NEJM Catalyst | Innovations in Care Delivery

ARTICLE

Machine Learning and the Pursuit of High-Value Health Care

Ishani Ganguli, MD, MPH, William J. Gordon, MD, MBI, Claire Lupo, Megan Sands-Lincoln, PhD, MPH, Judy George, PhD, Gretchen Jackson, MD, PhD, Kyu Rhee, MD, MPP, David W. Bates, MD, MS

Vol. 1 No. 6 | November — December 2020

DOI: 10.1056/CAT.20.0094

The United States faces unsustainable growth in health care costs despite limited return on this investment in the form of health outcomes, prompting efforts to improve the value of health care. Artificial intelligence has the potential to enable high-value decision-making from the perspectives of the patient, clinician, and health system, but it also could worsen value. This article assesses these opportunities and challenges, separates reality from hype, and proposes policy approaches for contemporary artificial intelligence — specifically, machine learning — to contribute to rather than detract from high-value care. The conclusion is that it is critical to consider value — in particular, the cost component — in all machine-learning work and to let humans and algorithms each do what they do best.

Much of the effort to curb unsustainable growth in U.S. health care spending has focused on reducing the estimated $76 to $101 billion spent on low-value care[1] — that is, medical services for which the potential for harm exceeds the potential for benefit given cost, available alternatives, and patient preferences. To date, most measures of low-value care, such as those named in the Choosing Wisely campaign,[2] have targeted clinical outcomes but not cost[3] and have used rule-based approaches in claims data to identify discrete tests or treatments, often among healthy patients. To realize broader opportunities to improve the value of health care across patient populations and clinical settings, with greater potential for savings,[1,4] we can apply emerging analytical tools to an increasingly rich array of data sources.

In this article, we explore how machine learning, defined as computational techniques that learn from examples rather than operating from predefined rules,[5] could improve the value of health care across a variety of settings, from a patient's decision to seek care, to patient–clinician decisions to pursue testing and treatment, to health-system decisions to promote value across populations. We focus on machine learning, rather than the broader concept of artificial intelligence, because the

capacity of contemporary computational techniques to learn from data and improve over time is particularly important in the age of large and rapidly growing data sets.

We define health care value as quality (including outcomes and patient experience) per unit of cost and focus on the sources of health care waste that are most relevant to care delivery; namely, unnecessary or inefficiently delivered services and missed prevention opportunities.[4] We address how machine-learning algorithms might learn from the actions, preferences, outcomes, and social determinants of health of many patients to promote high-value decisions for individual patients, clinicians, and health systems. At the same time, we emphasize that machine-learning applications may also generate low-value care and consider how to optimize value in all machine-learning work.

Machine Learning and High-Value Care

Most machine-learning methods fall into two categories: supervised and unsupervised. In supervised machine learning, a computer model mathematically maps a given set of inputs to known outcomes. The relationship between inputs and outcomes is formed during a "training" phase. The trained model is then tested on "new" data that the algorithm has never seen. Supervised machine learning underlies many risk prediction scores used in health care today — for example, those used to predict stroke risk (and therefore, the need for a blood thinner) for patients with atrial fibrillation or to predict mortality (and in turn, transplant eligibility) for patients with end-stage liver disease.[6] In contrast, when an outcome is not known, unsupervised machine learning can look for patterns in a data set[7] — for example, to create novel classifications of a disease state such as sepsis.[8]

The application of machine learning for the improvement of health care value has three key requirements. First, users must designate appropriate value measures as outcomes in these models. Although health care cost and quality measurements are notoriously subjective,[9,10] there is a large body of cost-effectiveness literature to guide cost and quality estimation, and results can be benchmarked against standardized thresholds of what we are willing to pay per quality-adjusted life year.[3]

> *"The second challenge is that, in their current state, machine-learning outputs may be unable to provide advice more specific than the fallback "talk to your doctor" or "report to the ED.""*

Second, developers and users must understand that the models are only as good as the data with which they are trained and calibrated. Models assessing health care value would benefit from high-quality, longitudinal, multimodal patient data across care settings and health systems (e.g., clinical and sociodemographic details, patient preferences, time use, and costs). Some of these data are more readily available than others. For instance, though some electronic health record (EHR) systems now allow clinicians to document certain social determinants of health in structured fields, these data are often missing or may be recorded in the format of free, unstructured text. Although natural language processing can be used to extract key concepts from free text and has made

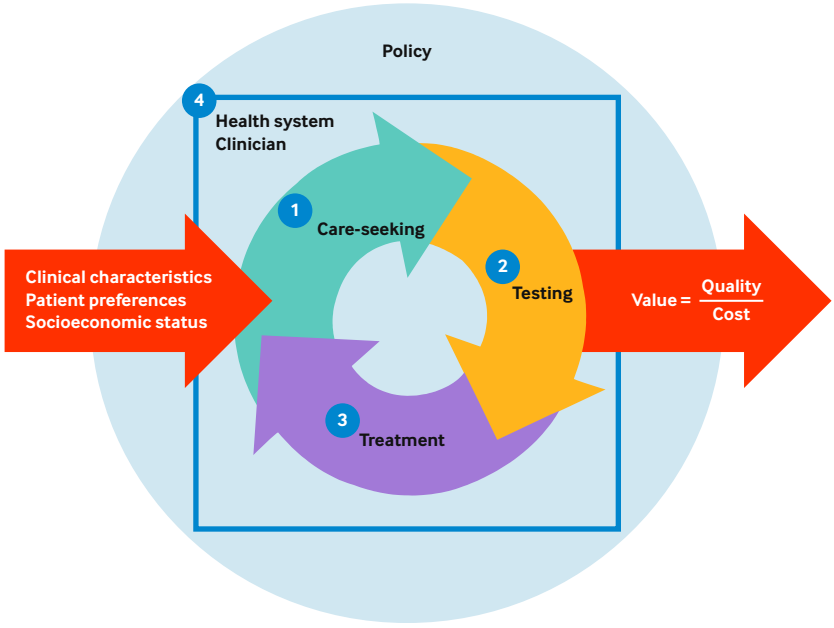
strides in recent years,[11,12] challenges remain in accurately structuring clinical free text.[13,14] Meanwhile, cost data from payers are usually available,[15] and sensors and wearables are enabling the collection of a large volume of behavioral, geographic, environmental, and other patient-generated data. The ability to incorporate multimodal data from disparate sources may help to overcome the shortcomings and biases that are present in any individual source.

Third, these models must be evaluated rigorously in real-world settings — that is, health systems that have invested in the necessary infrastructure, resources, and personnel to apply them effectively. To date, few studies have prospectively evaluated the implementation of machine learning (e.g., using a clinical end point instead of a statistical end point), and machine learning may worsen value if not deployed thoughtfully.[16,17] Privacy considerations are also paramount, and access to robust data may be limited by considerations related to patient preference, cybersecurity, and multi-institutional data sharing, among others. With these caveats in mind, we address each of the following use cases for machine-learning applications with cautious optimism (Figure 1).

FIGURE 1

Care-Seeking Across the Patient Journey

Diagram illustrating the potential role for machine-learning applications from the perspectives of the patient, clinician, and health system.



Source: The authors.

NEJM Catalyst (catalyst.nejm.org) © Massachusetts Medical Society

The Decision To Seek Care

Today, when patients develop a symptom such as heart palpitations, they may go online, ask a family member, or call their primary care office to decide whether, when, and from whom to seek

further medical care. Although several factors ought to play into this decision, patients rarely have the tools to integrate them systematically or with the appropriate weights. Machine-learning applications could assist laypeople in incorporating available information for decision-making. In this scenario, patients might seek guidance from a machine-learning–enabled Web- or application-based algorithm offered by their health system. Such a model might be used to consider inputs that are personal (e.g., demographics, medical history, prior laboratory and imaging results, genetics, social determinants of health, risk tolerance, or financial information) and system related (e.g., traffic-adjusted appropriateness of various care sites or cost considerations) to provide ranked differential diagnoses or management options. The model might in turn inform action plans such as Internet-supported self-care, a telemedicine visit, an in-person doctor's visit with their primary care physician or a specialist, or the ED. In this example, a model application might use the date (New Year's Eve), the patient's credit card history (recently purchased beer), general health (excellent), prior heart monitor results for the workup of palpitations (only premature atrial contractions), and a wearable-generated heart tracing (benign heart rhythm) to diagnose alcohol-induced palpitations, then offer reassurance and the suggestion to sleep it off.

Machine-learning approaches are moving in this direction, but they may instead worsen value, especially in the short term. For example, wearable devices such as the Apple Watch can detect the common arrhythmia atrial fibrillation (independent of symptoms) and prompt machine-learning–enabled guidance. The Apple Heart Study[18,19] evaluated this concept in adults 22 years of age and older under the premise that 700,000 Americans with atrial fibrillation remain undiagnosed at an incremental cost of $3.2 billion in potentially avoidable arrhythmia-induced strokes. Yet these wearables inevitably capture false-positive results, heightening anxiety and triggering inappropriate downstream utilization or overdiagnosis[20] at a competing downstream cost, especially among young adults with a low pretest probability of arrhythmia.[19,21-23] In kind, although the Apple Heart Study researchers designed the algorithm to mitigate false-positive results by basing conclusions on more than one heart tracing, most participants who were notified of an irregular pulse and underwent further evaluation did not end up receiving a diagnosis of atrial fibrillation. Although application-related adverse events were rare, nearly all of them were due to anxiety.

> *"Just as map applications allow the user to set start and end points as well as journey preferences (e.g., fastest route or avoid tolls), such an approach might allow patients and clinicians to set their priorities in the testing process within the bounds of diagnostic approaches that are available and cost-effective in that system."*

Machine-learning applications must overcome numerous challenges to promote high-value care-seeking at scale, two of which are highlighted here. First, model development depends on detailed, high-fidelity clinical outcomes data (e.g., the downstream impact of seeking care in a certain setting), which may be hard to find. Without sufficient insight, machine-learning algorithms based on prior behavior may perpetuate excess care-seeking of "the worried well," "cyberchondriacs," and those who have multiple, recurrent symptoms without clear cause.[24] It will be important to link

sensors to other relevant information, especially from the EHR (e.g., the "health records" section of the Health application for iPhone extracts EHR data from a variety of vendors), and to obtain cost and other relevant data from potential care sites.

The second challenge is that, in their current state, machine-learning outputs may be unable to provide advice more specific than the fallback "talk to your doctor" or "report to the ED."[25] On average, given liability concerns, we anticipate that direct-to-consumer machine-learning applications likely will generate more conservative advice (i.e., suggest more interventions) than will a clinician who knows the patient. Machine-learning applications also must explain the rationale for recommendations if patients are to trust them and must ensure that important results are shared with the patient's usual source of care.

The Decision To Test

Currently, decisions to choose screening or diagnostic tests are mostly made by clinicians on the basis of evidence that may not be specific to a patient's needs or circumstances, with the attendant consequences of overtesting, undertesting, and mistesting.[26] For example, a person might present to their primary care physician with a new, potentially concerning symptom such as a headache. The gamut of possible tests to rule out the serious causes is extensive, and many of the tests would be considered low value except in specific, relatively rare contexts (e.g., MRI is valuable only if the headache is associated with certain "red flag" symptoms). The evaluation process, including the sequence of testing and referrals, may be guided by what is available (as well as by financial and medicolegal incentives) rather than by what leads to the most efficient, highest-value outcome.

Ideally, clinicians could use machine learning to inform diagnostic flowcharts and then present these options through a point-of-care decision aid or clinical decision support tool. For example, on the basis of inputs such as patient history, clinical examination findings, and test characteristics, the tool could suggest how to arrive at the correct diagnosis in the shortest time or with the fewest total tests or the lowest risk of complications. Just as map applications allow the user to set start and end points as well as journey preferences (e.g., fastest route or avoid tolls), such an approach might allow patients and clinicians to set their priorities in the testing process within the bounds of diagnostic approaches that are available and cost-effective in that system. In addition, with appropriate explanatory capacity, these tools could aid clinicians in weighing short-term benefits against longer-term consequences, such as the cumulative radiation effects of repeated computed tomography scans that rule out unlikely appendicitis but could contribute to cancer years later. Such approaches also could be used to manage patient expectations throughout the process.

The example of reflex testing may provide a template for machine learning–based testing pathways. When a patient with suspected anemia undergoes a blood count panel test, the results might clearly demonstrate the anemia and its cause or might indicate the need for further tests. Although doctors often order all tests at once in the interest of expediency, a rule-based algorithm can prompt the laboratory to "reflex" certain follow-up tests on the basis of the prior result. Taking this a step further, machine learning could help to develop algorithmic approaches to more diverse clinical scenarios in which the rules are less clear.[27] For example, a machine-learning algorithm might help

to determine the optimum follow-up time for incidental findings on the basis of real-world data rather than according to the artificial constraints of the lunar calendar.[28]

Machine-learning applications are starting to improve testing decisions and processes. For example, Mullainathan and Obermeyer[29] applied machine learning to Medicare claims data to assess physician decision-making around testing for heart attack and reported high rates of both underuse and overuse of testing, which are potential targets for correction. Proof-of-concept studies have developed algorithms to detect skin and breast cancers,[15] which theoretically might be deployed in the primary care setting to improve diagnostic accuracy and avoid costly specialist referrals. A machine-learning algorithm to triage screening mammograms is already used routinely at Massachusetts General Hospital. The algorithm dramatically decreases the time it takes to read these tests and also improves specificity, which may lower the cost and burden of downstream testing.[30]

> *"Machine-learning algorithms based on multimodal data could better identify persistently high-cost patients and then learn from their experiences to suggest one strategy over another for a given patient."*

At the same time, machine-learning algorithms could worsen value to the extent that they prioritize certainty (i.e., finding a diagnosis) over other considerations. Specifically, machine learning may worsen the problem of overdiagnosis by making it faster, easier, and cheaper to label a pathology slide as showing cancer, for instance, even when there is no clear gold standard for the diagnosis.[20] By virtue of interpreting previously impenetrable data and making the results accessible to more people, machine learning could further increase downstream tests and treatments of limited utility (as was reported, e.g., when primary care physicians shared whole genome sequencing results with their patients[31]). Finally, lack of trust in an algorithm may erode patients' trust in their clinician.

The biggest hurdles in using machine learning to improve the value of testing are: (1) capturing clinically important outcomes and (2) integrating algorithms into clinical workflows to be used at point of care. For example, though much machine-learning work has focused on improving the interpretation of single tests or images, future efforts should examine outcomes from combinations or series of these tests and images to reflect the complexity of clinical decision-making. Researchers may also need to conduct socioenvironmental analyses of machine-learning models to determine effective implementation pathways.[32]

The Decision To Treat

Clinicians usually dictate treatment plans on the basis of factors such as practice norms and clinical trial results. Machine learning could be used to select and refine individualized treatments and decisions at the point of care. For instance, a patient who has been newly diagnosed with multiple sclerosis might meet with a neurologist to discuss treatment options such as steroids, biologics, or plasma exchange. A machine-learning algorithm not only could consider patient characteristics but also could "solve for" outcomes such as time to cure or stability, patient preferences, side effects, treatment burden, and cost to the patient and society. These algorithms could be repeated at

regular intervals to guard against the clinical inertia that can stymie improvement in patient outcomes.

There has been progress in the use of machine-learning algorithms to inform better treatment pathways. In one example, researchers used unsupervised machine learning based on EHR data to identify four unique phenotypes of patients with sepsis who might be best treated with four distinct approaches.[8] However, current models do not incorporate key factors such as patient out-of-pocket costs, toxicity, and quality of life.

Machine learning may also contribute to lower-value treatment decisions. First, if more clinicians have access to these algorithms, they may start more patients on a new treatment such as biologics earlier than the 10 years it usually takes for such treatments to be widely adopted, introducing new problems if cost is excluded from these valuations. Second, an algorithm that trains on data affected by sex and race biases (e.g., undertreatment of women after heart attacks or undertreatment of Black patients for pain) would perpetuate these biases, leading to worse outcomes in a manner that is more difficult to scrutinize than in analog patient–clinician interactions.[33] Fortunately, researchers are working to mitigate these biases by reducing reliance on automation, conducting follow-up studies, and building data sets from diverse populations.[34] Finally, machine-learning algorithms may be less accessible to historically underserved patients, worsening health outcome disparities to the extent that these algorithms are beneficial.

The Decision To Promote Value Across a Health System

Although most health care occurs in the myriad interactions between patients and clinicians, leaders of health care systems play a large role in shaping operations and policies to promote high-value care. Systems that take on value-based payment contracts (e.g., accountable care organizations [ACOs]) are especially incentivized to improve quality and lower cost for defined populations. Despite this incentive, they use relatively crude approaches to measure value across these populations or to identify the most effective strategies to improve value.[35] Ideally, ACOs might use machine learning to segment patients (on the basis of clinical, social, and other factors) into groups with similar needs to inform targeted interventions.[25] ACOs might also use machine learning to make real-time changes in care delivery — for example, when the price of a common medication goes up or as new patients are attributed to the ACO.

> *"In the Apple Heart study, the work of algorithms was closely intertwined with that of clinicians: notifications of irregular pulse prompted a telemedicine visit with a doctor, and if they observed urgent symptoms, they would ask the patient to go to an urgent care clinic or ED."*

In another example of a role for machine learning, ACOs and other health systems have adopted care-management programs to coordinate care for their highest-cost patients, although such

programs are themselves costly and have shown uneven benefits.[36] Traditional big-data analytical approaches to select patients for these programs have used claims, demographics, and prior care to predict future costs and use patterns. Such approaches overrepresent patients who are only transiently expensive and miss other factors (e.g., engagement level or personality traits) that may inform whether patients would benefit from a care-management program.[37] Machine-learning algorithms based on multimodal data could better identify persistently high-cost patients and then learn from their experiences to suggest one strategy over another for a given patient.[38-40]

Machine learning might also be used to better capture and compensate physician work in ways that promote high-value decisions[41] or, conversely, to identify clinicians who are more likely to make high-value decisions, given the limits of traditional prediction models for this purpose.[42] Finally, health systems can use machine learning to prioritize costly yet worthwhile programs and improvement efforts.[43]

Once again, machine-learning applications may worsen value: because provider organizations are incentivized to maximize their revenue, a model built primarily on billed claims may inflate or distort costs due to upcoding. As has been described above in the discussion on treatment decisions, population health program allocation is also rife with race and sex biases that algorithms may perpetuate.[33] Finally, health systems will see, and machine-learning algorithms will act on, only data for patients who have interacted with the health system. To address these potential biases, system leaders must be deliberate about wider community outreach and more expansive data collection.

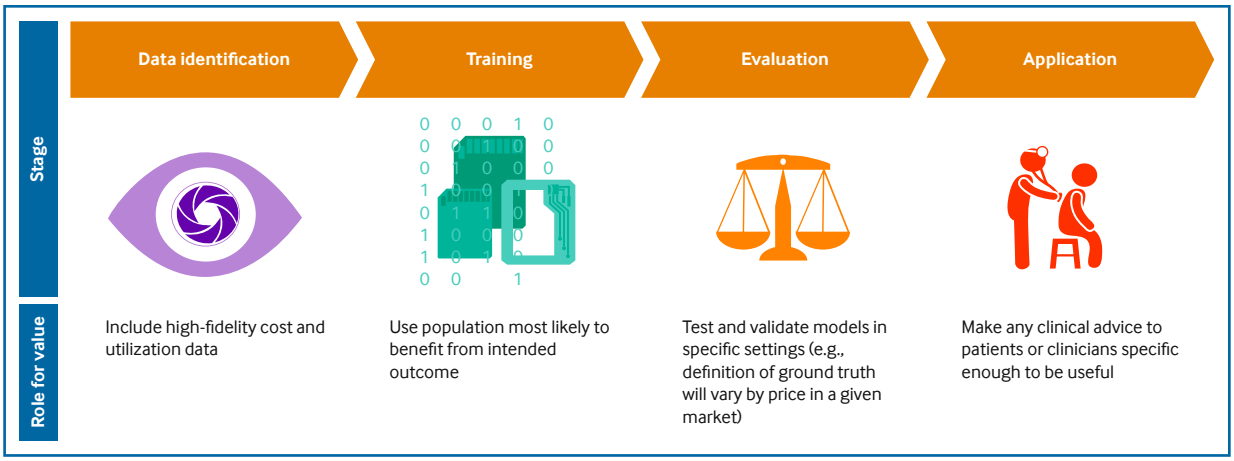
Takeaways

Because machine-learning implementation may worsen value as well as improve it, we propose two key takeaways for policymakers and practitioners of machine learning. First, we emphasize the importance of considering value writ large[1,4] — especially using readily available cost data — when developing and applying any machine-learning model (Figure 2). Researchers who then study the real-world impact of these algorithms should examine outcomes such as overdiagnoses and medical expenditures. Second, machine-learning applications should take advantage of what machines do best and what humans do best in the pursuit of high-value care (Figure 3). The recent National Academy of Medicine report cautions that in the short term, it is important to think of machine learning as human enabling rather than human replacing.[44] In the Apple Heart study, the work of algorithms was closely intertwined with that of clinicians: notifications of irregular pulse prompted a telemedicine visit with a doctor, and if they observed urgent symptoms, they would ask the patient to go to an urgent care clinic or ED. Similarly, in the Massachusetts General Hospital mammography example, machine-learning serves to enhance physician performance and efficiency in reviewing high volumes of images. To achieve the goal of high-value care, this principle means using machine learning to compensate for our human flaws, such as inattention, fatigue, and our imperfect knowledge base, while also letting clinicians exercise their human strengths, such as clinical intuition and the ability to build trust.

FIGURE 2

From Prediction to Patient: How Value Should Inform Machine Learning

Diagram illustrating how value should inform machine learning across the stages of creating and applying an algorithm.

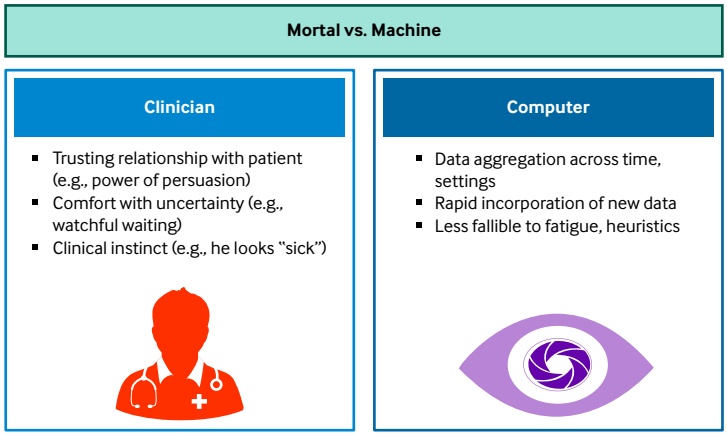


Source: The authors.

FIGURE 3

Mortal vs. Machine: Comparative Advantage in Promoting Value

Diagram illustrating the comparative advantages of humans and machine-learning algorithms in promoting value.



Source: The authors.

Ishani Ganguli, MD, MPH
Assistant Professor of Medicine, Harvard Medical School, Boston, Massachusetts, USA

Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, Massachusetts, USA

William J. Gordon, MD, MBI
Instructor in Medicine, Harvard Medical School, Boston, Massachusetts, USA

Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, Massachusetts, USA

Medical Director, Health Innovation Platform, Mass General Brigham, Boston, Massachusetts, USA

Claire Lupo
Research Assistant, Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, Massachusetts, USA

Megan Sands-Lincoln, PhD, MPH
Center for AI Research and Evaluation, IBM Watson Health, Cambridge, Massachusetts, USA

Judy George, PhD
Research Health Data Scientist, IBM Watson Health, Cambridge, Massachusetts, USA

Gretchen Jackson, MD, PhD
Vice President and Chief Science Officer, IBM Watson Health, Cambridge, Massachusetts, USA

Associate Professor of Surgery, Pediatrics, and Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee, USA

Kyu Rhee, MD, MPP
Vice President and Chief Health Officer, IBM Corporation and IBM Watson Health, Cambridge, Massachusetts, USA

David W. Bates, MD, MS
Professor of Medicine, Harvard Medical School, Boston, Massachusetts, USA

Division Chief, Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, Massachusetts, USA

Disclosures: Ishani Ganguli discloses consulting fees from Haven Healthcare and Blue Cross Blue Shield of Massachusetts. William J. Gordon, Claire Lupo, Megan Sands-Lincoln, Judy George, Gretchen Jackson, and Kyu Rhee have nothing to disclose. David W. Bates discloses consulting fees from EarlySense and CDI-Negev, Ltd, as well as equity from Valera Health, CLEW Medical, MDClone, and AESOP Technology.

References

1. Shrank WH, Rogstad TL, Parekh N. Waste in the US health care system: Estimated costs and potential for savings. JAMA 2019;322:1501-9 https://doi.org/10.1001/jama.2019.13978.

2. Morden NE, Colla CH, Sequist TD, Rosenthal MB. Choosing wisely—the politics and economics of labeling low-value services. N Engl J Med 2014;370:589-92 https://doi.org/10.1056/NEJMp1314965.

3. Pandya A. Adding cost-effectiveness to define low value care. JAMA 2018;319:1977-8 https://doi.org/10.1001/jama.2018.2856.

4. Institute of Medicine. Best care at lower cost: the path to continuously learning health care in America. Washington: The National Academies Press, 2013. Accessed August 17, 2020. https://www.nap.edu/catalog/13444/best-care-at-lower-cost-the-path-to-continuously-learning https://doi.org/10.17226/13444.

5. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019;380:1347-58 https://doi.org/10.1056/NEJMra1814259.

6. Beam AL, Kohane IS. Big data and machine learning in health care. JAMA 2018;319:1317-8 https://doi.org/10.1001/jama.2017.18391.

7. Deo RC. Machine learning in medicine. Circulation 2015;132:1920-30 https://doi.org/10.1161/CIRCULATIONAHA.115.001593.

8. Seymour CW, Kennedy JN, Wang S, et al. Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. JAMA 2019;321:2003-17 https://doi.org/10.1001/jama.2019.5791.

9. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. Health Aff (Millwood) 2008;27:759-69 https://doi.org/10.1377/hlthaff.27.3.759.

10. Frakt AB. Determining value and price in health care. JAMA 2016;316:1033-4 https://doi.org/10.1001/jama.2016.10922.

11. Ohno-Machado L. Realizing the full potential of electronic health records: The role of natural language processing. J Am Med Inform Assoc 2011;18:539 https://doi.org/10.1136/amiajnl-2011-000501.

12. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? J Biomed Inform 2009;42:760-72 https://doi.org/10.1016/j.jbi.2009.08.007.

13. Chapman WW, Nadkarni PM, Hirschman L, D'Avolio LW, Savova GK, Uzuner O. Overcoming barriers to NLP for clinical text: The role of shared tasks and the need for additional creative solutions. J Am Med Inform Assoc 2011;18:540-3 https://doi.org/10.1136/amiajnl-2011-000465.

14. Afzal N, Sohn S, Abram S, et al. Mining peripheral arterial disease cases from narrative clinical notes using natural language processing. J Vasc Surg 2017;65:1753-61 https://doi.org/10.1016/j.jvs.2016.11.031.

15. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25:44-56 https://doi.org/10.1038/s41591-018-0300-7.

16. Morse KE, Bagley SC, Shah NH. Estimate the hidden deployment cost of predictive models to improve patient care [published correction appears in Nat Med 2020;26:803]. Nat Med 2020;26:18-9 https://doi.org/10.1038/s41591-019-0651-8.

17. Heaven WD. Google's medical AI was super accurate in a lab. Real life was a different story. MIT Technology Review. April 27, 2020. Accessed July 24, 2020. https://www.technologyreview.com/2020/04/27/1000658/google-medical-ai-accurate-lab-real-life-clinic-covid-diabetes-retina-disease/.

18. Turakhia MP, Desai M, Hedlin H, et al. Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: The Apple Heart Study. Am Heart J 2019;207:66-75 https://doi.org/10.1016/j.ahj.2018.09.002.

19. Perez MV, Mahaffey KW, Hedlin H, et al. Apple Heart Study Investigators. Large-scale assessment of a Smartwatch to identify atrial fibrillation. N Engl J Med 2019;381:1909-17 https://doi.org/10.1056/NEJMoa1901183.

20. Adamson AS, Welch HG. Machine learning and the cancer-diagnosis problem—no gold standard. N Engl J Med 2019;381:2285-7 https://doi.org/10.1056/NEJMp1907407.

21. Lowres N, Neubeck L, Salkeld G, et al. Feasibility and cost-effectiveness of stroke prevention through community screening for atrial fibrillation using iPhone ECG in pharmacies. The SEARCH-AF study. Thromb Haemost 2014;111:1167-76 https://doi.org/10.1160/TH14-03-0231.

22. Ganguli I, Lupo C, Mainor AJ, et al. Prevalence and cost of care cascades after low-value preoperative electrocardiogram for cataract surgery in fee-for-service Medicare beneficiaries. JAMA Intern Med 2019;179:1211-9 https://doi.org/10.1001/jamainternmed.2019.1739.

23. Kale MS, Korenstein D. Overdiagnosis in primary care: framing the problem and finding solutions. BMJ 2018;362:k2820 https://doi.org/10.1136/bmj.k2820.

24. Barsky AJ, Orav EJ, Bates DW. Distinctive patterns of medical care utilization in patients who somatize. Med Care 2006;44:803-11 https://doi.org/10.1097/01.mlr.0000228028.07069.59.

25. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. Health Aff (Millwood) 2014;33:1123-31 https://doi.org/10.1377/hlthaff.2014.0041.

26. O'Sullivan JW, Albasri A, Nicholson BD, et al. Overtesting and undertesting in primary care: A systematic review and meta-analysis. BMJ Open 2018;8:e018557 https://doi.org/10.1136/bmjopen-2017-018557.

27. Hoffmann G, Bietenbeck A, Lichtinghagen R, Klawonn F. Using machine learning techniques to generate laboratory diagnostic pathways—a case study. J Lab Precis Med 2018;3:58 http://jlpm.amegroups.com/article/view/4401 https://doi.org/10.21037/jlpm.2018.06.01.

28. Ganguli I, Simpkin AL, Lupo C, et al. Cascades of care after incidental findings in a US national survey of physicians. JAMA Netw Open 2019;2:e1913325 https://doi.org/10.1001/jamanetworkopen.2019.13325.

29. Mullainathan S, Obermeyer Z. A machine learning approach to low-value health care: wasted tests, missed heart attacks and mis-predictions. Updated August 2020. Accessed August 16, 2020. https://www.nber.org/papers/w26168?utm_campaign=ntwh&utm_medium=email&utm_source=ntwg5.

30. Yala A, Schuster T, Miles R, Barzilay R, Lehman C. A deep learning model to triage screening mammograms: A simulation study. Radiology 2019;293:38-46 https://doi.org/10.1148/radiol.2019182908.

31. Vassy JL, Christensen KD, Schonman EF, et al.MedSeq Project. The impact of whole-genome sequencing on the primary care and outcomes of healthy adult patients: A pilot randomized trial. Ann Intern Med 2017;167:159-69 https://doi.org/10.7326/M17-0188.

32. Beede E, Baylor E, Hersch F, et al. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. New York: Association for Computing Machinery, 2020:1-12 https://doi.org/10.1145/3313831.3376718.

33. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366:447-53 https://doi.org/10.1126/science.aax2342.

34. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. JAMA Intern Med 2018;178:1544-7 https://doi.org/10.1001/jamainternmed.2018.3763.

35. Ganguli I, Ferris TG. Accountable care at the frontlines of a health system: Bridging aspiration and reality. JAMA 2018;319:655-6 https://doi.org/10.1001/jama.2017.18995.

36. Baker JM, Grant RW, Gopalan A. A systematic review of care management interventions targeting multimorbidity and high care utilization. BMC Health Serv Res 2018;18:65 https://doi.org/10.1186/s12913-018-2881-8.

37. Ganguli I, Orav EJ, Weil E, Ferris TG, Vogeli C. What do high-risk patients value? Perspectives on a care management program. J Gen Intern Med 2018;33:26-33 https://doi.org/10.1007/s11606-017-4200-1.

38. Joynt KE, Figueroa JF, Beaulieu N, Wild RC, Orav EJ, Jha AK. Segmenting high-cost Medicare patients into potentially actionable cohorts. Healthc (Amst) 2017;5:62-7 https://doi.org/10.1016/j.hjdsi.2016.11.002.

39. Ganguli I, Thompson RW, Ferris TG. What can five high cost patients teach us about healthcare spending? Healthc (Amst) 2017;5:204-13 https://doi.org/10.1016/j.hjdsi.2016.12.004.

40. Colbert J, Ganguli I. To identify patients for care management interventions, look beyond big data. Health Affairs Blog. April 19, 2016. Accessed July 20, 2020. https://www.healthaffairs.org/do/10.1377/hblog20160419.054528/full/.

41. Rajkomar A, Yim JW, Grumbach K, Parekh A. Weighting primary care patient panel size: A novel electronic health record-derived measure using machine learning. JMIR Med Inform 2016;4:e29 https://doi.org/10.2196/medinform.6530.

42. Schwartz AL, Jena AB, Zaslavsky AM, McWilliams JM. Analysis of physician variation in provision of low-value services. JAMA Intern Med 2019;179:16-25 https://doi.org/10.1001/jamainternmed.2018.5086.

43. Horwitz LI, Kuznetsova M, Jones SA. Creating a learning health system through rapid-cycle, randomized testing. N Engl J Med 2019;381:1175-9 https://doi.org/10.1056/NEJMsb1900856.

44. Matheny M, Israni ST, Auerbach A, et al. Artificial intelligence in health care: the hope, the hype, the promise, the peril. National Academy of Medicine. 2019. Accessed December 20, 2019. https://nam.edu/artificial-intelligence-special-publication/.