An Administrative Data Algorithm to Predict Inpatient Death at One Tertiary Cancer Center

Elizabeth B. Lamont, MD, MS,[1-3] David W. Bates, MD, MSc, [1,2] Rachelle E. Bernacki, MD, MS, [2,4] and Adam Wright, PhD[1,2]

1 Division of General Internal Medicine, Brigham and Women's Hospital, Boston, MA

2 Department of Medicine, Harvard Medical School, Boston, MA

3 Massachusetts General Hospital Cancer Center, Boston, MA

4 Department of Psychosocial Oncology and Palliative Care, Dana-Farber Cancer Institute, Boston, MA

Corresponding Author:

Elizabeth Lamont, MD, MS
Brigham and Women's Hospital
Division of General Internal Medicine
One Brigham Circle
Boston, MA 02120

Key Words: Advanced Cancer, Death in Hospital, Big Data, Clinical Prediction Models, Advance Care Planning, Medicare Costs in the Late Six Months of Life

Counts: Abstract: 342 (max 350) Body of Paper: 2907 (max 3000)

Pages 25

Tables/Figures: max 5 tables and/or figures (max)

SUPPLEMENTAL APPENDIX A: Tables (2) and Figure (1)

Three Key Points: (1) Questions: While most peoples' preferred place of death (ppd) is their own home, ~25% of Medicare beneficiaries still die in acute care hospitals. We are studying patients who've died in our hospital to isolate risk factors. **(2) Findings:** From 11,614 patients with solid tumors who appeared in a single hospital's cancer registry and died between 1999-2013, we determined several risk factors for death in our hospital. **(3) Meaning**: A risk assessment tool could help physicians to prioritize (ppd discussions (a part of broader advance care planning) for patients at an elevated risk.

Abstract

Importance – Despite almost universal agreement that the end of life is a time during which medical care should be directed at alleviation of symptoms and provided at home if possible, one quarter to one half of age-eligible Medicare beneficiaries still die in acute care hospitals.

Objective – In this single institution study, we evaluated a decedent cohort of patients with histories of solid tumors to determine patient attributes associated with risk of subsequent inpatient death at our institution.

Design – Using retrospective data, we divided the decedent cohort (N=11,614) into training (n=5,807) and testing patient cohorts (n=5,807). Using logistic regression to model inpatient death with the training cohort, we then applied it to the testing cohort.

Setting – A large tertiary care hospital in Boston, MA.

Participants – An inception cohort of all 11,614 patients who had died between 1999-2013 with histories of breast, lung, pancreatic, or prostate cancer in the hospital tumor registry.

Main Outcome and Measure – Inpatient death at our hospital vs. death in any other setting.

Results – Overall, 12% (1,461/11,614) of the decedent cohort died as inpatients at our hospital. In the training cohort, the attributes that portended for terminal hospitalization were primarily prior admissions for management of symptoms stemming from progressive cancer. Admission for hypercalcemia, embolism, ascites, supportive care, and blood transfusions were among the predictors. The model was robust when evaluated using the test cohort. Further the interval between the discharge date from the penultimate admission and the admission date for the terminal had a median value of 48 days and an interquartile range of 13 to 211 days.

Conclusions - We found prior admissions for symptoms of progression of advanced cancer were strong predictors of subsequent terminal hospitalizations, with half of terminal admissions occurring within the seven weeks following discharge from the preceding admission. Most of the predictive conditions are expected sequelae of advanced cancer and manageable at home via palliative care

providers. For patients at high risk for terminal readmission, the ambulatory post-admission follow-up visit may be the last moment to determine prefer place of death if nothing else.

Background

End of life care in the United States is poorly organized. Among Medicare patients who have expressed wishes for symptom-guided treatment in their homes, ~ 25%-50% die in acute care hospitals. [1-9] While it is of concern that advanced care wishes are not being honored, [10,11] the explanations for this are disparate and complicated. Prior work has shown that many physicians struggle with both prognostic accuracy and communication. [12,13] Relatedly, many patients undergoing anti-cancer therapy mistakenly believe the treatment is intended to cure them or extend their lives when their oncologists' intent is exclusively to relieve symptoms. [14,15]

Another financial paradox is that while Medicare pays for many high-cost therapies that offer little symptomatic or survival benefit for patients with advanced solid tumors, it compensates poorly for critically important, ethical time-intensive, and high-quality patient/provider advance care planning conversations and which would ultimately pay for themselves forty times over. [16-18] That is, Medicare shift costs of several thousand dollars per day related to patient care to patients and their families under their current hospice benefit. Medicare pays hospices $126/day to provide full patient care to patients with expected survivals of six months or less and who wish to die in their homes (or nursing homes). One hundred and twenty-six dollars typically covers twice per week nursing home visits; 24-hour telephone availability; emergency visits 24 hours/day; and provision of medications. In comparison, Medicare pays hospitals more than forty times that amount (~$5,000.00) per day for provision of full patient care including pharmaceuticals and physical therapy with round the clock nurses and physicians for the same patients. Families who opt for hospice must fill this expensive daily gap with their own time, working in shifts, finding friends to help, or hire or professional care givers. [19] This is not economically possible for some families and provides another example of healthcare inequalities by income. [20-22]

Fortunately, two significant health policy events have shown a bright light on the many of the problem in end of life care in the US including, perverse incentives: (1) the Institute of Medicine's (IOM) report "Dying in America: Improving Quality and Honoring Individual Preferences Near End of

Life" [23] and (2) an addition to the Affordable Care Act (ACA's) reimbursing high-quality advance care planning discussions between patients and their providers. [24] Both to heed the IOM's call for more practical research aimed at improving end of life care and to support providers informational needs in responding to the ACA policy, we sought to identify factors associated with in-hospital death at our hospital with the future goal of stratifying patient risk in real-time so that physicians may prioritize advance care planning discussions among those patients who appear at highest risk for future in-hospital death and who's preferred place of death is home.

Methods

Data Sources and Cohort Development

The inception cohort consisted of all 11,614 patients with breast cancer, lung cancer, pancreatic exocrine cancer, and prostate cancer who had been diagnosed, evaluated, and/or treated shortly after diagnosis at our hospital (appeared in the hospital cancer registry) and are known to have died between 1/1/1999-12/31/2013 according to the social security administration. We chose these dates because (A) we estimated we needed 15 years of data for adequate power to detect meaningful differences between those patients who died at the hospital vs any other setting and (B) the generally accurate social security administration death dates are at a four-year lag, making 2013 data available in 2017. [25,26] We linked information from the hospital cancer registry to several hospital databases including those holding clinical data in fixed (orders, dates of care, diagnosis) and free text (e.g., discharge summaries, notes) form.

An important methodological issue relates to our not wanting to conflate a general inpatient hospice (GIP) death with a non-GIP death. Through evaluation of the "Encounter file", we could identify 62 individuals who were discharged to "an inpatient hospice facility" and were readmitted the same day to our hospital with a palliative care physician as the attending (a sign that they may have been "admitted" to a GIP bed). For sixty percent of these people, a specific alpha numeric code unique to the hospital appeared during the putative GIP stay. We dropped all 62 patients from the

analytic sample given our high suspicion for GIP and did not rely on the code. This left us with a final analytic sample of 11,614 patients.

Outcome Variable

Inpatient death at our hospital was the key outcome variable. The SSA data does not provide site of death, it provides the date of death. Characterization of the death as occurring while an inpatient at our hospital was determined through evaluation of the *discharge disposition* variable in our hospital's data. Results therefore are applicable to our hospital and our quality improvement efforts, and should not be generalized to other treatment centers or patients.

Predictor Variables/Covariates

Covariates that were included in the initial model included demographics, tumor site and stage at diagnosis; tumor relapse; clinical encounters including hospital admissions, indications for admission, length of stay, and inpatient therapies.

Analytic Cohorts

After the elements of the data (N=11,614) were complete, we divided the it in a random fashion at the patient-level into two equal parts using the STATA command "runiform". This resulted in two numerically equal groups of patients, Group 1 (n=5,807) and Group 2 (n=5,807) From Group 1 patients, we developed the risk model using logistic regression. We then applied the model to Group 2 patients to determine if the risk model was valid beyond the patients from which it was created. We were asking, when the original model was applied other patients (i.e., Group 2 patients) were original attributes still important clinically and statistically?

DATA ANALYSES

Analytic Approach

The logistic regression risk model expresses risk of inpatient death at our hospital as an “odds ratio”, as distinct from a “risk ratio”. It is the quotient resulting from the probability of the event occurring over 1- probability it does not occur (i.e., p/(1-p)). [27,28] Because our outcome variable represented only 12% of the total cohort and can be considered rare the OR will approximate the RR. [29]

Using the training cohort's data, we developed the model evaluating the importance of all covariates listed in Table 1. These covariates were chosen based on prior medical literature and clinical experience. We then evaluated the generalizability of the model by applying it to Group 2 patients (n=5,807), the testing sample. We compared results statistically (e.g., model fit) to those of the Group 1, the training sample. [27,28]

Score Variable

We used the logistic regression results to create a score measuring risk of terminal hospitalization using traditional methods. [30,31] Specifically, we transformed the model's ORs into natural logs (coefficients), then divided all coefficients by the absolute value of the smallest coefficient. The smallest coefficient took on the score value of one as to be expected and ORs of zero had coefficients equal to zero, and score equal to zero. ORs less than one had negative coefficients and score values which we maintained given that an OR less than one is ***protective*** of the outcome.

Power calculations:

The decedent cohort consisted of N=11,614 individuals with histories of breast, lung, pancreas, or prostate cancer. For our training and testing groups (n=5,807). We are adequately powered ($\beta=0.9$) to detect a two-sided ($\alpha=0.25$) two percent difference in the outcome proportion which is anchored at 0.12 (Figure 1). [32]

The Institutional Review Board (IRB) approved the. All analyses were undertaken with STATA ML 14.2, College Station, Texas.

Results

Univariate/Bivariate Results

The analytic cohort (N=11,614) consisted of patients with histories of the following malignancies: breast cancer (n=2,236), lung cancer (n=4,550), pancreatic cancer (n=2,006) or prostate cancer (n=2,822) who had received care at our hospital at the very least surrounding the time of their initial diagnosis and first treatment and subsequently died between 1999-2013. The median age at the time of cancer diagnosis was 69. Fewer than 10% of patients were below 50 years of age at diagnosis. Only 36% (4,187/11,614) of the cohort patients were admitted at our hospital during the observation period, 30 days following the cancer diagnosis during the years of interest (i.e., 1999-2013).

Eighty six percent of the sample were white (n= 10,015), approximately two and one-half percent were African American (n=266), and the next most frequent race/ethnicities were Asian (1.2%, n=139) and Hispanic (1.2%, n=133). Most of the sample had been either married (56%) or widowed (18%). While there was no direct measure of socioeconomic status, the payer variable was informative with 313 patients (2.7%) receiving Free Care, Mass Health, or Medicaid as the entirety or part of their insurance; 2,356 patients (20%) with Medicare insurance as the entirety or part of their insurance; and 8,945 (77%) with private insurance as the entirety or part of their insurance.

One third were known to be free of cancer and 66% were known to have relapsed or have metastatic cancer at diagnosis. The time from cancer diagnosis to death was widely distributed, ranging from days to 43 years. While the entire sample was known to have died, only 1,461 (12%) of the cohort were known to have died at our hospital.

To calculate hospital health care use, we removed the final hospitalization from the cohort for all patients, agnostic to outcome status, before evaluating for associations between clinical care and inpatient death. We were interested in characteristics that might portend for a terminal hospitalization but not be derived from it as we are hoping these analytics will prevent terminal hospitalizations. Looking at the date of diagnosis plus thirty days, we found that 43% of patients (n= 4,922) had

hospitalizations during this post-diagnosis period and the median length of stay was five days, with an interquartile range of three days (25%) and 8 days (75%). We found that nearly 60% of patients had been admitted in the two years prior to their cancer diagnosis.

Among the important differences between those who died in our hospital and those who died elsewhere were the duration of oncologic illness, with those patients who died as inpatients at our hospital having had a shorter median duration of illness compared to those who died in any other setting. Additionally, in-hospital death varied by tumor site with nine percent of pancreatic patients dying at our hospital vs 17% of lung cancer patients, 12% breast and 11% prostate cancer patients ($p<0.01$). Table 1 shows all comparisons evaluated.

Results from the Mann-Whitney the rank-sum test for unpaired data was consistent with the null hypothesis that the two groups of patients studied (testing (Group I) and training (Group 2) sample) were drawn from the same population. [33]

For Group I patients who were observed to be hospitalized during our study period (n=2088), the median time from discharge from their penultimate hospitalization to death at our hospital during their terminal readmission was 142 days. The interquartile range was 34 days and 528 days for the patients in group 1 who were hospitalized during the observation period.

Logistic Regression –

Split Sample Training and Testing Samples Without Replacement

Our training model has seven terms (plus the four sites of primary cancer). Table 2 contains the results of our final model that analyzed data from Group 1 patients (i.e., the training cohort). Table 2 also shows how the model performed with an entirely new group of patients (i.e., Group 2 – testing cohort). When the out of sample group was studied with the training model, seven attributes in the in the test set appeared very similar in direction and magnitude to the training set -- though not all were statistically significant. Additionally, the diagnosis of pancreatic cancer was positive correlated with

inpatient death, while both prostate and lung cancer were apparently protective of inpatient death. In the testing model, the directionality of the seven factors remained clinically important and most statistically significant. As shown in Table 3 we were unable to reject the null hypothesis that the models are different. Table 4 shows the sensitivity (3.5%) and specificity (99.3%) as well as the positive predictive value (40%), negative predictive value (87.8%). Figure 2 shows overlapping area under the receiver operator curves for Group 1 and Group 2, each with values of 0.76 consistent with good fit in in the training model.

Table 2 show the following attributes as strongly associated with subsequent terminal admission to our hospital: (1) for red blood cell transfusion (OR 30.5, 95% CI: 2.56-3.64), (2) through the emergency room (OR 2.98, 95% CI: 1.51, 11.9); (3) for symptom management (OR 2.05 95% CI: 0.57-0.75); (4) for pancreatic cancer management OR 1.56, 95% CI:1.09-1.73; (5) five or more times since cancer diagnosis; (6) for ascites (OR 12.01, 95% CI: 1.11-129.77, coefficient from test model as training model had all positive observations associated with inpatient death and therefore dropped out of the model); (7) for pulmonary or cerebral embolism (OR 2.34, 95% CI:1.24-4.43), and (8) for hypercalcemia. One variable, prostate cancer had an odds ratio that suggested it was protective against subsequent terminal admission (OR 0.74, 95% CI: 0.57-0.75).

Risk Score

We bundled risk scores into categories with similar risks of subsequent terminal admissions at our hospital. We observed a monotonic association with rising risk category number (as well as raw risk score) and risk of terminal admission (Table A2). The evaluate whether the training and testing sample appeared to be A two-sample student's t-test applied to the training and testing samples revealed that the median values of the groups' risk scores were statistically indistinguishable (m1 3.0, m2 3.0, p=0.36, SD1 4.0 SD2 4.0, p=0.36) However, the students t-test did a good job at distinguishing those with a terminal hospitalization in our hospital on the horizon vs. those who did not based on either continuous risk score and categorical terminal hospitalization score. A chi2 test of

proportions revealed a shift toward higher score category among those who subsequently were admitted to our hospital for a terminal hospitalization. Additionally, the Kruskal-Wallis equality-of-populations rank test (i.e., null hypothesis was that within each of the three risk categories the rate of observed subsequent terminal admissions were equal) was associated with a chi2 value of 338.890 with two degrees of freedom where $p<0.001$ allowing us to reject the null hypothesis. The ROC curves in Figure 2 support model fit as well.

Discussion

Among age eligible Medicare patients who make their wishes known that their preferred place of death is their home, 25-40% are still dying in acute care hospitals. [10,34,35] Given that trying to support Medicare is straining our country economically and is fully expected to lead to issues of parity, this mis-assortment of location of death is both a morally and monetarily tragic irony. Not only are individuals unable to realize a dying wish, but because of that our nation needlessly becomes much further in debt per year from this as hospitalizations are the single most expensive component of Medicare. [36,37] If somehow, we could help patients who truly do not want to be hospitalized to be where they'd like to be at the end of life, we might all benefit by (1) catalyzing patient autonomy at the end of life and (2) saving CMS and by extension the country an enormous amount of money per year.

We use large and disparate data to identify patient phenotypes that may be associated with risk for death in an acute care hospital or "terminal hospitalization". [38] With an accurate model, a next step would include sharing results with providers (and potentially patients) – opening an avenue for patients and their providers to begin to discuss advance care plans. Arguably, until then, admissions with the primary indications of "symptom control" in newly diagnosed patients with historically rapidly progressive malignancy should perhaps flip a switch, either metaphorical or real as such patients are showing that their disease is progressing given their need for hospital admission for symptom management. For such patients, a steady state of their disease is not yet estimable given the newness of their diagnosis, particularly if they are stage four (i.e., metastatic at diagnosis) and these clinical clues may supplant the need to know the steady state as there may not be one. These

results are generalizable to other breast, lung, pancreatic, and prostate cancer patients at other urban NCI-designated cancer centers during the period of interest. We must be care not to overstep the external validity beyond this group.

Our results appear to show two anomalies which may be a function of cancer referral center bias: the transfusion finding and the low terminal hospitalization rate (12%). Patients often travel beyond their immediate neighborhood and medical catchment area to receive care from NCI designated cancer centers. NCI cancer centers have well stocked blood banks and frequently transfuse patients to help with symptoms or stabilize patients following bleeds due to progressive cancer (e.g., vessel erosion, procedures). Once their cancers have progressed, they may decide to receive care closer to home in a local hospital, nursing home, hospice, or at home.

This research has limitations. While we were cognizant of the risk of conflating GIP deaths with non-GIP deaths, our ability to correctly assign GIP patients used a reason-based algorithm and was not unambiguously documented in the data we reviewed. Additionally, it may be that many of the other patients who did not die at our hospital may have died at other hospitals closer to home.

Conclusion

We found 12% (1,461/11,614) of patients with histories of breast, pancreas, prostate, and lung cancer who received at least initial care at our hospital subsequently expired there as inpatients over a 15-year period. Using disparate hospital data sources, we found a cluster of patient attributes most closely associated with progressive cancer portended for subsequent terminal admissions. Future work will evaluate the extent to which these putative risk factors define patients' risk a priori. If they do, we should all focus on how to expeditiously relay risk information to patients and their oncologists to be certain to honor patients' end of life wishes including preferred place of death.

Table 1. Characteristics of Patients with Histories of Cancer Who Died, Either in Our Hospital or Not in Our Hospital, N= 11,614

Attribute	Inpatient Death + (N=1,409)	Inpatient Death – (N=10,205	p-value
Age at Diagnosis	66.4	67.1	p<0.01
Age 65 Plus	0.71	0.78	p<0.01
Male	0.55	0.54	p=0.75
Race			p<0.01
White	0.88	0.86	
AA	0.03	0.02	
Hispanic	0.01	0.01	
Asian	0.02	0.01	
Native American	0.00	<0.01	
Other	<0.01	<0.01	
Not Reported	0.04	0.09	
Refused	<0.01	<0.01	
Marital Status			p<0.01
Married	0.55	0.57	
Divorced	0.09	0.06	
Widowed	0.17	0.18	
Separated	0.01	0.01	
Single	0.15	0.13	
Unknown/Other	0.03	0.05	
Payor			p<0.01
Mass Health/Free Care/ Medicaid	0.04	0.03	
Medicare	0.19	0.21	
Private Insurance	0.77	0.77	
Tumor Site			p<0.01
Breast	0.17	0.19	
Lung	0.51	0.38	
Pancreas	0.15	0.18	
Prostate	0.17	0.25	
Tumor Stage			p<0.01
0	<0.01	<0.01	
1	0.16	0.21	
2	0.22	0.28	
3	0.20	0.19	
4	0.41	0.29	
Hospitalizations Predating Cancer Diagnosis	0.08	0.46	p<0.01
Embolism	2.40	0.72	p<0.01
RBC Transfusion	0.64	0.29	p<0.01
Preterminal Admission for Supportive Care	0.21	0.03	p<0.01
Ascites	0.34	0.03	p<0.01
Hypercalcemia	0.04	<0.01	p<0.01
Admit from ER	0.18	0.08	p<0.01
>=5 Prior Admissions	0.29	0.09	p<0.01

Legend: Description of cohort according to whether they experienced a terminal hospitalization at our hospital.

Table 2. Predictors of In-Hospital Death in a Training Decedent Cohort with Cancer, n=5807

	Training	n=5807	
Patient Predictor	**OR**	**95%**	**CI**
Clinical Variables			
Admission for Ascites	6.21	1.11	129.77
Admitted via ED	1.95	1.54	2.46
Admission for Embolism (cerebral or pulmonary)	2.41	1.24	4.43
Admission for Hypercalcemia	6.34	0.79	51.0
Admission for Symptom Management	1.87	1.44	2.41
Five or More Admissions Since Cancer Diagnosis	2.31	1.88	2.83
RBC Transfusions	3.08	2.58	3,67
Tumor Site			
Breast Cancer	1.00	Referent	1.00
Lung Cancer	0.82	061	1.11
Pancreatic Cancer	1.41	1.11	1.77
Prostate Cancer	0.74	0.56	0.98
Constant	0.06	0.05	0.07

Legend**:** The table above shows the magnitude and strength of associations between eleven patient factors and the risk of death in-hospital. In this training model, most predictors are statistically and apparently clinically important predictors of death in-hospital. OR= odds ratio; ED=Emergency Department; CI=Confidence Interval.

Table 3. Comparison in Likelihood Ratios Between the Training and Testing Samples

Likelihood-ratio test
(Assumption: model_2 nested in model_1)

LR chi2(1) = -298.41
Prob > chi2 = 1.0000

Akaike's information criterion and Bayesian information criterion

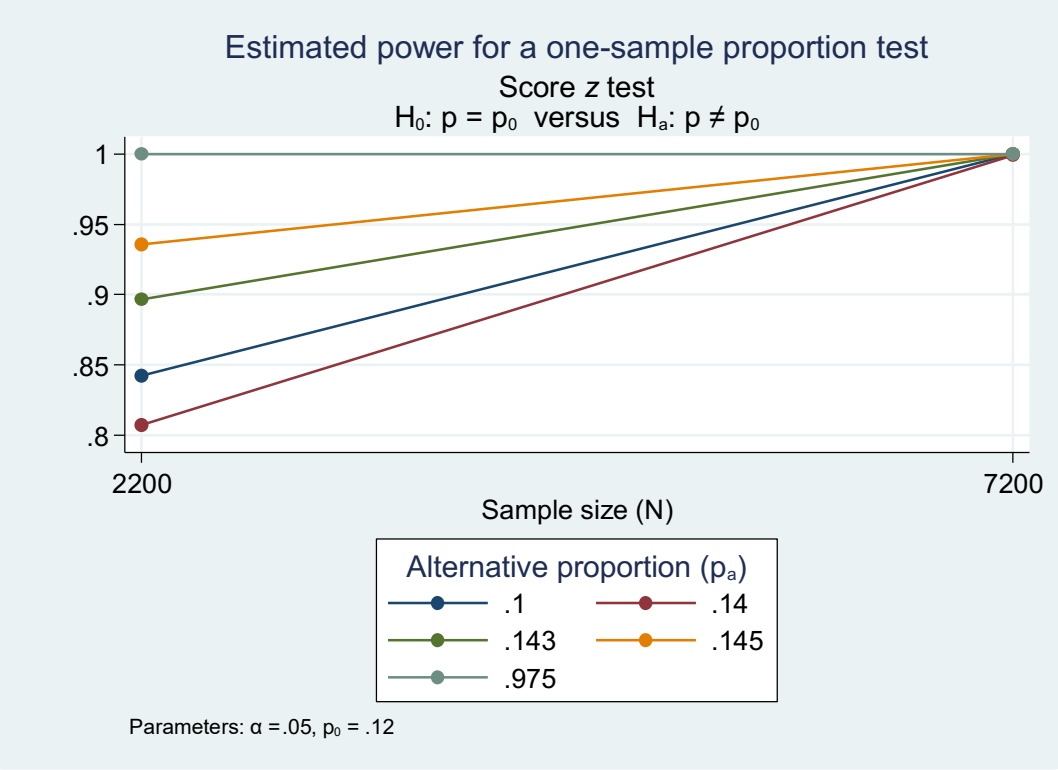
Model	Obs	ll (null)	ll (model)	df	AIC	BIC
Model_1	5,802	-2120.436	-1868.219	10	3756.438	3823.098
Model_2	5,807	-2260.965	-2017.422	11	4056.845	4130.18

Note: N=Obs used in calculating BIC; see [R] BIC note.

Name	Command	Depvar	npar	Title
Model_1	logistic	expired	11	Model Developed using Training Patients
Model_2	logistic	expired	11	Model Evaluated on Testing Patient Sample

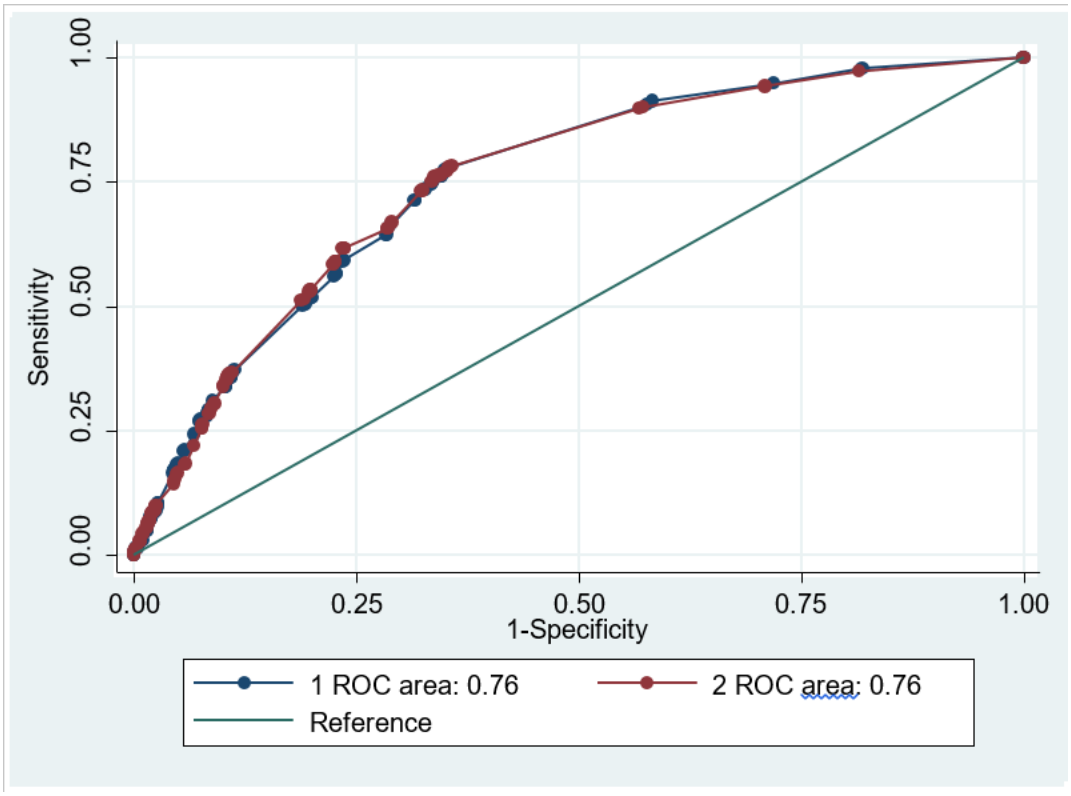
Legend: A single model was developed with 5807 patients termed the training sample. Then the model was evaluated for external validity by applying it to a separate sample of 5807 patients. Here we compare the model containing training patients to the same model that contains testi patients. The likelihood-ratio test suggests we cannot reject the null hypothesis that the nested model 2 is true.

Figure 1. Power Curve for Training and Testing Sample Sizes, n=5,807



Legend: Power curve for the training and testing samples of 5807 patients with lower limit approximating half the smallest tumor subset (patients with pancreatic cancer, n=1000) and half the upper limit approximating the aggregated sample of patients with one of four tumors (N=5,500)

Figure 2. Overlapping ROC Curves from the Same Prediction Model: Training (n=5807) vs Testing (n=5807) Samples



Legend: The areas under the curve for the logistic regression modeling the odds of terminal admission conditional on pre-terminal admission covariates. The curves are nearly superimposed on each other and along with analyses suggest that the training model and the test model are measuring similar outcomes through similar predictors.

Table A1. Regression-Based Score Estimating Risk of Inpatient Death in Training Sample

Patient Predictor	**Training**	**n=5807**	
	OR	**Coefficient**	**Score**
Clinical Variables			
Admission for Ascites	6.21	1.83	9.0
Admission for Embolism (cerebral or pulmonary)	2.41	0.88	4.5
Admission for Hypercalcemia	6.34	1.85	9.0
Admission for Symptom Management	1.87	0.63	3.0
Admit for RBC Transfusions	3.08	1.12	5.3
Admitted via ED	1.95	0.67	3.5
Five or More Admissions Since Cancer Diagnosis	2.31	0.84	4.0
Tumor Site			
Breast Cancer	1.00	0	0
Lung Cancer	0.82	-0.20	-1
Pancreatic Cancer	1.41	0.34	1.7
Prostate Cancer	0.74	-0.30	-1.5

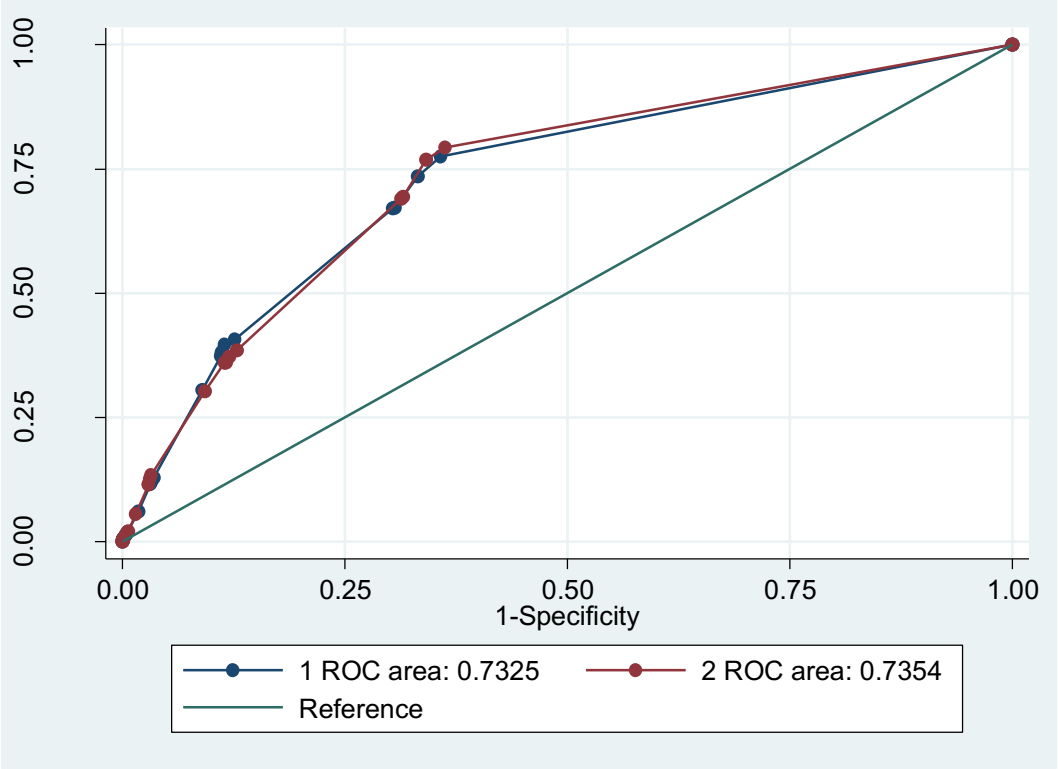
Legend: Table contains all variables from the training model estimating in-hospital death at one hospital. The score values were calculated by taking the natural log of the ORs which yields coefficients. We then divided each coefficient by the absolute value of the smallest coefficient to obain each variable's score. Lung cancer had the smallest coefficient, 0.20.Negative coeficients were associated with negative partital scores as they correspond to ORs less than one which are associated with ***lesser risk*** of admission for subsequent terminal hospitalization. and the positive value on their coefficient and score in Table A1 are consistent with this. However, most ORs were greater than one which means that they are associated a ***higer risk*** of admission for terrminal hospitalizations and the positive value on their coefficient and score in Table A1 are consistent with this.

Table A2. Risk Score Predicting Subsequent Terminal Admission to Our Hospital Hospitalized Patients from Group 1 (n=5807) and Group 2 (n=5807)

Groupings of Raw Prediction Score	Risk Scores Transformed into Risk Categories	Observed Risk of Subsequent Inpatient Death by Risk Score (proportion)	Hospitalized Group1 N	Observed Risk of Subsequent Inpatient Death by Risk Score (proportion)	Hospitalized Group2 N
0	1	0.05	3,429	0.05	3,383
1-5	2	0.22	2,107	0.23	2160
>5	3	0.34	271	0.38	264
Total			5,807		5,807

Legend: The chart shows the manner in which we aggregated raw patient-level risk scores and the observed risk of the patient's next hospitalization being a terminal admission. Additionally, data and analyses are divided by group (i.e., training and testing). On visual inspection, the training and testing groups have similar rates of terminal hospitalization by transformed score category. Further, as described in the prose of the paper in results, students' t-tests with raw score and transformed categorical score, correlation, and chi2 test of proportion all suggest that they are from indistinguishable populations – supporting the lack of over fitting described in formal-goodness-of-fit analyses.

Figure 1A. Graphic Relationship Between Risk Score and Observed Risk of Death



Legend: The areas under the curve for the regression-based predictive risk score category estimating subsequent terminal hospitalization at one hospital vs. any other place of death for training and testing cohorts. Results visually are numerically close and the curves appear similar, both consistent with good model fit.

REFERENCES

1. Steinhauser KE, Christakis NA, Clipp EC, McNeilly M, McIntyre L, Tulsky JA. Factors considered important at the end of life by patients, families, physicians, and other care providers. *JAMA* 2000; 284(19): 2476-2482.

2. Steinhauser KE, Clipp EC, McNeilly M, Christakis NA, McIntyre LM, Tulsky JA. In search of a good death: observations of patients, families, and providers. *Annals of Internal Medicine* 2000; 132(10): 825-832.

3. Pritchard, R. S., Fisher, E. S., Teno, J. M., Sharp, S. M., Reding, D. J., Knaus, W. A., Wennberg, J. E., Lynn, J. and SUPPORT Investigators. Influence of Patient Preferences and Local Health System Characteristics on the Place of Death. 1998 *Journal of the American Geriatrics Society*, 46: 1242–1250.

4. Tang ST. When death is imminent: Where terminally ill patients with cancer prefer to die and why. *Cancer Nurs* 26(3):245–51. 2003.

5, Cohen, J et al. "Differences in Place of Death between Lung Cancer and COPD Patients: A 14-Country Study Using Death Certificate Data." *NPJ Primary Care Respiratory Medicine* 27 (2017): 14. *PMC*.

6. Gott M, Seymour J, Bellamy G, Clark D, Ahmedzai S. Older people's views about home as a place of care at the end of life*. Palliative Medicine* 2004; 18(5): 460-467.

7. Brazil K, Howell D, Bedard M, Krueger P, and Heidebrecht C. Preferences for place of care and place of death among informal caregivers of the terminally ill. *Palliative Medicine* 2005; 19: 492-499.

8. Hall MJ, Levant S, DeFrances CJ. Trends in inpatient hospital deaths: National Hospital Discharge Survey, 2000–2010. NCHS data brief, no 118. Hyattsville, MD: National Center for Health Statistics. 2013.

9. Teno JM, Gozalo PL, Bynum JP, Leland NE, Miller SC, Morden NE, Scupp T, Goodman DC, Mor V.Change in end-ˇof-ˇlife care for Medicare beneficiaries: site of death, place of care, and health care transitions in 2000, 2005, and 2009*. JAMA* 2013; 309(5):470-7.

10. SUPPORT Principle Investigators. A controlled trial to improve care for seriously ill hospitalized patients. *JAMA* 1995; 274: 1591-1598

11.Gorina Y, Pratt LA, Kramarow EA, Elgaddal N. Hospitalization, readmission, and death experience of noninstitutionalized Medicare fee-for-service beneficiaries aged 65 and over. National health statistics reports; no 84. Hyattsville, MD: National Center for Health Statistics. 2015.

12. Christakis NA, Lamont EB. Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort study. *British Medical Journal* 2000;320(7233):469-72.

13. Lamont EB, Christakis NA. Prognostic disclosure to patients with cancer near the end of life. *Annals of Internal Medicine* 2001;134(12):1096-105.

14. Lennes IT, Temel JS, Hoedt C, Meilleur A, Lamont EB. Predictors of newly diagnosed cancer patients' understanding of the goals of their care at initiation of chemotherapy. *Cancer* 2013; 119(3): 691-9.

15. Weeks JC, Catalano PJ, Cronin A, Finkelman MD, Mack JW, Keating NL, et al. Patients' expectations about effects of chemotherapy for advanced cancer. *N Engl J Med*. 2012;367:1616–25.

16. Keary S1, Moorman SM2.Patient-physician end-of-life discussions in the routine care of Medicare beneficiaries. *J Aging Health*. 2015;27(6):983-1002

17. Auriemma CL, Chen L, Olorunnisola M, Delman A, Nguyen CA2, Cooney E, Gabler NB, Halpern SD. Public opinion regarding financial incentives to engage in Advance Care Planning and complete Advance Directives. *Am J Hosp Palliat Care*. 2016 1:104.

18. Bradley CJ, Yabroff KR, Warren JL, Zeruto C, Chawla N, Lamont EB. Trends in the treatment of metastatic colon and rectal cancer in elderly patients. *Medical Care.* 2016; 54:490-7.

19. McArtney JI, Broom A, Kirby E, Good P, Wootton J, and Adams J. Locating care at the end of life: burden, vulnerability, and the practical accomplishment of dying. *Sociology of Health and Illness*. 2015. 38(3):479.

20 Barclay J, Kuchibhatala M, Tulsky JA, Johnson KS. Association of hospice patients' income and care level with place of death. *JAMA internal medicine*. 2013;173(6):450-456.

21. Gallo WT, Baker MJ, Bradley EH. Factors associated with home versus institutional death among cancer patients in Connecticut. *J Am Geriatr Soc*. 2001 Jun;49(6):771-7.

22. Gruneir A, Mor V, Weitzen S, Truchil R, Teno J, Roy J. Where people die: a multilevel approach to understanding influences on site of death in America. *Med Care Res Rev*. 2007 Aug;64(4):351-78

23. Institute of Medicine of the National Academies. Dying in America: Improving Quality and Honoring Individual Preferences Near End of Life. The National Academy Press. Washington, DC. 2015.

24. **https://www.jointcommission.org/assets/1/23/Quick_Safety_Issue_Fifteen_July_20151.PDF**, accessed last 06/24/2017.

25. Cowper DC, Kubal JD, Maynard C, Hynes DM. A primer and comparative review of major US mortality databases. *Ann Epidemiol*. 2002 Oct;12(7):462-8.

26. Lash TL, Silliman RA. A comparison of the National Death Index and Social Security Administration databases to ascertain vital status. *Epidemiology*. 2001 12:259-61.

27. Hosmer, DW, Lemeshow, S. Applied Logistic Regression. Second ed. New York: Wiley 2000.

28. Breslow NE, Day NE. Statistical Methods in Cancer Research, Lyon, France: International Agency for Research on Cancer; 1980.

29. Cummings P. The Relative Merits of Risk Ratios and Odds Ratios. *Arch Pediatr Adolesc Med.*2009;163(5):438-445.

30. Ture, M, Kurt, I, Kurum, T, et al. Comparing classification techniques predicting essential hypertension. Expert Systems with Applications. 2005; 29: 583-588.

31. Bates DW, Cook EF, Goldman L, Lee TH. Predicting Bacteremia in Hospitalized Patients: A Prospectively Validated Model. *Ann Intern Med*. 1990; 113: 495-500.

32. StataCorp. 2013. Chapter Power and Sample Size. Stata: Release 14. Statistical Software. College Station, TX: StataCorp LP.

33. Mann HB, Whitney DR. On a test whether one of two random variables is stochastically larger than the other. Annals of Mathematical Statistics. 1947;18: 50-60.

34. Tang ST, McCorkle R. Determinants of place of death for terminal cancer patients. Cancer Invest2001;19(2):165–180**.**

35. Groth-Juncker A, McCusker J. Where do elderly patients prefer to die? Place of death and patient characteristics of 100 elderly patients under the care of a home health care team. J Am Geriatr Soc 1983;31(8):457–461.

3**6**. Yabroff RK, Lamont EB, Mariotto A, Warren JL, Toper M, Meekins A, Brown ML. Cost of care for elderly cancer patients in the United States. Journal of the National Cancer Institute 2008; 100:1-12.

37. Warren JL, Yabroff KR, Meekins A, Topor M, Lamont EB, Brown ML. Evaluation of trends in the cost of initial cancer treatment. Journal of the National Cancer Institute 2008;100(12):888-97.

38. Bates D, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Aff. 2014; 33: 71123-1131.