MAYO CLINIC

REVIEW

Clinical Practice Guidelines: A Primer on Development and Dissemination

M. Hassan Murad, MD, MPH

Abstract

Trustworthy clinical practice guidelines should be based on a systematic review of the literature, provide ratings of the quality of evidence and the strength of recommendations, consider patient values, and be developed by a multidisciplinary panel of experts. The quality of evidence reflects our certainty that the evidence warrants a particular action. Transforming evidence into a decision requires consideration of the quality of evidence, balance of benefits and harms, patients' values, available resources, feasibility of the intervention, acceptability by stakeholders, and effect on health equity. Empirical evidence shows that adherence to guidelines improves patient outcomes; however, adherence to guidelines is variable. Therefore, guidelines require active dissemination and innovative implementation strategies.

 ■ Mayo Clin Proc. 2017;92(3):423-433

Clinical practice guidelines are systematically developed statements that intend to assist clinicians and patients in making decisions about appropriate health care in specific circumstances.[1] Guidelines aim to improve the quality of patient care by encouraging interventions of proven benefit and discouraging the use of ineffective or potentially harmful interventions; to reduce unnecessary variation in practice; to lessen disparities; to empower patients; and to influence public policy.[2]

Production of guidelines has skyrocketed during the past 30 years. Currently, the library of the Guidelines International Network has 6187 documents from 76 countries, and the National Guideline Clearinghouse in the United States has 2017 guideline summaries.[3,4] Guidelines are critical for developing disease performance measures and defining high-value care.

This primer includes a description of the modern approach to developing guidelines; the criteria for trustworthy guidelines; advice on how clinicians can appraise, interpret, and implement practice recommendations; challenges and limitations of guidelines; and a future research agenda to address current knowledge gaps.

HISTORICAL PERSPECTIVE

Until the 1970s, medical actions were indirectly regulated through the training and credentials granted by medical schools or state authorities; however, such credentialing proved to be an insufficient guarantee of quality.[5] Further standardization and organization of the medical profession necessitated the development of guidelines. Guidelines in their current form started in the 1970s and were primarily based on the consensus of expert panels (eg, the National Institutes of Health Consensus Development Program).[6] Experts recommended management approaches they have used in their practice and cited references they recalled or were able to identify without an explicit systematic search. With the emergence of evidence-based medicine as a principle for decision making in the 1980s and coining of the term in 1991,[7] more rigorous approaches for guideline development have emerged. The next generation of guidelines would emphasize research evidence over opinion and base recommendations on the design of studies contributing evidence on benefits and harms of interventions. For example, the American College of Cardiology/American Heart Association used ratings of A, B, and C that were exclusively dependent on study design (level A, multiple randomized trials or meta-analyses; level B, a single trial or nonrandomized studies; and level C, consensus opinion of experts, case studies, or standard of care).[8] In 2000, many different guideline systems existed, which

For editorial comment, see page 327

From the Division of Preventive Medicine, Mayo Clinic, Rochester, MN.

ARTICLE HIGHLIGHTS

- Trustworthy clinical practice guidelines require a systematic review to select the best available evidence and should explicitly evaluate the quality of evidence.
- Factors that reduce the quality of evidence are risk of bias, indirectness, inconsistency, imprecision, and likelihood of publication and reporting bias.
- Transforming evidence into a decision requires consideration of the quality of evidence, balance of benefits and harms, patients' values, resources, feasibility, acceptability, and equity.
- Empirical evidence shows that guidelines improve patient outcomes; however, guidelines require active dissemination and innovative implementation strategies.

was confusing for stakeholders. The 6 most prominent systems had low reproducibility of judgments and did not fit the needs of all stakeholders.[9]

Subsequently, it became apparent that study design (a surrogate for the risk of bias) was insufficient.[9] Studies with the same design (eg, randomized trials) can have high risk of bias or low risk of bias. Factors other than risk of bias affect certainty in the evidence (eg, precision of estimates of effect, consistency of effect across studies). Moreover, factors other than evidence (eg, patient values) affect decision making.[10]

Considering these challenges and to unite many of these frameworks, the GRADE approach (Grading of Recommendations, Assessment, Development, and Evaluation) was developed in 2003. It advanced guideline methodology further by providing a framework for rating the quality of evidence based on 8 distinct domains (as opposed to study design only). The construct of the quality of evidence was then defined to reflect the extent of our confidence that the estimate of an effect is adequate to support a particular decision or recommendation.[11] An evidence to decision framework based on 8 criteria (as opposed to intuitive or global judgment) was provided by GRADE to assist in developing actions based on evidence.[12] Empirical evaluation showed that recommendations with a rating of A (based on multiple randomized trials or meta-analyses) would have been rated using GRADE as high, moderate, low, and very low.[13] Thus, this newer approach uncovered additional factors affecting the quality of evidence that were otherwise implicit. Judgments made using GRADE had good reproducibility and reliability compared with intuitive or global judgments and were consistent among raters (interrater reliability of 0.72)[14] even when panel members had short training in GRADE (two 1-hour didactic sessions).[15] GRADE was adopted by more than 100 organizations and has become, to some extent, the gold standard (when other systems are used,[16,17] they usually depend on GRADE domains and components).

In 2011, the National Academy of Medicine (formerly the Institute of Medicine) published criteria for trustworthy guidelines that greatly overlapped with GRADE (in emphasis on the systematic review process, rating the quality of evidence and strength of recommendations, and considering nonevidence factors).[18] These criteria were highly disseminated and cited, and they motivated guideline developers to improve the rigor of guidelines (Table 1). Similar criteria were also produced by the Guidelines International Network,[19] the National Institute for Health and Clinical Excellence in the United Kingdom,[20] and the World Health Organization.[21]

EVALUATING THE QUALITY OF EVIDENCE

A systematic review is a mechanism to reduce the risk of biased selection of evidence and should be conducted once the scope and preliminary questions of the guideline are determined. Meta-analysis may or may not be appropriate, but a systematic review is always needed. In the context of a guideline, the quality of evidence (also called certainty in the evidence, strength of the evidence, and confidence in the effect estimates) reflects the extent of our confidence that the estimates of an effect are adequate to support a particular decision or recommendation.[11] A good starting place to determine this confidence depends on study design (higher confidence in randomized trials and lower confidence in observational studies). There are, however, other factors that can moderate this initial rating of confidence. We are less confident of the effect of an intervention on benefits and harms (1) if the evidence is derived from studies with methodological limitations

Mayo Clin Proc. ■ March 2017;92(3):423-433 ■ http://dx.doi.org/10.1016/j.mayocp.2017.01.001
www.mayoclinicproceedings.org

(high risk of bias), (2) if the evidence is indirect (eg, the studies addressed populations or outcomes different from those targeted by the guideline), (3) if the evidence is inconsistent or heterogeneous (ie, the effect is different across studies, which can be evaluated qualitatively or statistically using the I^2 statistic and other tests of heterogeneity), (4) if the estimates of effect were imprecise (ie, the CI includes appreciable benefits and harms, making the decision different across the boundaries of the CI), or (5) if there was an indication of publication or reporting bias (ie, studies with positive results were more likely to be published than were studies with negative results or no effect, which may be detected statistically using funnel plots and other tests).

Conversely, there are scenarios that increase confidence in evidence derived from observational studies (eg, when a large effect is noted). These factors are considered the domains of the construct of quality of evidence and are used to reach a final rating of the quality of evidence (after starting with a rating based on study design). Table 2 includes examples to illustrate how these domains are applied.

FROM EVIDENCE TO RECOMMENDATION

The beginnings of evidence-based medicine have demonstrated a great desire to base decision making on evidence (as opposed to the previous approach of depending on expert opinion). However; evidence-based practitioners realized very quickly that evidence alone is insufficient for decision making. Hence, the second principle of evidence-based medicine acknowledged that decisions should also consider several nonevidence factors.[30] For example, chemotherapy can extend survival in lymphoma but causes serious adverse effects. Variation in patient values can lead to different recommendations (eg, making a stronger recommendation for chemotherapy for younger patients and making a weaker recommendation for older patients who may favor quality of life over extending survival).[31] Guideline developers can obtain information on patient values from surveys, from qualitative studies, or by engaging patients in the process of guideline development.[32,33] Frameworks for engaging patients and other stakeholders in guideline development have been proposed, although little comparative evidence is available to support one preferred approach.[34,35]

TABLE 1. Criteria for Trustworthy Guidelines

- Be based on a systematic review of the literature
- Be developed by a knowledgeable, multidisciplinary panel of experts and representatives from key affected groups
- Consider important patient subgroups and patient values and preferences
- Be based on an explicit and transparent process that minimizes distortions, biases, and conflicts of interest
- Provide a clear explanation of the logical relationships between alternative care options and health outcomes
- Provide ratings of the quality of evidence and the strength of recommendations
- Be reconsidered and revised as appropriate when important new evidence warrants modifications of recommendations

This Table was adapted from the National Academy of Medicine/formerly the Institute of Medicine.[18]

Table 3 presents factors to be considered when making a recommendation as described in the recently developed evidence to decision framework.[12] It is important to recognize that some of these factors may not be critical for decision making in certain situations. For example, when the balance of benefits and harms is clear, patient values may be less relevant. Similarly, cost may be critical for some decisions but not for others.

DETERMINING THE STRENGTH OF A RECOMMENDATION

The National Academy of Medicine and other authorities[18-21] have indicated that a key criterion of a trustworthy guideline is that a recommendation given to a patient or a population should have a strength attached to it. For example, in GRADE, recommendations are labeled as strong or weak (also called conditional or discretionary). This strength reflects the extent to which we can be confident that the desirable effects of an intervention outweigh the undesirable effects.[36] An example of a strong recommendation is recommending antiviral therapy to patients with immune-active chronic hepatitis B infection (where the benefits clearly outweigh the harms); whereas only a weak recommendation was given for using antiviral agents in those with compensated liver disease and low-level viremia.[37]

Strong recommendations usually require high-quality evidence. Less commonly, a

TABLE 2. Examples of the Domains Used to Assess the Quality of Evidence (Also Called Certainty in Evidence)

Domain (explanation)	Example
Quality of evidence is rated down	
Methodological limitations *(What is the extent of bias in the available studies?)*	A meta-analysis of 14 randomized trials showed that perioperative insulin infusion reduced mortality in patients undergoing surgery (relative risk, 0.69; 95% CI, 0.51-0.94). However, less than half of the trials concealed allocation sequence from the investigator enrolling patients; less than half of the trials blinded outcome assessors, and most trials did not report on those lost to follow-up.[22] These limitations suggest increased risk of bias that leads to rating down the quality of evidence.
Indirectness *(Does the available evidence fit the population and interventions of interest?)*	Trials of angiotensin-converting enzyme inhibitors in patients with diabetes can designate death, incidence of end-stage renal disease, or proteinuria as outcomes. Trials using death and incidence of end-stage renal disease provide the highest-quality evidence for decision making. Conversely, trials that use proteinuria as a surrogate for clinical outcomes produce indirect evidence that leads to rating down the quality of evidence.[23]
Inconsistency *(Do the results substantially differ across published studies?)*	A meta-analysis of 22 studies showed that medical students who learned using a self-directed approach had a moderate increase in knowledge compared with those who learned using traditional didactic curricula. The increase in knowledge was estimated to be 0.42 SD (95% CI, 0.14-0.70 SD). However, the effect was very inconsistent across studies (I^2=94%) and suggested important heterogeneity beyond what was expected by chance.[24] This heterogeneity leads to rating down the quality of evidence.
Imprecision *(Would our decision differ across the boundaries of the CI?)*	At 30 d and compared with endarterectomy, carotid stenting was associated with a nonsignificant reduction in the risk of death (relative risk, 0.61; 95% CI, 0.27-1.37).[25] If the lower boundary of CI was to represent the truth, this would mean that stenting reduced death by 73%. If the upper boundary of the CI was to represent the truth, this would mean that stenting increased death by 37%. Because the CI included appreciable benefit and harm, this evidence is considered imprecise, which leads to rating down the quality of evidence.
Publication bias *(Are there unpublished studies that show less impressive results than published ones?)*	Data on 74% of patients enrolled in trials evaluating the antidepressant reboxetine remained unpublished. Published data overestimated the benefit of reboxetine vs placebo by up to 115% and underestimated harm.[26] This is an example of publication bias in which sponsors of trials chose the data that are most impressive to publish, which leads to rating down the quality of evidence.
Quality of evidence is rated up	
Large effect *(Relative risk >2.0 or <0.5 can be used to define a large effect.)*	Meta-analysis of observational studies showed that infants with a front sleeping position had increased risk of sudden infant death syndrome compared with a back sleeping position (odds ratio, 4.1; 95% CI, 3.1-5.5). The large effect (4 times increased likelihood) increases certainty that a strong association exists and can lead to rating up the quality of evidence.[27]
Dose-response effect *(The higher the dose, the larger the effect.)*	An observational study of 2154 patients with septic shock showed a strong relationship between the delay in effective antimicrobial initiation and in-hospital mortality (adjusted odds ratio, 1.12 for each additional hour delay). This dose-response relationship increases certainty that a strong association exists and can lead to rating up the quality of evidence.[28]
Plausible confounding strengthens the association *(Despite a likely confounder that would weaken the association, the effect remains significant.)*	Observational studies have documented lower mortality rates in not-for-profit hospitals compared with for-profit hospitals. This is despite confounding by the fact that sicker patients usually are hospitalized in not-for-profit hospitals (thus, not-for-profit hospitals would be expected to have higher mortality rates). Despite this, it was observed that not-for-profit hospitals had lower mortality rates. This residual confounding strengthens the observed association and can lead to rating up the quality of evidence.[29]

Mayo Clin Proc. ■ March 2017;92(3):423-433 ■ http://dx.doi.org/10.1016/j.mayocp.2017.01.001
www.mayoclinicproceedings.org

strong recommendation can be justified with low-quality evidence in 5 paradigmatic situations (Table 4).[38]

WHERE TO FIND GUIDELINES?

The National Guideline Clearinghouse is a public resource that is supported by the Agency for Healthcare Research and Quality. Guidelines indexed in this register have to meet certain criteria, particularly, being based on a systematic review of the evidence and documenting an assessment of the benefits and harms of the recommended care and alternative care options.[4] Another way to find guidelines is to search websites of professional societies relevant to a particular topic. Guidelines can also be found through bibliographic database searches (eg, MEDLINE). The search can use a database filter (eg, in PubMed, there is a publication type called *guideline*), a controlled vocabulary (eg, Medical Subject Heading terms in PubMed), or text words (eg, *guideline, consensus, recommendation,* or *standards*). Incorporation of guideline recommendations in order sets and decision support tools in electronic medical records reduces the need for searching for guidelines but hides the rationale and rigor of guideline development from end users.

HOW TO JUDGE THE QUALITY OF A GUIDELINE?

There are more than 20 tools available to appraise a guideline.[39] Some tools focus on implementation.[40] The AGREE (Appraisal of Guidelines, Research, and Evaluation) Collaboration developed the AGREE II, which is the most validated and extensively used tool. The AGREE II is a generic instrument that aims to assess the process of guideline development and reporting and has 23 items grouped into 6 domains (scope and purpose, stakeholder involvement, rigor of development, clarity of presentation, applicability, and editorial independence).[41] The use of this tool is helpful for researchers, journal editors, and policy makers but not for a busy clinician reading a recently published guideline. A simplified approach is suggested for clinicians[42] that focuses on only a few questions (Was the recommendation based on the best available evidence and is that evidence reported and easily understood? Were patients' values, preferences, and resources considered? Was the strength of recommendations appropriate? Was the influence of conflicts of interest minimized?). These questions address the major criteria of a trustworthy guideline.

TABLE 3. Factors to Consider When Making a Recommendation

1. Is the question that the guideline is addressing a priority?
2. Balance of benefits and harms
a. How substantial are the desirable anticipated effects?
b. How substantial are the undesirable anticipated effects?
c. Does the balance between desirable and undesirable effects favor the intervention or the alternative?
3. What is the overall quality of the evidence?
4. Is there important uncertainty or variability in how much patients value the outcomes?
5. Resources
a. How large are the resource requirements (costs)?
b. What is the certainty in the evidence of resource requirements (costs)?
c. Is the intervention cost-effective?
6. What is the effect on health equity?
7. Is the intervention acceptable to patients, caregivers, and health care professionals?
8. Is implementing the intervention feasible for patients, caregivers, and health care professionals?

This Table was adapted from *BMJ*.[12]

IMPLEMENTATION

Implementation From the Perspective of a Clinician-Patient Dyad

The key information needed to implement a recommendation includes a clear description of the population to whom a recommendation is appropriate, the baseline risk of this population, the quality of evidence, and the strength of the recommendation. An example of a recommendation that contains these 4 pieces of information is presented in Table 5.

When the recommendation is strong, clinicians should offer the intervention to almost all their eligible patients and convey high confidence that the benefits outweigh the harms. Extensive discussion or exploration of values is not always needed.

Conversely, if the recommendation is weak or conditional, clinicians need to be more discretionary in applying the recommendation and offer additional discussion to elicit patients' values, preferences, cost implications, comorbidities, and clinical and social context.[42] Weak recommendations imply that although the recommended action is appropriate for

TABLE 4. Making a Strong Recommendation When the Quality of Evidence Is Low

Paradigmatic situation	Example
Low-quality evidence suggests benefit in a life-threatening situation	Recommendation to increase the glucocorticoid dosage in congenital adrenal hyperplasia during a febrile illness
Low-quality evidence suggests benefit and high-quality evidence suggests harm or a very high cost	Recommendation against screening with whole-body computed tomographic scan
Low-quality evidence suggests equivalence, but high-quality evidence suggests harm with one alternative	Recommendation for laparoscopic over open adrenalectomy in patients with unilateral primary aldosteronism
High-quality evidence suggests equivalence, but low-quality evidence suggests harm with one alternative	Recommendation for methyldopa over angiotensin-converting enzyme inhibitors for hypertension in pregnancy
Low-quality evidence suggests harm in a critical outcome that is valued much more than any of the benefits	Recommendation against testosterone replacement in men with prostate cancer

most patients, for some of them, the alternative action may be more appropriate. For example, a strong recommendation for intranasal glucocorticosteroid use in allergic rhinitis means that clinicians should prescribe it to most patients and just perhaps provide educational material to the patient about how to use it. On the other hand, implementing a weak recommendation for intransal antihistamines requires more discussion during which clinicians need to convey the lack of high-quality evidence in this setting and explore patient preferences, costs, and other factors affecting the decision to start treatment.[44]

Shared decision-making tools are highly desired in the setting of conditional recommendations and uncertain evidence. These tools can present all the plausible options to patients, increase their knowledge about the condition and the options, and enhance their awareness of their own risk of various outcomes. A meta-analysis of 115 studies showed that decision aids increased patients' knowledge, the proportion of patients with accurate risk perceptions, and the likelihood of patients choosing an option congruent with their values; reduced decisional conflict; and increased visit time by a median of only 2.6 minutes (range, 8 minutes shorter to 28 minutes longer).[45] Developing decision aids requires patient engagement in the design and testing process. The most helpful decision aids are the ones that encourage conversation during the clinical encounter and activate patients as opposed to those that merely provide education.[46]

TABLE 5. An Example of a Recommendation That Contains the 4 Elements Required for Proper Implementation (Description of the Population, Baseline Risk, Quality of Evidence, and Strength of Recommendation)

Recommendation: "For acutely ill hospitalized medical patients at increased risk of thrombosis who are bleeding or at high risk for major bleeding, we suggest the optimal use of mechanical thromboprophylaxis with graduated compression stockings.

(Weak recommendation, low quality evidence)"[43]

The recommendation is followed by 2 tools that help the clinician estimate the risk of thrombosis and bleeding in hospitalized medical patients.

Implementation From the Perspective of Health Care Systems

The strength of a recommendation is an essential factor that guides implementation of a guideline from the perspective of policy makers. Strong recommendations imply high confidence that benefits outweigh harms and that the recommended act is appropriate for almost all patients, is feasible, is acceptable to patients and providers, and does not exacerbate inequity. Such recommendations may be candidates for quality improvement projects that ensure that most patients would receive the intervention. Process measures can be structured to monitor health care delivery. Barriers and facilitators should also be explored in the context of health care delivery. Some of these recommendations may be appropriate for pay for performance initiatives and quality metrics.[42] Clearly, lower-grade (weak, discretionary, conditional) recommendations are not candidates for incorporation into quality improvement projects and quality metrics (unless the metric was to measure that most patients were engaged in relevant discussion and were introduced to all the available treatment options). Such recommendations remain, however, the best course of action and from a coverage standpoint are what most patients should be able to receive.

From a health system perspective, a guideline could be adopted or adapted. Adaptation is a process that may be needed to improve implementation in a targeted practice or geographic location. The process includes

Mayo Clin Proc. ■ March 2017;92(3):423-433 ■ http://dx.doi.org/10.1016/j.mayocp.2017.01.001
www.mayoclinicproceedings.org

identification of a guideline, followed by appraisal and then modification based on the feedback of local stakeholders, local disease prevalence or baseline risk, effect modifiers (eg, antimicrobial drug resistance), availability of resources, and cultural variations.[47]

In addition to integration of guidelines as decision support tools in electronic health records and initiating quality improvement projects, dissemination and implementation can be enhanced using point-of-care technology. An example is a mobile device application that provides a multilayered presentation for point-of-care use. Starting with the recommendation (the top layer), clinicians can on-demand access a rationale for and key information about the recommendation, the quality of evidence, balance between desirable and undesirable consequences, values and preferences, and resource considerations.[48] Based on clinicians' time, interest, and expertise, they can choose which layers to view. User testing led to extensive revisions of this application, and most stakeholders expressed overall satisfaction with the final format.[48]

Do Guidelines Improve Patient Outcomes?

Adherence to guidelines lowered the risk of hospitalization in patients with chronic heart failure across several European countries.[49] A study of patients with a new diagnosis of primary breast cancer showed that the greater the number of violations in guideline adherence, the lower the survival.[50] In patients with hospital-acquired or ventilator-associated pneumonia, guideline-adherent initial intravenous antibiotic drug therapy was clinically superior, saved more lives, and was less expensive than non–guideline-adherent therapy.[51] Guideline-adherent antithrombotic prophylaxis in patients discharged with atrial fibrillation was also associated with lower all-cause and cardiovascular mortality.[52] Numerous other examples exist to document better outcomes with guideline adherence.

Improving Guideline Uptake

Although adherence to guidelines improves outcomes, nonadherence to guidelines was shown to be frequent[53,54] and was more likely to affect patients who were older or had cancer or other comorbidites.[52] Low adherence to guidelines reinforces that passive diffusion of evidence is unreliable.[55] Four cluster trials showed that active interventions are required to enhance guideline uptake (eg, educational workshops, paper-based educational materials, order forms, and reminders).[56] Multifaceted interventions targeting different barriers were more likely to be effective than single interventions.[55] Qualitative research also demonstrated the need for building a culture that enabled guideline implementation.[57] An example is the successful implementation of performance measures for community-acquired pneumonia in several academic institutions in New Jersey. Success was attributed to the wide engagement of stakeholders; forming multidisciplinary teams championed by a nurse practitioner; making changes based on local input; dissemination via workshops, grand rounds, and mailings; and collecting data on process measures.[58]

LIMITATIONS AND FUTURE NEEDS

A large proportion of published guidelines were reported to have limited rigor and did not meet all the criteria for a trustworthy guideline.[59-61] Patients or methodologists were not included in the guideline development process in most guidelines (71% and 86%, respectively).[62] Guidelines also tend to address the common or average patient. For example, 1 study demonstrated the absence of incorporating the impact of multiple chronic conditions, sociopersonal context, and patient preferences in 29%, 39%, and 57% of a sample of guidelines, respectively.[62] This, indeed, suggests that recommendations fit the needs of the average patient. True practice of individualized medicine requires more nuanced recommendations, which necessitates advances in genomics and the basic sciences to produce reliable individualized risk prediction and response to therapy estimates (which is not the case at the present time). The current research evidence is based on assessing average treatment effects in a group of patients that is not conducive to individualized practice. The current evidence, however, can be leveraged by conducting N-of-1 trials (crossover trials in single individuals to evaluate treatment effectiveness in chronic conditions)[63] and rigorous subgroup analyses in trials and meta-analyses. Subgroup analyses can produce effect estimates stratified

TABLE 6. Future Needs in the Field of Practice Guidelines

1. Better methodology to engage patients and obtain their perspective. Current methods of patient engagement often lead to tokenistic engagement.[33] Randomized trials to test engagement methods have been performed and seem to be feasible.[71]
2. Reliable mechanism to update guidelines and incorporate new evidence. Although frameworks for updating guidelines have been suggested,[72] only 53% of surveyed guideline developers had a formal procedure for deciding when a guideline becomes out of date[73] and less than a third of methodology handbooks included an approach for updating guidelines.[68]
3. Collaborative initiatives to avoid contradictory recommendations by different professional societies. In 2011, five independent guidelines reported recommendations for the management of carotid artery stenosis.[70] Recommendations from these committees differed. The ACC/AHA suggested that stenting and endarterectomy were equal options for symptomatic patients, whereas the Society for Vascular Surgery and the Australasian Guidelines recommended endarterectomy as a first choice and suggested stenting as an alternative only in patients with high anatomical or perioperative risk.[70]
4. Data sharing to facilitate conducting individual patient meta-analyses. Study-level meta-analysis, which is very frequently performed, produces subgroup estimates that are highly subject to ecological bias. In comparison, individual patient meta-analysis can produce reliable estimates of effect in patient subgroups; which allows individualized recommendations. However, individual patient meta-analysis requires data sharing by trialists.

ACC/AHA = American College of Cardiology/American Heart Association.

by patient characteristics (effect modifiers) and can help with individualizing care. However, spurious findings are the norm in the current practice of subgroup analysis.[64] Individual patient meta-analysis can offer reliable individualized estimates but requires data sharing by trialists and remains quite uncommon.

In terms of validity and reproducibility of judgments made by guideline developers, the only available literature is on the GRADE approach. Although quality of evidence ratings had good reproducibility and reliability,[14,15] it remains by necessity a matter of judgment.[65] Aside from reliability, the benefit of using discrete domains of the quality of evidence and specific criteria for the evidence to decision framework is the transparency of the approach. Transparency allows guideline users to understand the rationale for the judgments made, even if they disagree with the perspective taken.[65]

Clinical practice guidelines become quickly outdated, with 1 of 5 recommendations being out of date after 3 years.[66] Recommendations that depended on multiple randomized trials were less likely to change.[67] The concept of a *living guideline* (ie, a guideline with a mechanism for real-time updating) has been proposed[68] but not implemented. The *British Medical Journal* has launched a rapid recommendation initiative to compensate for outdated guidelines.[69]

Another challenge that undermines guideline integrity and uptake is inconsistency in recommendations across guidelines evaluating the same evidence on the same condition.[70] Conflicting recommendations confuse guideline users and undermine evidence-based practice. A summary of key challenges and future needs is presented in Table 6.

Theoretical harms of guidelines include leaving insufficient room for clinicians to tailor care, adversely affecting insurance coverage for certain interventions, increasing health care costs; deskilling clinicians, and subjecting clinicians to litigation or unfavorable professional judgments by auditors and managers.[2,74] Nevertheless, surveys of primary care physicians showed that they had overall positive attitudes toward guidelines and their effectiveness; only a minority was concerned that guidelines may be used to set performance-related pay, reduce clinical freedom, or stifle innovation.[75-77]

CONFLICTS OF INTEREST

It is well-known that conflicts of interest affect the opinions and recommendations of experts and can compromise guideline validity.[78] Intellectual conflicts are common but remain implicit and are rarely addressed (compared with financial ones). Forming guideline panels without conflicted experts (who are more likely to have the most expertise and knowledge in a particular field) may lead to guidelines devoid of expertise and may impair guideline credibility and uptake. Therefore, if excluding conflicted panelists was not possible in certain situations, reliable strategies for managing these conflicts would be critical to ensure guideline credibility. Recent strategies that were shown to be effective were (1) tasking conflicted experts with interpreting and commenting on evidence but not with voting on recommendations; (2) appointing a nonconflicted methodologist as the panel chair or co-chair with equal authority as content experts; and (3) providing explicit criteria for determining intellectual and financial conflicts.[79]

Mayo Clin Proc. ■ March 2017;92(3):423-433 ■ http://dx.doi.org/10.1016/j.mayocp.2017.01.001
www.mayoclinicproceedings.org

CONCLUSION

Trustworthy clinical practice guidelines require a systematic review to select the best available evidence and should rigorously evaluate the quality of evidence and incorporate nonevidence factors to transform evidence into a decision. Empirical evidence shows that guidelines improve patient outcomes; however, adherence to guidelines is variable. Therefore, guidelines require active dissemination and innovative implementation strategies.

Abbreviations and Acronyms: **ACC/AHA** = American College of Cardiology/American Heart Association; **AGREE** = Appraisal of Guidelines, Research, and Evaluation; **GRADE** = Grading of Recommendations, Assessment, Development, and Evaluation

Correspondence: Address to M. Hassan Murad, MD, MPH, Division of Preventive Medicine, Mayo Clinic, 200 First St SW, Rochester, MN 55905 (Murad.mohammad@mayo.edu).

REFERENCES

1. Field M, Lohr K. *Clinical Practice Guidelines: Directions for a New Program, Institute of Medicine*. Washington, DC: National Academy Press; 1990.
2. Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. *BMJ*. 1999;318(7182):527-530.
3. International Guideline Library. http://www.g-i-n.net/library/international-guidelines-library. Accessed September 24, 2016.
4. National Guideline Clearinghouse. https://www.guideline.gov. Accessed September 21, 2016.
5. Weisz G, Cambrosio A, Keating P, Knaapen L, Schlich T, Tournay VJ. The emergence of clinical practice guidelines. *Milbank Q*. 2007;85(4):691-727.
6. Woolf SH. Practice guidelines, a new reality in medicine, II: methods of developing guidelines. *Arch Intern Med*. 1992;152(5):946-952.
7. Guyatt G. Evidence-based medicine [editorial]. *ACP J Club*. 1991;114:A16.
8. Tricoci P, Allen JM, Kramer JM, Califf RM, Smith SC Jr. Scientific evidence underlying the ACC/AHA clinical practice guidelines. *JAMA*. 2009;301(8):831-841.
9. Atkins D, Eccles M, Flottorp S, et al; GRADE Working Group. Systems for grading the quality of evidence and the strength of recommendations, I: critical appraisal of existing approaches. *BMC Health Serv Res*. 2004;4(1):38.
10. Guyatt G, Jaeschke R, Wilson M, Montori V, Richardson W. What is evidence-based medicine? In: Guyatt G, Rennie D, Meade M, Cook D, eds. *Users' Guide to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. 3rd ed. New York, NY: McGraw-Hill Education/Medical; 2014:7-14.
11. Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines, 3: rating the quality of evidence. *J Clin Epidemiol*. 2011;64(4):401-406.
12. Alonso-Coello P, Schunemann HJ, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices, 1: introduction. *BMJ*. 2016;353:i2016.
13. Murad MH, Altayar O, Bennett M, et al. Using GRADE for evaluating the quality of evidence in hyperbaric oxygen therapy clarifies evidence limitations. *J Clin Epidemiol*. 2014;67(1):65-72.
14. Mustafa RA, Santesso N, Brozek J, et al. The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. *J Clin Epidemiol*. 2013;66(7):736-742; quiz 742e1-742e5.
15. Kumar A, Miladinovic B, Guyatt GH, Schunemann HJ, Djulbegovic B. GRADE guidelines system is reproducible when instructions are clearly operationalized even among the guidelines panel members with limited experience with GRADE. *J Clin Epidemiol*. 2016;75:115-118.
16. Faraday M, Hubbard H, Kosiak B, Dmochowski R. Staying at the cutting edge: a review and analysis of evidence reporting and grading; the recommendations of the American Urological Association. *BJU Int*. 2009;104(3):294-297.
17. Methods and processes US Preventive Services Taskforce website. http://www.uspreventiveservicestaskforce.org/Page/Name/methods-and-processes#recommendation-process. Accessed August 22, 2016.
18. Institute of Medicine. *Clinical Practice Guidelines We Can Trust*. Washington, DC: National Academies Press; 2011.
19. Qaseem A, Forland F, Macbeth F, Ollenschlager G, Phillips S, van der Wees P. Guidelines International Network: toward international standards for clinical practice guidelines. *Ann Intern Med*. 2012;156(7):525-531.
20. Hill J, Bullock I, Alderson P. A summary of the methods that the National Clinical Guideline Centre uses to produce clinical guidelines for the National Institute for Health and Clinical Excellence. *Ann Intern Med*. 2011;154(11):752-757.
21. World Health Organization. *World Health Organization Handbook for Guideline Development*. 2nd ed. Geneva, Switzerland: World Health Organization Press; 2014.
22. Gandhi GY, Murad MH, Flynn DN, et al. Effect of perioperative insulin infusion on surgical morbidity and mortality: systematic review and meta-analysis of randomized trials. *Mayo Clin Proc*. 2008;83(4):418-430.
23. Jun M, Turin TC, Woodward M, et al. Assessing the validity of surrogate outcomes for ESRD: a meta-analysis. *J Am Soc Nephrol*. 2015;26(9):2289-2302.
24. Murad MH, Coto-Yglesias F, Varkey P, Prokop LJ, Murad AL. The effectiveness of self-directed learning in health professions education: a systematic review. *Med Educ*. 2010;44(11):1057-1068.
25. Murad MH, Flynn DN, Elamin MB, et al. Endarterectomy vs stenting for carotid artery stenosis: a systematic review and meta-analysis. *J Vasc Surg*. 2008;48(2):487-493.
26. Eyding D, Lelgemann M, Grouven U, et al. Reboxetine for acute treatment of major depression: systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials. *BMJ*. 2010;341:c4737.
27. Gilbert R, Salanti G, Harden M, See S. Infant sleeping position and the sudden infant death syndrome: systematic review of observational studies and historical review of recommendations from 1940 to 2002. *Int J Epidemiol*. 2005;34(4):874-887.
28. Kumar A, Roberts D, Wood KE, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med*. 2006;34(6):1589-1596.
29. Devereaux PJ, Choi PT, Lacchetti C, et al. A systematic review and meta-analysis of studies comparing mortality rates of private for-profit and private not-for-profit hospitals. *CMAJ*. 2002;166(11):1399-1406.
30. Murad MH, Montori VM, Guyatt GH. Incorporating patient preferences in evidence-based medicine. *JAMA*. 2008;300(21):2483; author reply 2483-2484.
31. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924-926.
32. MacLean S, Mulla S, Akl EA, et al. Patient values and preferences in decision making for antithrombotic therapy: a

systematic review: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012; 141(2 suppl):e1S-e23S.
33. Domecq JP, Prutsky G, Elraiyah T, et al. Patient engagement in research: a systematic review. *BMC Health Serv Res*. 2014; 14:89.
34. Armstrong MJ, Rueda JD, Gronseth GS, Daniel Mullins C. Framework for enhancing clinical practice guidelines through continuous patient engagement. *Health Expect*. 2017;20(1):3-10.
35. Legare F, Boivin A, van der Weijden T, et al. Patient and public involvement in clinical practice guidelines: a knowledge synthesis of existing programs. *Med Decis Making*. 2011;31(6):E45-E74.
36. Andrews J, Guyatt G, Oxman AD, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. *J Clin Epidemiol*. 2013;66(7): 719-725.
37. Terrault NA, Bzowej NH, Chang KM, et al. AASLD guidelines for treatment of chronic hepatitis B. *Hepatology*. 2016;63(1): 261-283.
38. Brito JP, Domecq JP, Murad MH, Guyatt GH, Montori VM. The Endocrine Society guidelines: when the confidence cart goes before the evidence horse. *J Clin Endocrinol Metab*. 2013; 98(8):3246-3252.
39. Vlayen J, Aertgeerts B, Hannes K, Sermeus W, Ramaekers D. A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. *Int J Qual Health Care*. 2005;17(3):235-242.
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. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. *CMAJ*. 2010;182(18):E839-E842.
42. Neumann I, Santesso N, Akl EA, et al. A guide for health professionals to interpret and use recommendations in guidelines developed with the GRADE approach. *J Clin Epidemiol*. 2016; 72:45-55.
43. Kahn SR, Lim W, Dunn AS, et al. Prevention of VTE in nonsurgical patients: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141(2 suppl):e195S-e226S.
44. Brozek JL, Bousquet J, Baena-Cagnani CE, et al. Allergic Rhinitis and its Impact on Asthma (ARIA) guidelines: 2010 revision. *J Allergy Clin Immunol*. 2010;126(3):466-476.
45. Stacey D, Legare F, Col NF, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev*. 2014;1:CD001431.
46. Hargraves I, LeBlanc A, Shah ND, Montori VM. Shared decision making: the need for patient-clinician conversation, not just information. *Health Aff (Millwood)*. 2016;35(4):627-629.
47. Schunemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: 13. Applicability, transferability and adaptation. *Health Res Policy Syst*. 2006;4:25.
48. Kristiansen A, Brandt L, Alonso-Coello P, et al. Development of a novel, multilayered presentation format for clinical practice guidelines. *Chest*. 2015;147(3):754-763.
49. Komajda M, Lapuerta P, Hermans N, et al. Adherence to guidelines is a predictor of outcome in chronic heart failure: the MAHLER survey. *Eur Heart J*. 2005;26(16):1653-1659.
50. Wockel A, Kurzeder C, Geyer V, et al. Effects of guideline adherence in primary breast cancer—a 5-year multi-center cohort study of 3976 patients. *Breast*. 2010;19(2):120-127.
51. Wilke M, Grube RF, Bodmann KF. Guideline-adherent initial intravenous antibiotic therapy for hospital-acquired/ventilator-associated pneumonia is clinically superior, saves lives and is cheaper than non guideline adherent therapy. *Eur J Med Res*. 2011;16(7):315-323.
52. Proietti M, Nobili A, Raparelli V, et al. Adherence to antithrombotic therapy guidelines improves mortality among elderly patients with atrial fibrillation: insights from the REPOSI study. *Clin Res Cardiol*. 2016;105(11):912-920.
53. Geahchan N, Basile M, Tohmeh M; DIONYS registry. Venous thromboembolism prophylaxis in patients undergoing abdominal and pelvic cancer surgery: adherence and compliance to ACCP guidelines in DIONYS registry. *Springerplus*. 2016;5(1):1541.
54. Tinkle MB, Tigges BB, Boursaw B, McFarlane DR. Adherence to the Women's Preventive Services Guidelines in the Affordable Care Act. *J Obstet Gynecol Neonatal Nurs*. 2016; 45(6):813-824.
55. Grimshaw JM, Shirran L, Thomas R, et al. Changing provider behavior: an overview of systematic reviews of interventions. *Med Care*. 2001;39(8 suppl 2):II2-II45.
56. Flodgren G, Hall AM, Goulding L, et al. Tools developed and disseminated by guideline producers to promote the uptake of their guidelines. *Cochrane Database Syst Rev*. 2016;8: CD010669.
57. Sinuff T, Cook D, Giacomini M, Heyland D, Dodek P. Facilitating clinician adherence to guidelines in the intensive care unit: a multicenter, qualitative study. *Crit Care Med*. 2007; 35(9):2083-2089.
58. Gross PA, Aho L, Ashtyani H, et al. Extending the nurse practitioner concurrent intervention model to community-acquired pneumonia and chronic obstructive pulmonary disease. *Jt Comm J Qual Saf*. 2004;30(7):377-386.
59. 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.
60. Bancos I, Cheng T, Prokop LJ, Montori VM, Murad MH. Endocrine clinical practice guidelines in North America: a systematic assessment of quality. *J Clin Epidemiol*. 2012;65(5):520-525.
61. Gorman SK, Chung MH, Slavik RS, Zed PJ, Wilbur K, Dhingra VK. A critical appraisal of the quality of critical care pharmacotherapy clinical practice guidelines and their strength of recommendations. *Intensive Care Med*. 2010; 36(10):1636-1643.
62. Wyatt KD, Stuart LM, Brito JP, et al. Out of context: clinical practice guidelines and patients with multiple chronic conditions: a systematic review. *Med Care*. 2014;52(suppl 3): S92-S100.
63. Vohra S. N-of-1 trials to enhance patient outcomes: identifying effective therapies and reducing harms, one patient at a time. *J Clin Epidemiol*. 2016;76:6-8.
64. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ*. 2010;340:c117.
65. Murad MH, Mustafa R, Morgan R, Sultan S, Falck-Ytter Y, Dahm P. Rating the quality of evidence is by necessity a matter of judgment. *J Clin Epidemiol*. 2016;74:237-238.
66. Martinez Garcia L, Sanabria AJ, Garcia Alvarez E, et al. The validity of recommendations from clinical guidelines: a survival analysis. *CMAJ*. 2014;186(16):1211-1219.
67. Neuman MD, Goldstein JN, Cirullo MA, Schwartz JS. Durability of class I American College of Cardiology/American Heart Association clinical practice guideline recommendations. *JAMA*. 2014;311(20):2092-2100.
68. Vernooij RW, Sanabria AJ, Sola I, Alonso-Coello P, Martinez Garcia L. Guidance for updating clinical practice guidelines: a systematic review of methodological handbooks. *Implement Sci*. 2014;9:3.
69. Siemieniuk RA, Agoritsas T, Macdonald H, Guyatt GH, Brandt L, Vandvik PO. Introduction to BMJ Rapid Recommendations. *BMJ*. 2016;354:i5191.
70. Paraskevas KI, Mikhailidis DP, Veith FJ. Comparison of the five 2011 guidelines for the treatment of carotid stenosis. *J Vasc Surg*. 2012;55(5):1504-1508.

71. Nilsen ES, Myrhaug HT, Johansen M, Oliver S, Oxman AD. Methods of consumer involvement in developing healthcare policy and research, clinical practice guidelines and patient information material. *Cochrane Database Syst Rev*. 2006;3:CD004563.
72. Shekelle P, Eccles MP, Grimshaw JM, Woolf SH. When should clinical guidelines be updated? *BMJ*. 2001;323(7305):155-157.
73. Alonso-Coello P, Martinez Garcia L, Carrasco JM, et al. The updating of clinical practice guidelines: insights from an international survey. *Implement Sci*. 2011;6:107.
74. Timmermans S, Mauck A. The promises and pitfalls of evidence-based medicine. *Health Aff (Millwood)*. 2005;24(1):18-28.
75. Siriwardena AN. Clinical guidelines in primary care: a survey of general practitioners' attitudes and behaviour. *Br J Gen Pract*. 1995;45(401):643-647.
76. Alexander PE, Li SA, Tonelli M, Guyatt G. Canadian primary care physicians' attitudes toward understanding clinical practice guidelines for diabetes screening. *Can J Diabetes*. 2016;40(6):580-585.
77. Lunyera J, Jonassaint C, Jonassaint J, Shah N. Attitudes of primary care physicians toward sickle cell disease care, guidelines, and comanaging hydroxyurea with a specialist. *J Prim Care Community Health*. 2017;8(1):37-40.
78. Wang AT, McCoy CP, Murad MH, Montori VM. Association between industry affiliation and position on cardiovascular risk with rosiglitazone: cross sectional systematic review. *BMJ*. 2010;340:c1344.
79. Guyatt G, Akl EA, Hirsh J, et al. The vexing problem of guidelines and conflict of interest: a potential solution. *Ann Intern Med*. 2010;152(11):738-741.