Envisioning Watson As a Rapid-Learning System for Oncology

By Jennifer L. Malin, MD, PhD

WellPoint, Thousand Oaks; and Jonsson Comprehensive Cancer Center and David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA

We are on the cusp of transformational change in health care, in general and in the field of oncology, specifically. The last four decades have witnessed remarkable advances in the treatment of cancer and significant improvements in outcomes, with mortality rates for most cancers declining.[1] By deciphering the molecular code that underlies the development of cancer, therapeutic advances long hoped for can finally be realized. However, this increased understanding of the molecular basis of disease is accompanied by a dramatic increase in the complexity of the treatment paradigms for cancer. A recent study identified 35 significantly mutated genes in breast cancer, many of which may be targets for new therapies.[2] As we begin to tailor our treatments to all of the different genetic varieties of malignancy, the process of determining the most appropriate cancer therapy will become exponentially more complex and strain the abilities of even the most seasoned clinician.

In parallel with the genomic revolution, advances in information technology are producing data at a rate heretofore unanticipated. To capitalize on this substantial progress and further drive innovation, the Institute of Medicine and others are advocating rapid-learning health care, wherein data obtained during the course of routine clinical practice are available in real time for comparative-effectiveness research, quality improvement, safety, and generation of new hypotheses for investigation.[3,4] An article in this issue of *Journal of Oncology Practice* (*JOP*) describes the efforts of the American Society of Clinical Oncology to develop CancerLinQ, a rapid-learning system for oncology.[5] However, a key challenge to fully realizing the rapid-learning health care system is that most information captured by electronic health records today is unstructured, requiring new information technology (IT) tools to access these data.[6]

Developing IT Applications for Oncology Using IBM Watson

Since the 1960s, visionaries have anticipated computers providing clinicians with intelligent real-time consultative services at the bedside.[7] However, decades later, this vision has yet to be realized. Recent advances in IT, including natural language processing, cognitive computing, and machine learning, promise to finally make the so-called e-consultant in medicine a reality.

Named after IBM founder Thomas J. Watson, the Watson computer system was built by a team of IBM scientists with the goal of creating a cognitive computing system that could respond to the evolving big-data informational challenges of today. To demonstrate the capabilities of the system, the team set an objective of the system winning *Jeopardy!*, which it did in 2011, beating two national champions of the game in a nationally televised show.[8,9] To achieve this breakthrough, the scientists who developed Watson had to integrate natural language processing, information retrieval, knowledge representation, machine learning, and a general reasoning paradigm and be able to respond in less than 5 seconds.[9]

The powerful image of Watson winning *Jeopardy!* engenders a mental picture of a computer system that is omniscient and, once primed with the world's medical literature, will be able to answer any question posed to it at lightning speed. However impressive such a system would be, it would not have much clinical utility at the bedside, where the real challenges are to first decompose a complex problem into a series of clinical decisions (eg, whether to proceed directly to a resection or first perform an endobronchial ultrasound biopsy followed by mediastinoscopy on an 82-year-old patient with a small peripheral lung cancer who had a small inferior myocardial infarction 5 months ago but still bicycles to work every day). In contrast to answering a simple question, providing a response regarding treatment options for a patient requires Watson to analyze all of the relevant clinical information about a patient case and then identify those attributes that are relevant to the clinical decision at hand. Although a physician could present a patient case to Watson, much like a trainee would to an attending, having Watson identify the relevant clinical attributes from the source clinical data would provide additional value because this summary could also be used by the oncologist during his or her initial consultation with the patient. Researchers at the Health Information Technology Research Laboratory at the University of Sydney are developing a similar capability for a tool for intensivists; the Intelligent Clinical Notes System extracts pertinent information from patients' medical records regarding their clinical course before transfer to the intensive care unit.[10] Developing a clinical summary for oncologists requires Watson to analyze potentially large volumes of clinical information, which can include years of physician notes, laboratory data, diagnostic imaging, and pathology reports, most of which, even in electronic health records, are stored as free text that can only be accessed with sophisticated natural language processing.

Once the question or patient case has been understood, the next step is for Watson to identify a broad set of candidate treatment options. In this early phase of developing Watson for oncology, treatment guidelines are being used to define the set of candidate treatment options. A physician considers a number of factors when selecting the best treatment option for a patient, including effectiveness, toxicity, comorbid conditions and their potential impact on the benefit and harm of the treatment, patient preferences, and health plan policies. Similarly, the algorithms used by Watson to rank treatment options as more or less optimal for a specific patient will need to weigh all of these factors. Much of the data that a clinician uses to weigh these factors comes for the medical literature; however, where evidence is lacking, he or she relies on experience and guidance from clinical experts. IBM has a dedicated team to seek out sources of cancer care knowledge—text books, medical journals (including *Journal of Clinical Oncology* and *JOP*), and clinical trial databases—and evaluate licensing rights from publishers, allowing content ingestion into the Watson corpus of knowledge. To date, more than 1 million individual articles of evidence have been ingested from hundreds of sources. For Watson to use this medical corpus, we are developing a system that again uses the natural language processing capabilities of Watson. To analyze the evidence Watson must evaluate: first, the relevancy of an article to a patient case (ie, are the patients described in the article similar to the patient in the query); second, the quality of the evidence (ie, an article reporting results of a randomized clinical trial is higher-quality evidence than one summarizing a case series); and third, the sentiment of the article (ie, does it provide evidence that is positive or negative about a particular treatment).

Watson uses machine learning to determine how to weigh the clinical factors in a patient case so as to identify treatment options from the data that have been ingested. The process of developing and tuning the algorithms used by Watson to provide decision support is referred to as training. Similar to medical school students and residents learning during rounds with experienced attending physicians, the performance of Watson improves as it learns from various forms of evidence and data. Watson is provided with training patient cases, which, with input from experts at Memorial Sloan-Kettering Cancer Center, help it tune its algorithms and improve the confidence levels associated with its responses. Initial efforts to date have focused primarily on metastatic non–small-cell lung cancer and early-stage breast cancer. This is an iterative process that repeats over many cycles, through which Watson continues to learn and improve its capabilities, improving its accuracy with each cycle.

Envisioning the Future

What might oncology look like in the future when it is supported by tools like Watson? Before an initial consultation, a patient's entire electronic health record is processed by Watson so that when the oncologist has his or her first encounter with the patient, all of the pertinent clinical data are already concisely summarized for review. During this process, Watson identifies any discrepancies in the medical record and highlights them for the physician to clarify when he or she conducts his or her own assessment of the patient. In addition, as Watson searches for treatment options for the specific patient, the computer will identify diagnostic tests that would enable a more refined set of treatment options and propose those to the clinician. One can envision a scenario where all of this occurs before the first visit so that all necessary diagnostic testing is obtained in advance of the initial consultation. Using the same algorithms, Watson then advises the oncologist when a test he or she was considering ordering would not impact clinical decision making, thereby improving health care efficiency. Finally, after the oncologist has completed his or her assessment of the patient and added any additional pertinent information to the medical record, Watson retrieves the most-applicable treatment options given the individual patient's tumor characteristics, overall health, and preferences, with links to the supporting the evidence. The clinician reviews the proposed treatment options and discusses them with the patient. Once a therapy has been selected, the clinician responds to Watson with the choice, and the information is directly relayed to the patient's health plan for immediate preauthorization. Watson captures as data the treatment selected and eventually the patient's outcomes, allowing it to also serve as the engine for a rapid-learning health care system.

Acknowledgment

Presented at the ASCO Quality Care Symposium, San Diego, CA, November 30-December 1, 2012.

Author's Disclosures of Potential Conflicts of Interest

Although all authors completed the disclosure declaration, the following author(s) and/or an author's immediate family member(s) indicated a financial or other interest that is relevant to the subject matter under consideration in this article. Certain relationships marked with a "U" are those for which no compensation was received; those relationships marked with a "C" were compensated. For a detailed description of the disclosure categories, or for more information about ASCO's conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors.

Employment or Leadership Position: Jennifer L. Malin, WellPoint (C) **Consultant or Advisory Role:** None **Stock Ownership:** Jennifer L. Malin, WellPoint **Honoraria:** None **Research Funding:** None **Expert Testimony:** None **Other Remuneration:** None

Jennifer L. Malin, MD, PhD, WellPoint, 1 WellPoint Way, Thousand Oaks, CA 91362-3893; e-mail: jennifer.malin@wellpoint.com.

DOI: 10.1200/JOP.2013.001021

References

1. Vogelzang NJ, Benowitz SI, Adams S, et al: Clinical cancer advances 2011: Annual report on progress against cancer from the American Society of Clinical Oncology. J Clin Oncol 30:88-109, 2012

2. Cancer Genome Atlas Network: Comprehensive molecular portraits of human breast tumours. Nature 490:61-70, 2012

3. Abernethy AP, Etheredge LM, Ganz PA, et al: Rapid-learning system for cancer care. J Clin Oncol 28:4268-4274, 2010

4. Institute of Medicine: Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. Washington, DC, National Academies Press, 2012

5. Sledge GW, Hudis CA, Swain SM, et al: ASCO's approach to a learning health care system in oncology. J Oncol Pract 9:145-148, 2012

6. Academy Health: Early glimpses of the learning healthcare system: The potential impact of health IT. http://www.academyhealth.org/files/publications/HIT4AKPotential.pdf

7. Schwartz WB: Medicine and the computer: The promise and problems of change. N Engl J Med 283:1257-1264, 1970

8. Ferrucci D, Brown E, Chu-Carroll J, et al: Building Watson: An overview of the DeepQA project. http://www.aaai.org/ojs/index.php/aimagazine/article/view/2303/2165

9. Kroeker KL: Weighing Watson's impact. CACM 54:13-15, 2011

10. Patrick J, Li M: High accuracy information retrieval and information extraction system for electronic clinical notes. Presented at the Fourth Australasian Workshop on Health Informatics and Knowledge Management, Brisbane, Queensland, Australia, January 18-21, 2010