Artificial Intelligence and Clinical Decision Making: The New Nature of Medical Uncertainty

Vinyas Harish, Felipe Morgado, Ariel D. Stern, PhD, and Sunit Das, MD, PhD

Abstract

Estimates in a 1989 study indicated that physicians in the United States were unable to reach a diagnosis that accounted for their patient's symptoms in up to 90% of outpatient patient encounters. Many proponents of artificial intelligence (AI) see the current process of moving from clinical data gathering to medical diagnosis as being limited by human analytic capability and expect AI to be a valuable tool to refine this process. The use of AI fundamentally calls into question the extent to which uncertainty in medical decision making is tolerated. Uncertainty is perceived by some as fundamentally undesirable and thus, for them, optimal decision making should be based on minimizing uncertainty. However, uncertainty cannot be reduced to zero; thus, relative uncertainty can be used as a metric to weigh the likelihood of various diagnoses being correct and the appropriateness of treatments. Here, the authors make the argument, using as examples the experiences of 2 AI systems, IBM Watson on *Jeopardy* and Watson for Oncology, that medical decision making based on relative uncertainty provides a better lens for understanding the application of AI to medicine than one that minimizes uncertainty. This approach to uncertainty has significant implications for how health care leaders consider the benefits and trade-offs of AI-assisted and AI-driven decision tools and ultimately integrate AI into medical practice.

Medical diagnosis has a singular goal: to identify the set of subjective and objective findings (symptoms and signs) that demarcate a patient's illness to correctly identify a disease or diseases. Although accurate diagnosis is central to effective patient care, clinicians often fail in this process. In fact, not receiving a diagnosis is a significant cause of patient dissatisfaction with medical providers.[1,2] Mitigating uncertainty is integral to addressing both the clinical needs of patients and their anxieties over their condition. Thus, considering the nature of uncertainty in medical decision making can be valuable in attempting to improve diagnoses.

Diagnosis is a complex cognitive task that involves logical reasoning and pattern recognition.[3,4] Richardson and Wilson describe the process of diagnosis as involving 2 essential steps.[5] First, the clinician enumerates the diagnostic possibilities and estimates their relative likelihood. Second, the clinician incorporates new information to update the relative probabilities, rules out certain possibilities, and, ultimately, chooses the most likely diagnosis. Thus, with each new finding, the clinician moves from one probability (the pretest probability) to another probability (the posttest probability) to arrive at a diagnosis.

Uncertainty also has implications downstream from the process of diagnosis. The subsequent delivery of care involves an educated prediction of what will happen to the patient in the future given his or her present condition (i.e., the prognosis), and how treatment or prevention might alter the natural progression of the disease. Fuller and Flores have described this process as involving 2 distinct inferences: generalizing risk from a study population to the target-patient population of interest, followed by a patient-specific estimation of the probability that a given individual falls within the target population.[6]

In our view, for clinicians to successfully harness the power of artificial intelligence (AI) as an integral part of the clinical decision-making framework, they should learn to see uncertainty as a relative measure rather than an absolute value that must be minimized. To support this claim, we examine how a popular class of AI methods (deep learning) process complex decision-making tasks through a case study of IBM's Watson and how the translation of these methods to medical decision making has exposed significant considerations around uncertainty.

The Promise of AI

At its core, AI is a tool for using pattern recognition to make predictions. Thus, AI has been leveraged in certain administrative and scheduling domains of medicine; for instance, automated reminders for patients to refill their prescriptions can promote medication adherence.[7] Certain approaches such as deep learning have improved AI's predictive performance on increasingly complex datasets. This ability has enabled the use of AI in domains of medicine traditionally limited to human experts, such as diagnosis and treatment.[8]

AI proponents believe that diagnosis is hindered by humans' analytic capabilities and expect AI to refine the analytic process.[9] This early optimism has perhaps been most significantly realized in areas of medicine dominated by imaging. In a 2017 article in the journal *Nature*, a multidisciplinary group from Stanford University developed a convolutional neural network that performed comparably to 21 board-certified dermatologists on a recognition task designed to differentiate cancers

Please see the end of this article for information about the authors.

The authors have informed the journal that they agree that both Vinyas Harish and Felipe Morgado completed the intellectual and other work typical of the first author.

Correspondence should be addressed to Sunit Das, St. Michael's Hospital, Division of Neurosurgery, 30 Bond St., Toronto, ON M5B 1W8, Canada; telephone: (416) 864-5548; email: sunit.das@utoronto.ca.

Acad Med. 2021;96:31–36.
First published online August 25, 2020
doi: 10.1097/ACM.0000000000003707

from benign seborrheic keratoses and nevi.[10] A similar algorithm has since been developed that was able to outperform 58 dermatologists in identifying malignant melanomas and properly segregating these cancers from benign lesions.[11] In neurology, Siddhartha Mukherjee has written about the ability of AI to identify early signs of stroke on computed tomography scans, which could have profound implications for early intervention and consequently improving patient outcomes.[12] Finally, a recent study from China described a natural language processing system (i.e., a long short-term memory network, a type of deep learning approach) that integrated multifaceted clinical data from 1.3 million pediatric electronic health records to diagnose a wide range of childhood diseases across multiple organ systems; the performance of this system was comparable to that of experienced physicians.[13]

The first 2 examples above represent visual pattern recognition and image analysis. In both applications, the goal of applying AI has been to develop an algorithm that not only builds on human clinical knowledge but also identifies patterns and features invisible to humans. Yet both cases diverge from the diagnostic process typical of most other clinical situations in which subjective evidence (e.g., the patient's experience of illness) informs a clinician's understanding, gathering, and interpretation of objective data (i.e., clinical exam findings, lab tests, imaging). The third example is notable because it encapsulates much of the diagnostic process.

In daily practice, clinicians often need to make decisions in spite of, rather than because of, inconclusive evidence. Another challenge to medical decision making brought forward by Redelmeier and Shafir is the difficulty of weighing multiple alternatives in a given situation, what they call a cognitive bias.[14] They found that family physicians who were presented with more than 1 choice of medication were less likely to prescribe any medication. Clinical encounters in which a preliminary diagnosis remains elusive are common. It was reported in a 1989 study that in nearly 90% of outpatient patient encounters in the United States, physicians were unable to reach an organic diagnosis that accounted for their patient's symptoms.[15] A more recent review (2017) suggests that unexplained symptoms account for 10% to 15% of all general practice consultations.[16] These percentages equate to a large absolute number of patients living with the frustration and confusion of unexplained symptoms. The burden of diagnostic uncertainty is less well studied in the inpatient setting; however, one study found that 22% of patients with pneumonia presented with enough diagnostic uncertainty that the external clinicians reviewing these cases after the fact said they would have delayed antibiotic treatment.[17]

Technology advocates have argued that failures of clinical diagnostics are the result of the limits of human cognition, and, as such, are an opportunity to enhance medical care through the introduction of tools such as AI.[18,19] Providing tools for clinical decision support that incorporate detailed information from a patient's entire medical record to a physician grappling with multiple possibilities could go a long way to resolving the cognitive bias Redelmeier and Shafir present. However, proponents of such an approach might be disappointed to learn that an AI diagnostician generates probabilities rather than discrete answers.

The use of AI fundamentally calls into question the extent to which we tolerate uncertainty in medical decision making. Some view uncertainty as undesirable and argue that optimal decision making is based on the minimization of uncertainty. Yet medical decision making is extraordinarily complex; one study suggested that 45 factors can influence the diagnostic process.[20] Even with AI to help clinicians weigh the likelihood of various diagnoses (and the usefulness of various treatments) against one another, it is not possible to reduce diagnostic uncertainty to zero. We believe that successful integration of AI into the clinical decision-making framework requires clinicians to handle uncertainty as a relative measure rather than an absolute value to minimize. To explore our claim, we use a case study of IBM's Watson to examine how such systems process complex decision-making tasks. We also look at how translating these tasks to medical decision making has exposed significant considerations around uncertainty.

Watson and *Jeopardy*

In January 2011, during a 3-day special event on the television program *Jeopardy*, IBM's AI system, Watson, competed against 2 former show champions. The Watson team's aim was to develop a sophisticated query machine that could process natural language to answer questions.[21,22] Because most human questions are not neatly defined for the discrete logic of a computer operating system, Watson had to process the human-phrased question the *Jeopardy* host asked into a set of search aims, find bodies of knowledge that contained information relevant to the query (e.g., Wikipedia pages, newspaper articles, academic papers, patent files), identify relevant information, and synthesize an answer that most likely satisfied the query and that humans could understand.

The Watson system followed an approach called DeepQA.[21] DeepQA has 4 basic steps: The computer (1) analyzes the question to determine whether different interpretations exist; (2) searches multiple databases and generates thousands of possible answers; (3) scores possible answers on the basis of learned relationships between words and phrases using a collection of algorithms; and (4) weights, ranks, and presents the answers in order of decreasing confidence. If Watson's highest-ranking answer surpassed a confidence threshold, it would attempt to answer the host's question.

Notably, Watson's "thinking" process did not mirror how a human *Jeopardy* contestant processes questions. While both humans and Watson take confidence-driven approaches, only Watson explicitly incorporated confidence as a quantifiable and objective metric. Watson had to proceed in this manner because, unlike humans, it associates all potentially related concepts from raw data with each question. Humans, on the other hand, have an immediate instinct for whether they know the correct answer. This intuitive confidence is a subjective experience for a human contestant. AI approaches such as DeepQA therefore function in a way that is fundamentally different from human intelligence.[23,24] In pop culture, the friction between human and nonhuman reasoning is often highlighted to provide comedic relief. AI-powered characters such as Data in Star Trek and C-3PO in Star Wars baffle their human counterparts by offering logical advice that never lands well because it sidesteps the emotional gravity of a situation. As AI moves from science fiction into scientific fact and medical practice, we must reconcile these differences in reasoning approaches.

By the end of its 3-game *Jeopardy* run, Watson had defeated its human competitors by a considerable margin.[25] While this result was impressive, Watson's most memorable moment for some came during the final round when it responded "Toronto" to a question about American cities. In this instance, Watson's probabilistic answering design prevented it from excluding any solutions with total certainty, leading to an incorrect (albeit low-confidence) conclusion that the audience knew was obviously incorrect.

This anecdote exemplifies why the public may be uncomfortable with an AI system functioning under uncertainty. For a system to wield decision-making power, one must accept that the AI system will eventually draw incorrect inferences and that humans using intuition will see these incorrect inferences as blatantly obvious.

Watson for Oncology

Watson's mistaken inference on *Jeopardy* may serve as a canary in the coal mine for overzealous promises about the use of AI in medicine. While DeepQA performed exceptionally well in the context of a game show, it still demonstrated behavior under uncertainty that called into question its readiness for use in critical systems.[26] Despite this foreshadowing, IBM identified medicine, and oncology in particular, as an early market opportunity for Watson.

Watson for Oncology is a recommendation engine that digests massive amounts of medical literature and patient information to suggest treatment approaches for cancer patients. The system was envisioned to save doctors time and empower them to achieve better outcomes for their patients. During its development, Watson's growing abilities were likened to those of medical professionals at different stages of their training. In 2011, researchers at the University of Maryland and Columbia University trained Watson on Medline, PubMed, and medical textbooks, then tested Watson with questions from the United States Medical Licensing Exam (USMLE) and the New England Journal of Medicine's clinicopathological puzzlers. One researcher proclaimed that Watson was at that point "as good as the smartest second-year medical student."[27] In 2012, Watson passed the USMLE after more training at the Cleveland Clinic Lerner College of Medicine of Case Western Reserve University.[28] Watson then did its "residency" in oncology at Memorial Sloan Kettering Cancer Center (MSKCC) in late 2012, learning about best practices for treating lung, prostate, and breast cancers.[29] Finally, in October 2013, Watson was trained on MD Anderson Cancer Center's extensive leukemia database—subspecializing much like a clinical fellow at a top academic cancer center.[30]

In June 2017, IBM's CEO Ginni Rometty announced that Watson would be able to diagnose and treat "what causes 80% of the cancer in the world."[31] This bold statement led to raised eyebrows among medical journalists given that the MD Anderson Cancer Center had ended its partnership with Watson just a few months earlier.[32] In the months that followed, multiple news outlets released their postmortems on Watson for Oncology.[33–35] After journalists conducted interviews with physicians, AI experts, and company executives, they blasted IBM for "[turning] the marketing engine loose" without acknowledging the complexity and nuance of cancer treatment—and, by extension, the limitations of the company's product.

One clear concern was the external validity, or generalizability, of Watson's recommendations. While IBM used its experience at MSKCC as a selling point, this approach—having Watson get the majority of its training from one institution in New York City and the patients there—introduced a type of bias that legal scholars have termed "contextual bias."[36] Doctors in other parts of the world reported lower concordance between the treatments they recommended and Watson's than the concordance reported for the diagnoses of U.S. doctors and Watson. They claimed that the recommendations Watson had learned from MSKCC oncologists may not be appropriate or relevant for their patients (who may, for example, be drastically different from the generally affluent New Yorkers served by MSKCC). Further, recommendations that disproportionately place more weight on American studies when surveying international literature may be less relevant for international practitioners and at risk of propagating a sort of medical ethnocentrism. Even in cases where Watson was found to return relevant results, Watson's input was estimated to have changed the course of care in only 2% to 10% of cases globally (between 1,680 and 8,400 patients).[37] Finally, there is no published research on whether Watson for Oncology improved survival for the patients it has "treated."[37] It is clear that recommendations from AI systems will need to be channeled through and vetted by local requirements, resources, and expertise before they are followed in patient care.

The Limits of Certainty

Pressure to adopt deep learning–based decision support systems (DL-DSS) like Watson for Oncology will become more pervasive in diagnostics and treatment. This evolution should compel clinicians, regulators, and policymakers to seek to understand why uncertainty is intrinsic to these systems. The U.S. Food and Drug Administration has added clarity to the software provision in the 21st Century Cures Act as to which types of clinical decision support software are no longer under its jurisdiction. However, as of April 2020, no policy discussions have taken place around the role uncertainty has in DL-DSS designed to be used and administered by humans.[38–41] Further, legal scholars have already begun to argue for more adaptive regulatory approaches that would require developers to "disclose information underlying their algorithms."[42]

Perhaps more importantly, the clinical adoption of AI may be a reflection of how intrinsic uncertainty is to medicine. As Sir William Osler once said, "Medicine is the science of uncertainty and the art of probability."[43] The "science of uncertainty" is what has driven interest in DL-DSS, such as IBM's Watson for Oncology, and is what makes the application of such systems to medicine so appealing. Clinicians must reckon with and ultimately accept the fact that no diagnosis is certain, which is why they synthesize differential diagnoses. The calculated probabilities of DL-DSS must, in practice, be reconciled with the intuition of expert clinicians if we are to understand differences in how recommendations emerge. What does it mean to be 76% confident that a patient has myelodysplastic syndrome? We do not assume the intuition of expert clinicians, acquired over many years of experience, could generate such precise measurements of confidence.

To make matters more challenging, the complexity inherent in creating a treatment plan can be even greater than that of arriving at a diagnosis. While diagnostic decision support systems can be verified with gold standards for accuracy, there may not be a gold standard for a therapeutic plan.[38] Treatment decisions are driven not only by a diagnosis but also by a patient's other biomedical comorbidities, biopsychosocial factors, patient preferences, and systems-level constraints about what therapies can be offered. Empirical approaches dominate, especially when there appears to be more than one right answer.[44] Thus, experts may independently generate similarly effective but markedly different treatment plans. In certain areas of medicine, such as oncology, a disagreement between experts may be leveraged to foster best practices. In academic cancer centers, a "tumor board," composed of a range of experts in various disciplines (e.g., medical oncology, surgical oncology, radiology, radiation oncology, and pathology), meets to discuss their most challenging cases. One study found that a referral to a multidisciplinary tumor board led to changes in recommendations for surgical management in 52% of breast cancer patients studied.[45] But when multiple treatment options are similarly effective, how can a DL-DSS evaluate what the best course of therapy is? Should these systems be considered a nonhuman member of a multidisciplinary board of experts? In this area, and others, AI will create new opportunities and raise new ethical and practical challenges.

If clinicians are to practice medicine alongside a DL-DSS, we must ask difficult epistemological questions about uncertainty—and continuously do so as technology evolves. Further, introduction to these technologies—including their benefits and limitations—should be a prioritized focus of future medical training and continuing medical education. A number of aspects of these topics should be tackled. Trainees could be taught about the folly and dangers of excessive diagnostic testing, a lesson which campaigns such as Choosing Wisely have sought to communicate.[46] Trainees may also benefit from acquiring a deeper understanding of Bayesian statistics, which are commonly used in AI algorithms. Bayesian approaches shift conclusions away from the use of frequentist statistics—methods that have historically been taught in medical schools and which view findings as either statistically significant or not—and toward probability distributions, which can be updated as information accrues.

Trainees must be able to critique AI systems for bias, much like they critique observational and experimental studies in the medical literature.[47] This critical stance is especially important because algorithms can affect the care of millions of patients. This bias was notably demonstrated in recent work where a widely used commercial algorithm to identify patients with complex needs was found to use health costs as a proxy for needs, biasing against Black patients.[48] Competencies in clinical reasoning relevant to uncertainty must also be incorporated into assessments. As Cooke and Lemay point out, integrating uncertainty into the evaluation of clinical reasoning for medical trainees is still in its infancy.[44] They conclude that embracing uncertainty and acknowledging the presence of more than one right answer go hand in hand and recommend including these competencies in trainee assessments. Regardless of how it is done, if clinicians are to use AI to aid in diagnosis and therapeutic selection, clinicians and health policymakers must come to terms with what it means to accept human and algorithmic uncertainty as a cornerstone of medicine.

Whether or not one believes in the promise and value of IBM's Watson for Oncology, experiences with this system serve as a poignant reminder that AI's role as a diagnostic or therapeutic aid needs to be scrutinized and evaluated in a multitude of contexts. Like many other clinical decision support tools, AI is built on a scaffolding of statistics and probability. A recent study revealed that two-thirds of doctors surveyed self-reported as not being confident in their understanding of tests and probability.[49,50] Complicating the reliance of AI systems on statistics is a caution from AI experts that many modern deep learning techniques are black boxes: Even the creators of these algorithms cannot fully explain their behavior.[19] Clinicians, patients, payers, and regulators may be understandably concerned by an inability to fully understand AI processes: What ethical role can AI have in diagnostic and therapeutic processes if its recommendations are inscrutable to human experts?

Ultimately, Watson for Oncology serves as a meaningful case study to help those interested in applying AI to medical applications temper their expectations because of diagnostic and therapeutic uncertainty. Humans may not be able to fully comprehend the inner workings of AI algorithms, and there are countless ways in which these algorithms can be imbued with imperfections and biases. Responsible clinicians must therefore endeavor to acknowledge these biases; openly discuss them with regulators, colleagues, and patients alike; and ensure that the messages of pundits and marketing agencies do not compromise the Hippocratic principles underlying their practice.

Funding/Support: None reported.

Other disclosures: S. Das is on the scientific board of the NICO Corporation and is a consultant for AbbVie and Guidepoint.

Ethical approval: Reported as not applicable.

V. Harish is a fourth-year MD–PhD student, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada; ORCID: https://orcid.org/0000-0001-6364-2439.

F. Morgado is a fourth-year MD–PhD student, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada; ORCID: https://orcid.org/0000-0003-3000-9455.

A.D. Stern is associate professor, Technology and Operations Management Unit, Harvard Business School, Harvard University, Cambridge, Massachusetts; ORCID: https://orcid.org/0000-0002-3586-1041.

S. Das is associate professor, Department of Surgery, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada; ORCID: https://orcid.org/0000-0002-2146-4168.

References

1 Burton C, McGorm K, Richardson G, Weller D, Sharpe M. Healthcare costs incurred by patients repeatedly referred to secondary medical care with medically unexplained symptoms: A cost of illness study. J Psychosom Res. 2012;72:242–247.
2 Morriss R, Kai J, Atha C, et al. Persistent frequent attenders in primary care: Costs, reasons for attendance, organisation of care and potential for cognitive behavioural therapeutic intervention. BMC Fam Pract. 2012;13:39.

3 Sox HC, Higgins MC, Owens DK. Medical Decision-Making. Hoboken, NJ: Wiley-Blackwell; 2013.
4 Glass R. Diagnosis, A Brief Introduction. Oxford, UK: Oxford University Press; 1996.
5 Richardson WS, Wilson MC. The process of diagnosis. In: Guyatt G, Rennie D, Meade MO, Cook DJ, eds. Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice. 3rd ed. New York, NY: McGraw-Hill; 2015.
6 Fuller J, Flores LJ. The risk GP model: The standard model of prediction in medicine. Stud Hist Philos Biol Biomed Sci. 2015;54:49–61.
7 Mira JJ, Navarro I, Botella F, et al. A Spanish pillbox app for elderly patients taking multiple medications: Randomized controlled trial. J Med Internet Res. 2014;16:e99.
8 Sanders SF, Terwiesche M, Gordon WJ, Stern AD. How artificial intelligence is changing health care delivery. NEJM Catalyst. https://catalyst.nejm.org/health-care-ai-systems-changing-delivery. Published 2019. Accessed November 24, 2019.
9 Kite-Powell J. See how artificial intelligence can improve medical diagnosis and healthcare. Forbes. https://www.forbes.com/sites/jenniferhicks/2017/05/16/see-how-artificial-intelligence-can-improve-medical-diagnosis-and-healthcare/#39d6d5d06223. Published 2017. Accessed November 24, 2019.
10 Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542:115–118.
11 Haenssle HA, Fink C, Schneiderbauer R, et al; Reader study level-I and level-II Groups. Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol. 2018;29:1836–1842.
12 Mukherjee S. A.I. versus M.D. What happens when diagnosis is automated? The New Yorker. https://www.newyorker.com/magazine/2017/04/03/ai-versus-md. Published 2017. Accessed November 24, 2019.
13 Liang H, Tsui BY, Ni H, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med. 2019;25:433–438.
14 Redelmeier DA, Shafir E. Medical decision making in situations that offer multiple alternatives. JAMA. 1995;273:302–305.
15 Kroenke K, Mangelsdorff AD. Common symptoms in ambulatory care: Incidence, evaluation, therapy, and outcome. Am J Med. 1989;86:262–266.
16 Johansen ML, Risor MB. What is the problem with medically unexplained symptoms for GPs? A meta-synthesis of qualitative studies. Patient Educ Couns. 2017;100:647–654.
17 Metersky ML, Sweeney TA, Getzow MB, Siddiqui F, Nsa W, Bratzler DW. Antibiotic timing and diagnostic uncertainty in Medicare patients with pneumonia: Is it reasonable to expect all patients to receive antibiotics within 4 hours? Chest. 2006;130:16–21.
18 Naylor CD. On the prospects for a (deep) learning health care system. JAMA. 2018;320:1099–1100.
19 Hinton G. Deep learning—A technology with the potential to transform health care. JAMA. 2018;320:1101–1102.
20 Croskerry P. Adaptive expertise in medical decision making. Med Teach. 2018;40:803–808.
21 Ferrucci D, Brown E, Chu-Carroll J, et al. Building Watson: An overview of the DeepQA project. AI Magazine. 2010;31:59–79.
22 International Business Machines Corporation. The DeepQA research team. https://researcher.watson.ibm.com/researcher/view_group.php?id=2099. Published 2016. Accessed November 24, 2019.
23 Kwok K. From perception to cognition: Towards human-understanding and human centricity in AI. Slides presented at: Fujitsu AI Conference; 2017. https://www.fujitsu.com/sg/Images/Perception to Cognition by Dr Kenneth Kwok .pdf. Accessed November 24, 2019.
24 Laird J, Lebiere C, Rosenbloom P. A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. AI Magazine. 2017;38:13–26.
25 Gabbatt A. IBM computer Watson wins Jeopardy clash. The Guardian. https://www.theguardian.com/technology/2011/feb/17/ibm-computer-watson-wins-jeopardy. Published 2011. Accessed November 24, 2019.
26 Knight JC. Safety critical systems: Challenges and directions. In: Proceedings of the 24th International Conference on Software Engineering. ICSE 2002. 2002:547–550.
27 Upbin B. IBM's Watson now a second-year medical student. Forbes. https://www.forbes.com/sites/bruceupbin/2011/05/25/ibms-watson-now-a-second-year-med-student/#3ee192b3300a. Published 2011. Accessed November 24, 2019.
28 Lohr S. I.B.M.'s Watson goes to medical school. Bits by The New York Times. https://bits.blogs.nytimes.com/2012/10/30/i-b-m-s-watson-goes-to-medical-school. Published 2012. Accessed November 24, 2019.
29 Memorial Sloan Kettering. Memorial Sloan Kettering Cancer Center and IBM will collaborate on powerful new medical technology. https://www.mskcc.org/blog/mskcc-and-ibm-will-collaborate-powerful-new-medical-technology. Published 2012. Accessed November 24, 2019.
30 International Business Machines Corporation. MD Anderson taps IBM Watson to power "moon shots" mission aimed at ending cancer, starting with leukemia. https://www-03.ibm.com/press/us/en/pressrelease/42214.wss. Published 2013. Accessed November 24, 2019.
31 Gurdus E. IBM "woke up the A.I. world," CEO Ginni Rometty says. CNBC Mad Money. https://www.cnbc.com/2017/06/20/ibm-woke-up-the-a-i-world-ceo-ginni-rometty-says.html. Published 2017. Accessed November 24, 2019.
32 Herper M. MD Anderson benches IBM Watson in setback for artificial intelligence in medicine. Forbes. https://www.forbes.com/sites/matthewherper/2017/02/19/md-anderson-benches-ibm-watson-in-setback-for-artificial-intelligence-in-medicine/#7bbfd8843774. Published 2017. Accessed November 24, 2019.
33 Ross C, Swetlitz I. IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close. STAT News. https://www.statnews.com/2017/09/05/watson-ibm-cancer. Published 2017. Accessed November 24, 2019.
34 Hernandez D, Greenwald T. IBM has a Watson dilemma. The Wall Street Journal. https://www.wsj.com/articles/ibm-bet-billions-that-watson-could-improve-cancer-treatment-it-hasnt-worked-1533961147. Published 2018. Accessed November 24, 2019.
35 Freedman DH. A reality check for IBM's AI ambitions. MIT Technology Review. https://www.technologyreview.com/s/607965/a-reality-check-for-ibms-ai-ambitions. Published 2017. Accessed November 24, 2019.
36 Price WN. Medical AI and contextual bias. Harvard J Law Technol. 2019. https://ssrn.com/abstract=3347890. Accessed November 24, 2019.
37 Ramsey L. Here's how often IBM's Watson agrees with doctors on the best way to treat cancer. Business Insider. https://www.businessinsider.com/ibm-watson-for-oncology-at-asco-concordance-2017-6. Published 2017. Accessed November 24, 2019.
38 Shortliffe EH, Sepulveda MJ. Clinical decision support in the era of artificial intelligence. JAMA. 2018;320:2199–2200.
39 Center for Devices and Radiological Health. Digital Health Innovation Action Plan. Silver Spring, MD: CDRH, U.S. Food and Drug Administration; 2017. https://www.fda.gov/media/106331/download. Accessed November 24, 2019.
40 Center for Devices and Radiological Health. Software as a Medical Device (SAMD): Clinical Evaluation. Silver Spring, MD: CDRH, U.S. Food and Drug Administration; 2017. https://www.fda.gov/media/100714/download. Accessed November 24, 2019.
41 U.S. Food and Drug Administration. Clinical and Patient Decision Support Software. Silver Spring, MD; FDA, U.S. Health and Human Services, 2017. https://www.fda.gov/media/109618/download. Accessed November 24, 2019.
42 Price WN. Regulating black-box medicine. Mich Law Rev. 2017;116:421–474. http://michiganlawreview.org/wp-content/uploads/2017/12/116MichLRev421_Price.pdf. Accessed November 24, 2019.
43 Osler W. Sir William Osler: Aphorisms from His Bedside Teachings and Writings. In: Bean RB, Bean WB, eds. Springfield, IL: Charles C. Thomas; 1968.
44 Cooke S, Lemay JF. Transforming medical assessment: Integrating uncertainty into the evaluation of clinical reasoning in medical education. Acad Med. 2017;92:746–751.
45 Newman EA, Guest AB, Helvie MA, et al. Changes in surgical management resulting from case review at a breast cancer multidisciplinary tumor board. Cancer. 2006;107:2346–2351.

46 Choosing Wisely. Advancing the Choosing Wisely Campaign in Clinical Practices and Communities. Philadelphia, PA: ABIM Foundation; 2014. http://www.choosingwisely.org/wp-content/uploads/2014/10/Choosing-Wisely-Grant-Report.pdf. Accessed November 24, 2019.

47 Courtland R. Bias detectives: The researchers striving to make algorithms fair. Nature News Feature. https://www.nature.com/articles/d41586-018-05469-3. Published 2018. Accessed November 24, 2019.

48 Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366:447–453.

49 Krouss M, Croft L, Morgan DJ. Physician understanding and ability to communicate harms and benefits of common medical treatments. JAMA Intern Med. 2016;176:1565–1567.

50 Morgan D. What the tests don't show. The Washington Post. https://www.washingtonpost.com/news/posteverything/wp/2018/10/05/feature/doctors-are-surprisingly-bad-at-reading-lab-results-its-putting-us-all-at-risk/?noredirect=on&utm_term=.30ab561336db. Published 2018. Accessed November 24, 2019.

Teaching and Learning Moments

Cause of Life

Shortly after walking into the workroom on my first morning as a resident on the autopsy service, a staff member told me about a voicemail from a patient with questions about autopsy. I figured the caller was a patient's next of kin wanting more information about our process before deciding on legal consent for a loved one, a common conversation on the autopsy service.

My colleague played the message, and I heard a soft voice asking to discuss the details of her own autopsy following her death. We sat there stunned, wondering if we had heard her correctly. Was she really calling about her own autopsy?

I called her back right away to get the full story. She explained her medical history and that she was ready to end her suffering. She was in her 70s with a long history of chronic obstructive pulmonary disease. She also shared with us that she planned to donate her brain and one lung to research and through the consenting process for that was presented with the possibility of an autopsy. "Well, what's an autopsy?" she asked.

I answered her question and explained that an autopsy is also an incredibly powerful learning tool for trainees to learn about diseases and for physicians to understand how an intervention may or may not affect clinical outcomes in patients. She was thrilled. Not wanting to miss a chance for others to learn from her, the patient decided she wanted an autopsy.

Prior to our conversation, this patient's encounters with the health care system were primarily through home visits with her family physician, whose detailed documentation told me not only about the patient's suffering but also how she lived. They had met regularly in her home, every other month for years. And frequently, at the end of a progress note, one or two sentences would detail a conversation about quality of life, the burden of her illness, or her fear of becoming a burden to others. She knew that the state of Vermont allowed patients to follow a process outlined in a law known as Act 39 to legally end their own life with the assistance of a physician.

In the last few months, the patient had deteriorated clinically. She was unable to perform many of her daily activities, and shortness of breath from simply changing her posture was making self-care more of a challenge. She did not have a date for performing Act 39, but she had seen enough of where her disease was taking her. As she told her family physician, "I just want it to be ready when I'm ready."

A couple days after our phone call, I learned that the patient had performed Act 39 and that I could expect her remains to arrive later that day. I felt numb after getting the news, and I struggled to believe that this was the person I had just talked to on the phone. The autopsy team and I even checked for a pulse at multiple anatomic sites before beginning the autopsy, something I cannot recall doing in any other case. We were playing catch-up trying to accept what the patient had made peace with long ago.

As I reflect on this case, I think about the special role of autopsy, giving us a profound and unique view of suffering and humanity. Where else might we witness the next step on a patient's journey? Where else might a patient find meaning and purpose in allowing current and future healers to learn from what she endured? We are challenged to understand both the reason for a person's suffering and that she was much more than a suffering person.

I also think about the amazing family physician who served this patient. She had taken so much time and care to understand her and what she did and did not want from her life. In the end, she had followed her patient's wishes.

Finally, I wonder about those suffering who may decide that they do not wish to continue living under those circumstances. If I learned that someone suffering had taken her own life without the assistance of a physician, would I feel the same empathy I do now? Would I take the time to share my experiences and feelings about the case with others? I have not found all the answers, but I am thankful someone called with the right question.

William Oliver Humphrey, MD

W.O. Humphrey is a second-year resident, Pathology and Laboratory Medicine, University of Vermont Medical Center, Burlington, Vermont; email: william.humphrey@uvmhealth.org; Twitter: @TheDoctorIsIll.

An Academic Medicine Podcast episode featuring this article is available wherever you get your podcasts.