BMJ Open

Development and validation of 15-month mortality prediction models: a retrospective observational comparison of machine-learning techniques in a national sample of Medicare recipients

Gregory D Berg,[1] Virginia F Gurley[2]

¹Analytics, AxisPoint Health, Westminster, Colorado, USA
²Medical Affairs, AxisPoint Health, Westminster, Colorado, USA

Correspondence to
Dr Gregory D Berg;
gregory.d.berg@icloud.com

ABSTRACT

Objective The objective is to develop and validate a predictive model for 15-month mortality using a random sample of community-dwelling Medicare beneficiaries.

Data source The Centres for Medicare & Medicaid Services' Limited Data Set files containing the five per cent samples for 2014 and 2015.

Participants The data analysed contains de-identified administrative claims information at the beneficiary level, including diagnoses, procedures and demographics for 2.7 million beneficiaries.

Setting US national sample of Medicare beneficiaries.

Study design Eleven different models were used to predict 15-month mortality risk: logistic regression (using both stepwise and least absolute shrinkage and selection operator (LASSO) selection of variables as well as models using an age gender baseline, Charlson scores, Charlson conditions, Elixhauser conditions and all variables), naïve Bayes, decision tree with adaptive boosting, neural network and support vector machines (SVMs) validated by simple cross validation. Updated Charlson score weights were generated from the predictive model using only Charlson conditions.

Primary outcome measure C-statistic.

Results The c-statistics was 0.696 for the naïve Bayes model and 0.762 for the decision tree model. For models that used the Charlson score or the Charlson variables the c-statistic was 0.713 and 0.726, respectively, similar to the model using Elixhauser conditions of 0.734.

The c-statistic for the SVM model was 0.788 while the four models that performed the best were the logistic regression using all variables, logistic regression after selection of variables by the LASSO method, the logistic regression using a stepwise selection of variables and the neural network with c-statistics of 0.798, 0.798, 0.797 and 0.795, respectively.

Conclusions Improved means for identifying individuals in the last 15 months of life is needed to improve the patient experience of care and reducing the per capita cost of healthcare. This study developed and validated a predictive model for 15-month mortality with higher generalisability than previous administrative claims-based studies.

Strengths and limitations of this study

► The data contains over 2.8 million Centres for Medicare & Medicaid Services (CMS) beneficiaries which makes it the largest sample of data to produce an end-of-life predictive model.

► The data included the entire five per cent sample of the US Medicare population, which increases its generalisability by including people under the age of 65, people enrolled in an health maintenance organisation and dually eligible beneficiaries.

► Various methods including machine learning algorithms are tested including naïve Bayes, decision tree with adaptive boosting, logistic regression with least absolute shrinkage and selection operator selection of variables, neural networks and support vector machines.

► The CMS data did lack pharmacy claims which limits the model's ability to select drug information and interactions.

INTRODUCTION
Background

Despite 80% of people stating they would prefer to die at home, only 20% actually do[1]. Increasing the delivery of advanced care planning and palliative care to individuals nearing the last year of life is likely to result in greater adherence to patient and caregiver treatment and setting of care preferences, reduced use of low-value services and lower healthcare costs.[2-4] The scope of these missed opportunities is significant in that there were over 2.7 million deaths in the USA in 2015 which represented just under 1% of the US population and an age-adjusted death rate of 733.1 per 100 000.[5] At the same time, only one in three US adults having any type of advance directive for end-of-life care[6] and as a result, many individuals receive unwanted life-prolonging services in care settings that

differ from the intensity and setting of care they wish to receive[7 8].

Analytics-based indicators identifying patients that are likely to be nearing the last year of life hold the potential to help clinicians increase the delivery of advanced care planning, palliative and end-of-life care. Accurate predictive models using machine learning techniques is one such tool for providing advanced care planning indicators to aid clinicians' in the identification of patients likely to benefit from advanced care planning. For example, in recognition of the benefits of advanced care planning, the state of California Senate Bill 1004 requires the Department of Healthcare Services to 'establish standards and provide technical assistance for Medi-Cal managed care plans to ensure delivery of palliative care services.'[9 10] Such technical assistance may take the form of a predictive model-derived indicator placed in a patient's medical record alerting clinicians when a person is likely to benefit from initiation of advanced care planning services.

Predictive models and their translation into a patient-level indicator in the medical record are meant to support professional clinical judgement and trigger consideration of when to initiate a shared decision-making process involving treating clinicians, the patient and caregivers.[11]

In addition to helping maximise the quality of life for end-of-life people, predictive models may also be used to identify gaps between risk adjustment and higher probability of entry into a costly spending period. Payers know that end-of-life care is often expensive, as evidenced by findings that the last year of life accounts for almost 28% of total Medicare spending - amounting to over $50 000 total payments per person per year for decedents.[12] Higher spending during the last week of life was found to be associated with a decreased quality of death.[13] Spending patterns in the last year of life show that costs do not skyrocket at the end of life but, instead, can be described by four distinct cost patterns: high persistent, moderate persistent, progressive and late rise. Almost half of people were classified as high persistent,[14] indicating that payers and providers should not target just the last few months of life, but rather attend to missed opportunities for advanced care planning and cost avoidance when patients are nearing the last year of life. This is precisely where medical record indicators derived from predictive model algorithms can help clinicians identify those who can most benefit from advanced care planning services.

Previous algorithms

The most widely known algorithms that use administrative medical claims data to predict mortality are from Charlson et al[15] in 1987, which used a weighted sum of the presence or absence of one of 19 conditions with weights updated in 1993 by Romano et al[16] and Elixhauser et al[17] in 1988, which used 30 conditions.

A more recent review of the literature for all-cause mortality was conducted by Yourman et al [18] in 2012, who found six mortality predictive models for community-dwelling older adults. Although many studies use surveys of either providers or patients instead of administrative claims, which are more expensive to administer compared with administrative claims, one study by Gagne et al[19] in 2013 did use administrative claims. This study combined the conditions from the Romano et al[16] implementation of Charlson index and the van Walraven et al[20] implementation of the Elixhauser system into one model to predict 1 year mortality. The Gagne results had modest improvements over the Charlson and Elixhauser models run separately. When run separately, the c-statistic using the Charlson index was 0.778 and using the Elixhauser system was 0.772. However, combined the c-statistic was 0.788, showing improvements when combining models.

Other examples using administrative Medicare claims include one by Hamlet et al[21] in 2010, to predict 1 year mortality for people with heart failure and/or diabetes and were all high risk as part of a Medicare Health Support pilot, instead of a random sample of all Medicare beneficiaries. Also, Schneeweiss et al[22] in 2003 had a restricted Medicare population from New Jersey and Pennsylvania limiting its generalisability but did perform well with c-statistics ranging between 0.70 and 0.80.

Beyond the epidemiology and health services research perspective, mortality prediction in the computer science community includes machine learning models based on data from hospitals, intensive care units (ICUs), unstructured text data from ECGs and electronic medical records. Xu et al (2017) reviewed the literature from engineering and computer science perspectives which in prediction of ICU mortality using many of the same methods in this paper.[23] Other models from a hospital mortality perspective include logistic regressions, neural networks, random forests and decision trees, naïve Bayes and support vector machines.[24 25 26 27]

Objectives

The objective of this study was to develop and validate a predictive model for 15-month mortality using a random sample of community-dwelling Medicare beneficiaries who might benefit from end-of-life services and to update the Charlson score weights using a national sample data set. The predictive model uses administrative claims rather than surveys, which enables it to be run on a larger population and at different times while not being dependent on physicians or members filling out a survey.

METHODS
Data

The Centres for Medicare & Medicaid Services (CMS) makes Limited Data Set (LDS) files available to researchers. As such no ethical board approval was needed. Although the LDS files contain beneficiary-level health information, they do not contain specific direct identifiers as defined in the Health Insurance Portability and Accountability Act Privacy Rule. This analysis used the CMS' five per cent LDS samples for 2014 and 2015

which contain administrative claims information at the beneficiary level, including diagnoses, procedures and beneficiary demographic data.

Administrative claims and beneficiary data for calendar year 2014 was used to create the explanatory variables used in the predictive models. Beneficiary data for calendar year 2015 was used to create the outcome variable (death indicator) used in the predictive models. Explanatory variables were derived from the 2014 data, in order to be temporally antecedent to the outcome variable, which was derived from 2015 to 2016 beneficiary data.

Sample size

The initial population was 2.8 million beneficiaries in 2014 and 2.9 million beneficiaries in 2015. The following three steps were used to create a final analytical data set.

First, in creating an analytical data set, beneficiaries were required to be in both years' data sets, which resulted in 2.7 million beneficiaries with data in both 2014 and 2015. This restriction ensured that beneficiaries were eligible for Medicare at the time of the prediction, 31 December, 2014, and that each member had up to a year of previous Medicare eligibility (2014) to be used for explanatory variable generation when predicting future death between January 2015 and March 2016.

Second, the 2.7 million beneficiaries were randomly split into a training and a validation data set with 1.35 million beneficiaries in each. This method of simple cross validation, or hold-out cross validation, was chosen given the abundance of data. The validation data set was used to ensure that overfitting did not occur. Overfitting occurs when a model with enough complexity predicts very well in the training data set but does not predict well in a validation data set. Overfitting would generate spurious relationships that are not generalisable to another data set. It is the validation data set that checks that the estimated model is generalisable and not spurious.

Third, people with one or more hospice claims in the last 120 days before prediction were excluded. However, further analysis revealed that about 64 per cent of people with a hospice claim prior to prediction actually expired in the subsequent 15 months. Since this was not near 100% as expected, a second analytical data set was created which included people with one or more hospice claims.

This study included the entire five per cent sample of the US Medicare population, which increases its generalisability by including a random sample of all Medicare beneficiaries. Included are people under the age of 65, people enrolled in an health maintenance organisation (HMO), and people who participated in a state buy-in (duals who are dually eligible for Medicare and Medicaid). The motivation for including all CMS beneficiaries is to develop a predictive model from all CMS community-dwelling beneficiaries rather than restrict to a certain state, age or demographic characteristic to enhance the model's generalisability.

Patient and public involvement

The present work does not include original patient data but is based on a five per cent sample of the US Medicare population available for purchase by the CMS. Therefore, no patients or public were involved in this study.

Outcome variable

CMS uses three sources to determine the dates of death for beneficiaries, including Medicare claims, online data submitted by family members and benefit information collected from the Railroad Retirement Board and the Social Security Administration with over 99% of death dates being validated.[28] These dates of death in the beneficiary data are used to calculate a binary death indicator which is the outcome variable in the predictive models to show if the person died during the 15 month time period between January 2015 through March 2016.

Explanatory variables

Included in the annual LDS beneficiary data are monthly indicators. One such indicator shows whether or not a person had monthly HMO coverage. Another indicator shows whether or not a person was dually eligible for Medicare and Medicaid which is also known as the person participating in a state buy-in defined as the state paying for the federal Medicare premium for those who are unable to afford it. Both the state buy-in and HMO coverage are used as explanatory variables in the predictive model.

The explanatory variables used to predict death included over 568 potential variables, including a constant term, in the categories from the 2014 data listed in table 1.

Statistical methods

Eleven different models were estimated. Six types of machine learning classification models were used to predict mortality: naïve Bayes, decision tree with adaptive boosting, logistic regression, neural network, least absolute shrinkage and selection operator (LASSO) and support vector machines. For each model, the area under the curve was calculated as the c-statistic. A c-statistic is a standard measure of goodness of fit for binary outcomes. In clinical terms the c-statistic gives the probability that a randomly selected beneficiary who expired will have a higher risk score of mortality than a beneficiary who had not expired[29]. Generally, c-statistics between 0.7 and 0.8 indicate a good model and values over 0.8 indicate a strong model.[30]

Predictive models

The first model included age and gender as the only explanatory variables and was used as a baseline model. Second, a model with the age and gender categories plus the Charlson score was estimated. Third, a model with the age and gender categories plus the Charlson condition indicators was estimated. This model was used to derive new Charlson weights to be compared with the original weights and the Romano adaptation of the Charlson weights. Fourth, a model with the age and gender

Table 1 Explanatory variables

Variable category	Number of variables	Explanation of variables
Charlson condition indicators	17	0/1 indicator for the presence of the Charlson condition (ICD-9 and ICD-10 diagnoses)
Charlson score	1	
AHRQ's clinical classification software condition indicators	286	0/1 indicator for the presence of each of the condition categories identified by the AHRQ software (ICD-9 and ICD-10 diagnoses)
AHRQ's clinical classification software procedure indicators	231	0/1 indicator for the presence of each of the procedure categories identified by the AHRQ software (ICD-9 and ICD-10 procedures)
Age and gender categories	20	10 age categories for each gender
Medical service utilisation groups	8	
Long-term care indicator	1	
Hospice indicator	1	This was determined as the presence of a hospice claim in the last 120 days before 31 December, 2014
Palliative care indicator	1	
HMO indicator.	1	This was determined as six or more months of HMO coverage from the CMS LDS data

CMS, Centres for Medicare & Medicaid Services; HMO, health maintenance organisation; ICD, International Classification of Diseases; ICD-9, ICD – ninth revision; ICD-10, ICD - tenth revision; LDS, Limited Data Set.

categories plus the Elixhauser condition indicators was estimated. Fifth, a model with all 568 explanatory variables was estimated. Sixth, a stepwise model was estimated to reduce the 568 explanatory variables into a more parsimonious set of variables. With over 1.3 million observations in the training data set, the entry p value into the stepwise model was 0.05 and the staying p value into the model was set at 0.001.

The seventh model estimated was a LASSO model. As with the stepwise model, the LASSO model was estimated to reduce the 568 explanatory variables into a more parsimonious set of variables. For estimation, the SAS procedure HPGENSELECT was used with the LASSO option. The eighth model estimated was a neural network model. The neural network used one hidden layer with 64 hidden neurons with a skip layer. This was the limit imposed in SAS Enterprise Miner 14.1 in terms of the hidden layer and hidden neurons. The ninth model estimated was a naïve Bayes model with additive (Laplacian) smoothing using the e1071 package in R. The tenth model was a decision tree with adaptive boosting, with 10 trials for boosting, using the C50 package in R. Lastly, a support vector machine (SVM) algorithm was estimated in SAS Enterprise Miner 14.1. SAS Enterprise Miner could not estimate a SVM model with the full sample size of over 1.3 million observations. As such, a 10% sample of the original training and validation data sets were used.

Two sets of the above nine models were estimated corresponding to the two different analytical data sets that included and excluded people with a hospice claim in the last 120 days prior to prediction.

In addition, decile risk bands of the predicted probability of death were calculated. The first risk band are those people who have a predicted probability between 0 and 0.1. The second risk band are those people who have a predicted probability between 0.1 and 0.2, etc. The stepwise predictive model was used to derive these predicted probabilities using the validation data set and used to compare actual death rates by risk band.

Charlson weights

Schneeweiss (2003) updated the original Charlson weights from 1983 using the Romano 1993 reclassification of conditions for New Jersey Medicare. The updated weights here follow the same method, where weights were increased by 1 for each increase of 0.3 in the logistic regression estimates (the log of the ORs). The updated ORs here are from a logistic regression using the age/gender variables in addition to the Charlson conditions as adapted by Romano. Table 2 scoring was used to generate updated weights using the national sample of Medicare beneficiaries of 1.35 million members.

To facilitate visually looking at the average Charlson scores by risk band, the stepwise regression model was scored against the validation data set to obtain risk bands.

Table 2 Updated Charlson weight rules

Weight	Rule
0	$0<=\ln(OR)<=0.15$
1	$0.15<\ln(OR)<=0.45$
2	$0.45<\ln(OR)<=0.75$
3	$0.75<\ln(OR)<=1.05$
4	$1.05<\ln(OR)<=1.35$
5	$1.35<\ln(OR)<=1.65$
6	$1.65<\ln(OR)<=1.95$

Table 3 Descriptive statistics from the training and validation data sets

Training data set	Validation data set	Variable description
1 357 989	1 356 245	Sample size
4.9%	4.9%	Per cent with a death date listed (2015 through first quarter of 2016)
29.5%	29.5%	Per cent with six or more months of HMO in 2014
16.7%	16.7%	Per cent with six or more months of state buy in for 2014 (duals)
1.56	1.55	Average Charlson score per person
0.6%	0.6%	Per cent of people with a hospice claim in the last 120 days of 2014
Annualised utilisation rates per 1000		
250.3	252.0	Ambulatory/surgery visits
249.9	251.6	Ancillary visits
1456.5	1452.3	Diagnostic services visits
1318.3	1316.0	Emergency department visits
318.0	315.9	Inpatient admissions
221.0	220.4	Non-acute visits
7749.5	7746.4	Outpatient visits
Demographics		
54.3%	54.5%	Per cent Female
45.7%	45.5%	Per cent Male
87.1%	87.0%	Per cent age 60 and over
69.8	69.9	Average age
Charlson conditions (percentage of population in data set)		
2.9%	2.9%	Myocardial infarction
7.3%	7.3%	CHF
9.4%	9.4%	PVD
8.5%	8.5%	Cerebrovascular disease
2.6%	2.6%	Dementia
14.1%	14.1%	Chronic pulmonary disease
2.7%	2.7%	Connective tissue disease-rheumatic disease
0.9%	0.9%	PUD
3.1%	3.1%	Mild liver disease
18.4%	18.4%	Diabetes without complications
6.1%	6.1%	Diabetes with complications
1.0%	1.1%	Paraplegia and hemiplegia
7.7%	7.6%	Renal disease
7.8%	7.7%	Cancer
0.3%	0.3%	Moderate or severe liver disease
0.9%	0.9%	Metastatic carcinoma
0.3%	0.3%	AIDS/HIV

CHF, congestive heart failure; HMO, health maintenance organisation; PUD, peptic ulcer disease; PVD, peripheral vascular disease.

Then, the average Charlson score using the Schneeweiss (2003) Medicare weights and updated weights using the CMS national sample data is calculated for each risk band.

RESULTS
Descriptive statistics
Table 3 shows the sample size for the training and validation data sets of over 1.35 million Medicare beneficiaries, which was derived from a 50–50 random split of beneficiaries. The random split into training and validation data sets can be seen by the nearly identical descriptive statistics between each data set. All of the variables, except the per cent with a death date as noted, are calculated from calendar year 2014 data. The per cent with a death date used the 2015 demographic data, which also included the first quarter of 2016, giving a death date up to 15 months after the baseline.
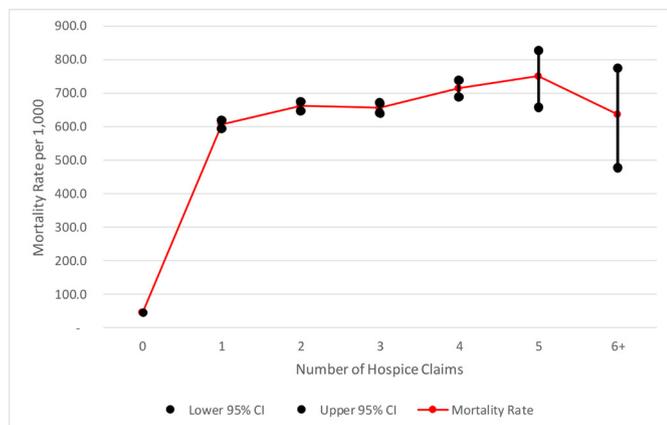
Figure 1 Mortality rate per 1000 and 95% CI by number of hospice claims. The red line is the mortality rate for all people at various levels of hospice claims. The black vertical lines represent the 95% CI for the mortality rate.

Not surprising, given higher female longevity compared with males, there are more females than males, represented by roughly 54 per cent female in each data set. Eighty-seven per cent of all members are over the age of 60, with the largest percentages having the Charlson conditions of diabetes or chronic pulmonary disease. The per cent of beneficiaries with a death date listed in the subsequent 15 months is 4.9% for both data sets, which is the outcome variable that is used in the development of the predictive models.

Figure 1 shows the mortality rate for people with different numbers of hospice claims in the last 120 days. Not surprisingly, people with zero hospice claims had the lowest 15-month mortality rate of 4.6%. However, the 15-month mortality rate for people that had hospice claims was unexpected. Overall, for the people with hospice claims, the mortality rate was 64%.

Predictive model results

Table 4 shows the c-statistics for the nine models that were estimated. Each set of 11 models was estimated using both
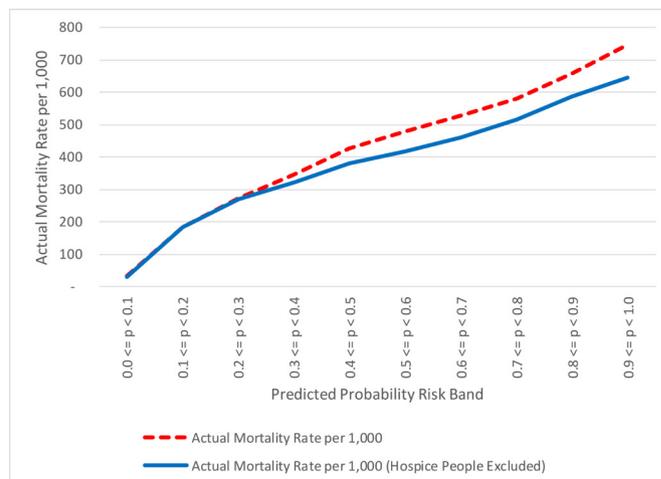

Figure 2 Actual mortality rate per 1000 by predicted risk band. Risk bands are deciles of the predicted mortality. Both predicted and actual mortality were calculated using the validation data set. The dashed line is the mortality rate for all people whereas the solid line is the mortality rate for all people excluding those with hospice.

the entire training data set, as well as a training data set that excluded people with one or more hospice claims in the 120 days preceding the prediction. For all models, except the SVM and decision tree with adaptive boosting, the c-statistic from the training and validation data sets is within one-half a per cent of each other, which shows that the generalisation error did not rise due to overfitting the training data.

With a c-statistic of 0.5 indicating random chance of correctly classifying mortality, the baseline model that used only age and gender did not show much improvement over chance, with a c-statistic of 0.547. The naïve Bayes model improved over the age gender baseline model with a c-statistic of 0.696. The models that used either the Charlson score, the Charlson conditions or the Elixhauser conditions improved over the age gender

Table 4 C-statistic (area under the curve) for estimated models

Model description	C-statistic all people	C-statistic excluding hospice people
Logistic regression: age+gender only	0.547	0.549
Naïve Bayes	0.696	0.675
Logistic regression: Charlson score	0.713	0.701
Logistic regression: Charlson conditions	0.726	0.714
Logistic regression: Elixhauser conditions	0.734	0.724
Decision tree with adaptive boosting	0.762	0.744
Support vector machine	0.788	0.773
Neural network	0.795	0.780
Logistic regression: stepwise regression	0.797	0.783
Logistic regression: LASSO	0.798	0.784
Logistic regression: all variables	0.798	0.784

LASSO, least absolute shrinkage and selection operator.

Table 5 Charlson conditions and scoring weights

Conditions from the Romano adaptation of the Charlson index	Original Charlson (1987) weights	Schneeweiss (2003) weights	CMS national sample OR estimates	Assigned CMS national sample weights
Myocardial infarction	1	1	1.11	0
Congestive heart failure	1	2	2.58	3
Peripheral vascular disease	1	1	1.48	1
Cerebrovascular disease	1	1	1.19	1
Dementia	1	3	5.07	5
Chronic pulmonary disease	1	2	1.47	1
Connective tissue disease-rheumatic disease	1	0	0.95	0
Peptic ulcer disease	1	0	1.09	0
Mild liver disease	1	2	1.16	1
Diabetes without complications	1	1	0.89	0
Diabetes with complications	2	2	1.03	0
Paraplegia and hemiplegia	2	1	1.48	1
Renal disease	2	3	1.69	2
Cancer	2	2	1.32	1
Moderate or severe liver disease	3	4	2.74	3
Metastatic carcinoma	6	6	5.72	6
AIDS/HIV	6	4	0.94	0

CMS, Centres for Medicare & Medicaid Services.

baseline model, with a c-statistic of 0.713, 0.726 and 0.734, respectively, indicating a good predictive model with the c-statistic over 0.7 as did the decision tree with adaptive boosting with a c-statistic of 0.762. The support vector machine algorithm had the next best c-statistic of 0.788 in the validation data set.

The four models that performed the best were the neural network (c-statistic=0.795), the logistic regression using a stepwise selection of variables (c-statistic=0.797), the logistic regression using LASSO selection of variables (c-statistic=0.798) and the logistic regression using all variables (c-statistic=0.798). The online supplementary list shows the ORs for the 154 variables retained in the stepwise model. The LASSO model retained 401 variables and is thus close to the full model of 568 variables. All four of these models had a c-statistic over 0.79 indicating that all of them have very good predictive power and could be used in predicting mortality in the next 15 months.

In clinical review of variables from the logistic regression which used stepwise selection of variables, those highly associated with 15-month mortality with ORs over 2.0 include:

▶ Presence of a hospice claim in the last 120 days before the prediction,
▶ Age and gender groups, and
▶ Various cancers.

In addition, people who were dually eligible for Medicare and Medicaid had a higher risk of mortality, with an OR of 1.18. People who were in an HMO for at least 6 of the 12 months prior to prediction had a statistically

significant and higher risk of mortality with an OR of 1.50. In total, 154 of the 338 variables were selected as significant predictors of 15-month mortality.

Figure 2 shows that the actual mortality rate for each risk band for both models, one which includes people with hospice claims and the other which excludes people
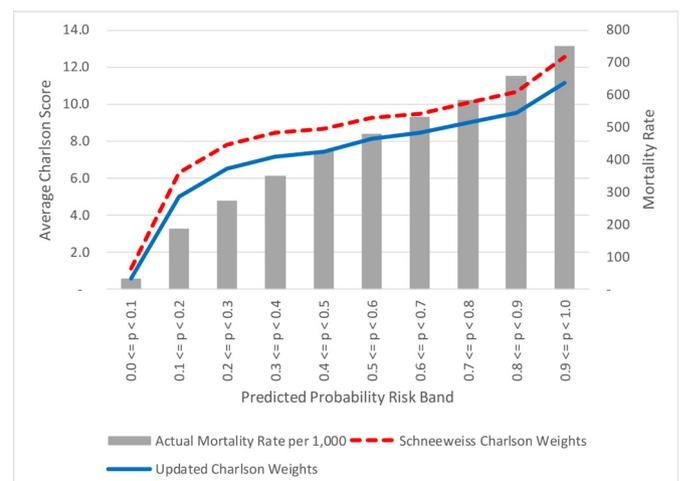

Figure 3 Average Charlson scores and actual mortality rate by predicted risk band. Risk bands are deciles of the predicted mortality. Both Charlson scores and actual mortality were calculated using the validation data set. The dashed line is the average Charlson score using the Schneeweiss weights whereas the solid line is the average Charlson score using the updated weights. The bars represent the mortality rate for all people.

with hospice claims. In both cases, the mortality rate increases in a nearly linear fashion from a rate of 32 to 748 per 1000 people between the lowest and highest risk bands for the model, which includes hospice claimants, and from a rate of 31 to 645 for the model which excludes hospice claimants. As the c-statistic shows, this alternate view shows that the model does a good job at discriminating between mortality. The people with a predicted mortality between 0 and 0.1 have a very low death rate of 32 per 1000 people and the people with a predicted mortality between 0.9 and 1.0 have a very high death rate of 748 per 1000 people.

Updated Charlson weights

Table 5 shows the updated Charlson weights using the national sample of Medicare beneficiaries of 1.35 million members. Some of the biggest changes from the Schneeweiss weights to the updated weights here are dementia, rising from a weight of three to a weight of five with the CMS national sample data. Also, the AIDS/HIV weight previously was four and is now weighted 0 with the CMS national sample data from this study, indicating that AIDS/HIV now does not increase the likelihood of mortality enough to warrant a Charlson weight. Given the treatment advances, the original 1987 weight of six, moved down to four in 2003 and is now zero.

Figure 3 shows the average Charlson scores by risk band using the Schneeweiss and updated weights. Both show that with an increasing risk band, the average Charlson score increases in a nearly linear fashion. The Schneeweiss range from a score of 1.1 to 12.5, between the lowest and highest risk bands, while the updated weights range from a score of 0.6 to 11.2. The updated weights follow the same upward sloping pattern as the Schneeweiss weights but with the line shifted downwards. This is because of the 17 condition categories only two increased, while nine weights decreased and six weights stayed the same. Only congestive heart failure and dementia increased in weight with congestive heart failure increasing by one and dementia increasing by two.

Although the c-statistic for the models using the Charlson score or the Charlson conditions was 0.713 and 0.726, respectively, this simple model with updated weights still predicts 15-month mortality fairly well with a c-statistic over 0.7.

DISCUSSION

Factors to consider when evaluating predictive model algorithms include accuracy and generalisability.[18] Both calibration (the degree to which predicted death match observed death) and discrimination (how well those who die are distinguished from those who don't die) are part of accuracy. Within the category of generalisability, both geographical (different geographic locations are used) and spectrum (diverse disease states and trajectories) are part of generalisability. In terms of calibration, the models here show a match between predicted and observed death from figure 2. In

terms of discrimination, while the above described predictive models did not produce significantly different c-statistics from previously developed models. However, in terms of generalisability, both geographical and spectrum, these new models do provide much greater clinical utility based on generalisability across geographical areas and the full spectrum of disease and full range of health statuses. So, although these models have similar discriminative ability, given the broader disease categories and locations, this is seen as highly favourable to these new predictive models. This is in contrast to previous algorithms discussed in the introduction which were focused on a particular geography, used surveys instead of administrative claims, had small samples or were focused on particular conditions limiting their generalisability.[18] Generalisability is a key requirement for the application of predictive analytics indicators toward enhanced clinical judgements of which patients are likely to benefit from the initiation of advanced care planning and subsequent transition to palliative and hospice care.

Given the key clinical eligibility criteria for transition into hospice is a physician's assessment that death is probable in the ensuing 6 months, it was assumed that individuals with hospice claims had already been identified as being in the last year of life. However, the 64% 15-month mortality rate finding in these individuals indicates the difficulty of predicting near-term mortality. It was expected that this rate would be near 100%, and thus, people with a hospice claim should be excluded from a mortality predictive model since having a hospice claim would be a perfect predictor and not of value. Given the observed mortality rate was not near 100%, two sets of models were estimated: one set with all people, including those with a hospice claim, and another excluding those with a hospice claim. Also of note, as the number of hospice claims increases, the mortality rate does not increase, but, instead, is fairly steady at 64%. This calls into question the accuracy of clinical judgement alone when referring individuals to hospice care since over one-third of people with six or more hospice claims prior to prediction do not die within 15 months.

Unlike previous research, which found a lower rate of mortality for people in a Medicare Advantage plan[31], this study found a positive correlation between enrolment in an HMO for at least 6 of the 12 months prior to prediction and a statistically significant and higher risk of mortality. While not an indication of causation, one of several possible explanations for this correlation could be a higher likelihood of attention to resource utilisation and advanced directives in Medicare Advantage plans, and thus greater avoidance of aggressive end-of-life treatments. Conversely, the correlation between Medicare Advantage enrolment and higher mortality rates could be an indication of differential provision of medically necessary care. Further investigation of this relationship is warranted.

Improved identification of patients nearing the transition into the last year of life has become increasingly important as ageing populations increase. The models from this study can be used from a clinical perspective for identifying potential transition or initiation of advanced

care planning for palliative and advanced illness. Further, the models can be used from a payer perspective for risk adjustment. Compared with previous administrative claims-based studies that were limited to geographical areas and risk levels, this study used a CMS national sample of beneficiaries, giving these results higher generalisability than previous models. In addition, with updated Charlson weights from an updated data set and from a more generalizable result, the comorbidity scores can be used to save time and resources when calculating a mortality score. When claims data are not available, or in small populations, a Charlson score would still be a useful predictor of mortality. The c-statistic for the Charlson condition model was over 0.7 in the validation data indicating that, even with the limited conditions, this model still performs well. Six Charlson conditions now have a weight of zero as compared with the weights by Schneeweiss which had two Charlson conditions with a weight of zero. The updated weights with a weight of zero had previously had weights of one or two. Diabetes with complications had previously had a weight of two but now has a weight of zero. The largest difference was with HIV/AIDS which had a Schneeweiss weight of four but now a weight of zero which most likely reflects the change in treatment and mortality of this condition over the last 15 years.

One criticism of a logistic regression is that prediction may be difficult when the decision boundary is highly non-linear which is overcome with neural networks can have highly non-linear decision boundaries. However, this does not appear to be a limitation with the current logistic regression predictive models given the similar c-statistics as that of the neural network. Alternatively, one criticism of the neural network and SVM model approaches is the 'black box' interpretability of these models. That is, it cannot be known which explanatory variable is important in classifying the outcome variable and by what magnitude as with an OR in a logistic regression. As such, this type of model does not lend itself to clinical review of the explanatory variables.

REFERENCES

1. Stanford School of Medicine. Where do Americans die? https://palliative.stanford.edu/home-hospice-home-care-of-the-dying-patient/where-do-americans-die/ (Accessed 29 Jun 2018).
2. Wang SY, Hsu SH, Huang S, *et al*. Longer hospice stays associated with decreased end-of-life spending for patients in regions with high expenditures. *Health Affairs* 2017;36:328–36.
3. Meier DE, Back AL, Berman A, *et al*. A National Strategy For Palliative Care. *Health Aff* 2017;36:1265–73.
4. Bynum JPW, Austin A, Carmichael D, *et al*. High-cost dual eligibles' service use demonstrates the need for supportive and palliative models of care. *Health Aff* 2017;36:1309–17.
5. Centers for Disease Control and Prevention. CDC wonder. https://wonder.cdc.gov/ (Accessed 8 Aug 2017).
6. Yadav KN, Gabler NB, Cooney E, *et al*. Approximately one in three us adults completes any type of advance directive for end-of-life care. *Health Aff* 2017;36:1244–51.
7. Fried TR, Bradley EH, Towle VR, *et al*. Understanding the treatment preferences of seriously ill patients. *N Engl J Med* 2002;346:1061–6.
8. Brown atlas of dying. http://www.chcr.brown.edu/dying/brownatlas.htm (Accessed 29 Jun 2018).
9. Department of Health Care Services. Palliative care and medi-cal managed care. http://www.dhcs.ca.gov/formsandpubs/Documents/MMCDAPLsandPolicyLetters/APL2017/APL17-015.pdf (Accessed 29 Jun 2018).
10. An act to add Section 14132.75 to the Welfare and Institutions Code, relating to health care. http://www.leginfo.ca.gov/pub/13-14/bill/sen/sb_1001-1050/sb_1004_bill_20140925_chaptered.htm (Accessed 29 Jun 2018).
11. Barnato AE. Challenges in understanding and respecting patients' preferences. *Health Aff* 2017;36:1252–7.
12. Spending in the last year of life and the impact of hospice on medicare outlays (Updated August 2015). http://www.medpac.gov/docs/default-source/contractor-reports/spending-in-the-last-year-of-life-and-the-impact-of-hospice-on-medicare-outlays-updated-august-2015-.pdf?sfvrsn=0 (Accessed 29 Jun 2018).
13. Zhang B, Wright AA, Huskamp HA, *et al*. Health care costs in the last week of life: associations with end-of-life conversations. *Arch Intern Med* 2009;169:480–8.
14. Davis MA, Nallamothu BK, Banerjee M, *et al*. Identification of four unique spending patterns among older adults in the last year of life challenges standard assumptions. *Health Aff* 2016;35:1316–23.
15. Charlson ME, Pompei P, Ales KL, *et al*. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373–83.
16. Romano PS, Roos LL, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. *J Clin Epidemiol* 1993;46:1075–9.
17. Elixhauser A, Steiner C, Harris DR, *et al*. Comorbidity measures for use with administrative data. *Med Care* 1998;36:8–27.
18. Yourman LC, Lee SJ, Schonberg MA, *et al*. Prognostic indices for older adults: a systematic review. *JAMA* 2012;307:182–92.
19. Gagne JJ, Glynn RJ, Avorn J, *et al*. A combined comorbidity score predicted mortality in elderly patients better than existing scores. *J Clin Epidemiol* 2011;64:749–59.
20. van Walraven C, Austin PC, Jennings A, *et al*. A modification of the elixhauser comorbidity measures into a point system for hospital death using administrative data. *Med Care* 2009;47:626–33.
21. Hamlet KS, Hobgood A, Hamar GB, *et al*. Impact of predictive model-directed end-of-life counseling for Medicare beneficiaries. *Am J Manag Care* 2010;16:379–84.
22. Schneeweiss S, Wang PS, Avorn J, *et al*. Improved comorbidity adjustment for predicting mortality in Medicare populations. *Health Serv Res* 2003;38:1103–20.
23. Xu J, Zhang Y, Zhang P, *et al*. Data Mining on ICU Mortality Prediction Using Early Temporal Data: A Survey. *Int J Inf Technol Decis Mak* 2017;16:117–59.
24. Almardini M, Ras ZW. *A supervised model for predicting the risk of mortality and hospital readmissions for newly admitted patients*: Foundations of Intelligent Systems Conference, 2017.
25. Harrison E, Chang M, Hao Y, *et al*. *Using machine learning to predict near-term mortality in cirrhosis patients hospitalized at the University of Virginia health system*: 2018 Systems and Information Engineering Design Symposium (SIEDS) Conference, 2018.

26. Sahni N, Simon G, Arora R. Development and validation of machine learning models for prediction of 1-year mortality utilizing electronic medical record data available at the end of hospitalization in multicondition patients: a proof-of-concept study. *J Gen Intern Med* 2018;33:921–8.

27. Krishnan G, Kamath SS. *A supervised learning approach for ICU mortality prediction based on unstructured electrocardiogram text reports*: Natural Language Processing and Information Systems Conference, 2018.

28. Research Data Assistance Center. Death Information in the Research Identifiable Medicare Data. https://www.resdac.org/resconnect/articles/117 (Accessed 23 Feb 2018).

29. C-Statistic: definition, examples, weighting and significance. http://www.statisticshowto.com/c-statistic/ (Accessed 19 Jun 2018).

30. Hosmer DW, Lemeshow S. *Applied logistic regression*. 2nd ed. New York: Wiley.

31. Beveridge RA, Mendes SM, Caplan A, *et al*. Mortality differences between traditional medicare and medicare advantage: a risk-adjusted assessment using claims data. *Inquiry* 2017;54:1–8.