original reports

System for High-Intensity Evaluation During Radiation Therapy (SHIELD-RT): A Prospective Randomized Study of Machine Learning–Directed Clinical Evaluations During Radiation and Chemoradiation

Julian C. Hong, MD, MS[1,2,3]; Neville C. W. Eclov, PhD[3]; Nicole H. Dalal, MD[4]; Samantha M. Thomas, MS[5,6]; Sarah J. Stephens, MD[3]; Mary Malicki, MSN, ACNP[3]; Stacey Shields, ANP-BC[3]; Alyssa Cobb, RN, BSN[3]; Yvonne M. Mowery, MD, PhD[3,6]; Donna Niedzwiecki, PhD[5,6]; Jessica D. Tenenbaum, PhD[5]; and Manisha Palta, MD[3,6]

abstract

PURPOSE Patients undergoing outpatient radiotherapy (RT) or chemoradiation (CRT) frequently require acute care (emergency department evaluation or hospitalization). Machine learning (ML) may guide interventions to reduce this risk. There are limited prospective studies investigating the clinical impact of ML in health care. The objective of this study was to determine whether ML can identify high-risk patients and direct mandatory twice-weekly clinical evaluation to reduce acute care visits during treatment.

PATIENTS AND METHODS During this single-institution randomized quality improvement study (ClinicalTrials.gov identifier: NCT04277650), 963 outpatient adult courses of RT and CRT started from January 7 to June 30, 2019, were evaluated by an ML algorithm. Among these, 311 courses identified by ML as high risk (> 10% risk of acute care during treatment) were randomized to standard once-weekly clinical evaluation (n = 157) or mandatory twice-weekly evaluation (n = 154). Both arms allowed additional evaluations on the basis of clinician discretion. The primary end point was the rate of acute care visits during RT. Model performance was evaluated using receiver operating characteristic area under the curve (AUC) and decile calibration plots.

RESULTS Twice-weekly evaluation reduced rates of acute care during treatment from 22.3% to 12.3% (difference, −10.0%; 95% CI, −18.3 to −1.6; relative risk, 0.556; 95% CI, 0.332 to 0.924; $P = .02$). Low-risk patients had a 2.7% acute care rate. Model discrimination was good in high- and low-risk patients undergoing standard once-weekly evaluation (AUC, 0.851).

CONCLUSION In this prospective randomized study, ML accurately triaged patients undergoing RT and CRT, directing clinical management with reduced acute care rates versus standard of care. This prospective study demonstrates the potential benefit of ML in health care and offers opportunities to enhance care quality and reduce health care costs.

J Clin Oncol 38. © 2020 by American Society of Clinical Oncology

ASSOCIATED CONTENT

Data Supplement
Protocol

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on August 7, 2020 and published at ascopubs.org/journal/jco on September 4, 2020: DOI https://doi.org/10.1200/JCO.20.01688

The Duke Endowment had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

INTRODUCTION

An estimated 650,000 patients with cancer receive systemic therapy or radiation therapy (RT) annually in the United States.[1] Among these, 10%-20% of patients undergoing outpatient RT or chemoradiation (CRT) will require acute care with an emergency department (ED) visit or hospital admission because of symptoms from treatment, disease, or comorbidities.[2,3] This can affect outcomes, patient quality of life and preferences, and costs to patients and the health care system, making it a priority to the Centers for Medicare & Medicaid Services (CMS).[2-5] Early identification and intervention may prevent such events.[6,7]

Artificial intelligence (AI) and machine learning (ML) have garnered enthusiasm for their potential to provide accurate predictions, with the goal of directing interventional strategies. We previously developed an ML algorithm utilizing electronic health record (EHR) data, which retrospectively demonstrated strong predictive ability to identify patients at high risk for acute care,[8] comparing favorably to other models for this complex problem.[9,10] Despite growing retrospective medical ML literature,[11,12] prospective evaluation remains tremendously limited, primarily to diagnostic fields.[13-16] Despite the need for prospective interventional trials,[17] few have been published.[18] There are even fewer studies investigating the use of ML to improve clinical

CONTEXT

Key Objective

Can machine learning based on routine pretreatment electronic health record data direct supplemental care to reduce acute care (emergency visits and hospitalization) during outpatient cancer radiotherapy and chemoradiation?

Knowledge Generated

In this prospective randomized study, machine learning accurately identified outpatient treatment courses as high risk for acute care. Machine learning–identified high-risk patients randomly assigned to supplemental clinical evaluations during treatment had reduced acute care rates.

Relevance

Machine learning with electronic health data can accurately triage patients and guide intervention to decrease acute care.

outcomes, which may be related to challenges in clinical implementation and viability of systematic management strategies. The need to prospectively demonstrate the value of these technologies is underscored by historical findings, such as the potential worsening of mammographic detection with computer-aided diagnosis.[19]

Randomized quality improvement (QI) studies offer opportunities to create a learning health care system to improve care delivery.[20] In certain scenarios, this may represent a platform for evaluating the utility of ML. SHIELD-RT (System for High Intensity Evaluation During Radiation Therapy) was a prospective, randomized QI study evaluating the benefit of ML to triage high-risk patients and direct supplemental clinical evaluation to potentially decrease acute care required during outpatient RT or CRT. To our knowledge, this study represents one of the first to leverage health care ML prospectively to direct a randomized intervention.

Patients and Methods

This single-institution randomized study was approved as a prospective QI project by the Duke University Medical Center Institutional Review Board (Pro00100647) and registered on ClinicalTrials.gov (NCT04277650). All adult outpatient RT courses with or without concurrent systemic therapy (chemotherapy or immunotherapy) started from January 7, 2019, to June 30, 2019, at the Duke Cancer Center were included, with the exception of total body irradiation (as these patients are planned for admission for hematopoietic stem-cell transplantation). The study protocol is available in the Data Supplement (online only).

For the randomized component, the ML algorithm was run weekly to identify high-risk patients who had started RT in the current week, defined as those with $\geq$ 10% risk of acute care (ED visit or hospital admission) during treatment. This threshold was predetermined based on clinical judgment, the Youden cut point in retrospective development,[8] and available resources. Patients had to have ongoing RT/CRT planned for the following week to be eligible. Two physicians in the department opted out of randomization before study initiation.

Data Processing and Machine Learning

Our ML pipeline was previously described, and source code is available online.[8] The algorithm aggregates a patient's pretreatment EHR history via the Duke Enterprise Data Unified Content Explorer (DEDUCE)[21] and cancer treatment plan (ie, RT prescription, concurrent chemotherapy) and utilizes gradient-boosted trees (GBTs) to predict the likelihood (on a continuous scale from 0-1) that a patient will require acute care during treatment. Gradient boosting is a supervised ML technique that uses data with predictors and explicit outputs (ie, classification of whether an acute care event occurred). Multiple weak predictive models are generated, with subsequent models fitting pseudo-residuals. This study used XGBoost, which has gained popularity because of its success in ML challenges.[22] GBTs balance predictive accuracy with interpretability by identification of important variables. One of the strengths of decision trees is their ability to accommodate missing values at the time of prediction.

For this study, the algorithm was retrained on the entire original development cohort, treated from January 2013 to December 2016 and locked for the duration of the study. As previously published, no individual variables dominated, and top predictive factors were broad, including treatment parameters, encounter history, vitals, age, and laboratories (Data Supplement). Of note, the interpretation of variable importance can be challenging because of correlated factors.

For pragmatic reasons, the pipeline was implemented on Fridays for courses started during the current week. New patient starts were identified via Aria (Varian Medical Systems, Palo Alto, CA). A manual process was required to verify eligibility and confirm that patients had initiated RT as outpatients and not yet completed RT before random assignment (N.C.W.E., J.C.H.). High-risk patients who completed a short course of RT within a calendar week before the weekly algorithm run were thus not eligible for

random assignment, as they would not have been able to complete the intervention. The algorithm was independently rerun by two investigators to verify risk output (N.C.W.E., Y.M.M.).

Randomization and Intervention

ML-identified high-risk courses were randomized in a 1:1 fashion to standard-of-care management (clinical evaluation at least once weekly) or mandatory twice-weekly evaluation. During RT, patients are seen weekly by the treating radiation oncologist as standard of care. These visits are problem focused and involve a directed history and physical examination with appropriate symptomatic management. The supplemental visit followed the same structure and did not require additional studies, such as laboratory studies or imaging. As a QI study, informed consent was not required. However, patients randomly assigned to twice-weekly visits were informed of their computational identification as high risk and the purpose of the additional evaluations. They were given the opportunity to ask questions at each visit. Patients in either arm could also have additional ad hoc clinical evaluations as deemed appropriate by the treating physician. Visits with other providers continued per standard of care, and referrals were based on the clinical judgment of the treating physician.

Random assignment was performed on a per-course basis, and patients undergoing multiple eligible courses were randomly assigned separately for each course. A permuted block randomization schema with a block size of six was generated using SAS 9.4 (Cary, NC) by the associate statistician. No stratification variables were used. All eligible patients were entered into REDCap, which performed random assignment.[23] Courses randomized to intervention were relayed to study team members responsible for seeing patients for their supplemental visits, which were initiated in the second week of treatment. The team consisted of attending physicians, resident physicians, advanced practice providers, and nurse clinicians. When possible, a primary radiation team member was responsible for supplemental evaluations. As in standard weekly evaluations, the mandatory second visits were performed by a single clinician during a treatment course unless unavailable. Assignment to mandatory twice-weekly evaluation was unblinded to the study team and patients. However, ML identification of high-risk patients randomly assigned to control was blinded.

Statistical Methods

The primary end point of the randomized component was the rate of acute care visits during RT, defined as unplanned ED visits or hospital admissions (planned admissions for procedures or chemotherapy were not included). Target sample size was initially pragmatically determined to be 202 courses because of concerns regarding feasibility of sustained ML deployment. Given seamless integration of ML implementation, the study was amended on May 2, 2019, to increase the power to facilitate a hypothesis more consistent with prior retrospective studies.[6] An interim analysis was not performed at the time of amendment. The final sample size was determined to be 314 treatment courses to test the hypothesis of decreased acute care visits in the high-intensity evaluation arm. This was designed with 80% power to detect a difference between 20% and 10% of patients requiring acute care visits in the control and intervention arms, respectively, with a one-sided significance level of 0.05. Secondary end points included rate of acute care during RT and the 15 days after treatment, rate of missed intervention evaluations, and reasons for acute care. Reasons for acute care were grouped based on those designated as potentially preventable by CMS: anemia, nutrition (including dehydration), diarrhea, emesis, infections (including fever, pneumonia, and sepsis), nausea, neutropenia, and pain.[5] An intent-to-treat analysis was used.

The overall performance of the ML model and clinician prediction were evaluated by plotting the receiver operating curve (ROC) and decile calibration plots of risk score and true event probability. For courses on the intervention arm, clinician predictions were collected during the first mandatory supplemental evaluation. These predictions were unblinded to knowledge that the patient had been assigned to intervention with an ML risk of at least 10%. Clinicians documented time spent after each mandatory supplemental evaluation.

Although acute care visits outside the Duke system were documented and reviewed in the routine and supplemental clinical evaluations for courses that underwent randomization, nonrandomized courses were retrospectively assessed and limited to encounters within the Duke system.

Patient and treatment course characteristics were summarized with No. (%) and median (interquartile range, IQR) for categorical and continuous variables, respectively. Differences between groups were tested using χ^2 or Fisher's exact tests for categorical variables and t tests or analysis of variance for continuous variables. ROC curves were created, and area under the curve (AUC) values were estimated based on ML prediction. Decile calibration plots were created for ML risk versus true event probability and clinician risk estimate. No adjustments were made for multiple testing. All statistical analyses were conducted using SAS version 9.4 (SAS Institute, Cary, NC).

RESULTS

Between January 7, 2019, and June 30, 2019, 963 treatment courses were assessed by ML (Fig 1, Table 1). The study completed accrual with 361 courses identified by ML as high risk and 314 courses undergoing planned randomization. Among the remaining 47 ML high-risk

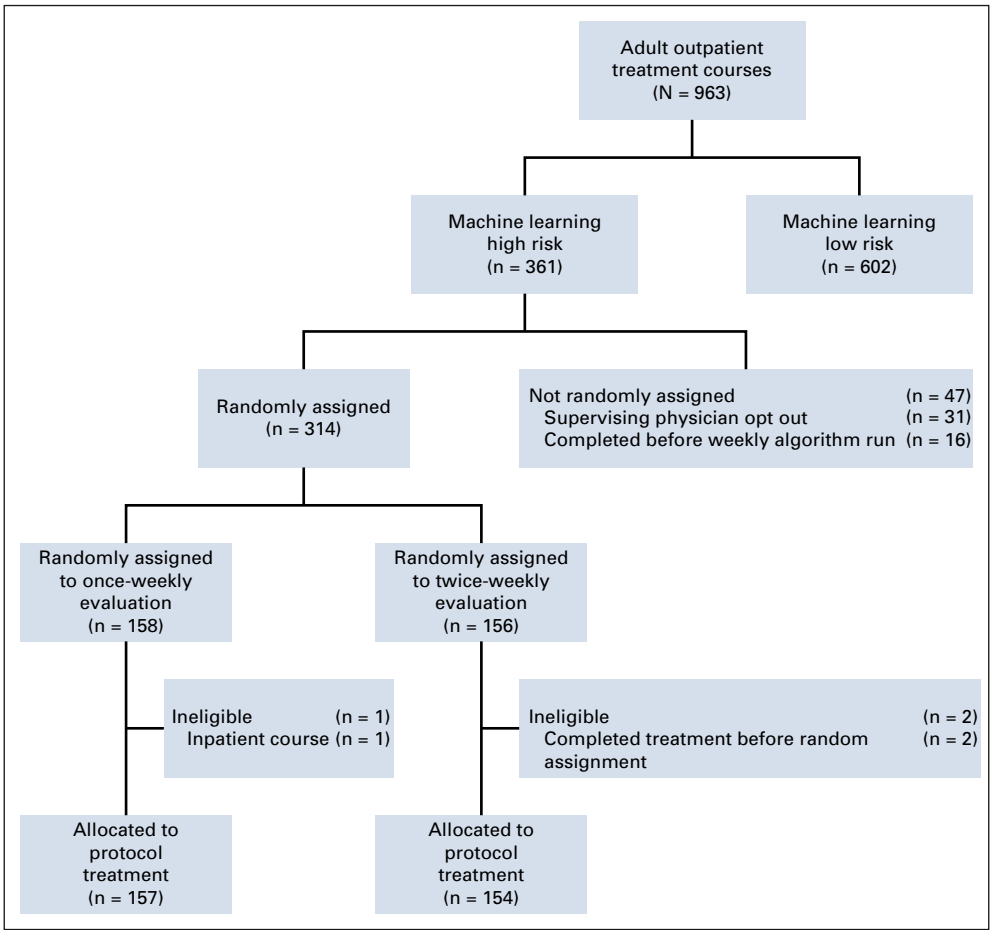

FIG 1. CONSORT diagram.

courses, 31 courses were treated by physicians who had opted out of the randomized study before initiation, and 16 had completed treatment before the weekly algorithm run. Three ineligible courses were randomized, with one initiated in the inpatient setting and another two completed before ML screening and randomization. These patients underwent standard care. There were 311 total eligible and analyzable courses (Table 2). The Duke analytics platform was offline during 3 separate weeks because of independent technical reasons, and courses started during these weeks were neither assessed nor included.

ML prediction was comparable between the standard (median, 18.8%) and intervention arms (19.6%). The randomized arms were also similar in baseline characteristics, including age, sex, race, ethnicity, marital status, cancer diagnosis, and treatment characteristics (Table 2). Rates of concurrent systemic therapy were comparable across the standard (42.7%) and interventional arms (47.4%). Radiation technique, a correlate of treatment intent, was also comparable. Days elapsed on RT were similarly distributed for the standard (median, 35 days) and interventional arms (34 days; Data Supplement). The most common diagnoses in the randomized cohort were GI (21.5%) and primary brain (16.4%) malignancies.

Randomized Study Outcomes

Among ML high-risk patients, supplemental clinical evaluation demonstrated a decrease in the primary end point (aggregate ED visits or unplanned hospital admission during radiotherapy), with an event rate of 12.3% compared with 22.3% (difference, −10.0%; 95% CI, −18.3 to −1.6; relative risk, 0.556; 95% CI, 0.332 to 0.924; $P = .02$; Table 3). This difference remained when the 15 days after treatment were included (22.1% *v* 32.5%; difference, −10.4%; 95% CI, −20.2 to −0.6; relative risk, 0.68; 95% CI, 0.468 to 0.987; $P = .04$; Table 3). This effect was seen across a broad spectrum of primary diagnoses (Data Supplement). For ML low-risk courses (< 10% risk of acute care), 2.7% resulted in an acute visit. Overall, 79.7% (444 of 557) of study-mandated supplemental evaluations were completed.

The most common reasons for individual acute care visits during radiation were neurologic (18.4%), nutritional (11.8%), and other treatment complication (11.8%; Table 4). Among these, 30.3% met CMS criteria as potentially preventable. The proportion of CMS-designated preventable visits was greater in the standard (35.3%) than in the intervention arm (20%).

There was a median of no missed supplemental visits (IQR, 0-1) during a median of three (IQR, 1-5) additional visits per

TABLE 1. Major Characteristics for Entire Cohort

Characteristic	All Courses (N = 963)	High Risk (n = 361)	Low Risk (n = 602)
Distinct patients[a]	917	352	581
ML risk, %	7.2 (2.8-15.1)	18.8 (13.6-26.2)	3.6 (1.7-6.4)
Age at prediction, years	65 (56-72)	64 (55-72)	66 (57-72)
Sex			
Female	472 (49)	154 (42.7)	318 (52.8)
Male	491 (51)	207 (57.3)	284 (47.2)
Race			
White	681 (70.7)	251 (69.5)	430 (71.4)
Black or African American	220 (22.8)	93 (25.8)	127 (21.1)
Other	32 (3.3)	10 (2.8)	22 (3.7)
Unknown	30 (3.1)	7 (1.9)	23 (3.8)
Ethnicity			
Hispanic	18 (1.9)	8 (2.2)	10 (1.7)
Non-Hispanic	892 (92.6)	339 (93.9)	553 (91.9)
Unknown	53 (5.5)	14 (3.9)	39 (6.5)
Marital status			
Married or life partner	625 (64.9)	232 (64.3)	393 (65.3)
Single	146 (15.2)	64 (17.7)	82 (13.6)
Divorced or legally separated	91 (9.4)	35 (9.7)	56 (9.3)
Widowed	70 (7.3)	27 (7.5)	43 (7.1)
Unknown	31 (3.2)	3 (0.8)	28 (4.7)
Disease site			
Bone metastases	131 (13.6)	44 (12.2)	87 (14.5)
Primary brain cancer	82 (8.5)	53 (14.7)	29 (4.8)
Brain metastases	156 (16.2)	31 (8.6)	125 (20.8)
Breast cancer	123 (12.8)	7 (1.9)	116 (19.3)
GI cancer	108 (11.2)	75 (20.8)	33 (5.5)
Genitourinary cancer	92 (9.6)	28 (7.8)	64 (10.6)
Gynecologic cancer	29 (3)	18 (5)	11 (1.8)
Head and neck cancer	45 (4.7)	32 (8.9)	13 (2.2)
Respiratory/intrathoracic cancer	122 (12.7)	44 (12.2)	78 (13)
Other cancer	364 (37.8)	139 (38.5)	225 (37.4)
Concurrent treatment			
Chemotherapy	160 (16.6)	147 (40.7)	13 (2.2)
Immunotherapy	11 (1.1)	7 (1.9)	4 (0.7)
Radiation technique			
2D or 3D conformal RT	367 (38.1)	135 (37.4)	232 (38.5)
Intensity-modulated RT or volumetric modulated arc therapy	331 (34.4)	221 (61.2)	110 (18.3)
Stereotactic body RT/stereotactic radiosurgery	261 (27.1)	5 (1.4)	256 (42.5)
Total skin irradiation	4 (0.4)	0 (0)	4 (0.7)
Days on treatment	14 (6-35)	32 (14-42)	9 (4-23)

NOTE. Data are presented as median (interquartile range) or No. (%). Percentages may not add up to 100 because of rounding or missing values.

Abbreviations: 2D, two-dimensional; 3D, three-dimensional; ML, machine learning; RT, radiotherapy.

[a]A total of 917 distinct patients were included in this study, but random assignment was done on a per-course basis, so a specific patient may appear in multiple arms if multiple course of RT were administered.

TABLE 2. Major Characteristics for Randomized Cohort

Characteristic	All (N = 311)	Randomly Assigned: Control (n = 157)	Randomly Assigned: Intervention (n = 154)
Distinct patients[a]	305	156	152
ML risk, %	19.4 (14.2-26.6)	18.8 (14.2-26.9)	19.6 (14.3-25.9)
Age at prediction, years	64 (55-72)	64 (55-72)	64 (56-72)
Sex			
Female	128 (41.2)	63 (40.1)	65 (42.2)
Male	183 (58.8)	94 (59.9)	89 (57.8)
Race			
White	222 (71.4)	120 (76.4)	102 (66.2)
Black or African American	72 (23.2)	31 (19.7)	41 (26.6)
Other	10 (3.2)	4 (2.5)	6 (3.9)
Unknown	7 (2.3)	2 (1.3)	5 (3.2)
Ethnicity			
Hispanic	8 (2.6)	3 (1.9)	5 (3.2)
Non-Hispanic	290 (93.2)	146 (93)	144 (93.5)
Unknown	13 (4.2)	8 (5.1)	5 (3.2)
Marital status			
Married or life partner	207 (66.6)	104 (66.2)	103 (66.9)
Single	50 (16.1)	24 (15.3)	26 (16.9)
Divorced or legally separated	31 (10)	15 (9.6)	16 (10.4)
Widowed	20 (6.4)	12 (7.6)	8 (5.2)
Unknown	3 (1)	2 (1.3)	1 (0.6)
Disease site			
Bone metastases	33 (10.6)	21 (13.4)	12 (7.8)
Primary brain cancer	51 (16.4)	23 (14.6)	28 (18.2)
Brain metastases	25 (8)	12 (7.6)	13 (8.4)
Breast cancer	5 (1.6)	4 (2.5)	1 (0.6)
GI cancer	67 (21.5)	31 (19.7)	36 (23.4)
Genitourinary cancer	23 (7.4)	12 (7.6)	11 (7.1)
Gynecologic cancer	18 (5.8)	7 (4.5)	11 (7.1)
Head and neck cancer	32 (10.3)	20 (12.7)	12 (7.8)
Respiratory/intrathoracic cancer	41 (13.2)	23 (14.6)	18 (11.7)
Other cancer	109 (35)	63 (40.1)	46 (29.9)
Concurrent treatment			
Chemotherapy	140 (45)	67 (42.7)	73 (47.4)
Immunotherapy	6 (1.9)	3 (1.9)	3 (1.9)
Radiation technique			
2D or 3D conformal RT	105 (33.8)	52 (33.1)	53 (34.4)
Intensity-modulated RT or volumetric modulated arc therapy	204 (65.6)	103 (65.6)	101 (65.6)
Stereotactic body RT/stereotactic radiosurgery	2 (0.6)	2 (1.3)	0 (0)
Days on treatment	34 (15-43)	35 (14-43)	34 (15-43)

NOTE. Data are presented as median (interquartile range) or No. (%). Percentages may not add up to 100 because of rounding or missing values.

Abbreviations: 2D, two-dimensional; 3D, three-dimensional; ML, machine learning; RT, radiotherapy.

[a]A total of 305 distinct patients were randomly assigned in this study, but random assignment was done on a per-course basis, so a specific patient may appear in multiple arms if multiple course of RT were administered.

TABLE 3. Acute Care Visits

Visit	All Courses (N = 963)	Randomly Assigned			% Difference (95% CI)	Relative Risk (95% CI)	P[a]
		All (n = 311)	Control (n = 157)	Intervention (n = 154)			
Admission or ED visit							
During RT	76 (7.9)	54 (17.4)	35 (22.3)	19 (12.3)	−10 (−18.3 to −1.6)	0.553 (0.332 to 0.924)	.02
During RT + 15 days	131 (13.6)	85 (27.3)	51 (32.5)	34 (22.1)	−10.4 (−20.2 to −0.6)	0.68 (0.468 to 0.987)	.04
Admission							
During RT	48 (5)	34 (10.9)	24 (15.3)	10 (6.5)	−8.8 (−15.6 to −1.9)	0.425 (0.21 to 0.858)	.01
During RT + 15 days	92 (9.6)	59 (19)	36 (22.9)	23 (14.9)	−8 (−16.7 to 0.7)	0.651 (0.406 to 1.046)	.07
ED visit							
During RT	44 (4.6)	25 (8)	15 (9.6)	10 (6.5)	−3.1 (−9.1 to 3)	0.68 (0.315 to 1.466)	.32
During RT + 15 days	77 (8)	40 (12.9)	25 (15.9)	15 (9.7)	−6.2 (−13.6 to 1.2)	0.612 (0.336 to 1.115)	.10

NOTE. Data are presented as No. (%) unless otherwise noted. Percentages may not add up to 100 because of rounding or missing values. Differences were estimated as Intervention % − Control %. Relative risks were estimated as Intervention %/Control %.

Abbreviations: ED, emergency department; RT, radiation therapy.

[a]χ^2 or Fisher's exact *P* values for categorical variables and *t* test *P* values for continuous variables for comparison of two groups (control *v* intervention).

course. The distribution approximates the expected once-weekly supplemental visits (Data Supplement). Visits required a median of 5 minutes of clinician time (IQR, 5-10 minutes).

ML Performance

Overall, ML demonstrated good prospective predictive performance. Binary discrimination of high and low risk demonstrated AUC of 0.820 across all courses (Data Supplement). For patients who underwent standard once-weekly clinical evaluation, discrimination was higher (AUC, 0.851; Data Supplement). Calibration demonstrated a similar observation; ML prediction versus observed acute care rate decile CIs overlapped with the diagonal reference, particularly for patients undergoing standard management (Data Supplement).

Concordant with the primary end point results, ML estimated risk was greater than the observed rate of acute care in those undergoing intervention, although mostly within the 95% CI (Data Supplement).

Clinician Predictions

Clinicians completed predictions for 145 of 154 intervention courses (94%). With knowledge that ML had identified patients with at least 10% risk, clinician predictions had a narrow distribution centered around a median of 10% (IQR, 5%-15%). In addition, there were wide CIs at most deciles (Data Supplement). Among seven patients assigned a 0% risk of acute care, 14.3% had an acute care visit during RT.

DISCUSSION

This study represents one of the first prospective evaluations, to our knowledge, investigating ML-directed clinical intervention, demonstrating that ML can be feasibly incorporated into a clinical setting and accurately triage and direct supportive care, reducing the rate of acute care events during outpatient RT and CRT. This has potential to improve clinical outcomes and reduce health care costs. We also present randomized data supporting the benefit of more frequent clinical evaluations during treatment.

At present, there are few prospective interventional studies of ML in health care. Randomized interventional data are limited, including a recent study applying ML to reduce intraoperative hypotension.[18] High-quality data are necessary to demonstrate the prospective accuracy and value of the expanding number of published ML algorithms, particularly given limitations of retrospective validation.[17,24,25] Despite their general proliferation, the value of AI and ML algorithms in medicine remains generally unverified, and their deployment remains accordingly limited. This study prospectively demonstrated predictive accuracy of our model with one application where AI and ML can improve patient care. In addition, our study demonstrates an application of randomized QI studies to enable a learning health care system.[20]

Our study addresses the critical clinical problem of decreasing acute care during cancer treatment, central to delivering high-quality cancer care and a point of emphasis for CMS.[4,5] Importantly, we tied our model to a systematic interventional strategy streamlined into the care pathway. This approach may minimize additional cognitive burden to health care providers.[26,27] It is possible that the intervention may also benefit low-risk patients, although our study demonstrates the effectiveness of ML to appropriately triage and optimize resources. On the basis of available resources, individual practices could opt for different

TABLE 4. Reasons for Acute Care Visits During Radiation Therapy

Reason for Acute Care Visit	All Randomly Assigned (N = 76)	Randomly Assigned	
		Control (n = 51)	Intervention (n = 25)
Distinct courses[a]	54	35	19
Reason for visit			
Cardiac	4 (5.3)	3 (5.9)	1 (4)
Fatigue	1 (1.3)	1 (2)	0 (0)
GI	3 (3.9)	3 (5.9)	0 (0)
Infection	7 (9.2)	6 (11.8)	1 (4)
Liver	2 (2.6)	2 (3.9)	0 (0)
Nausea/vomiting	4 (5.3)	3 (5.9)	1 (4)
Neurologic	14 (18.4)	10 (19.6)	4 (16)
Neurologic deficits	6 (7.9)	4 (7.8)	2 (8)
Headaches	2 (2.6)	2 (3.9)	0 (0)
Altered mental status	5 (6.6)	3 (5.9)	2 (8)
Seizures	1 (1.3)	1 (2)	0 (0)
Nutrition	9 (11.8)	7 (13.7)	2 (8)
Pain	3 (3.9)	2 (3.9)	1 (4)
Renal	2 (2.6)	1 (2)	1 (4)
Respiratory	4 (5.3)	3 (5.9)	1 (4)
Thrombotic	3 (3.9)	2 (3.9)	1 (4)
Other treatment complication	9 (11.8)	3 (5.9)	6 (24)
Other	8 (10.5)	4 (7.8)	4 (16)
Preventable reason for visit	23 (30.3)	18 (35.3)	5 (20)

NOTE. Data are presented as No. (%).

Abbreviations: ED, emergency department; RT, radiation therapy.

[a]Of 54 randomized courses with at least one admission or ED visit during RT, 76 admissions or ED visits during RT were observed.

prediction thresholds for identifying high-risk patients. This strategy may also be appropriate in other settings, including during systemic therapy, the subject of many prediction models.[9,10,28,29]

We also evaluated reasons for visits to determine if certain types of acute care were preventable. CMS-designated preventable diagnoses during RT were less common with intervention, although more data may be required to understand which acute care visits are truly preventable.

This study was performed at a single institution, potentially affecting generalizability. It is likely that individual institutions will require distinct models, given variations in patient populations, clinical practice, and supportive resources. Moreover, this approach may not universally generalize. However, given increasing standardization in EHRs and adoption of common data models, it is likely that many institutions will be able to deploy and validate a similar approach. We previously made our code available online.[8] Our group is currently assessing reproducibility at a second institution. Ongoing quality assurance and recalibration are also imperative; we verified our results during this study, but over time, data and practice can shift because of phenomena such as distributional shift, automation bias, and complacency.[25]

Although we were able to demonstrate the benefit of ML-directed supportive care, our study was designed for feasibility, and we opted for a control arm representing the standard-of-care management (once weekly visits, with additional visits per clinician discretion). Randomly assigning all patients to undergo ML evaluation or not would have required a substantially larger sample size to detect a decrease in the lower event rate (7.9%) across all patients in this study. Nevertheless, our study remains one of the few prospective ML studies broadly, and patients prospectively identified as low risk by ML were appropriately triaged, with a 2.7% event rate. To ensure that the control arm was the current standard of care, we were unable to pragmatically randomly assign patients to mandatory evaluations on the basis of predictions generated by other approaches, such as simpler logistic regression models or clinician predictions. Our retrospective analysis had previously demonstrated superior predictions with GBTs compared with Least Absolute Shrinkage and Selection Operator (LASSO) regularized logistic regression.[8] It is

possible that a simpler model may offer sufficient predictive power to reproduce these results, with the benefits of greater interpretability and generalizability. Nevertheless, the prospective metrics of our model compare favorably to our retrospective validation and, although potentially biased by knowledge of ML identification, clinician predictions were loosely calibrated in the intervention arm. Moreover, standard management allowed clinicians to perform ad hoc supplemental evaluations, which should also reflect clinician risk assessment. Combining physician and ML predictions may warrant future investigation, as human-in-the-loop studies may yield greater accuracy.[30] The advantage of our study, however, is that automated evaluation removes human variability and potential bias and minimizes additional cognitive load on providers. Moreover, this approach can also ensure consistent use of resources on the basis of the limitations of a specific clinical setting.

Finally, the mechanism by which mandatory twice-weekly evaluation reduced acute care is not well defined. Standard weekly evaluations are primarily intended to enable clinicians to manage symptoms related to treatment or otherwise. More routine interaction may provide reassurance, continuity, and overall care management or enable opportunities for detection and early intervention. It is possible that additional visits facilitated reduced acute care through medication management, outpatient hydration, additional laboratory or imaging studies, and/or consultation to other services. Although knowledge that a patient was identified as "high risk" by the ML algorithm may affect a patient or clinician's decision to pursue acute care, this would typically be expected to raise awareness and lead to increased acute care in the interventional arm. However, our study remained positive, demonstrating decreased acute care despite the lack of blinding in the interventional arm. We ultimately selected an intervention that has been adopted at other centers to enable a systematic protocol.[6] Prior before-after studies in ML and digital health may have been limited in clinical impact without systematic interventional strategies.[31,32]

Given the significant reduction in acute care, we are evaluating the cost impact of our intervention, as acute care visits make up nearly half of the financial burden of cancer patients.[2] Currently, on-treatment clinical evaluations are included in the costs of a radiotherapy course, and reducing acute care would yield cost savings for payers. We are also planning randomized evaluations comparing mandatory supplemental evaluations on the basis of ML versus clinician risk estimates. Additional areas of active investigation include integration of natural language processing[33] and patient-reported outcomes[34] and assessment of generalizability.[35]

In conclusion, machine learning based on electronic health records can be used to generate accurate and clinically actionable predictions. We demonstrate the ability of machine learning to direct clinical management through additional supportive evaluation and to decrease the rate of acute care during outpatient cancer radiotherapy and chemoradiotherapy.

AFFILIATIONS

[1]Department of Radiation Oncology, University of California, San Francisco, San Francisco, CA
[2]Bakar Computational Health Sciences Institute, University of California, San Francisco, San Francisco, CA
[3]Department of Radiation Oncology, Duke University, Durham, NC
[4]Department of Medicine, University of California, San Francisco, San Francisco, CA
[5]Department of Biostatistics and Bioinformatics, Duke University, Durham, NC
[6]Duke Cancer Institute, Duke University, Durham, NC

CORRESPONDING AUTHOR

Julian C. Hong, MD, MS, Department of Radiation Oncology and Bakar Computational Health Sciences Institute, University of California, San Francisco, 1825 Fourth St, Suite L1101, San Francisco, CA 94158; Twitter: @julian_hong; @DukeCancer, @UCSF; e-mail: julian.hong@ucsf.edu.

SUPPORT

Supported by the Duke Endowment and the Duke Department of Radiation Oncology.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST AND DATA AVAILABILITY STATEMENT

Disclosures provided by the authors and data availability statement (if applicable) are available with this article at DOI https://doi.org/10.1200/JCO.20.01688.

AUTHOR CONTRIBUTIONS

Conception and design: Julian C. Hong, Neville C. W. Eclov, Mary Malicki, Yvonne M. Mowery, Donna Niedzwiecki, Jessica D. Tenenbaum, Manisha Palta
Administrative support: Manisha Palta
Provision of study material or patients: Sarah J. Stephens, Mary Malicki, Stacey Shields, Alyssa Cobb, Manisha Palta
Collection and assembly of data: Julian C. Hong, Neville C. W. Eclov, Nicole H. Dalal, Sarah J. Stephens, Mary Malicki, Stacey Shields, Alyssa Cobb, Manisha Palta
Data analysis and interpretation: Julian C. Hong, Samantha M. Thomas, Sarah J. Stephens, Yvonne M. Mowery, Donna Niedzwiecki, Jessica D. Tenenbaum, Manisha Palta
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

REFERENCES

1. Halpern MT, Yabroff KR: Prevalence of outpatient cancer treatment in the United States: Estimates from the Medical Panel Expenditures Survey (MEPS). Cancer Invest 26:647-651, 2008
2. Jairam V, Lee V, Park HS, et al: Treatment-related complications of systemic therapy and radiotherapy. JAMA Oncol 5:1028-1035, 2019
3. Waddle MR, Chen RC, Arastu NH, et al: Unanticipated hospital admissions during or soon after radiation therapy: Incidence and predictive factors. Pract Radiat Oncol 5:e245-e253, 2015
4. Phillips CM, Deal K, Powis M, et al: Evaluating patients' perception of the risk of acute care visits during systemic therapy for cancer. J Oncol Pract 16:e622-e629, 2020
5. Centers for Medicare & Medicaid Services: Admissions and Emergency Department (ED) Visits for Patients Receiving Outpatient Chemotherapy. https://cmit.cms.gov/CMIT_public/ViewMeasure?MeasureId=2929
6. Terzo L, Fleming M, Yechoor A, et al: Reducing unplanned admissions: Focusing on hospital admissions and emergency department visits for patients with head and neck cancer during radiation therapy. Clin J Oncol Nurs 21:363-369, 2017
7. Handley NR, Schuchter LM, Bekelman JE: Best practices for reducing unplanned acute care for patients with cancer. J Oncol Pract 14:306-313, 2018
8. Hong JC, Niedzwiecki D, Palta M, et al: Predicting emergency visits and hospital admissions during radiation and chemoradiation: An internally validated pretreatment machine learning algorithm. JCO Clin Cancer Inform 10.1200/CCI.18.00037
9. Brooks GA, Uno H, Aiello Bowles EJ, et al: Hospitalization risk during chemotherapy for advanced cancer: Development and validation of risk stratification models using real-world data. JCO Clin Cancer Inform 3:1-10, 2019
10. Daly B, Gorenshteyn D, Nicholas KJ, et al: Building a clinically relevant risk model: Predicting risk of a potentially preventable acute care visit for patients starting antineoplastic treatment. JCO Clin Cancer Inform 4:275-289, 2020
11. Tomašev N, Glorot X, Rae JW, et al: A clinically applicable approach to continuous prediction of future acute kidney injury. Nature 572:116-119, 2019
12. Rajkomar A, Oren E, Chen K, et al: Scalable and accurate deep learning with electronic health records. NPJ Digit Med 1:18, 2018
13. Lindsey R, Daluiski A, Chopra S, et al: Deep neural network improves fracture detection by clinicians. Proc Natl Acad Sci USA 115:11591-11596, 2018
14. Abràmoff MD, Lavin PT, Birch M, et al: Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit Med 1:39, 2018
15. Wang P, Liu X, Berzin TM, et al: Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): A double-blind randomised study. Lancet Gastroenterol Hepatol 5:343-351, 2020
16. Gong D, Wu L, Zhang J, et al: Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): A randomised controlled study. Lancet Gastroenterol Hepatol 5:352-361, 2020
17. Challener DW, Prokop LJ, Abu-Saleh O: The proliferation of reports on clinical scoring systems: Issues about uptake and clinical utility. JAMA 321:2405-2406, 2019
18. Wijnberge M, Geerts BF, Hol L, et al: Effect of a machine learning–derived early warning system for intraoperative hypotension vs standard care on depth and duration of intraoperative hypotension during elective noncardiac surgery: The HYPE randomized clinical trial [Internet]. JAMA 323:1052-1060, 2020
19. Lehman CD, Wellman RD, Buist DSM, et al: Diagnostic accuracy of digital screening mammography with and without computer-aided detection. JAMA Intern Med 175:1828-1837, 2015
20. Horwitz LI, Kuznetsova M, Jones SA: Creating a learning health system through rapid-cycle, randomized testing. N Engl J Med 381:1175-1179, 2019
21. Horvath MM, Winfield S, Evans S, et al: The DEDUCE Guided Query tool: Providing simplified access to clinical data for research and quality improvement. J Biomed Inform 44:266-276, 2011
22. Chen T, Guestrin C: XGBoost: A scalable tree boosting system. arXiv.org cs.LG:785-794, 2016
23. Harris PA, Taylor R, Thielke R, et al: Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 42:377-381, 2009
24. Angus DC: Randomized clinical trials of artificial intelligence. JAMA 10.1001/jama.2020.1039 [epub ahead of print on February 17, 2020]
25. Challen R, Denny J, Pitt M, et al: Artificial intelligence, bias and clinical safety. BMJ Qual Saf 28:231-237, 2019
26. Osheroff JA, Teich JM, Middleton B, et al: A roadmap for national action on clinical decision support. J Am Med Inform Assoc 14:141-145, 2007
27. Schneider DF: Making machine learning accessible and actionable for clinicians. JAMA Netw Open 2:e1917362, 2019
28. Hurria A, Mohile S, Gajra A, et al: Validation of a prediction tool for chemotherapy toxicity in older adults with cancer. J Clin Oncol 34:2366-2371, 2016
29. Grant RC, Moineddin R, Yao Z, et al: Development and validation of a score to predict acute care use after initiation of systemic therapy for cancer. JAMA Netw Open 2:e1912823, 2019
30. Patel BN, Rosenberg L, Willcox G, et al: Human–machine partnership with artificial intelligence for chest radiograph diagnosis. NPJ Digit Med 2:111, 2019
31. Connell A, Montgomery H, Martin P, et al: Evaluation of a digitally-enabled care pathway for acute kidney injury management in hospital emergency admissions. NPJ Digit Med 2:67, 2019
32. Downing NL, Rolnick J, Poole SF, et al: Electronic health record-based clinical decision support alert for severe sepsis: A randomised evaluation. BMJ Qual Saf 28:762-768, 2019
33. Hong JC, Tanksley J, Niedzwiecki D, et al: Accuracy of a natural language processing pipeline to identify patient symptoms during radiation therapy. Int J Radiat Oncol 105:S70, 2019 (suppl)
34. Basch E, Deal AM, Kris MG, et al: Symptom monitoring with patient-reported outcomes during routine cancer treatment: A randomized controlled trial. J Clin Oncol 34:557-565, 2016
35. Obermeyer Z, Powers B, Vogeli C, et al: Dissecting racial bias in an algorithm used to manage the health of populations. Science 366:447-453, 2019

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

System for High-Intensity Evaluation During Radiation Therapy (SHIELD-RT): A Prospective Randomized Study of Machine Learning–Directed Clinical Evaluations During Radiation and Chemoradiation

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/jco/authors/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Julian C. Hong
Patents, Royalties, Other Intellectual Property: Inventor on pending patent, "Systems and methods for predicting acute care visits during outpatient cancer therapy"

Neville C. W. Eclov
Employment: Adrenas Therapeutics (I)
Stock and Other Ownership Interests: Adrenas Therapeutics (I)
Travel, Accommodations, Expenses: Adrenas Therapeutics (I)

Samantha M. Thomas
Consulting or Advisory Role: AbbVie

Stacey Shields
Stock and Other Ownership Interests: Pfizer

Donna Niedzwiecki
Patents, Royalties, Other Intellectual Property: Inventor on pending patent, "Systems and methods for predicting acute care visits during outpatient cancer therapy"

Jessica D. Tenenbaum
Patents, Royalties, Other Intellectual Property: Inventor on pending patent, "Systems and methods for predicting acute care visits during outpatient cancer therapy"

Yvonne M. Mowery
Honoraria: Oakstone, UpToDate
Research Funding: Merck (Inst)

Manisha Palta
Employment: Duke University
Consulting or Advisory Role: Voxelmetric, Syntactx
Research Funding: Merck (Inst)
Patents, Royalties, Other Intellectual Property: UptoDate: Annual royalties for being a section author. Inventor on pending patent, "Systems and methods for predicting acute care visits during outpatient cancer therapy"

No other potential conflicts of interest were reported.