

## VIEWPOINT

# Clinical Decision Support in the Era of Artificial Intelligence

**Edward H. Shortliffe, MD, PhD**

Biomedical Informatics, Columbia University, New York, New York; and Biomedical Informatics, Arizona State University, Phoenix.

**Martin J. Sepúlveda, MD, ScD**

Retired from IBM Research, Watson Research Laboratory, Yorktown Heights, New York.

**Clinicians and researchers** have long envisioned the day when computers could assist with difficult decisions in complex clinical situations. The first article on this subject appeared in the scientific literature about 60 years ago,<sup>1</sup> and the notion of computer-based clinical decision support has subsequently been a dominant topic for informatics research. Two recent Viewpoints in *JAMA* highlighted the promise of deep learning in medicine.<sup>2,3</sup> Such new data analytic methods have much to offer in interpreting large and complex data sets. This Viewpoint is focused on the subset of decision support systems that are designed to be used interactively by clinicians as they seek to reach decisions, regardless of the underlying analytic methodology that they incorporate.

With the evolution of digital and communication technologies plus innovative software methods, the ability to offer high-quality support to clinicians has resulted in impressive new capabilities and several commercial products. For example, many decision support tools are built into medical devices, creating new ways to visualize or interpret data that are provided

- Time is a scarce resource: A CDSS should be efficient in terms of time requirements and must blend into the workflow of the busy clinical environment.
- Complexity and lack of usability thwart use: A CDSS should be intuitive and simple to learn and use so that major training is not required and it is easy to obtain advice or analytic results.
- Relevance and insight are essential: A CDSS should reflect an understanding of the pertinent domain and the kinds of questions with which clinicians are likely to want assistance.
- Delivery of knowledge and information must be respectful: A CDSS should offer advice in a way that recognizes the expertise of the user, making it clear that it is designed to inform and assist but not to replace a clinician.
- Scientific foundation must be strong: A CDSS should have rigorous, peer-reviewed scientific evidence establishing its safety, validity, reproducibility, usability, and reliability.

Health care is a particularly challenging domain for decision support. A CDSS requires strong analytical capabilities that can function effectively in a domain where the understanding of causal mechanisms and relationships is still incomplete and where uncertainty, and an approach to managing it, is accordingly inevitable. A CDSS must provide valid support while simultaneously addressing the list of demanding requirements to help ensure a system's adoption by clinicians. For example,

effective decision support capabilities are often those that avoid additional data entry tasks, such as a CDSS that acquires the bulk of the data needed for analyzing a case through integration with an electronic health record (EHR). Today's EHRs have not made this easy because they generally lack the cross-platform transparency and standards that would be needed for a single CDSS to be tightly integrated with multiple EHR products or implementations.

Different decision-making tasks often pose different challenges for a CDSS. For example, a system designed to assist with clinical diagnosis is very different from one that is intended to assist with therapy planning. A CDSS for diagnosis can generally be built on linkages between clinical data and gold standards for accuracy (eg, biopsies, autopsies, biomolecular markers, or surgical findings). But in formulating a therapeutic plan, especially in complex settings, there is often no gold standard, and there may be disagreement, even among experts. For example, an early study evaluated a program designed to assist with the selection of antibiotic therapy

Despite the enthusiasm for exploring the potential of artificial intelligence and decision support in clinical settings, several complexities limit the ability to move ahead quickly.

to expert users. Artificial intelligence programs, which are increasingly based on a variety of machine learning and natural language processing methods, are especially prominent in these data interpretation and text mining settings.

Why, then, do clinical decision support systems (CDSSs) designed for direct interactive use by clinicians have challenges of credibility and adoption when the literature has been replete for 4 decades with studies that present computing systems demonstrating diagnostic accuracy that rivals the performance of expert clinicians?<sup>4,5</sup> The reasons are varied and reflect the realities and complexities of clinical practice. Biomedical informaticians have long understood those reasons, recognizing the spectrum of capabilities and characteristics that must be incorporated into a CDSS if it is to be accepted and integrated into routine workflow:

- Black boxes are unacceptable: A CDSS requires transparency so that users can understand the basis for any advice or recommendations that are offered.

## Corresponding

**Author:** Edward H. Shortliffe, MD, PhD, Biomedical Informatics, Columbia University, 272 W 107th St, 5B, New York, NY 10025 ([ted@shortliffe.net](mailto:ted@shortliffe.net); [ehs79@columbia.edu](mailto:ehs79@columbia.edu)).

for patients with meningitis prior to the identification of the specific infecting organism by the laboratory. Results showed that none of the infectious disease experts in the study (including the program) suggested a regimen that was judged acceptable more than 70% of the time (as determined by a separate expert panel that assessed all suggested therapies without knowing which was offered by the computer program).<sup>6</sup> This highlights that for many decisions there is no single “right answer.”

There are resulting implications for the design of studies to evaluate a CDSS that offers therapeutic advice rather than a diagnosis. The same can be said for any program that recommends actions (eg, how to evaluate a patient with a specific abnormality). If a therapeutic advice system and an expert clinician reach different conclusions about how to manage a specific case, it is not clear that either is “correct.” Both plans, or neither, may be reasonable. Studies must accordingly be designed not to determine what is correct but rather how well a system performs on a given task when (blindly) compared with how other clinicians, usually acknowledged experts, perform on the same task.

There are regulatory standards for analytic decision software that is implemented as part of a medical device, such as in closed-loop systems that interpret clinical data and treat a patient without intervention by human intermediaries (eg, processing data from sensors and then making automatic adjustments in medication infusions or ventilator settings). However, there are no such formal standards for decision support software that interacts directly with a clinician who then takes action, potentially influenced by that advice. The US Food and Drug Administration (FDA) first formulated informal policy on this in 1987<sup>7</sup> and, with ongoing discussion with the community, reaffirmed it in draft guidance to industry and FDA staff in 2017.<sup>8</sup>

How, then, should a CDSS be evaluated? It is generally accepted that a CDSS should provide rigorous evidence of its safety and reliability as well as scientific evidence that the advice provided is similar to or better than the standard of care. Furthermore, the CDSS should provide evidence that it has implemented processes to ensure that the system maintains the currency of its knowledge base and that it is safe to use. Demonstrating product safety requires (1) that there is evidence of a systematic process that identifies predictable errors during product development and then engineers them out; (2) that the developer has established procedures for dealing with unpredictable errors; and (3) that a monitoring system is deployed during use that identifies near-misses or other problems so as to inform product improvement. Clinical decision support systems are not perfect instruments and will have failures. But a CDSS must be designed to be fail-safe and to do no harm.<sup>9</sup> In addition, equal attention must be paid to evidence of a CDSS' ease of workflow integration and its usability as measured against the attributes previously noted, without which a CDSS will inevitably fail.

Despite the enthusiasm for exploring the potential of artificial intelligence and decision support in clinical settings, several complexities limit the ability to move ahead as quickly as some may predict. Ongoing work requires realistic recognition of the full set of capabilities that an effective CDSS requires. This includes research on how best to address some important lingering issues and carefully designed and peer-reviewed studies that develop the formal and sequenced evidence that is required in medicine when patients' lives and well-being are at stake. Valuable systems are already in development or use, and the potential for their effective integration into routine care settings has never been stronger.

## ARTICLE INFORMATION

**Published Online:** November 5, 2018.  
doi:10.1001/jama.2018.17163

**Conflict of Interest Disclosures:** Drs Shortliffe and Sepúlveda report having part-time roles as paid senior executive consultants with IBM Watson Health. This article was conceived and written without any involvement by that company.

## REFERENCES

1. Ledley RS, Lusted LB. Reasoning foundations of medical diagnosis; symbolic logic, probability, and value theory aid our understanding of how physicians reason. *Science*. 1959;130(3366):9-21. doi:10.1126/science.130.3366.9
2. Naylor CD. On the prospects for a (deep) learning health care system. *JAMA*. 2018;320(11):1099-1100. doi:10.1001/jama.2018.11103
3. Hinton G. Deep learning—a technology with the potential to transform health care. *JAMA*. 2018;320(11):1101-1102. doi:10.1001/jama.2018.11100
4. Miller RA, Pople HE Jr, Myers JD. Internist-1, an experimental computer-based diagnostic consultant for general internal medicine. *N Engl J Med*. 1982;307(8):468-476. doi:10.1056/NEJM198208193070803
5. Ting DSW, Cheung CYL, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*. 2017;318(22):2211-2223. doi:10.1001/jama.2017.18152
6. Yu VL, Fagan LM, Wraith SM, et al. Antimicrobial selection by a computer: a blinded evaluation by infectious diseases experts. *JAMA*. 1979;242(12):1279-1282. doi:10.1001/jama.1979.03300120033020
7. Young FE. Validation of medical software: present policy of the Food and Drug Administration. *Ann Intern Med*. 1987;106(4):628-629. doi:10.7326/0003-4819-106-4-628
8. US Food and Drug Administration. *Software as a Medical Device (SaMD): Clinical Evaluation*. December 8, 2017. <https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM524904.pdf>. Accessed October 1, 2018.
9. Fox J, Das S. *Safe and Sound: Artificial Intelligence in Hazardous Applications*. Cambridge, MA: MIT Press; 2000.