COMMENTARY

The Need to Systematically Evaluate Clinical Practice Guidelines

Allen F. Shaughnessy, PharmD, MMedEd, Lisa Cosgrove, PhD, and Joel R. Lexchin, MD

Clinical practice guidelines abound. The recommendations contained in these guidelines are used not only to make decisions about the care of individual patients but also as practice standards to rate physician "quality." Physicians' confidence in guidelines is based on the supposition that there is a rigorous, objective process for developing recommendations based on the best available evidence. Though voluntary standards for the development of guidelines exist, the process of guideline development is unregulated and the quality of many guidelines is low. In addition, the few tools available to assess the quality of guidelines are time consuming and designed for researchers, not clinicians. Few guidelines are evaluated, either before or after their dissemination, for their impact on patient outcomes. Just as with pharmaceuticals and other products that can affect patients for better or worse, perhaps it is time to develop more standardized ways to evaluate the development and dissemination of clinical practice guidelines to ensure a similar balance between risk and benefit. (J Am Board Fam Med 2016;29:644–648.)

The role of clinical practice guidelines has evolved over time. Early guidelines gathered dust on physicians' shelves; now many clinicians use them to guide their medical decisions. However, many feel buffeted by guidelines recommending different approaches to practice.[1] They see guidelines as cudgels used by insurance companies and other payers to incentivize or deter certain practices. Lawyers point to them in courtrooms. Practice guidelines, therefore, may shape medical practice and affect patient outcomes in many ways. Usually untested before or after implementation, guidelines can lead to misuse or overuse of medical services, resulting in harm. It is time to develop a process to ensure their safety and efficacy.

Discrepancy among practice guidelines has been documented[2] and is well known by clinicians trying to determine the best treatment goals, such as for patients with type 2 diabetes (should the HbA_{1c} goal be <6.5%?[3] 7% to 8%?[4]), or whom to screen for prostate cancer[5,6] or breast cancer.[7,8] For example, at 1 of our institutions, patients receive 2 letters following a normal mammogram: their family medicine physician tells them to follow-up in 2 years, based on guidelines from the US Preventive Services Task Force,[7] whereas the radiologist performing the screening recommends returning in 1 year, based on American College of Radiology[8] guidance. What are patients to think?

These guidelines are based on a synthesis and analysis of the same available evidence. Can disparate recommendations all provide equivalent results? Evidence is limited (Table 1) but suggests that some guidelines are more trustworthy than others.

Faith in the effectiveness of guidelines hinges on the premise that benefit and risks identified in con-

This article was externally peer reviewed.
Submitted 6 April 2016; revised 6 June 2016; accepted 13 June 2016.
From the Department of Family Medicine, Tufts University School of Medicine, Tufts University Family Medicine Residency at Cambridge Health Alliance, Boston, MA (AFS); the Department of Counseling and School Psychology, University of Massachusetts, Boston (LC); and the School of Health Policy and Management, York University, Toronto, Canada (JRL).
Funding: none.
Conflict of interest: none declared.
Corresponding author: Allen F. Shaughnessy, PharmD, MMedEd, Tufts University Family Medicine Residency, Cambridge Health Alliance, Malden, MA 41111 (E-mail: Allen.Shaughnessy@tufts.edu).

See Related Editorial on Page 642.

Table 1. Examples of Harm Associated with Guidelines

Guideline	Harmful Effect
American Thoracic Society guidelines for the treatment of pneumonia.[9]	Prescribing consistent with guidelines had a trend toward higher mortality and subsequent hospitalization.[10]
European Society of Cardiology guidelines for anticoagulant treatment of atrial fibrillation.[11]	Recommendation by this group result in a 3-fold higher rate of hemorrhage compared with guidelines from the American College of Chest Physicians.[12]
Guidelines from Australia, Canada, China, Europe, France, Japan, South Africa, United Kingdom, and United States recommending thrombolytic treatment from 3 to 4.5 hours after the onset of acute ischemic stroke.	Administering the thrombolytic alteplase 3 to 4.5 hours after the onset of stroke increases mortality 2%, without evidence of benefit.[13]
Draft guidelines on the use of electrodiagnostic testing for patients with low-back pain.	Given clinical vignettes, physicians *increased* ordering of tests when given specific guidance recommending against their use compared with physicians given nonspecific guidelines also recommending against routine use of the tests.[14]

trolled clinical trials will automatically translate to similar benefits and risks in the general population. This premise ignores the interpretive turn that occurs as guideline developers formulate a recommendation from the available clinical evidence, for developing guidelines is not solely a scientific endeavor. In moving from the evidence to a specific recommendation, guideline developers must weigh the benefit of an intervention against the likelihood of harm. With clear and explicit methods, and a multidisciplinary guideline development group, the biases that may result from this interpretive turn are minimized, but not eliminated. Thus there is an inevitable aspect of guideline development that makes it subject to value judgments and can be unconsciously colored by intellectual, professional, or financial conflicts of interest.[15]

This coloring can produce recommendations that result in misuse or overuse of medical care, particularly when guidelines are produced by speciality societies. As Quanstrum and Hayward[15] noted, "Although it is true that individual medical providers care deeply about their patients, the guild of health care professionals—including their specialty societies—has a primary responsibility to promote its members' interests . . . it is a fool's dream to expect the guild of any service industry to harness its self-interest and to act according to beneficence alone—to compete on true value when the opportunity to inflate perceived value is readily available." As the saying goes, never ask a barber if you need a haircut.

Clinical practice guidelines are produced by people—usually well-intentioned people who give their time and energy for the good of society. However, they are subject to the same cognitive biases as all humans.[16] These biases, which are typically not recognized by the person, can result in decisions colored by tunnel vision (job conditioning), "seeing what you want to see" (confirmation bias), decisions limited to the tools at hand (Maslow's hammer), or other inclinations that can affect judgment. These biases can be influenced by financial, professional, and intellectual conflicts of interest.[17]

There are >6000 guidelines from 96 groups in 76 countries in the Guidelines International Network database alone, suggesting that there is a "guideline industry" that is similar to that seen the early years of drug commerce. Medicines were initially unregulated until the Pure Food and Drug Act of 1906 established rudimentary standards for drug safety. Yet medicines did not have to demonstrate effectiveness until 1962, and "good manufacturing practices" were not required until even more recently.[18] Regulation of guidelines should not proceed at such a leisurely pace.

Although all guidelines have limitations in applicability (eg, due to the variety of clinical presentations encountered by clinicians), some important lessons can be learned by looking at the standards that have been developed by well-respected guideline producers. For example, in the United Kingdom, the National Institute for Health and Care Excellence has set up rigorous processes to minimize bias and to regulate the guideline development process.[19] In the United States, several organizations such as the Institute of Medicine,[20] the US Preventive Services Task Force, and the Guideline International Network[21] have developed standards for the production of practice guidelines. However, guideline producers infrequently follow these "good manufacturing practice" equiva-

lents.[22,23] As a result, these efforts have not prevented questionable or harmful guidelines from being disseminated, adopted by clinicians, or used in support of so-called quality measures.

Poor-quality guidelines are unlikely to help patient care, and may actually harm it. But even good-quality guidelines may not improve clinical outcomes since many factors are involved in changing outcomes. In addition, because guidelines are dynamic, they should be updated frequently; the Agency for Healthcare Research and Quality recently required that guidelines be updated at least every 5 years or they would not be included on the National Guideline Clearinghouse website. However, there are no such standardized time criteria for guidelines produced by specialty organizations. Therefore, just as pharmaceutical companies are required to conduct postmarketing surveillance, so too must there be a system to determine the effects of new practice guidelines on patient outcomes. Measures are needed to evaluate the safety and effectiveness of guidelines when they are used in practice.

This call for sound governance—whether it comes from within the profession or from a government agency—will be met with little enthusiasm, especially among groups with a vested interest. For guidelines from groups with direct or indirect support from pharmaceutical companies, the reason seems clear.[24] Likewise, insurance companies may be reluctant because they may be denying coverage based on faulty guidelines or supporting care that may be causing more harm than good.[25] Professional societies may be too invested in their guidelines to want to subject them to independent scrutiny.[26,27]

One way out of this dilemma could be to incorporate an analysis of effectiveness and safety into the guideline production process, as is currently recommended.[28] Guidelines would be regarded as preliminary until effectiveness and safety analyses are completed using data from real-world experiences. Communicating the preliminary nature of a guideline might help to discourage its widespread adoption until concrete evidence of good results is available.

Another approach is education, whereby physicians are taught how to identify trustworthy guidelines. Several tools exist,[29,30] and we are working on a new one that is easy to use and relevant to clinicians.[31]

The obvious questions then become, Where would the money come from to do the analysis? Who would be recruited to undertake the evaluation? In Australia, the National Health and Medical Research Council approves guidelines written by other groups.[32] In the United States, expanding the role of the Agency for Healthcare Research and Quality to include such oversight might be one possibility. An international approach could be taken, perhaps through expansion of the mission of the Cochrane Collaboration. Others have called for a public-private alliance, with independent panels studying the impact of guidelines, with funding and representation coming from a combination of governments, private foundations, and provider and payer groups.[15]

"Postmarketing surveillance" of clinical practice guidelines requires appropriate measurement tools. Practice-based research networks and other research collaboratives can collect outcomes data in practice. Payer databases can be used to identify changes in outcomes. The results of these analyses could be used to refine guidelines.

In the past, other producers of medical interventions—from medicine and device manufacturers to diagnostic tests developers—have opposed calls for oversight. Rigorous evaluation through the regulation of these industries, though, provides protection from harm associated with the misuse and overuse of their products. While more rigorous practice guideline evaluation will most likely be met with similar resistance, it should not follow the same timeline medicine regulation has taken; we need to start now. More evaluation and perhaps even regulation will mean less bias in guidelines, less confusion among physicians attempting to adhere to evidence-based practices, and more benefit for the end users: our patients.

References

1. Genuis SJ. The proliferation of clinical practice guidelines: professional development or medicine-by-numbers? J Am Board Fam Pract 2005;18:419–25.
2. Dobrow MJ, Goel V, Lemieux-Charles L, Black NA. The impact of context on evidence utilization: a framework for expert groups developing health policy recommendations. Soc Sci Med 2006;63:1811–24.
3. Garber AJ, Abrahamson MJ, Barzilay JI, et al. Consensus statement by the American Association of Clinical Endocrinologists and American College of Endocrinology on the comprehensive type 2 diabetes management algorithm - 2016 executive summary. Endocrine Pract 2016;22:84–113.

4. Redmon B, Caccamo D, Flavin P, et al. Diagnosis and management of type 2 diabetes mellitus in adults. Guideline summary. 16th ed. Revised July 2014. Available from: https://www.icsi.org/guidelines_more/catalog_guidelines_and_more/catalog_guidelines/catalog_endocrine_guidelines/diabetes/. Accessed April 4, 2016.
5. Moyer VA; US Preventive Services Task Force. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2012;157:120–34.
6. Carter HB, Albertsen PC, Barry MJ, et al. Early detection of prostate cancer: AUA guideline. J Urol 2013;190:419–26.
7. US Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2009; 151:716–26, W-236.
8. Lee CH, Dershaw DD, Kopans D, et al. Breast cancer screening with imaging: recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. J Am Coll Radiol 2010;7:18–27.
9. Niederman MS, Bass JB Jr, Campbell GD, et al. Guidelines for the initial management of adults with community-acquired pneumonia: diagnosis, assessment of severity, and initial antimicrobial therapy. American Thoracic Society. Medical Section of the American Lung Association. Am Rev Resp Dis 1993; 148:1418–26.
10. Gleason PP, Kapoor WN, Stone RA, et al. Medical outcomes and antimicrobial costs with the use of the American Thoracic Society guidelines for outpatients with community-acquired pneumonia. JAMA 1997;278:32–9.
11. Camm AJ, Lip GY, De Caterina R, et al. 2012 Focused update of the ESC guidelines for the management of atrial fibrillation: an update of the 2010 ESC guidelines for the management of atrial fibrillation–developed with the special contribution of the European Heart Rhythm Association. Europace 2012;14:1385–413.
12. Andrade AA, Li J, Radford MJ, Nilasena DS, Gage BF. Clinical benefit of American College of Chest Physicians versus European Society of Cardiology guidelines for stroke prophylaxis in atrial fibrillation. J Gen Intern Med 2015;30:777–82.
13. Alper BS, Malone-Moses M, McLellan JS, Prasad K, Manheimer E. Thrombolysis in acute ischaemic stroke: time for a rethink? BMJ 2015;350:h1075.
14. Shekelle PG, Kravitz RL, Beart J, Marger M, Wang M, Lee M. Are nonspecific practice guidelines potentially harmful? A randomized comparison of the effect of nonspecific versus specific guidelines on physician decision making. Health Serv Res 2000;34: 1429–48.
15. Quanstrum KH, Hayward RA. Lessons from the mammography wars. N Engl J Med 2010;363: 1076–9.
16. Kahneman D, Tversky A. On the reality of cognitive illusions. Psychol Rev 1996;103:582–91; discussion 592–6.
17. Guyatt G, Akl EA, Hirsh J, et al. The vexing problem of guidelines and conflict of interest: a potential solution. Ann Intern Med 2010;152:738–41.
18. Immel BK. A brief history of the GMPs for pharmaceuticals. Pharm Technol 2001:44–52.
19. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual. Published October 2014; updated April 2016. Available from: https://www.nice.org.uk/article/pmg20/chapter/14-Updating-guidelines. Accessed March 30, 2016.
20. Institute of Medicine. Clinical practice guidelines we can trust. Washington, DC: National Academies Press; 2011.
21. Schunemann HJ, Al-Ansary LA, Forland F, et al. Guidelines international network: principles for disclosure of interests and management of conflicts in guidelines. Ann Intern Med 2015;163:548–53.
22. Shaneyfelt TM, Centor RM. Reassessment of clinical practice guidelines: go gently into that good night. JAMA 2009;301:868–9.
23. Burda BU, Norris SL, Holmer HK, Ogden LA, Smith ME. Quality varies across clinical practice guidelines for mammography screening in women aged 40–49 years as assessed by AGREE and AMSTAR instruments. J Clin Epidemiol 2011;64:968–76.
24. Coyne DW. Influence of industry on renal guideline development. Clin J Am Soc Nephrol 2007;2:3–7; discussion 13–4.
25. Berg M, Horstman K, Plass S, Van Heusden M. Guidelines, professionals and the production of objectivity: standardisation and the professionalism of insurance medicine. Sociol Health Illn 2000;22:765–91.
26. Norris SL, Burda BU, Holmer HK, et al. Author's specialty and conflicts of interest contribute to conflicting guidelines for screening mammography. J Clin Epidemiol 2012;65:725–33.
27. Norris CL, Holmer HK, Ogden LA, Burda BU. Conflict of interest in clinical practice guideline development: a systematic review. PloS One 2011;6:e25153.
28. AGREE Next Steps Consortium. Appraisal of guidelines for research and evaluation II. AGREE II instrument. May 2009. Available from: http://www.agreetrust.org/wp-content/uploads/2013/06/AGREE_II_Users_Manual_and_23-item_Instrument_ENGLISH.pdf. Accessed September 29, 2016.
29. Lenzer J, Hoffman JR, Furberg CD, Ioannidis JP; Guideline Panel Review Working Group. Ensuring the integrity of clinical practice guidelines: a tool for protecting patients. BMJ 2013;347:f5535.
30. Semlitsch T, Blank WA, Kopp IB, Siering U, Siebenhofer A. Evaluating guidelines: a review of key quality criteria. Dtsch Arztebl Int 2015;112:471–8.

31. Cosgrove L, Shaughnessy AF. Development of a clinician-friendly guideline assessment tool. Guideline International Network Scientific Programme 2015; 2015; Amsterdam, the Netherlands.
32. Australian Government, National Health and Medical Research Council (August 10). Procedures and requirements for meeting the 2011 NHMRC standard for clinical practice guidelines. May 2011, version 1.1. Available from: https://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/cp133_nhmrc_procedures_requirements_guidelines_v1.1_120125.pdf. Accessed March 30, 2016.