

The Benefits and Concerns Surrounding the Automation of Clinical Guidelines

Samuel Cykert

Automated guidelines often improve outcomes when applied to simple clinical states. They are more effective when human-computer interaction and workflow changes are considered in implementation. "Alert fatigue" might lead to uneven implementation of guidelines. For complex patients with multiple illnesses, more research should be geared toward the structure and effect of guidelines. Evidentiary uncertainty and complicating comorbid conditions continue to require meticulous incorporation of patient values and physician judgment.

The use of clinical guidelines has always been controversial. The phrase "cookbook medicine" became a rallying cry against any standardized care or proscribed behaviors because of perceptions that such measures restricted physicians' use of their judgment and training. However, as the measurement of evidence-based outcomes became more doable and transparent, data revealed that unexpectedly low proportions of patients receive expected care and that large regional variations exist in cost and outcomes; this led to a groundswell of support from both the medical and patient communities to establish guidelines for circumstances where the evidence is strong and the answers to best tests and treatments can indeed be straightforward and standardized. The National Academy of Medicine's report *Crossing the Quality Chasm* added to this sentiment [1].

In an era that preceded the widespread adoption of electronic health records (EHRs), McGlynn and colleagues demonstrated that optimal care recommended by clinical guidelines occurred slightly more than half the time in adult primary care practices, thus showing that the ambulatory care environment was not immune from concerns about quality of care [2]. In this report, it did not matter whether the care was for prevention, diagnosis, or treatment—the limitations were the same. Given this backdrop, one of the anticipated benefits of EHRs was the opportunity to use automated protocols to help clinicians adhere more closely to recommended care. As early as 1995, however, Tierney and colleagues demonstrated that, in the case of congestive heart failure, incorporating complex guidelines could be difficult for several reasons: lack of specific electronic definitions for symptom severity and adverse events, difficulties accounting for concurrent drugs and comorbid ill-

nesses, and problems assessing the longitudinal timing for when tests and treatments should occur [3]. Therefore, rather than implementing wholesale guidelines, most EHR vendors and clinical organizations have taken a more conservative approach, using point-of-care reminders or computerized clinical decision support built into disease-specific templates.

With all these factors in play, the following questions should be applied to guideline use. First, if guidelines are created, what is the quality of the evidence on which they are based? Second, if the guidelines are to be incorporated into EHRs, are the algorithms defined sharply enough to create structured and measurable programming logic? Third, EHRs do not treat patients, providers do; how do you create the human-electronic interactions to assure that the guidelines are properly addressed and enacted? Fourth, what is a clinically meaningful way to measure the effect of guidelines? If guidelines are enacted and they do not move the process or outcome needle, then they simply become another administrative checkbox.

Regarding the first question, the Patient Protection and Affordable Care Act of 2010 mandates coverage of preventive services given an A or B evidence grade by the US Preventive Services Task Force (USPSTF) [4]. This highlights the standard evaluation system used to rate quality of evidence and states that every enacted guideline should at least meet the defined minimum of Grade B. [Editor's note: For more information about levels of evidence, refer to the sidebar by Lisa Edgerton on pages 240-241.]

The second question brings up much semantic confusion and controversy. As with Tierney's congestive heart failure case [3], how do you program the degree of shortness of breath that necessitates an increased diuretic dose? How much fatigue is needed to reduce a beta blocker dose? When have symptoms progressed enough to order another echocardiogram or perform a new ischemic work-up? The full guideline that constitutes the management of heart failure is almost impossible to program. However, rather than

Electronically published September 11, 2015.

Address correspondence to Dr. Samuel Cykert, 145 N Medical Dr, CB #7165, Chapel Hill, NC 27599 (samuel_cykert@med.unc.edu).

N C Med J. 2015;76(4):235-237. ©2015 by the North Carolina Institute of Medicine and The Duke Endowment. All rights reserved. 0029-2559/2015/76407

programming an intact end-to-end protocol, elements of the protocol that clearly impact outcomes—such as ACE inhibitors and beta blockers for patients with low ejection fractions, spironolactone for patients with class IV symptoms according to the New York Heart Association Functional Classification system, or a trigger to order a serum potassium within 1 week for a patient recently started on a potassium-retaining medication—can be separately programmed as point-of-care reminders.

On the other hand, the ordering of a mammogram can be fairly straightforward and simple to program. Given the latest USPSTF recommendations [5], a reminder system can be built where, for women aged 40–49 years, a shared decision-making reminder pops up; for women aged 50–74 years, a firm reminder with order set pops up at the appropriate time; and then, beginning at age 75 years, another shared decision-making signal comes up. Although these 2009 recommendations are slightly more complicated than past guidelines, the EHR output can easily adapt.

The third question delineated above is crucial. If EHR decision support is triggered and no workflow is assigned to enact the decision, then improvement does not happen. Even by implementing a thorough office workflow, what happens to recommended care if the patient does not have a scheduled appointment when specific adjustments, tests, or treatments are due? Optimum benefit is not obtained without population management principles (eg, creating lists of required care independent of office visits). Back to the mammography example, the shared decision making warning must trigger an actual discussion or it has no value. In the more straightforward 50–74-year-old group, if a nurse has a standing protocol to place a mammogram order for each woman who is electronically flagged on arrival, then mammogram rates will rise. However, optimal care will more likely be achieved if women in this age group receive automated reminders when their mammogram is due, whether or not they are in the office.

Finally, concerning question #4, if an organization does not measure what they have built in response to questions 1–3, then the medical staff and affected patients will not be able to assess whether their time, sweat, and costs led to better outcomes, thus restricting our ability to make adjustments that make guidelines more effective.

Does evidence show that automated protocols or other EHR supports work? Although this research in the modern EHR environment is relatively immature, there are positive signs. A systematic review published in 1998 showed beneficial results for preventive care, drug dosing, and treatment, although an improvement in diagnosis was less clear [6]. More recent systematic reviews show that computerized clinical decision support as a sole intervention modestly improves processes of care in chronic disease management [7–9]. Very specific areas have also been studied. Among patients with asthma, for example, significant improvements in prescriptions for controller medications and action

plans as well as reductions in exacerbations have been demonstrated [10, 11]. A recent systematic review of diabetes care showed small improvements in hemoglobin A_{1c} levels with similar reminders [12].

The experience of the Practice Support Program of the North Carolina Area Health Education Centers (AHEC) parallels a report by Cleveringa and colleagues [13]. Simply adding computer reminder capabilities can lead to a small improvement in diabetes outcomes. However, if onsite education and data feedback are added, then the improvements have greater clinical impact. In the AHEC example, the first 50 practices that met meaningful use criteria in the Practice Support Program increased the percentage of well-controlled diabetic patients (hemoglobin A_{1c} levels less than 7%) from 41% to 51%. When onsite practice facilitation and formal quality improvement (QI) workflow techniques, including data feedback, were added to the electronic reminder systems, the percentage of diabetic patients whose condition was well controlled rose to 60% (unpublished data). Although hemoglobin A_{1c} is an intermediate outcome, we know that the improvements experienced by the 26,000 patients in these practices will lead to large reductions in mortality and disease sequelae over the next 10 years [14].

Automated guidelines or related decision support can have significant limitations. Some guidelines supported by credible evidence cannot be applied to underrepresented groups. For instance, the recent USPSTF guidelines on lung cancer screening with low-dose computed tomography are based on evidence that includes few patients above the age of 65 years and contains multiple exclusions for those with complex medical illnesses [15]. Given the predominance in the lung cancer population of older heavy smokers with comorbid illnesses, simple triggers for screening in this population group will not suffice.

Another possible barrier to broad implementation of computerized decision support is the development of alert fatigue. Although there are no quantitative estimates of the extent and effect of alert fatigue in primary care, there have been occasional case reports of complications attributable to an ignored alert [16]. By propagating multiple electronic guidelines without appropriate prioritization, resultant alert fatigue could potentially lead to trivial alerts being addressed and the guidelines with the greatest potential impact being indiscriminately ignored.

Finally, guidelines tend to be based on single conditions in isolation. Given the aging population and the increasing prevalence of complex multisystem disease, following multiple single-disease protocols is likely to lead to polypharmacy and patient harm. More research is needed to create the best electronic care algorithms for patients with complex conditions; ultimately, such algorithms should prioritize and eliminate tests and/or medications, rather than compounding and multiplying the related risks and side effects in this vulnerable group.

In conclusion, automated guidelines for EHRs with distinct programmable definitions are extremely useful and improve outcomes when applied to simple procedures and clinical states. They are even more effective when significant attention is paid to human-computer interaction, workflow, and utilization of data feedback with QI and population management techniques. However, more research needs to be geared toward the effect of these guidelines in complex clinical states, particularly elderly populations with multisystem disease. Automated guidelines should be prioritized toward high-impact conditions in order to avoid overuse and alert fatigue, and the effectiveness of guidelines and workflows should constantly be assessed and adjusted when indicated. Although guidelines and computerized decision support will undoubtedly improve care for which standardization is well established, evidentiary gray areas and complicating comorbid conditions continue to require meticulous incorporation of patient values and physician judgment. *NCMJ*

Samuel Cykert, MD professor, Department of Medicine, University of North Carolina at Chapel Hill; director, Program on Health and Clinical Informatics, UNC School of Medicine; associate director, Medical Education, North Carolina Area Health Education Centers program; clinical director, Practice Support Services, North Carolina Area Health Education Centers program, Chapel Hill, North Carolina.

Acknowledgments

Potential conflicts of interest. S.C. has no relevant conflicts of interest.

References

- Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academies Press; 2001.
- McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. *N Engl J Med*. 2003;348(26):2635-2645.
- Tierney WM, Overhage JM, Takesue BY, et al. Computerizing guidelines to improve care and patient outcomes: the example of heart failure. *J Am Med Inform Assoc*. 1995;2(5):316-322.
- Koh HK, Sebelius KG. Promoting prevention through the Affordable Care Act. *N Engl J Med*. 2010;363(14):1296-1299.
- US Preventive Services Task Force. Screening for breast cancer: US Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2009;151(10):716-726.
- Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *JAMA*. 1998;280(15):1339-1346.
- Damiani G, Pinnarelli L, Colosimo SC, et al. The effectiveness of computerized clinical guidelines in the process of care: a systematic review. *BMC Health Serv Res*. 2010;10(1):2.
- Roshanov PS, Misra S, Gerstein HC, et al; CCDSS Systematic Review Team. Computerized clinical decision support systems for chronic disease management: a decision-maker-researcher partnership systematic review. *Implement Sci*. 2011;6(1):92.
- Shojania KG, Jennings A, Mayhew A, Ramsay C, Eccles M, Grimshaw J. Effect of point-of-care computer reminders on physician behaviour: a systematic review. *CMAJ*. 2010;182(5):E216-E225.
- Bell LM, Grundmeier R, Localio R, et al. Electronic health record-based decision support to improve asthma care: a cluster-randomized trial. *Pediatrics*. 2010;125(4):e770-e777.
- Kuhn L, Reeves K, Taylor Y, et al. Planning for action: the impact of an asthma action plan decision support tool integrated into an electronic health record (EHR) at a large health care system. *J Am Board Fam Med*. 2015;28(3):382-393.
- Jeffery R, Iserman E, Haynes R; CDSS Systematic Review Team. Can computerized clinical decision support systems improve diabetes management? A systematic review and meta-analysis. *Diabet Med*. 2013;30(6):739-745.
- Cleveringa FG, Gorter KJ, van den Donk M, van Gijssel J, Rutten GE. Computerized decision support systems in primary care for type 2 diabetes patients only improve patients' outcomes when combined with feedback on performance and case management: a systematic review. *Diabetes Technol Ther*. 2013;15(2):180-192.
- Holman RR, Paul SK, Bethel MA, Neil HA, Matthews DR. Long-term follow-up after tight control of blood pressure in type 2 diabetes. *N Engl J Med*. 2008;359(15):1565-1576.
- Moyer VA; US Preventive Services Task Force. Screening for lung cancer: US Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014;160(5):330-338.
- Carspecken CW, Sharek PJ, Longhurst C, Pageler NM. A clinical case of electronic health record drug alert fatigue: consequences for patient outcome. *Pediatrics*. 2013;131(6):e1970-e1973.