NEJM Catalyst | Innovations in Care Delivery

ARTICLE

Interpretable Machine Learning Models for Clinical Decision-Making in a High-Need, Value-Based Primary Care Setting

Surabhi Bhatt, MS, Adam Cohon, PhD, Jenna Rose, MHS, Natalia Majerczyk, Brian Cozzi, MS, Drew Crenshaw, Griffin Myers, MD, MBA

Vol. 2 No. 4 | April 2021
DOI: 10.1056/CAT.21.0008

While health care organizations are increasingly interested in using artificial intelligence (AI), there is a significant lack of literature on the real-world application of a risk stratification AI tool in primary care. Oak Street Health, a network of more than 80 primary care centers in medically underserved communities, successfully implemented a machine learning–based risk stratification tool across their organization that outperformed prior backward-looking approaches in identifying high-risk patients. The data science team collaborated with an interdisciplinary set of stakeholders to test, iterate, and implement the tool into clinical practice. Early feedback from Oak Street Health's primary care providers (physicians and nurse practitioners) and nonproviders (social workers) suggests that the display of top risk factors based on model predictions created a broadly interpretable and actionable risk stratification tool in caring for the highest-risk patients.

Recent advances in digitized health data have allowed state-of-the-art machine learning methods to be more easily applied to complex clinical data and outcomes such as in-hospital mortality, unplanned readmission, and extended hospital stays.[1] There are, however, very few examples in the literature of rigorous validation of models by clinical experts before application in a care setting. Still, a 2020 study demonstrated the utility of a mortality risk prediction model that was clinically validated before model integration into a hospital setting.[2]

Many health care organizations also struggle to implement artificial intelligence (AI) in a care delivery setting, with various challenges facing interdisciplinary teams when integrating machine

NEJM Catalyst is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from catalyst.nejm.org by Scott Drummond on April 19, 2021. For personal use only.
No other uses without permission. Copyright © 2021 Massachusetts Medical Society. All rights reserved.

learning tools into clinical workflows.[3] A common criticism of machine learning in health care is that it is not easily interpretable.[4-6] Additionally, most of the publications on real-world clinical implementation focus on outcomes related to specialty care such as cardiology,[7] neurology,[8] and image analysis.[9] There is little evidence demonstrating the utility of predictive models in supporting care delivery and clinical decision-making or in explaining how multidisciplinary care teams interpret and use model predictions.[10,11]

In primary care, much of the AI literature is focused on three areas: model design and implementation framework proposals,[12] highlighting key considerations for integration,[13,14] and emphasizing the importance of using AI in primary care.[15] The existing health care AI literature currently lacks information on the development and real-world implementation of primary care–focused AI models. Compared with acute and specialized settings, little is known about user engagement, acceptance, and interpretation of AI in primary care.[16,17]

Methods

Oak Street Health is a value-based, primary care organization for adults on Medicare that operates in medically underserved neighborhoods in more than 80 locations across 11 states: Illinois, Indiana, Michigan, Pennsylvania, Ohio, Rhode Island, Mississippi, North Carolina, New York, Texas, and Tennessee. Of the approximately 90,000 patients we serve, 86% have multiple chronic conditions, 40% have a behavioral health diagnosis, and approximately 50% struggle with one or more social risk factors. Approximately 40% of our patients are dually eligible for both Medicare and Medicaid, and approximately 66% of all patients are under a capitation agreement.

Our care model uses a multidisciplinary, evidence-based approach with a significant emphasis on strong patient–provider relationships. These relationships allow us to develop personalized care management and navigation plans to address the comprehensive needs of our patients. Oak Street Health's data infrastructure is instrumental in supporting these efforts across four domains: data intake, analytics and data science, purpose-built workflow applications, and the return of point-of-care data back into the data warehouse.

In mid-2019, the executive medical team outlined a strategic vision for risk stratification to better leverage our growing data assets, with a central goal of improving both patient care delivery and outcomes. Our selected approach uses data and clinical expertise to identify the medical, behavioral, and social needs of each patient and population. As a result, our machine learning–based tools aim to be both accurate and actionable so that they can inform evidence-based interventions, optimize the allocation of resources, and prioritize operational changes. As an example of our comprehensive approach, our tools may flag a patient as high risk of admission because of factors including their recent Patient Health Questionnaire-9 score, prompting the team to enroll the patient in the Oak Street Behavioral Health program. Similarly, a patient may be flagged for medical cost risk because of social determinants of health, initiating an assessment by a medical social worker (MSW).

NEJM Catalyst is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from catalyst.nejm.org by Scott Drummond on April 19, 2021. For personal use only.
No other uses without permission. Copyright © 2021 Massachusetts Medical Society. All rights reserved.

Table 1. Comparing the Old Logic, Using Retrospective Techniques, with the New Logic, Leveraging Predictive Modeling

Old logic: retrospective rules	New logic: predictive models
Three data points on each patient HCC RAF score Third-party medical costs in last 12 months Admissions in last 60 days	400+ data points on each patient, including: Claims Census EMR data
Binary All patients meeting absolute threshold are identified as high risk	Probabilistic All patients above percentile threshold are identified as high risk

HCC = Hierarchical Condition Category, RAF = Risk Adjustment Factor, EMR = electronic medical record. Source: The authors.

> *"In mid-2019, the executive medical team outlined a strategic vision for risk stratification to better leverage our growing data assets, with a central goal of improving both patient care delivery and outcomes. Our selected approach uses data and clinical expertise to identify the medical, behavioral, and social needs of each patient and population."*

Prior to 2019, risk stratification at Oak Street Health took into consideration only historical patient data and provider nominations. Patients were identified as high risk on the basis of reaching thresholds in previous medical costs, admissions, and diagnoses included in the Centers for Medicare & Medicaid Services' (CMS) Hierarchical Condition Category Risk Adjustment Factor score. Providers could also add patients on the basis of their clinical assessment of the patient's risk. Patients stayed on the registry until they no longer met the criteria or until they were removed by the care team. Although the risk stratification methodology did identify patients at higher risk of admission, medical costs, and mortality, there were limitations. Because only a few data points were leveraged, no distinction was made between various types of medical costs or admissions. All such utilization contributed equally to the patient's risk level. Perhaps more important, the algorithm provided little actionable information to the care teams. Other than flagging life-limiting illnesses, the risk stratification tool generated little clinically relevant information.

The previous models' shortcomings led to the formation of a provider advisory group with the goals of developing a more accurate and more actionable high-risk patient registry (Table 1).

The development and implementation of the machine learning–based risk stratification tool occurred in five primary stages: preimplementation model development; pilot; postpilot evaluation and improvement; deployment; and postdeployment evaluation and monitoring (Table 2).

Preimplementation Model Development

Given Oak Street Health's role as an at-risk provider, the provider advisory group decided to focus on patients' risk of an inpatient admission, projected 6-month third-party (i.e., nonprimary care) costs, and risk of mortality in the next 12 months as key outcomes in determining their overall risk. The data science team, building on guidance and factor engineering work previously provided by ClosedLoop.ai, built machine learning models to predict these three outcomes. In order to capture

NEJM Catalyst is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from catalyst.nejm.org by Scott Drummond on April 19, 2021. For personal use only.
No other uses without permission. Copyright © 2021 Massachusetts Medical Society. All rights reserved.

Table 2. Key Activities by Implementation Phase

Implementation phase	Key activities
Preimplementation model development, May–November 2019	Executive Medical Director vision-setting session
	Formation of provider advisory group
	Confirmation of model targets
	Development of models, including data quality adjustments, selection methodologies, and use of SHAP libraries
	Evaluation of model accuracy on the basis of quantitative assessment and clinical validation
Pilot, December 2019–January 2020	Training 14 care teams on tool and pilot expectations
	Weekly use of tool during high-risk patient review meetings
Postpilot evaluation and tool refinement, January–July 2020	Complete postpilot survey and debrief sessions
	Collection and analysis of pilot success metrics
	Endorsements from pilot participants and provider advisory group
	Tool improvements made in collaboration with application development team
Organization-wide deployment, August 2020	Delayed because of Covid-19 response and transition to virtual care model
	Approval from Project Management Office
	Introduction at center huddles
	Virtual trainings completed via prerecorded video and live Q&A
	Ongoing assistance and shadowing by data science team
Postdeployment evaluation and monitoring, Ongoing	Routine monitoring of model performance
	Response to end-user feedback (e.g., prediction fluctuations)
	End-user survey 2 months after deployment

SHAP = SHapley Additive exPlanations (an approach to explain the output of a machine learning model), Q&A = question and answer. Source: The authors.

the risk of our unique patient population, we used Oak Street Health's internal data to both build the models and test their accuracy on the outcomes of interest. Responding to concerns of racial bias related to algorithms that predict on the basis of health care costs without considering illness or access to care,[18] the medical cost predictor is always used in combination with other outcomes.

Once the outcomes were determined, our data science team trained two machine learning models to support risk stratification: a classification model to predict inpatient admissions and a regression model to predict 6-month medical costs. Both models were developed using the native XGBoost API.[19] Data for both these machine learning models came from a unified data warehouse that integrates data on all Oak Street patients from our common electronic medical record, from claims data provided by CMS's Blue Button and payer partners, and from several additional sources. Input factors ranged broadly from prescription fills to visit nonattendance to the number of skilled nursing facility visits. Outcomes were captured for the years 2017 to 2019, using claims data, and input measures were taken up to 12 months prior to the outcome. Testing data to evaluate model performance were collected both from a holdout set of the same data and from outcomes in quarter (Q) 4 2019 and Q1 2020. Model versions were scored on the area under the precision-recall curve and on their ability to correctly forecast admissions and costs among the patients with the highest predicted values. (Because admissions are rare events and costs are highly skewed in health care, there are well-known problems with using the standard receiver operating characteristic curves.)

Several considerations arose in model development. First, we had to account for data quality. Data from Oak Street Health's early years differed in structure, so we selected a 2017 start date for

NEJM Catalyst is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from catalyst.nejm.org by Scott Drummond on April 19, 2021. For personal use only.
No other uses without permission. Copyright © 2021 Massachusetts Medical Society. All rights reserved.

outcome measurement in training the models. In addition, individual-level claims data were unavailable from some of our payer partners, so historical predictions could not be tested for those patients. Second, inpatient admissions are statistical rare events, and costs are known to be highly skewed across patients and member months. For these reasons, we had to carefully sample the data and the treatment dates at which outcomes were measured.

> *"Responding to concerns of racial bias related to algorithms that predict on the basis of health care costs without considering illness or access to care, the medical cost predictor is always used in combination with other outcomes."*

To improve the utility of the predictions for providers, we generated reasons for each prediction. Using the SHapley Additive exPlanations (SHAP) library, we identified important factors.[20] Top factors are those that made the patient more likely to be high risk; bottom factors are those that are most influential in making a patient low risk. On the basis of provider advisory group feedback, we added to the display of top factors the specific events or diagnoses (i.e., feature values) that qualified the patient for that factor.

The provider advisory group supervised the development of the stratification models for pilot testing. They paid particular attention to model display, the number of patients selected as highest risk, clinical relevancy of the factors in determining risk, and model accuracy. To clinically validate model accuracy, each advisory group member was provided a list of predictions for high-risk patients under their care, which they reviewed to substantiate the machine learning predictions and top factors. Finally, our team worked closely with end users and providers to right-size the user interface to ensure that the outputs were easily interpretable and actionable.

Pilot

The completed tool was previewed and piloted for 2 months with 14 multidisciplinary teams (providers, scribes, registered nurses, MSWs, behavioral health specialists, and medical assistants) across several Oak Street Health locations. All pilot teams were trained on the tool and asked to use the tool during their weekly high-risk patient review meetings. A robust set of metrics was developed to evaluate the success of the pilot, including whether the tool was usable, technically feasible, and more accurate at identifying high-risk patients compared with the previous approach and its impact on patient care.

Postpilot Evaluation and Tool Refinement

All pilot participants completed a postpilot survey and a facilitated group debrief session to share qualitative feedback on the perceived accuracy and actionability of the tool as well as recommendations for improvements. On the basis of those postpilot survey results, 90% of the 14 teams rated the tool as easy to use or very easy to use, and 80% of teams said they were likely to

NEJM Catalyst is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from catalyst.nejm.org by Scott Drummond on April 19, 2021. For personal use only.
No other uses without permission. Copyright © 2021 Massachusetts Medical Society. All rights reserved.

recommend the tool to other teams. The data science team used input from this survey to improve the risk stratification tool prior to deployment. For example, end users suggested improving the interpretability of the model outputs by providing more easily accessible factor definitions and benchmarks.

Organization-Wide Deployment

While the deployment of the tool was put on hold for 2 months because of the Covid-19 pandemic, the data science team aligned with senior organizational leadership and the project management office to set a new go-live date of August 3, 2020. The additional time was used to further test and refine the tool and, in particular, the factor display. The data science team found continued high accuracy for mid-2020 outcomes even though the models were not trained on data from the pandemic period, further validating the decision to deploy more broadly.

Once the tool received approval from executive leadership teams, the data science team partnered with regional leadership to develop a comprehensive go-live communication and training plan using a staggered approach. Important consideration was given to minimizing disruption to workflows. The implementation rollout began at center huddles with regional leadership introducing the tool and explaining its impact on patient care, followed by a structured training and question-and-answer (Q&A) session at all Oak Street Health centers. Site-based champions were identified at each center to support regional needs and act as local experts on the tool. Additionally, a detailed playbook, tool guide, and frequently asked questions were created to ensure the tool was integrated effectively into weekly care team reviews of high-risk patients. All team members involved in these reviews have the same access to the tool and its predictions. Beyond this, the data science team was available for on-call troubleshooting for center teams.

Postdeployment Evaluation and Monitoring

The tool has been continuously monitored to evaluate its efficacy through postimplementation user adoption and model accuracy. Updated metric reporting was made widely available once the tool was deployed. The data science team also offered additional support to care teams via shadowing of weekly high-risk patient review meetings. Model predictions were automated to be generated daily and weekly, and the models will be retrained on a semiannual basis.

> *“A robust set of metrics was developed to evaluate the success of the pilot, including whether the tool was usable, technically feasible, and more accurate at identifying high-risk patients compared with the previous approach and its impact on patient care.”*

Additionally, 316 providers and 87 MSWs were surveyed 2 months after deployment (October 2 to October 14, 2020) via an anonymous online form. The data science team sent one initial invitation to complete the survey on October 2, together with two follow-up reminders, resulting in 127

NEJM Catalyst is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from catalyst.nejm.org by Scott Drummond on April 19, 2021. For personal use only.
No other uses without permission. Copyright © 2021 Massachusetts Medical Society. All rights reserved.

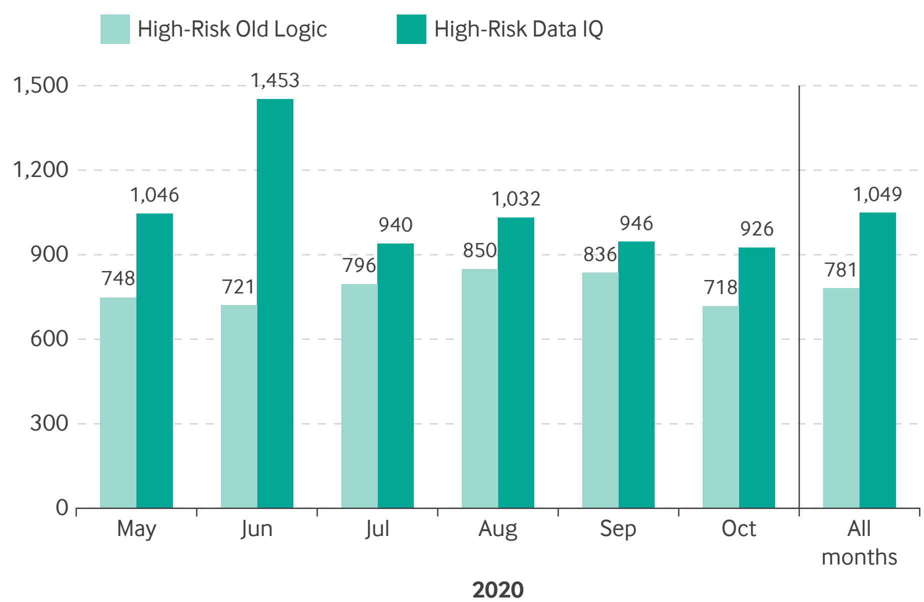
responses, for a total response rate of 31%. The survey asked care team members to identify by their role alone. All other responses were anonymous, and anonymity was preserved. The survey had 11 questions that assessed the following key domains:

- Accuracy of the predictive models in identifying high-risk patients
- Efficacy of the risk factors in identifying patient-specific risk of an adverse outcome
- Actionability of the risk factors, tooltips, and benchmarks
- Ability of risk factors to inform next steps
- Impact of the tool overall on reducing administrative burden of chart review
- Impact of the tool overall on improving patient care

FIGURE 1

High-Risk Patients Annualized Admissions per Thousand

On the basis of observed admission rates and real-world data, the new machine learning model, known as Data IQ, predicts 34% more admissions than the previous rules-based logic. Note that the old logic was deployed until August 2020 and then subsequently retired. The results can be in excess of 1,000 because patients can have more than one admission in a year.



Source: The authors.

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

NEJM Catalyst is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from catalyst.nejm.org by Scott Drummond on April 19, 2021. For personal use only.
No other uses without permission. Copyright © 2021 Massachusetts Medical Society. All rights reserved.

Results

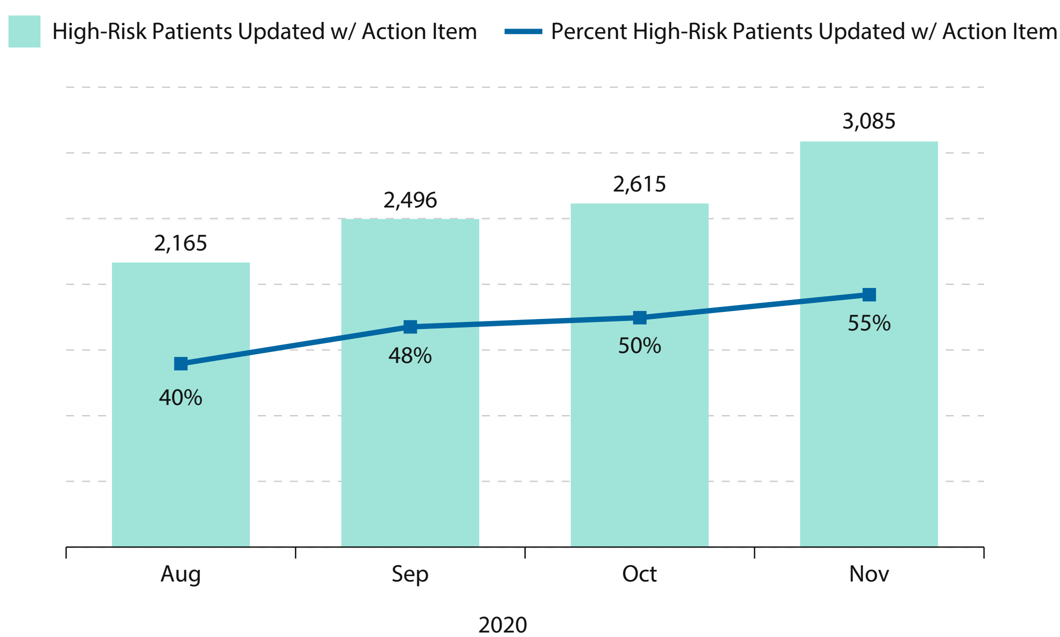
Since the machine learning tool was developed, Oak Street Health has observed a notable increase in admission rates among patients identified as high risk. To measure the improvement, we considered each patient's high-risk status at the start of the month, as assessed by the old methodology and as assessed by the new machine learning approach. We then observed the following month's admission rates for both groups using census admission data. Across the months studied, patients identified by the new approach had a 34% higher admission rate than patients identified by the old logic (Table 1, Figure 1). Please note that some patients were identified by both methodologies.

Together with the increase in accuracy, Oak Street Health has observed a steady increase in care team engagement with high-risk patient registries and tools. At the beginning of the deployment period, 40% of high-risk patients had an action item documented by the care team in the previous 30 days. Action items may include a referral to the internal behavioral health program, patient education, or care coordination with a specialist, among other activities. As of November 2020, this engagement metric had increased to 55% of high-risk patients with a recent action item documented (Figure 2).

FIGURE 2

Number of High-Risk Patients with Recent Action Item, Percent of Total High-Risk Panel

Care team engagement with high-risk patients increased in the months after our new risk stratification tool was deployed, in terms of both the number of high-risk patients assigned an action item by their care team and the percent of high-risk patients with a recent action item.



Source: The authors.

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

NEJM Catalyst is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from catalyst.nejm.org by Scott Drummond on April 19, 2021. For personal use only.
No other uses without permission. Copyright © 2021 Massachusetts Medical Society. All rights reserved.

Qualitative feedback on the risk stratification tool was collected from Oak Street Health providers and MSWs 2 months after deployment. The 127 responses included 79 providers (doctor of medicine [MD], doctor of osteopathic medicine [DO], physician assistants [PAs], and advanced practice nurses [APNs]) and 46 MSWs, as well as two “other” responses. Survey response rates for the two cohorts were 25% and 53%, respectively (Table 3).

The survey assessed clinician judgment of the tool’s accuracy. One-half of all respondents agreed or strongly agreed that the models identify patients who are at high risk of admission, mortality, or high medical cost. The majority of respondents (55%) also indicated that the models identify patients they would not have otherwise considered high risk. Regarding interpretability, 59% of respondents reported that the risk factors displayed are relevant in assessing risk of an adverse outcome. While we would like to see these numbers improve over time, we are encouraged that one-half of our providers perceived a benefit from the tool within 2 months of deployment, particularly given the documented challenges of deploying machine learning tools in health care.

The clinicians’ perceptions of actionability were also measured. Forty-nine percent of respondents agreed or strongly agreed that the tool is helpful with care planning and, in particular, highlighted the utility of the display of the top five risk factors identified by SHAP (61.4%). Interestingly, 38% of respondents agreed or strongly agreed that the tool saves time, while slightly fewer disagreed or strongly disagreed (33%). With additional investments in future iterations, we believe the majority of teams will find the tool to be a time saver.

Discussion

We first want to acknowledge that this work did not come without its challenges. The tool was deployed in the midst of the Covid-19 pandemic, while Oak Street Health was transitioning to a mostly virtual care model. Given the burden on care teams during this time, the data science team focused on minimizing undue strain from the tool rollout. We reformulated the high-need patient panel size to avoid dramatic increases when the new tool went live. The pandemic also required our team to conduct all predeployment care team trainings virtually using prerecorded videos, as well as to partner with regional leadership for the live Q&A. Our team provided virtual at-elbow support after deployment to ensure that care team members received all required technical and logistical support.

After deployment, care teams identified few concerns. Most important among them, care teams noticed week-over-week fluctuations in the high-need patient list after the tool was deployed, hindering care planning efforts. In response, the data science team identified a process for smoothing predictions. Last, lags in data receipt and ingestion continue to be an ongoing focus. For example, data on medical claims for the postdeployment period were not available in mid-February 2021, preventing us from measuring the machine learning–based tool’s accuracy in predicting medical costs.

NEJM Catalyst is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from catalyst.nejm.org by Scott Drummond on April 19, 2021. For personal use only.
No other uses without permission. Copyright © 2021 Massachusetts Medical Society. All rights reserved.

Table 3. Postdeployment Provider Survey of Data IQ Tool

Questions	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Data IQ is accurate at predicting who is at high risk of adverse outcomes. For example: being admitted in the next month, passing away in the next 12 months, being high med cost in the next 6 months.	3%	10%	37%	46%	4%
Data IQ helps me identify patients as high risk that I would have otherwise not considered.	3%	13%	29%	49%	6%
The risk factors identified by Data IQ are relevant to assessing a patient's risk of adverse outcomes.	1%	11%	29%	50%	9%
The risk factors, associated tool tips, and benchmarks are easily interpretable.	6%	19%	36%	35%	5%
The displayed risk factors are helpful in determining appropriate action items (i.e., next steps) in care planning for VIPs.	3%	18%	30%	43%	6%
The Data IQ tool reduces the amount of time we spend gathering information/reviewing the chart for VIP care planning.	6%	27%	29%	31%	7%
	Top five risk factors	**Tool tips**	**Benchmarks**	**Additional risk factors**	**None of the above**
Which of the following have been useful in care planning? (check all that apply)	61%	42%	19%	25%	23%
	Yes	**Unsure**	**No**		
Do you do anything differently in caring for your patients because of the information displayed through Data IQ?	28%	38%	34%		
	Never	**Rarely**	**Sometimes**	**Often**	**Very often**
How often does Data IQ highlight risk factors that are relevant for patient care but not previously considered?	4%	15%	62%	17%	2%
	1–2	**3–4**	**5–6**	**7–8**	**9–10**
How likely are you to recommend this tool to a colleague?	11%	13%	31%	37%	8%

Of the 127 Oak Street clinicians who responded to the survey, 50% agreed or strongly agreed that Data IQ — the internal name for our new machine learning–based tool — is accurate at identifying high-risk patients. A total of 59% of respondents agreed or strongly agreed that the risk factors are relevant for assessing risk. VIP = very important patient (a term used internally for high-risk patients). Source: The authors.

> *"At the beginning of the deployment period, 40% of high-risk patients had an action item documented by the care team in the previous 30 days ... As of November 2020, this engagement metric had increased to 55% of high-risk patients with a recent action item documented."*

We also failed in our initial attempts to build a predictive model for 12-month mortality risk. The first predictive model we created did not significantly outperform a rules-based algorithm developed from the medical literature on life-limiting illnesses. However, we have since revisited the mortality model, applying lessons from our recent data science efforts. Our second attempt appears to have improved accuracy, and we are planning to deploy it across Oak Street Health in Q1 2021.

NEJM Catalyst is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from catalyst.nejm.org by Scott Drummond on April 19, 2021. For personal use only.
No other uses without permission. Copyright © 2021 Massachusetts Medical Society. All rights reserved.

Our use of top factors for display is a new development in primary care. As noted previously, the SHAP algorithm generates top factors for each patient's predictions.[6] We believe the top factors serve three roles:

1. They help explain and build trust in the models' predictions.

2. They remind care team members of potentially overlooked aspects of the patients' conditions and history.

3. They generate discussion topics for care planning. This last element is particularly relevant in a multidisciplinary team setting, because factors spanning medical, social, and behavioral risks may contribute to a comprehensive view of the patient.

To further increase the actionability and value of top factors, we plan to enrich the predictions' display. By clearly delineating a factor's source, we aim to prompt further care team discussion and enable more accurate assessment of patients' risk. Keeping our multidisciplinary end users in mind, we are expanding our provider advisory group to include MSWs, behavioral health specialists, and registered nurses.

The most important factors, however, are not necessarily causal, and they may not provide the best information for care planning. Nothing in the SHAP algorithm compels the top factors to be the most medically relevant for patients' current conditions. We are planning further testing to combine top factors with other not-readily-available information about the patient's history and present condition.

While implementing a primary care machine learning–based tool that is broadly seen by clinicians as interpretable and actionable is an important first step, more work is needed to reach those providers and social workers who remain uncertain about its utility. We believe that continuing to adopt a collaborative approach to building AI tools and engaging clinical advisors, database engineers, user interface experts, and end users will allow us to improve the accuracy and actionability of our tools. We will also continue to integrate a variety of data sources into our models, from medical to social to behavioral, thereby generating various focal points for a multidisciplinary discussion.[21]

Surabhi Bhatt, MS
Senior Associate, Population Health, Oak Street Health, Chicago, Illinois, USA

Adam Cohon, PhD
Data Scientist, Population Health, Oak Street Health, Chicago, Illinois, USA

Jenna Rose, MHS
Vice President, Population Health, Oak Street Health, Chicago, Illinois, USA

Natalia Majerczyk
Data Analyst, Population Health, Oak Street Health, Chicago, Illinois, USA

NEJM Catalyst is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from catalyst.nejm.org by Scott Drummond on April 19, 2021. For personal use only.
No other uses without permission. Copyright © 2021 Massachusetts Medical Society. All rights reserved.

Brian Cozzi, MS
Data Scientist, Population Health, Oak Street Health, Chicago, Illinois, USA

Drew Crenshaw
Chief Population Health Officer, Population Health, Oak Street Health, Chicago, Illinois, USA

Griffin Myers, MD, MBA
Cofounder and Chief Medical Officer, Oak Street Health, Chicago, Illinois, USA

Acknowledgments

We thank Dr. Ali Khan, Dr. Bradley Banko, Dr. David Buchanan, Dr. Ebony Harpool, Dr. Julie Blankemeier, Dr. Julie Silverstein, Dr. Marisa Rogers, Dr. Mark Bustamante, Dr. Rupen Amin, Dr. Niraj Shah, Dr. Palav Shah, Dr. Amish Desai, Dr. Michele Mitchell, and all pilot participants, survey respondents, Oak Street Health care teams, Oak Street Population Health Director partners, Business Intelligence and Canopy Product & Engineering teams, and ClosedLoop.ai.

Disclosures: Surabhi Bhatt, Adam Cohon, Jenna Rose, Natalia Majerczyk, Brian Cozzi, Drew Crenshaw, and Griffin Myers have nothing to disclose.

References

1. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med 2018;1:18 https://www.nature.com/articles/s41746-018-0029-1 https://doi.org/10.1038/s41746-018-0029-1.

2. Brajer N, Cozzi B, Gao M, et al. Prospective and external evaluation of a machine learning model to predict in-hospital mortality of adults at time of admission. JAMA Netw Open 2020;3:e1920733 https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2760438 https://pubmed.ncbi.nlm.nih.gov/32031645/ https://doi.org/10.1001/jamanetworkopen.2019.20733.

3. Elish MC, Watkins EA. Repairing Innovation: A Study of Integrating AI in Clinical Care. Data & Society. September 2020. Accessed December 7, 2020. https://datasociety.net/wp-content/uploads/2020/09/Repairing-Innovation-DataSociety-20200930-1.pdf.

4. Fihn SD, Saria S, Mendonça E, et al. Deploying AI in clinical settings. In: Matheny M, Thadaney-Israni S, Ahmed M., Whicher D, eds. Artificial intelligence in health care: the hope, the hype, the promise, the peril. NAM Special Publication. Washington D.C.: National Academy of Medicine, 2019:145-79 https://nam.edu/wp-content/uploads/2019/12/AI-in-Health-Care-PREPUB-FINAL.pdf.

5. Li RC, Asch SM, Shah NH. Developing a delivery science for artificial intelligence in healthcare. NPJ Digit Med 2020;3:107 https://www.nature.com/articles/s41746-020-00318-y https://doi.org/10.1038/s41746-020-00318-y.

NEJM Catalyst is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from catalyst.nejm.org by Scott Drummond on April 19, 2021. For personal use only.
No other uses without permission. Copyright © 2021 Massachusetts Medical Society. All rights reserved.

6. Hilton CB, Milinovich A, Felix C, et al. Personalized predictions of patient outcomes during and after hospitalization using artificial intelligence. NPJ Digit Med 2020;3:51 https://www.nature.com/articles/s41746-020-0249-z https://doi.org/10.1038/s41746-020-0249-z.

7. Awan SE, Sohel F, Sanfilippo FM, Bennamoun M, Dwivedi G. Machine learning in heart failure: ready for prime time. Curr Opin Cardiol 2018;33:190-5 https://journals.lww.com/co-cardiology/Abstract/2018/03000/Machine_learning_in_heart_failure__ready_for_prime.12.aspx https://pubmed.ncbi.nlm.nih.gov/29194052/ https://doi.org/10.1097/HCO.0000000000000491.

8. Heo J, Yoon JG, Park H, Kim YD, Nam HS, Heo JH. Machine learning–based model for prediction of outcomes in acute stroke. Stroke 2019;50:1263-5 https://www.ahajournals.org/doi/10.1161/STROKEAHA.118.024293 https://doi.org/10.1161/STROKEAHA.118.024293.

9. Olczak J, Fahlberg N, Maki A, et al. Artificial intelligence for analyzing orthopedic trauma radiographs. Acta Orthop 2017;88:581-6 https://www.tandfonline.com/doi/full/10.1080/17453674.2017.1344459 https://pubmed.ncbi.nlm.nih.gov/28681679/ https://doi.org/10.1080/17453674.2017.1344459.

10. Tomašev N, Glorot X, Rae JW, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature 2019;572:116-9 https://www.nature.com/articles/s41586-019-1390-1 https://pubmed.ncbi.nlm.nih.gov/28681679/ https://doi.org/10.1038/s41586-019-1390-1.

11. Connell A, Black G, Montgomery H, et al. Implementation of a digitally enabled care pathway (Part 2): qualitative analysis of experiences of health care professionals. J Med Internet Res 2019;21:e13143 https://www.jmir.org/2019/7/e13143/ https://pubmed.ncbi.nlm.nih.gov/31368443/ https://doi.org/10.2196/13143.

12. Mistry P. Artificial intelligence in primary care. Br J Gen Pract 2019;69:422-3 https://bjgp.org/content/69/686/422 https://doi.org/10.3399/bjgp19X705137.

13. Liyanage H, Liaw S-T, Jonnagaddala J, et al. Artificial intelligence in primary health care: perceptions, issues, and challenges. Yearb Med Inform 2019;28:41-6 https://www.thieme-connect.de/products/ejournals/abstract/10.1055/s-0039-1677901 https://pubmed.ncbi.nlm.nih.gov/31022751/ https://doi.org/10.1055/s-0039-1677901.

14. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial intelligence and the implementation challenge. J Med Internet Res 2019;21:e13659 https://www.jmir.org/2019/7/e13659/ https://pubmed.ncbi.nlm.nih.gov/31293245/ https://doi.org/10.2196/13659.

15. Lin SY, Mahoney MR, Sinsky CA. Ten ways artificial intelligence will transform primary care. J Gen Intern Med 2019;34:1626-30 https://link.springer.com/article/10.1007/s11606-019-05035-1 https://pubmed.ncbi.nlm.nih.gov/31090027/ https://doi.org/10.1007/s11606-019-05035-1.

16. Shah NH, Milstein A, Bagley SC. Making machine learning models clinically useful. JAMA 2019;322:1351-2 https://jamanetwork.com/journals/jama/article-abstract/2748179 https://doi.org/10.1001/jama.2019.10306.

NEJM Catalyst is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from catalyst.nejm.org by Scott Drummond on April 19, 2021. For personal use only.
No other uses without permission. Copyright © 2021 Massachusetts Medical Society. All rights reserved.

17. Sendak MP, D'Arcy J, Kashyap S, et al. A path for translation of machine learning products into healthcare delivery. EMJ Innov January 27, 2020 [Online ahead of print] https://www.emjreviews.com/innovations/article/a-path-for-translation-of-machine-learning-products-into-healthcare-delivery/ https://pubmed.ncbi.nlm.nih.gov/28085085/ https://doi.org/10.33590/emjinnov/19-00172.

18. 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://science.sciencemag.org/content/366/6464/447 https://pubmed.ncbi.nlm.nih.gov/31649194/ https://doi.org/10.1126/science.aax2342.

19. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In Krishnapuram B, Shah M, eds. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. August 2016. New York: Association for Computing Machinery, 2016:785-94 https://dl.acm.org/doi/10.1145/2939672.2939785.

20. Lundberg S, Lee S-I. A unified approach to interpreting model predictions. In Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, eds. Advances in neural information processing systems 30 (NIPS 2017). San Diego, CA: Neural Information Processing Systems, 2017. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.

21. Schelling TC. The strategy of conflict. Cambridge, MA: Harvard University Press, 1960.

NEJM Catalyst is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from catalyst.nejm.org by Scott Drummond on April 19, 2021. For personal use only.
No other uses without permission. Copyright © 2021 Massachusetts Medical Society. All rights reserved.