

Hospitalization Risk During Chemotherapy for Advanced Cancer: Development and Validation of Risk Stratification Models Using Real-World Data

Gabriel A. Brooks, MD, MPH¹; Hajime Uno, PhD²; Erin J. Aiello Bowles, MPH³; Alexander R. Menter, MD⁴; Maureen O’Keeffe-Rosetti, MS⁵; Anna N.A. Tosteson, ScD¹; Debra P. Ritzwoller, PhD⁴; and Deborah Schrag, MD, MPH²

PURPOSE Hospitalizations are a common occurrence during chemotherapy for advanced cancer. Validated risk stratification tools could facilitate proactive approaches for reducing hospitalizations by identifying at-risk patients.

PATIENTS AND METHODS We assembled two retrospective cohorts of patients receiving chemotherapy for advanced nonhematologic cancer; cohorts were drawn from three integrated health plans of the Cancer Research Network. We used these cohorts to develop and validate logistic regression models estimating 30-day hospitalization risk after chemotherapy initiation. The development cohort included patients in two health plans from 2005 to 2013. The validation cohort included patients in a third health plan from 2007 to 2016. Candidate predictor variables were derived from clinical data in institutional data warehouses. Models were validated based on the C-statistic, positive predictive value, and negative predictive value. Positive predictive value and negative predictive value were calculated in reference to a prespecified risk threshold (hospitalization risk \geq 18.0%).

RESULTS There were 3,606 patients in the development cohort (median age, 63 years) and 634 evaluable patients in the validation cohort (median age, 64 years). Lung cancer was the most common diagnosis in both cohorts (26% and 31%, respectively). The selected risk stratification model included two variables: albumin and sodium. The model C-statistic in the validation cohort was 0.69 (95% CI, 0.62 to 0.75); 39% of patients were classified as high risk according to the prespecified threshold; 30-day hospitalization risk was 24.2% (95% CI, 19.9% to 32.0%) in the high-risk group and 8.7% (95% CI, 6.1% to 12.0%) in the low-risk group.

CONCLUSION A model based on data elements routinely collected during cancer treatment can reliably identify patients at high risk for hospitalization after chemotherapy initiation. Additional research is necessary to determine whether this model can be deployed to prevent chemotherapy-related hospitalizations.

Clin Cancer Inform. © 2019 by American Society of Clinical Oncology

INTRODUCTION

Hospitalizations are a common, burdensome, and potentially avoidable complication of cancer treatment.¹⁻³ Hospitalizations are particularly common among patients with advanced cancer,⁴ with both cancer-related symptoms and treatment-related toxicities as contributing factors.³ These hospitalizations are also costly, representing the largest single component of Medicare spending for patients with advanced cancer.⁵ Furthermore, there is emerging evidence that preempting or avoiding complications that lead to hospitalization may be associated with improvements in survival among patients receiving cancer treatment.^{6,7}

Cancer care providers and policymakers are increasingly recognizing the importance of developing new strategies to prevent hospitalizations. Alternative payment models, such as the Centers for Medicare and Medicaid Services Oncology Care Model,^{8,9} contain

financial incentives to enhance clinic-based, ambulatory care systems and prevent avoidable hospitalizations. A number of cancer-focused delivery system innovations have shown promise for reducing hospitalizations during cancer treatment, including patient navigation programs,¹⁰ proactive telephonic nursing,¹¹ and electronic systems for between-visit symptom reporting.^{6,12} Many of these interventions are resource intensive, and targeting outreach interventions to patients at high risk for hospitalization has been identified as a key strategy for preventing unplanned hospitalizations.¹³

Beyond the routine assessment of performance status, risk stratification tools for identifying patients at risk for hospitalization during cancer treatment are not commonly used. Hurria et al^{14,15} created a risk stratification tool for identifying patients at increased risk for grade-3-or-higher chemotherapy toxicity, an outcome that can often lead to hospitalization. That risk stratification

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

Accepted on February 11, 2019 and published at ascopubs.org/journal/cci on April 17, 2019; DOI <https://doi.org/10.1200/CCI.18.00147>

CONTEXT

Key Objective To develop and validate an approach for identifying patients at increased risk for hospitalization after initiation of chemotherapy for advanced cancer.

Knowledge Generated We describe the development and external validation of three risk stratification models designed to identify patients at increased risk for any-cause 30-day hospitalization after initiation of chemotherapy for advanced solid tumor malignancy. A two-variable model using pretreatment sodium and albumin levels stratified patients into a high-risk group (39% of patients, with a 30-day hospitalization risk of 24.2%) and a standard-risk group (61% of patients, with a 30-day hospitalization risk of 8.7%).

Relevance The two validated risk stratification models described here can be used to identify patients at increased risk for hospitalization during chemotherapy treatment. We encourage further study to confirm and update these risk stratification models in an evolving ambulatory practice environment. Studies testing approaches for preventing hospitalizations in high-risk patients are of particular importance.

tool was developed in an elderly patient population and requires geriatric assessment input variables that are not routinely collected in the course of clinical care. Separately, a proof-of-concept study showed that a model using observable patient characteristics provided good discrimination in predicting risk for chemotherapy-related hospitalizations.¹⁶ Here we describe the development and validation of a suite of novel risk stratification models—relying exclusively on routinely collected clinical and administrative data—to identify patients with advanced cancer at increased risk for hospitalization after chemotherapy initiation. An accurate, validated model using routinely collect data has the potential for wide dissemination and application within electronic medical records of US health care systems.¹⁷

PATIENTS AND METHODS

Overview

We sought to develop a suite of clinical prediction models to evaluate the risk of hospitalization after chemotherapy initiation among patients with advanced solid tumor malignancies. The primary study outcome was all-cause hospitalization within 30 days of chemotherapy treatment, and we developed models for two overlapping prediction periods. The first prediction period was the 30 days after the initial chemotherapy treatment day (day 1). The second prediction period was a 30-day interval starting on the day of the first follow-up treatment visit occurring 15 to 22 days after the first treatment day (day N); in most cases, this was the second chemotherapy visit. Candidate predictors were informed by a prior study¹⁶ and were restricted to data elements initially recorded as part of routine health care delivery (in electronic health records [EHRs] and/or health plan administrative data). Study conduct and reporting adhered to the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) statement.¹⁸

Patients and Data Sources

We assembled two patient cohorts for this study; both cohorts comprised patients receiving their first chemotherapy treatment for stage IV or recurrent solid tumor malignancy. Patients were drawn from three Kaiser Permanente (KP) regional health systems that are founding members of the Cancer Research Network (CRN)¹⁹: Colorado (KPCO), Northwest (KPNW), and Washington (KPWA, formerly known as Group Health Cooperative). The CRN is a consortium of integrated health care systems affiliated with the National Cancer Institute.²⁰ Each of these autonomous, community-based health systems uses an Epic EHR. Although the regions share clinical guidelines, patient care is coordinated at the community level.²¹

Data used in this study were derived primarily from the CRN-supported Virtual Data Warehouse (VDW). The VDW contains administrative, EHR, and other clinical data that have been extracted at each site.^{19,22} Cancer diagnoses were obtained from the Virtual Tumor Registry component of the VDW, which adheres to standards of the National Cancer Institute SEER program and the North American Association of Central Cancer Registries.²³ The KPCO and KPNW tumor registries (but not the KPWA registry) also collect information about cancer recurrence among patients initially diagnosed with American Joint Committee on Cancer stage I to III cancer. Oral and infused chemotherapy treatments were captured in the VDW infusion, pharmacy, and procedure files.²⁴ Institutional review board approval was obtained from all three sites.

The first of the two cohorts was used for model development and included patients from the KPCO and KPNW regions with stage IV or recurrent²⁵ cancer diagnosed between January 1, 2005, and December 31, 2013. We used a second, independent cohort for model validation; this cohort included patients from the KPWA region with stage IV cancer diagnosed between January 1, 2007,

and October 31, 2016. Chemotherapy use was captured during the 180 days after cancer diagnosis.^{26,27} Patients receiving chemotherapy for testicular cancer or chemoradiotherapy for head and neck cancer were excluded, because these treatments can be administered with curative intent, and our analyses were focused on patients receiving palliative-intent therapies. Patients receiving chemotherapy for all other solid tumor malignancies were included in cohort selection.

Candidate Predictors of Hospitalization

Candidate predictors included in model development were selected based on clinical expertise and prior evidence of their association with chemotherapy-related hospitalizations.¹⁶ The 15 candidate predictors were age, systolic blood pressure, diastolic blood pressure, pulse rate, creatinine clearance, sodium, calcium, albumin, total bilirubin, absolute neutrophil count, presence of leukopenia and/or thrombocytopenia, multiagent (*v* single-agent) chemotherapy, hospitalization in the 180 days before chemotherapy initiation (categorized as 0, 1, or ≥ 2), receipt of radiation therapy in the 30 days before chemotherapy initiation, and Charlson comorbidity score (calculated from VDW utilization data, excluding points for advanced cancer).²⁸⁻³⁰ Values for laboratory and vital sign data were the most recent recorded results from on or before the reference day of chemotherapy treatment. For the development cohort, all laboratory and vital sign data were from within a maximum of 30 days before the treatment date. For the validation cohort, laboratory data were from within a maximum of 13 days and vital sign data were from within a maximum of 90 days before the reference treatment date. Creatinine clearance was calculated using the Cockcroft-Gault equation, using serum creatinine and measured patient weight.³¹ Presence of leukopenia and/or thrombocytopenia was defined as a platelet count of fewer than 150,000 cells per microliter and/or a WBC count of fewer than 3,800 cells per microliter; inclusion of this candidate predictor was based on the findings of a prior study.¹⁶

Model Development

After identifying eligible patients for the development cohort, we used multiple imputation (five imputed data sets) to develop complete data sets with no missing predictor values. The entire development cohort was used to model the risk of hospitalization in the first 30 days after chemotherapy initiation. Modeling for the second prediction period included the subset of patients from the development cohort who had a return visit for a second chemotherapy treatment between 15 and 22 days after the initial treatment date ("visit two"), without an intervening hospitalization. We examined several multivariable logistic regression models to calculate the predicted probability of hospitalization within 30 days of the reference chemotherapy treatment date for each of the two prediction

periods (models of two to 15 variables). With the multivariable logistic regression models, we used the least absolute shrinkage and selection operator (LASSO) for variable selection,³² using one of the imputed data sets and employing a tuning parameter that maximized the 10-fold cross-validation estimate of the C-statistic. We then evaluated several unique models with five complete imputed data sets. After ranking models by their C-statistic, we selected three high-performing models for validation (two models from the first prediction period, and one model from the second prediction period). Because many models had similar discrimination characteristics, we used clinical judgment and the parsimony principle to select the final models for validation.

Model Validation

We used a complete case analysis for model validation, excluding patients from the external validation cohort if they had incomplete data for any of the predictor variables for the selected model. We applied the models fitted in the development cohort to the patients in the validation cohort, calculating the predicted 30-day hospitalization risk from each model. We evaluated three main measures of predictive performance: the C-statistic (ie, area under the curve), the observed hospitalization rate of patients above a high-risk cut point (positive predictive value), and the observed hospitalization rate of patients below the high-risk cut point (1-negative predictive value). We prospectively assigned the high-risk cut point as the risk threshold with 80% specificity in the development cohort. We additionally report the sensitivities and specificities observed in the validation cohort around this same cut point. We evaluated model calibration visually, using calibration plots.

RESULTS

The development cohort included 3,606 patients with a median age of 63 years; all patients were retained for model development, with imputation of missing data elements. The validation cohort included 634 evaluable patients with a median age of 64 years; an additional 387 eligible patients were excluded from the validation analysis because of missing data (complete case analysis). Lung cancer and colorectal cancer were the most common malignancies in both cohorts; platinum, taxanes, and fluoropyrimidines were the most common classes of chemotherapy agents. Additional demographic and clinical details of the study cohorts are listed in [Table 1](#), and descriptive statistics for values of candidate predictor variables are listed in [Table 2](#). The risk of hospitalization in the 30 days after chemotherapy day 1 was 14.6% in the development cohort and 14.7% in the validation cohort. The 30-day hospitalization risk after chemotherapy day N (where N is the day of the first treatment visit between day 15 and day 22) was 10.5% in the development cohort and 10.0% in the validation cohort.

TABLE 1. Demographic and Clinical Characteristics of Patients Receiving Chemotherapy for Advanced Cancer

Characteristic	Development Cohort (n = 3,606), No. (%)	Validation Cohort (n = 634), No. (%)
Sex		
Male	1,662 (46.1)	311 (49.1)
Female	1,942 (53.9)	323 (50.9)
Age, years		
Median	63	64
22-49	462 (12.8)	57 (9.0)
50-64	1,537 (42.6)	262 (41.3)
65-74	1,046 (29.0)	199 (31.4)
≥ 75	561 (15.6)	116 (18.3)
Race/ethnicity		
White	2,913 (80.8)	500 (78.9)
Black	118 (3.3)	32 (5.0)
Asian	90 (2.5)	47 (7.4)
Hispanic	217 (6.0)	24 (3.8)
Other or unknown	268 (7.4)	31 (4.9)
Site of primary cancer		
Lung	946 (26.2)	198 (31.2)
Colon or rectum	531 (14.7)	95 (15.0)
Upper GI tract	500 (13.9)	58 (9.1)
Breast	492 (13.6)	38 (6.0)
Other	1,137 (31.5)	245 (38.6)
Cancer stage		
IV	2,469 (68.5)	634 (100)
I-III, with recurrence	1,137 (31.5)	—
Chemotherapy agents*		
Platinum	1,966 (54.5)	420 (66.2)
Taxane	1,384 (38.4)	189 (29.8)
Fluoropyrimidine	926 (25.7)	168 (26.5)
Gemcitabine	409 (11.3)	88 (13.9)
Tyrosine kinase inhibitor	317 (8.8)	37 (5.8)

*Patients may receive chemotherapy agents from multiple classes.

Model development yielded three candidate models for predicting the 30-day hospitalization risk. We selected two candidate models for the day 1 period (the 30 days after day 1) and one model from the day N period (the 30 days after day N). Variables for the selected models are listed in Table 3; the parsimonious day 1 model and the day N model both contained the same two predictor variables: blood albumin level and blood sodium level (both entered as continuous variables).

Model discrimination for the two-variable (parsimonious) day 1 model yielded a C-statistic of 0.69 (95% CI, 0.62 to 0.75) when tested in the validation cohort. After dichotomization by the predefined risk cut point, 38.5% of

patients were assigned to the high-risk group and 61.5% of patients were assigned to the low-risk group. The 30-day hospitalization risk was 24.2% (95% CI, 18.9% to 30.1%) in the high-risk group and 8.7% (95% CI, 6.1% to 12.0%) in the low-risk group (Table 4). Model calibration was satisfactory, as shown in the calibration plot in Figure 1; however, the model seemed to overestimate hospitalization risk at lower levels of predicted risk. The nine-variable day 1 model exhibited modestly improved discrimination and calibration compared with the two-variable model (Table 4; Fig 1), although the larger variable set for this model rendered 83 patients inevaluable because of missing data. The day N prediction model showed moderate to poor discrimination and calibration, with a C-statistic of 0.59 (Table 4; Fig 1).

DISCUSSION

We describe the development and validation of three prediction models for estimating the risk of hospitalization in patients starting chemotherapy for advanced cancer. Two models for predicting hospitalization risk in the 30 days after chemotherapy initiation showed good discrimination and calibration. A third model for predicting hospitalization risk after a subsequent chemotherapy treatment visit was less successful. These models were developed and validated in independent cohorts, and the two validated models can be implemented with use of data that are already routinely collected in the course of clinical care. As examples of the output of the two-variable day 1 parsimonious model, consider two hypothetical patients with metastatic colon cancer, both women in their 60s. For a patient with serum albumin and sodium in the normal range (albumin, 4.0 g/dL; sodium, 140 mmol/L), the model predicts a 30-day hospitalization risk of 12%. For a second patient with a serum albumin level of 3.0 g/dL and a sodium level of 136 mmol/L, the predicted hospitalization risk is 23%. An interactive Web application and a downloadable spreadsheet implementing these models are available online.³³

Because the objective of our analysis was risk stratification rather than biologic inference, we did not explore the mechanisms by which our predictor variables were related to hospitalization risk. However, the predictors identified in our models have good face validity as indicators of physiologic status and prognosis in cancer. Low albumin levels are a recognized marker of both poor nutritional status and systemic inflammation, and albumin levels have been associated with advanced cancer survival in a number of cancer types.^{34,35} Mild to moderate hyponatremia (serum sodium < 136 or 131 mEq/L) has also been associated with adverse survival among patients with cancer.³⁶ Because albumin and sodium are associated with prognosis in cancer, and because they are routinely assessed and documented in EHRs during cancer treatment, these parameters are well suited for use in risk stratification models and clinical decision support tools.

TABLE 2. Descriptive Statistics for Candidate Predictor Variables

Candidate Variable	Derivation Cohort (n = 3,606)		Validation Cohort (n = 1,021)*	
	Median (IQR)†	No. (%) Missing	Median (IQR)†	No. (%) Missing
Age, years	63 (16)	0	64 (14)	0
Systolic BP, mm Hg	125 (23)	85 (2.4)	124.5 (26)	105 (10.3)
Diastolic BP, mm Hg	73 (14)	86 (2.4)		
Pulse rate, Hz	80 (21)	138 (3.8)		
Creatinine, mg/dL‡	0.89 (0.3)	406 (11.2)	0.8 (0.3)	272 (26.6)
Body weight, kg‡	76 (25)	70 (1.9)	74.8 (26)	25 (2.4)
Sodium, mmol/L	139 (5)	868 (24.1)	137 (6)	342 (33.5)
Calcium, mg/dL	9.3 (0.8)	847 (23.4)		
Albumin, g/dL	3.9 (0.8)	1250 (37.7)	3.7 (0.8)	350 (34.3)
Total bilirubin, mg/dL	0.5 (0.4)	361 (10.0)	0.4 (0.2)	346 (33.9)
WBC count, K/μL	7.7 (3.7)	0	8.3 (3.9)	252 (24.7)
ANC, K/μL	5.12 (3.14)	386 (10.7)	5.7 (3.3)	332 (32.5)
Platelet count, K/μL	281 (136)	375 (10.4)	287 (152)	258 (25.3)
Charlson comorbidity score†§		0		
0	1,685 (47)	—		
1	960 (27)	—		
≥ 2	961 (27)	—		
Multiagent chemotherapy†		0		
Yes	2,233 (62)	—		
No	1,373 (38)	—		
No. of hospitalizations in 6 months before chemotherapy†	—	0		0
No hospitalization	2,044 (57)	—	538 (53)	—
1	1,154 (32)	—	358 (35)	—
≥ 2	408 (11)	—	125 (12)	—
Radiation therapy in 30 days before chemotherapy†	—	0	—	0
Yes	199 (6)	—	172 (17)	—
No	3,407 (94)	—	849 (83)	—

NOTE. Candidate variables not retained in any of the selected risk stratification models were not assessed for patients in the validation cohort. Abbreviations: ANC, absolute neutrophil count; BP, blood pressure; mm Hg, millimeters of mercury; IQR, interquartile range.

*All eligible patients, inclusive of patients excluded from complete case validation analyses because of missing data.

†Categorical variables are shown as No. (%) of patients per category.

‡Creatinine clearance was calculated from creatinine, body weight, and sex using the Cockcroft-Gault formula.

§Charlson scores are modified to exclude points for metastatic cancer.

Although the concept of identifying patients at risk for toxicity during cancer treatment is not new, our model is one of few tools to focus on hospitalization risk, rather than chemotherapy-related adverse events more generally. Examples of models for predicting chemotherapy toxicity include the Chemotherapy and Aging Research Group model, which uses 11 input variables to predict the risk of grade-3-or-higher toxicity among adults age 65 years or older (validation C-statistic, 0.65),¹⁵ and the CRASH (Chemotherapy Risk Assessment Scale for High-Age Patients) score, which uses two distinct four-variable models

to predict the risk for hematologic and nonhematologic toxicities (independently) among adults ≥ 70 years of age (validation C-statistic, 0.65 for hematologic toxicity and 0.62 for nonhematologic toxicity).³⁷ Both of these toxicity models have been developed exclusively in elderly populations, and both require input data that are not otherwise collected as part of routine care. The PROACCT (PRediction Of Acute Care use during Cancer Treatment) model is a hospitalization model, using the same outcome of 30-day hospitalization that we report here. That model has been described in abstract form, using 11 variables and

TABLE 3. Logistic Regression Models to Predict Hospitalization Within 30 Days of Chemotherapy Delivery

Variable	OR (95% CI)	β Coefficient	P
Day 1 parsimonious model			
Albumin, g/dL	0.62 (0.51 to 0.77)	-0.472	< .001
Sodium, mmol/L	0.93 (0.90 to 0.95)	-0.075	< .001
Intercept	—	10.392	< .001
Day 1 expanded model			
Low platelets or WBC*	3.04 (1.18 to 7.82)	1.113	.021
Creatinine clearance, mL/min	0.99 (0.99 to 1.00)	-0.006	.001
Albumin, g/dL	0.75 (0.60 to 0.93)	-0.292	.01
ANC, K/ μ L	1.05 (1.02 to 1.08)	0.048	< .001
Sodium, mmol/L	0.94 (0.91 to 0.96)	-0.064	< .001
Total bilirubin, mg/dL	1.07 (1.02 to 1.14)	0.071	.01
Systolic BP, mm Hg	0.99 (0.99 to 1.00)	-0.007	.02
Hospitalizations in 6 months before chemotherapy†	1.21 (1.05 to 1.40)	0.193	.007
Radiation therapy in 30 days before chemotherapy	1.68 (1.17 to 2.40)	0.518	.005
Intercept	—	9.063	< .001
Day N model			
Albumin, g/dL	0.53 (0.38 to 0.73)	-0.635	< .001
Sodium, mmol/L	0.89 (0.85 to 0.93)	-0.114	< .001
Intercept	—	15.867	< .001

NOTE. The day 1 parsimonious and expanded models predict risk of hospitalization in the 30 days after initial chemotherapy treatment (day 1). The day N model predicts risk of hospitalization in the 30 days following a subsequent treatment visit occurring 15 to 22 days after day 1.

Abbreviations: ANC, absolute neutrophil count; BP, blood pressure; mm Hg, millimeters of mercury; OR, odds ratio.

*Platelets fewer than 150,000/ μ L and/or WBC count fewer than 3,800/ μ L.

†Categorized as 0, 1, or ≥ 2 hospitalizations in the prior 180 days.

TABLE 4. Discrimination and Risk Stratification Performance of Models for Classification of Hospitalization Risk Within the Validation Cohort

Parameter	Day 1 Parsimonious Model (n = 634)	Day 1 Expanded Model (n = 551)	Day N Model (n = 441)
C-statistic	0.69	0.71	0.59
95% CI	0.62 to 0.75	0.64 to 0.77	0.49 to 0.69
Threshold for risk stratification, %*	18.0	18.0	14.0
Patients in high-risk group, %	38.5	39.0	32.9
Patients in standard-risk group, %	61.5	61.0	67.1
30-day hospitalization risk for high-risk group, % (PPV)	24.2	25.6	15.2
95% CI	18.9 to 30.1	19.9 to 32.0	9.8 to 22.1
30-day hospitalization risk for low-risk group, % (1-NPV)	8.7	7.4	7.4
95% CI	6.1 to 12.0	4.9 to 10.8	4.7 to 11.0
Sensitivity, %	63	69	50
95% CI	53 to 73	57 to 79	35 to 65
Specificity, %	62	61	67
95% CI	58 to 65	57 to 65	63 to 71

NOTE. The day 1 parsimonious and expanded models predict risk of hospitalization in the 30 days after initial chemotherapy treatment (day 1). The day N model predicts risk of hospitalization in the 30 days following a subsequent treatment visit occurring 15 to 22 days after day 1.

Abbreviations: NPV, negative predictive value; PPV, positive predictive value.

*Risk thresholds associated with approximately 80% specificity in the development cohort.

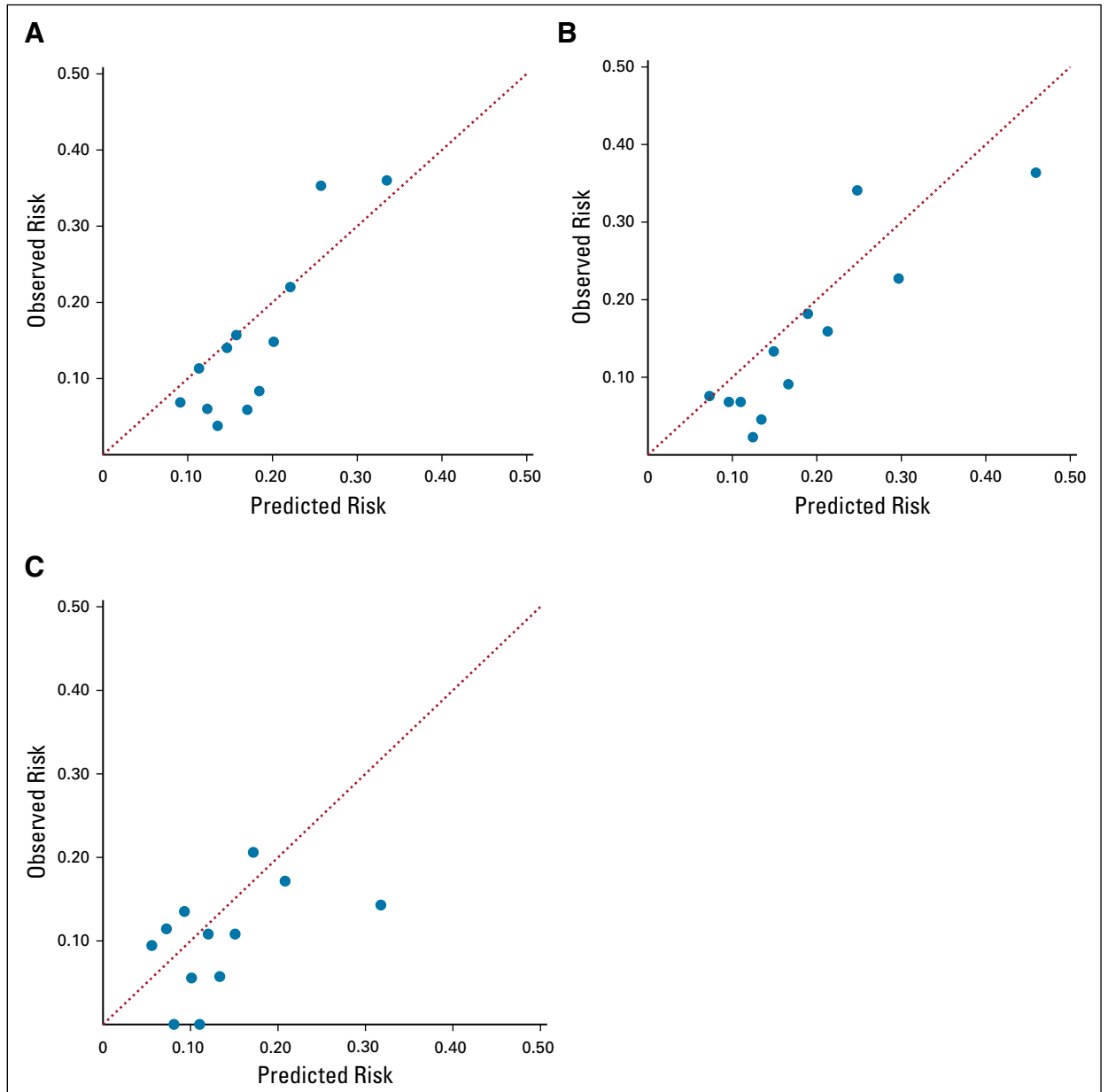


FIG 1. Calibration of models for classification of hospitalization risk within the validation cohort. (A) Day 1 parsimonious model; (B) day 1 expanded model; (C) day N model. Each point represents a nonoverlapping stratum of patients ordered by predicted risk for hospitalization. The diagonal line represents perfect calibration, where observed risk and predicted risk are identical. The day 1 parsimonious and expanded models predict risk of hospitalization in the 30 days after initial chemotherapy treatment (day 1). The day N model predicts risk of hospitalization in the 30 days following a subsequent treatment visit occurring 15 to 22 days after day 1.

yielding a validation C-statistic of 0.65.³⁸ A model to evaluate for risk of hospitalization or emergency department visits during a course of radiation treatment (with or without concurrent chemotherapy) was recently reported, using a machine learning approach.³⁹ That model was developed at a single institution using a split-sample development-validation approach and reported a C-statistic of 0.80. Intriguingly, this model also identifies serum albumin as an important predictor of acute care during cancer treatment.

The models reported here compare favorably with the CRASH and PROACCT models; however, each of these models has different strengths and target populations. Although the use of routinely collected data is an advantage of our approach, collection of patient-reported data is becoming easier and more common in the age of the smartphone. As electronic symptom reporting enters routine practice, it seems likely that risk stratification models will be improved by including patient-reported data. The machine learning model for predicting acute care during

radiation therapy is of considerable interest as well.³⁹ However, caution is warranted in confirming the generalizability of machine learning models in new clinical settings and across time.⁴⁰ Also, a black box problem can occur in machine learning models, where the contribution of individual predictors to a model prediction is opaque. This opacity can pose substantial implementation and interpretation challenges in real-world clinical settings. Regression models are relatively simple, transparent, and understandable and have a long history of use in medical research and practice.

What purpose can be served by models such as these? The most likely use for a model to predict hospitalization risk is to identify patients who may benefit from especially close follow-up after chemotherapy initiation. The reported success of both patient navigator programs¹⁰ and patient-reported outcome interventions^{6,12} has shown that proactive checkups between oncology clinic visits can prevent (or reduce) hospitalizations, presumably through early identification and management of clinical deterioration. Patients with elevated hospitalization risk could also be referred for concurrent palliative care, another intervention with evidence for preventing hospitalizations.⁴¹ By pairing a risk stratification tool with these and other supportive care interventions, the resources of those interventions can be targeted where they are most likely to serve a benefit, enhancing the efficiency of care delivery.

A potential limitation to the generalizability of our findings is that we conducted our study among patients insured through integrated health plans in the United States, as

distinct from patients with fee-for-service health insurance. Indeed, the 30-day hospitalization risk of approximately 15% that was seen in these cohorts is somewhat lower than the hospitalization risk observed in other studies, which has often exceeded 20% over the 30 days after chemotherapy initiation.^{7,38} Therefore, it is likely that the risk threshold for separating high- and low-risk subgroups will need to be reevaluated in distinct clinical settings. Another limitation of our study is that our findings predate the age of immunotherapy. Checkpoint inhibitor immunotherapies are transforming advanced-cancer treatment, and these therapies carry adverse effect profiles distinct from those of traditional chemotherapy agents. Whether our models will retain their risk stratification properties in patients treated with checkpoint inhibitor therapies is unknown. However, cytotoxic chemotherapies remain a key component of treatment of many advanced cancers, including sometimes in combination with checkpoint inhibitors.

Given the limitations noted, we support efforts to validate these models in a range of care delivery settings and patient populations. Validation studies will be facilitated by the relative ease of use of our models, particularly the parsimonious two-variable model. Beyond additional validation efforts, cancer care delivery research will benefit greatly from studies that test the utility of validated risk stratification models for guiding the use of supportive care interventions. Until the findings of risk stratification tools can be shown to translate into tangible benefits for patients, such as fewer hospitalizations or reduced symptom burden, the potential of such tools will remain unfulfilled.

AFFILIATIONS

¹Geisel School of Medicine, Lebanon, NH

²Dana-Farber Cancer Institute, Boston, MA

³Kaiser Permanente Washington Health Research Institute, Seattle, WA

⁴Kaiser Permanente Colorado, Denver, CO

⁵Kaiser Permanente Center for Health Research, Portland, OR

CORRESPONDING AUTHOR

Gabriel A. Brooks, MD, MPH, Geisel School of Medicine, 1 Medical Center Drive, Lebanon, NH 03756; e-mail: gabriel.a.brooks@hitchcock.org.

SUPPORT

Supported by Grant No. U24C171524 from the National Cancer Institute to the Cancer Research Network (Lawrence H. Kushi) and by a Career Development Award from the Cancer Conquer Foundation (G.A.B.).

AUTHOR CONTRIBUTIONS

Conception and design: Gabriel A. Brooks, Alexander R. Menter, Debra P. Ritzwoller, Deborah Schrag

Financial support: Deborah Schrag

Administrative support: Deborah Schrag

Provision of study material or patients: Erin J. Aiello Bowles, Debra P. Ritzwoller, Deborah Schrag

Collection and assembly of data: Gabriel A. Brooks, Erin J. Aiello Bowles, Maureen O'Keefe-Rosetti, Debra P. Ritzwoller, Deborah Schrag

Data analysis and interpretation: Gabriel A. Brooks, Hajime Uno, Erin J. Aiello Bowles, Alexander R. Menter, Anna N.A. Tosteson, Deborah Schrag

Manuscript writing: All authors

Final approval of manuscript: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. 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/cci/author-center.

Gabriel A. Brooks

Consulting or Advisory Role: Abt Associates, CareCentrix

Hajime Uno

Consulting or Advisory Role: Shionogi, ION Pharma

Deborah Schrag

Consulting or Advisory Role: *Journal of the American Medical Association*

Research Funding: Pfizer

Other Relationship: *Journal of the American Medical Association*

No other potential conflicts of interest were reported.

ACKNOWLEDGMENT

Jeff Holzman (Kaiser Permanente Colorado) and Rebecca Ziebell (Kaiser Permanente Washington) provided programming and data analytic support for this analysis.

REFERENCES

- Manzano JG, Luo R, Elting LS, et al: Patterns and predictors of unplanned hospitalization in a population-based cohort of elderly patients with GI cancer. *J Clin Oncol* 32:3527-3533, 2014
- Rocque GB, Barnett AE, Illig LC, et al: Inpatient hospitalization of oncology patients: Are we missing an opportunity for end-of-life care? *J Oncol Pract* 9:51-54, 2013
- Brooks GA, Abrams TA, Meyerhardt JA, et al: Identification of potentially avoidable hospitalizations in patients with GI cancer. *J Clin Oncol* 32:496-503, 2014
- Whitney RL, Bell JF, Tancredi DJ, et al: Hospitalization rates and predictors of rehospitalization among individuals with advanced cancer in the year after diagnosis. *J Clin Oncol* 35:3610-3617, 2017
- Brooks GA, Li L, Uno H, et al: Acute hospital care is the chief driver of regional spending variation in Medicare patients with advanced cancer. *Health Aff (Millwood)* 33:1793-1800, 2014
- Basch E, Deal AM, Dueck AC, et al: Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. *JAMA* 318:197-198, 2017
- Brooks GA, Austin AM, Uno H, et al: Hospitalization and survival of Medicare patients treated with carboplatin plus paclitaxel or pemetrexed for metastatic, nonsquamous, non-small cell lung cancer. *JAMA Netw Open* 1:e183023, 2018
- Kline RM, Bazell C, Smith E, et al: Centers for Medicare and Medicaid Services: Using an episode-based payment model to improve oncology care. *J Oncol Pract* 11:114-116, 2015
- Centers for Medicare and Medicaid Services: Oncology care model. <https://innovation.cms.gov/initiatives/oncology-care/>
- Rocque GB, Pisu M, Jackson BE, et al: Resource use and Medicare costs during lay navigation for geriatric patients with cancer. *JAMA Oncol* 3:817-825, 2017
- Hoverman JR, Neubauer MA, Jameson M, et al: Three-year results of a Medicare advantage cancer management program. *J Oncol Pract* 14:e229-e237, 2018
- 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
- Handley NR, Schuchter LM, Bekelman JE: Best practices for reducing unplanned acute care for patients with cancer. *J Oncol Pract* 14:306-313, 2018
- Hurria A, Togawa K, Mohile SG, et al: Predicting chemotherapy toxicity in older adults with cancer: A prospective multicenter study. *J Clin Oncol* 29:3457-3465, 2011
- 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
- Brooks GA, Kansagra AJ, Rao SR, et al: A clinical prediction model to assess risk for chemotherapy-related hospitalization in patients initiating palliative chemotherapy. *JAMA Oncol* 1:441-447, 2015
- Parikh RB, Kakad M, Bates DW: Integrating predictive analytics into high-value care: The dawn of precision delivery. *JAMA* 315:651-652, 2016
- Moons KG, Altman DG, Reitsma JB, et al: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Ann Intern Med* 162:W1-W73, 2015
- National Cancer Institute: Cancer Research Network. <https://crn.cancer.gov/>
- Chubak J, Ziebell R, Greenlee RT, et al: The Cancer Research Network: A platform for epidemiologic and health services research on cancer prevention, care, and outcomes in large, stable populations. *Cancer Causes Control* 27:1315-1323, 2016
- Ritzwoller DP, Carroll NM, Delate T, et al: Patterns and predictors of first-line chemotherapy use among adults with advanced non-small cell lung cancer in the cancer research network. *Lung Cancer* 78:245-252, 2012
- Ross TR, Ng D, Brown JS, et al: The HMO Research Network Virtual Data Warehouse: A public data model to support collaboration. *EGEMS (Wash DC)* 2:1049, 2014
- North American Association of Central Cancer Registries: NAACCR Strategic Management Plan: 2011-2016. <https://www.naacccr.org/wp-content/uploads/2016/11/NAACCR-Strategic-Management-Plan-2011-2016.pdf>
- Carroll NM, Burniece KM, Holzman J, et al: Algorithm to identify systemic cancer therapy treatment using structured electronic data. *JCO Clin Cancer Inform* 1:1-9, 2017
- Hassett MJ, Ritzwoller DP, Taback N, et al: Validating billing/encounter codes as indicators of lung, colorectal, breast, and prostate cancer recurrence using 2 large contemporary cohorts. *Med Care* 52:e65-e73, 2014
- Aiello Bowles EJ, Tuzzio L, Ritzwoller DP, et al: Accuracy and complexities of using automated clinical data for capturing chemotherapy administrations: Implications for future research. *Med Care* 47:1091-1097, 2009
- Ritzwoller DP, Carroll N, Delate T, et al: Validation of electronic data on chemotherapy and hormone therapy use in HMOs. *Med Care* 51:e67-e73, 2013
- Charlson ME, Pompei P, Ales KL, et al: A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chronic Dis* 40:373-383, 1987
- Deyo RA, Cherkin DC, Ciol MA: Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol* 45:613-619, 1992
- Klabunde CN, Potosky AL, Legler JM, et al: Development of a comorbidity index using physician claims data. *J Clin Epidemiol* 53:1258-1267, 2000
- Cockcroft DW, Gault MH: Prediction of creatinine clearance from serum creatinine. *Nephron* 16:31-41, 1976
- Tibshirani R: Regression shrinkage and selection via the Lasso. *J Roy Stat Soc B Met* 58:267-288, 1996
- GAB Oncology: Risk stratification models. <http://www.gaboncology.com/models/>
- Kuroda D, Sawayama H, Kurashige J, et al: Controlling Nutritional Status (CONUT) score is a prognostic marker for gastric cancer patients after curative resection. *Gastric Cancer* 21:204-212, 2018
- Okada S, Yamazaki S, Kaiga T, et al: Impact of nutritional status in the era of FOLFOX/FIRI-based chemotherapy. *World J Surg Oncol* 15:162, 2017
- Castillo JJ, Glezerman IG, Boklage SH, et al: The occurrence of hyponatremia and its importance as a prognostic factor in a cross-section of cancer patients. *BMC Cancer* 16:564, 2016

37. Extermann M, Boler I, Reich RR, et al: Predicting the risk of chemotherapy toxicity in older patients: The Chemotherapy Risk Assessment Scale for High-Age Patients (CRASH) score. *Cancer* 118:3377-3386, 2012
38. Grant RC, Moineddin R, Yao Z, et al: Predicting acute care use following initiation of systemic therapy for solid tumors. *J Clin Oncol* 36, 2018 (suppl; abstr 6)
39. 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* 2:1-11, 2018
40. Rose S: Machine learning for prediction in electronic health data. *JAMA Netw Open* 1:e181404, 2018
41. Temel JS, Greer JA, Muzikansky A, et al: Early palliative care for patients with metastatic non-small-cell lung cancer. *N Engl J Med* 363:733-742, 2010

