Tim Sweeney
tes17@stanford.edu
CS 229 2013 - Final Project

Prediction of outcomes in mild TBI in the NTDB

Traumatic brain injury (TBI) accounts for >1.3 million ED visits and >2750,000 hospitalizations each year[1]. Although guidelines exist for how to triage and treat these patients based on the severity of TBI, there is no clear consensus statement on which patients need intensive care unit (ICU) admission, and wide variability in practice. Many hospital guidelines currently suggest that all patients with intracranial hemorrhage of any severity be observed in ICU due to risk of decompensation and possible need for intervention. Several studies have tried to make prediction models of outcomes after head trauma[2-7]. In particular, some single-center datasets have shown that certain classes of low-risk patients have low probabilities of decompensation, and so may be better observed on the hospital floor in order to conserve hospital resources. For instance, Nishijima et al used binary recursive partitioning on the UC Davis TBI-ICU data set and found that four parameters (abnormal mental status (GCS < 15), non-isolated head injury, age > 65 years, and swelling/shift on CT) was 98% sensitive and 50% specific with a ROC AUC of 0.74[8]. Washington et al. showed that age >65 years, anticoagulation therapy, frontal contusions and high volume intracranial hemorrhage were significant predictors of decompensation, but did not evaluate their model[6]. Overall, no consensus prediction formula exists. Thus a robust formula for clinical decision support with regard to ICU admission for mild TBI patients is needed (since moderate-severe TBI is an easier call for ICU admission); only one study to date has looked at ICU admission in mild TBI. Here the goal was to find a robust score for prediction that is also sparse. Clinicians are highly unlikely to apply a complicated rule, especially if requires input to a computer in a 'black-box' model. Thus penalized linear regression was chosen, with a hypothesis that the Lasso penalty would be able to identify a sparse model for easy applicability.

Data Set:

The National Trauma Data Bank (NTDB) is a national dataset with millions of post-trauma hospitalized patients and several hundred variables. The NTDB has many variables not related to patient physiology (i.e., number of trauma surgeons at a given hospital, region code of place of injury) and does not contain as many physiologic parameters as some single-institution data sets. However, it does have many of the key physiologic parameters identified by prior studies as being significant predictors of outcome, as well as ICD-9 diagnostic and procedural codes.

Methods:

Patients from 2007-2011 with an ICD-9 code of either skull fracture (800.xx-801.xx, 803.xx-804.xx) or intracranial injury (851.xx-854.xx) were selected as the base dataset. Only patients with a Glasgow coma scale (GCS) of 14 or 15 (i.e., only mild head injuries) were included. Patients with an injury type other than "Blunt" were excluded.

All physiologic variables (age, gender, EMS and ED vital signs, GCS, and injury severity scores) were included as dependent variables. All TBI-related ICD-9 codes were included as dependent binary variables to either 3, 4, or 5 decimal places (i.e., the model matrix with 5-decimal ICD-9 codes was ~100 times larger and sparser than the 3-decimal ICD-9 matrix). For the 3-decimal ICD-9 codes, all fractures were grouped into two single variables: 'open fracture 800-804' and 'closed fracture 800-804' (plus 851-854), thus there were only six TBI variables in this case. The number of co-occurring non-TBI diagnosis codes was included as a continuous dependent variable (as an estimator of number of non-TBI injuries). Response variables tested were either: neurosurgery (yes/no), or complicated status (yes/no), where complicated status was defined as neurosurgery and/or a critical care intervention. Patients with missing physiologic data were thrown out (i.e. imputation was not done).

Critical care interventions were modelled after those defined by Nishijima et al. ICD-9 procedural codes were used to identify interventions which might reflect decompensated status. These included: vasopressor

use (00.17), interventional angiography (00.6x), invasive hemodynamic monitoring (89.6x), invasive mechanical ventilation of any duration (96.7x), blood transfusion of any component (99.0x), and conversion of arrhythmia, including CPR (99.6x). The presence of any number of these interventions was coded as a binary variable. Undergoing neurosurgery (01.xx-02.xx) was coded as a separate binary variable. The 'complicated' response variable was coded as the presence of either or both a critical care intervention or neurosurgery.

Penalized logistic regression (using package 'glmnet' in R) was carried out at an alpha of 0.95 (i.e., Lasso-weight Elastic-net parameter)[9]. Regression was run in 6 models, where each of 3, 4, or 5-decimal ICD-9 TBI codes as dependent variables (along with all other physiologic variables) were tested against both neurosurgery-only and 'complicated' response variables. All variables (continuous and categorical) were scaled, in accordance with Lasso-penalized regression[10]. The data were split into training (80%) and test (20%) sets. Five-fold cross validation was used. Error curves were constructed to test for bias in the model. ROC curves were constructed for three different values of the penalty lambda (effectively, different numbers of included variables) for each model. Coefficients were extracted at a lambda for which ~8-10 variables were included.

To test for presence of nonlinear effects, soft-margin support vector machines (with C=1,5,20) using a Gaussian kernel were made using package 'kernlab' in R. Both response variables were tested.

Results:

A total of ~250,000 patients were included for analysis. Of these, 1.5% underwent neurosurgery, 6.0% had a critical care intervention, and 7.1% had one or both, and so were 'complicated'. Regression using 4-decimal ICD-9 codes and yielded models with best error rates of ~2% for neurosurgery-only and ~8% for 'complicated' outcomes .The models were evaluated for several different values of lambda, yielding predictors with different numbers of variables (Figure 1A). ROC curves for the 4-decimal ICD-9 model showed greater attainable sensitivity and AUROC in the neurosurgery-only outcome than for the 'complicated' outcome (Figure 1B). Coefficients were pulled out of the models, and showed that neurosurgery alone was most strongly predicted by open skull fracture. The joint 'complicated' variable was also predicted by physiologic variables such as age and injury severity (Table 1).

I next tested the 3-decimal ICD9 codes (for which all fractures had been further collapsed into variables open/closed). ROC curves were similar between the 4-decimal and 3-decimal models (Figure 1B). Moreover, the coefficients for both response variables were (by design) enriched for physiologic parameters compared to ICD9 codes (Table 1).

Table 1: Logistic regression models

4-Decimal ICD9 TBI Model				**3-Decimal ICD9 TBI Model**			
Neurosurgery only:		Neurosurgery + Critical intervention:		Neurosurgery only:		Neurosurgery + Critical intervention:	
Variable	Coefficient	Variable	Coefficient	Variable	Coefficient	Variable	Coefficient
803.9 – open fx	21.64973	803.9 – open fx	241.2097	Any Open Fx	0.0591327	ISS	0.19117386
804.9 – open fx	1.32474	804.6 – open fx	164.9622	ISS	0.0394562	Age	0.14474897
800.6 – open fx	0.066791	Age	0.256611	852.xx	0.0262511	Any Open Fx	0.04984469
800.7 – open fx	0.066659	ISS	0.239259	854.xx	0.0248537	# non-TBI injuries	0.04206349
852.2 – SDH	0.054994	# non-TBI injuries	0.084876	Closed Fx	0.0219145	EMS_sbp	-0.0234376
800.8 – open fx	0.044459	852.2 – SDH	0.045855	# non-TBI injuries	-0.0099927	ED_sbp	-0.0151327
800.9 – open fx	0.036282	EMS_sbp	**-0.042775**	EMS_gcseye	-0.0003407	ED_pulse	0.01240867
800.5 – open fx	0.034661	ED_sbp	**-0.036347**	ED_rr	-0.0004717	852.xx - SDH	0.01092237
ISS	0.034608	800.7 – open fx	0.035693			EMS_pulse	0.00856837
801.7 – open fx	0.021792	800.5 – open fx	0.032838			ISS - AIS	0.00604703
801.9 – open fx	0.021248	ED_pulse	0.026772			EMS_gcsmot	-0.0059931

Figure 1 – A. . Output for linear regression for 4-decimal model and 'complicated' response variable. Upper left, coefficient plot for range of lambda (each line is one variable; when the coefficient becomes nonzero, it moves off the x-axis). ROC curves constructed from the model at three different values of lambda (and hence differing numbers of included variables).

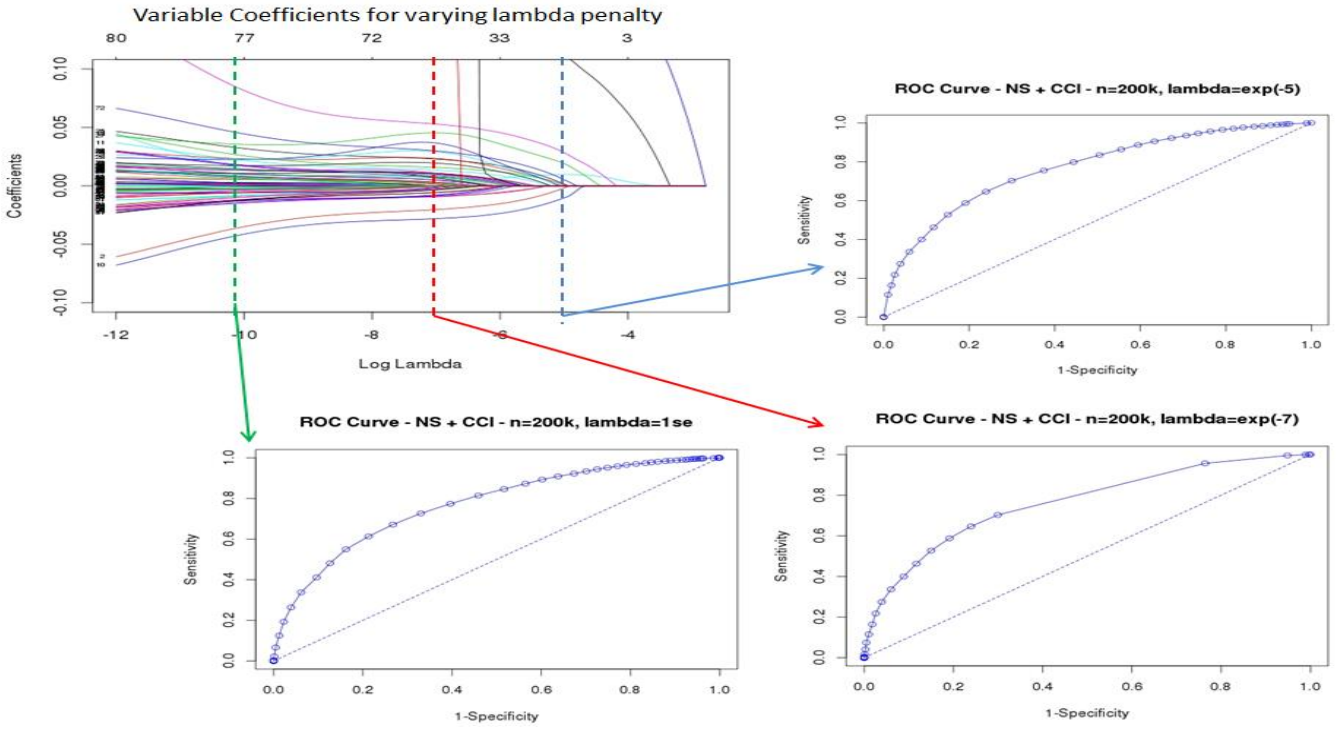
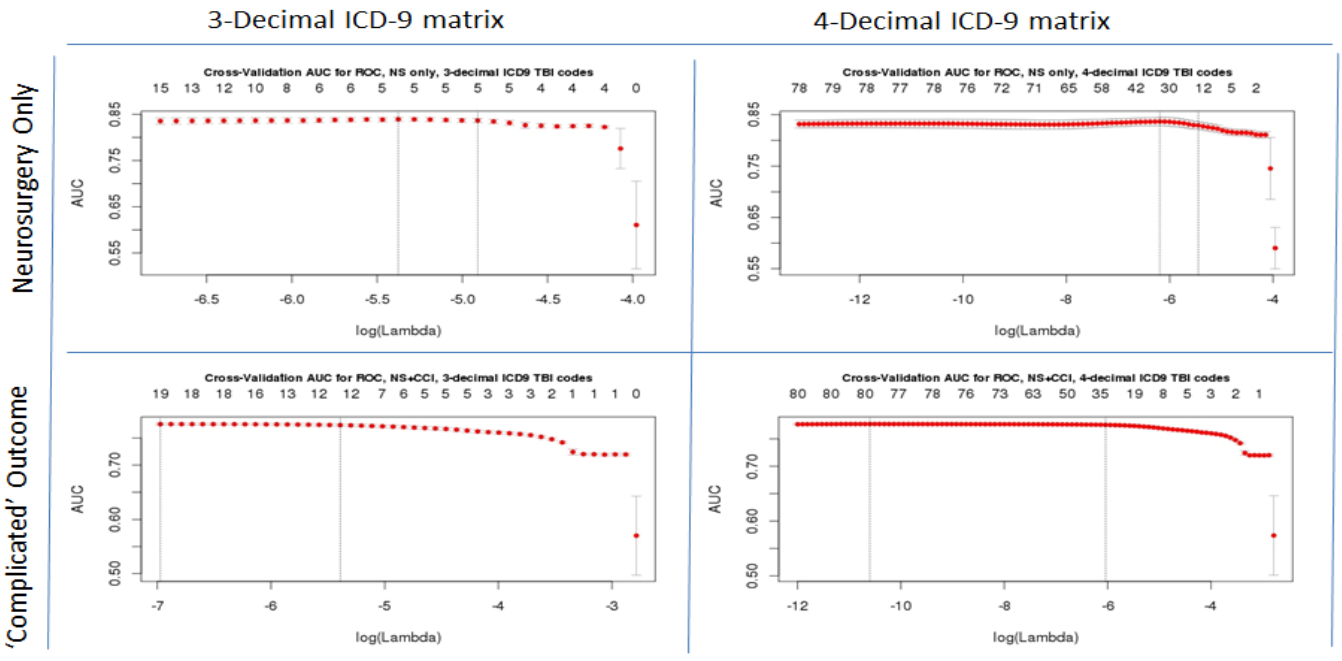


Figure 1 – B. Cross-validation models were used to compute areas under ROC curves (here abbreviated AUC) for a range of lambda penalty, Across the top of each plot is the number of variables for the given AUC.



Five-decimal ICD-9 code matrices were tested and showed similar error rates to the 4-decimal ICD codes, but with many more dependent variables included for each regression model. They were thus not studied further, and the data is not shown.

Error curves were constructed and showed that for both response variables, the model suffers from a high-bias problem, in that test and training error rates converge at an unfortunately high rate (see Figure 2; note that initial error is low because there the response variable is infrequent). For this reason, soft-margin SVMs were constructed, to see if a non-linear Gaussian kernel could better fit the data. SVMs were tried with a C allowance of 1, 5, and 20; error rates were best at C=5. As expected, training set error rates were lower for SVMs than for linear regression; however, test error rates were similar for both methods (neurosurgery-only response: training error: 0.79%, test error: 1.47%; 'complicated' response: training error: 5.07%, test error: 7.52%).

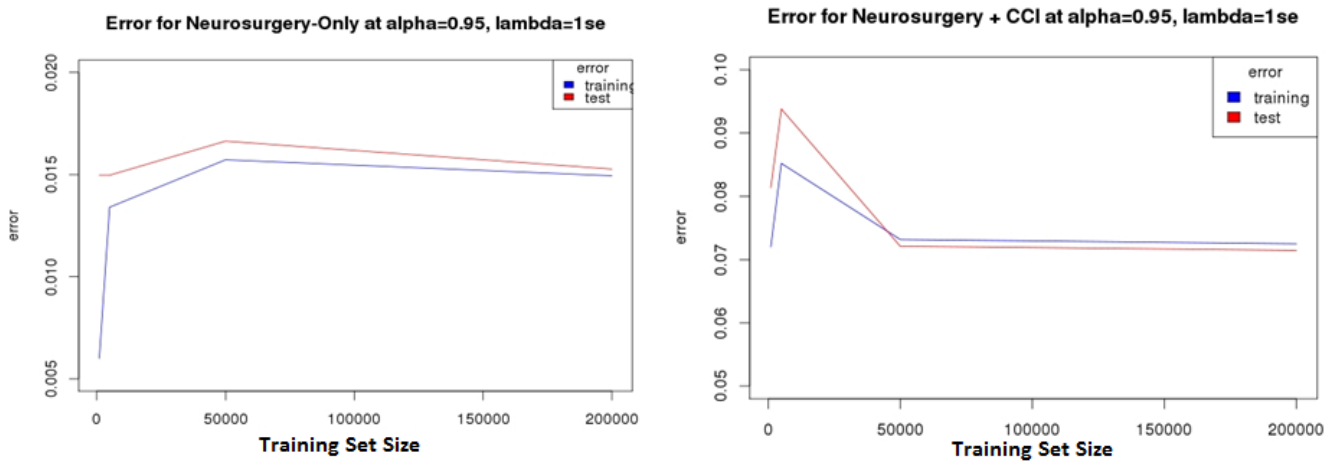


Figure 3. Error vs training set size for the 4-decimal ICD9 models for the two response variables.

Conclusions:

The goal of this project was to produce a sparse model for predicting which patients with mild TBI will go on to require care that would require an ICU stay. There is no hard definition of what it means to require intensive care; in clinical practice, reasons for admission to the ICU are varied, and often up to clinician preference and local factors. Thus, modelling this outcome is difficult and prone to bias. Here I modelled two different response variables: either ever undergoing neurosurgery, or a 'complicated' variable, which is a composite of neurosurgery and various ICD9-coded procedures that indicate the need for a higher level of care. Need for neurosurgery is a narrow response, but one which is easily defined and should be captured well by the NTDB. Using the combined neurosurgery + critical care intervention response variable is likely a better predictor of needing ICU care, but there is likely bias in how or whether the comparatively minor procedures like invasive monitoring and blood transfusion are captured in NTDB.

Two different dependent variable sets were used in prediction, as it was not clear *de novo* to what extent ICD-9 TBI code specificity would be required by the model. There was little difference in the predictive power (as judged by AUROC) between the models with collapsed 3-decimal ICD-9 codes and the models with sparser 4-decimal ICD-9 codes. Since the 3-decimal ICD-9 code versions led to sparser prediction scores, I will plan to proceed with that model in further work.

The predictive power of this model cannot be judged by simple error rates, as the positive outcomes are rare; thus a model that predicted no patient would ever go for neurosurgery would still be correct 98.5% of the time. For this reason, I used AUROC as a better judge of predictive power. In clinical application, the most important goal is a high sensitivity. It could be disastrous to have a patient decompensate outside the ICU, whereas if we could safely decrease the number of unnecessary admissions to the ICU by even 50%, the cost

savings could be large. Thus, in practical application, a cutoff for this prediction score would be chosen for a sensitivity of >90%, even at the expense of a very poor specificity (<30%).

Sparse prediction models built with L1-penalized regression performed well in this project, with AUROCs for models with 5-10 variables performing nearly as well as those using all variables available. The variables in the predictive models for a 'complicated' outcome are, not surprisingly, more enriched for non-TBI parameters (higher ISS, higher rates of other injuries, lower BP, higher pulse). Similarly, the neurosurgery-only response variable is highly dependent on the presence of open fractures and intracranial hemorrhage, which does not seem to add anything new to our understanding.

There is still significant error in the regression models constructed here. Error-estimation curves show a high-bias problem with both responses. Soft-margin SVMs with Gaussian kernels failed to show any improvement in test set error. Thus the bias that is present in the current model is unlikely to be from a simple nonlinear relationship of the present data, but rather from confounders not present in the NTDB.

Comparing our model directly to previous models like that of Nishijima et al is impossible because of the differences in the variables present in each dataset, and the nature of the differences between a single-center trial like theirs and a national database like the NTDB. Still, it is encouraging that age and non-head injury appear to be important in both models. Notably GCS was a large part of the Nishijima score, but here we conditioned on only mild GCS (14 or 15), so the magnitude of this variable was comparatively small. Interestingly, when I applied Nishijima's cutoff rules (without CT findings, variable not present) to the present dataset, it failed to accurately partition the patients into complicated/uncomplicated groups. The failure of that model in the NTDB highlights how different the two data sets are.

This study had several weaknesses. First, no risk adjustment for hospital type was done here. In a subanalysis I found that trauma accreditation level of a given hospital is significantly associated with our response variable. However, inclusion of trauma level of a given hospital in a patient prediction score doesn't make intuitive sense. Further work should be in trying to identify a patient-level factor that can capture this bias. The second weakness is in the building of the 'complicated' response variable. It is highly likely that minor procedures are not all captured by ICD-9 codes, and, moreover, that they may be missing in nonrandom ways. Further, these criteria do not capture all patients that could potentially benefit from a higher level of nursing care and monitoring.

Here I have shown an early potential application of penalized logistic regression to a large national dataset (NTDB) in prediction of mild TBI outcome. These preliminary models do show some predictive power for the constructed response variables, and the coefficients selected make intuitive sense. Better response variables and corrections for bias need to be built into the models in the next phase of work.

References

1. Coronado VG, Xu L, Basavaraju SV, et al. Surveillance for traumatic brain injury-related deaths--United States, 1997-2007. *MMWR Surveill Summ.* May 2011;60(5):1-32.

2. Hukkelhoven CW, Steyerberg EW, Habbema JD, Maas AI. Admission of patients with severe and moderate traumatic brain injury to specialized ICU facilities: a search for triage criteria. *Intensive Care Med.* Jun 2005;31(6):799-806.

3. Kuppermann N, Holmes JF, Dayan PS, et al. Identification of children at very low risk of clinically-important brain injuries after head trauma: a prospective cohort study. *Lancet.* Oct 2009;374(9696):1160-1170.

4. Steyerberg EW, Mushkudiani N, Perel P, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. *PLoS Med.* Aug 2008;5(8):e165; discussion e165.

5. Nishijima DK, Sena MJ, Holmes JF. Identification of low-risk patients with traumatic brain injury and intracranial hemorrhage who do not need intensive care unit admission. *J Trauma.* Jun 2011;70(6):E101-107.

6. Washington CW, Grubb RL. Are routine repeat imaging and intensive care unit admission necessary in mild traumatic brain injury? *J Neurosurg.* Mar 2012;116(3):549-557.

7. Nishijima DK, Sena M, Galante JM, et al. Derivation of a Clinical Decision Instrument to Identify Adult Patients With Mild Traumatic Intracranial Hemorrhage at Low Risk for Requiring ICU Admission. *Ann Emerg Med.* Dec 2013.

8. Nishijima DK, Shahlaie K, Echeverri A, Holmes JF. A clinical decision rule to predict adult patients with traumatic intracranial haemorrhage who do not require intensive care unit admission. *Injury.* Nov 2012;43(11):1827-1832.

9. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw.* 2010;33(1):1-22.

10. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med.* Feb 1997;16(4):385-395.