.uspreventiveservicestaskforce.org/uspstf/about-uspstf. Accessed July 19, 2020.
13. Cochrane. 2020. Available at: https://www.cochrane.org. Accessed July 19, 2020.
14. Papanicolaou GN. The diagostic value of vaginal sears in carcinoma of the uterus. A J Obstet Gynecol 1941;42:193–206.
15. Papanicolaou GN, Traut HF. Diagnosis of uterine cancer by the vaginal smear. New York: Commonwealth Fund; 1943.
16. Ayre JE, Chevalier PM, Ayre WB. A comparative study of vaginal and cervical cornification in human subjects. J Clin Endocrinol Metab 1947;7:749–52.
17. Waxman AG. Guidelines for cervical cancer screening: history and scientific rationale. Clin Obstet Gynecol 2005;48:77–97.
18. Breslow L. A history of cancer control in the US with emphasis on the period 1946-1971. Los Angeles (CA): University of California at Los Angeles School of Public Health; 1977.
19. Eddy D. Cervical Cancer Screening. CA Cancer J Clin 1980;30:215–23.
20. Gusberg SB. The Role of Guidelines for Cancer Screening. In: Mettlin C, Murphy GP, editors. Progress in clinical and biological research: issues in cancer screening and communication. New York: Alan R. Liss, Inc.; 1982. p. 25–32.
21. Lockhart-Mummery JP, Dukes C. The precancerous changes in the rectum and colon. Surg Gynecol Obstet 1927;36:591–6.
22. Gilbertsen VA, Nelms JM. The prevention of invasive cancer of the rectum. Cancer 1978;41:1137–9.
23. Day E. The cancer detection examination. CA Cancer J Clin 1961;11:103–6.
24. Greegor DH. Diagnosis of large-bowel cancer in the asymptomatic patient. JAMA 1967;201:943–5.
25. Greegor DH. Detection of silent colon cancer in routine examination. CA Cancer J Clin 1969;19:330–7.
26. American Cancer Society. Summary of recommendations for early detection. Cancer 1974;33:1759–61.
27. Strax P, Venet L, Shapiro S. Value of mammography in reduction of mortality from breast cancer in mass screening. Am J roentgenol radium Ther Nucl Med 1973; 117:686–9.
28. National Cancer Institute. National Institutes of Health/National Cancer Institute consensus development meeting on breast cancer screening: issues and recommendations. J Natl Cancer Inst 1978;60:1519–21.
29. Dodd GD. American cancer society guidelines on screening for breast cancer. An overview. Cancer 1992;69:1885–7.
30. Eddy D. ACS report on the cancer-related health checkup. CA Cancer J Clin 1980;30:193–240.
31. Institute of Medicine. Finding what works in health care: standards for systematic reviews. Washington, DC: National Academies Press; 2011.
32. Institute of Medicine. Clinical practice guidelines we can Trust. Washington, DC: National Academies Press; 2011.
33. Brawley O, Byers T, Chen A, et al. New American cancer society process for creating trustworthy cancer screening guidelines. JAMA 2011;306:2495–9.

34. Institute of Medicine. Conflicts of interest in medical research, education, and practice. Washington, DC.: National Academies Press; 2009.
35. Guirguis-Blake J, Calonge N, Miller T, et al. Current processes of the U.S. preventive services task force: refining evidence-based recommendation development. Ann Intern Med 2007;147:117–22.
36. Atkins D, Eccles M, Flottorp S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches the GRADE working group. BMC Health Serv Res 2004;4:38.
37. Available at: http://www.gradeworkinggroup.org/. Accessed July 10, 2020.
38. Siu AL, U.S. Preventive Services Task Force. Screening for breast cancer: U.S. preventive services task force recommendation statement. Ann Intern Med 2016;164:279–96.
39. Oeffinger KC, Fontham ET, Etzioni R, et al. Breast cancer screening for women at average risk: 2015 guideline update from the American cancer society. JAMA 2015;314:1599–614.
40. Shekelle P, Eccles MP, Grimshaw JM, et al. When should clinical guidelines be updated? BMJ 2001;323:155–7.
41. Jue JJ, Cunningham S, Lohr K, et al. Developing and testing the agency for healthcare research and quality's national guideline clearinghouse extent of adherence to trustworthy standards (NEATS) instrument. Ann Intern Med 2019;170:480–7.
42. ECRI guidelines Trust. ECRI guidelines Trust, 2020. Available at: https://guidelines.ecri.org. Accessed July 24, 2020.
43. Brouwers MC, Kerkvliet K, Spithoff K, et al. The AGREE reporting checklist: a tool to improve reporting of clinical practice guidelines. BMJ 2016;352:i1152.
44. Qaseem A, Lin JS, Mustafa RA, et al. Clinical guidelines committee of the american college of p. screening for breast cancer in average-risk women: a guidance statement from the american college of physicians. Ann Intern Med 2019;170:547–60.
45. Qaseem A, Crandall CJ, Mustafa RA, et al. Clinical guidelines committee of the american college of p. screening for colorectal cancer in asymptomatic average-risk adults: a guidance statement from the american college of physicians. Ann Intern Med 2019;171:643–54.
46. Ransohoff DF, Pignone M, Sox HC. How to decide whether a clinical practice guideline is trustworthy. JAMA 2013;309:139–40.
47. Kung J, Miller RR, Mackowiak PA. Failure of Clinical Practice Guidelines to Meet Institute of Medicine Standards: Two More Decades of Little, If Any, Progress. Arch Intern Med 2012;172(21):1–6.