

# A Data Mining Framework for Improving Student Outcomes on Step 1 of the United States Medical Licensing Examination

Jimmy Clark, PhD

jimmy.clark@bcm.edu

## PROBLEM STATEMENT

- Prior research used multivariate regression models to indicate correlates of pass and fail outcomes
- Low national failure rate (4%) provides small sample for research
- Individual student outcomes are still unknown

## RESEARCH GOAL

- Use data mining techniques to create a model which predicts Step 1 outcomes at the student level
- Use over/under sampling techniques to increase sample size
- Determine the point in the BCM preclinical curriculum when intervention programs can be offered

## METHODS

- Medical school students matriculating from 2013 to 2015 (n=514) were extracted from the BCM student information system
- The dataset included MCAT scores, undergraduate GPA, final course grades from 25 courses taken in the preclinical years, and pass/fail results from the comprehensive basic sciences examination and Step 1
- To increase the failed sample size, passing scores within one standard deviation were considered failing which increased the number of failed observations from 2 to 19, consistent with the national failure rate of 4%
- Over and under sampling techniques were used to address the imbalance of pass and fail observations
- Eight experiments were run with predictor variables presented in a stepwise fashion to match the progression of preclinical courses
- The CART algorithm was used for each of the experiments, using prediction accuracy and model effectiveness (F1) as the measure of success

## RESULTS

- The model created with preadmission variables alone had an accuracy rate of 90.4% and an F1 of 0.22
- The second experiment, using preadmission variables and final grades for the first block of preclinical courses, had an improved accuracy of 95.9% and F1 of 0.57
- Model accuracy and F1 peaked with this experiment then dropped for the remaining experiments
- Final grades from the first block of courses best identified at-risk students, specifically the Foundations Basic to the Science of Medicine course as the best predictor of Step 1 outcomes, with students with a final course grade lower than 85.35 predicted to fail Step 1

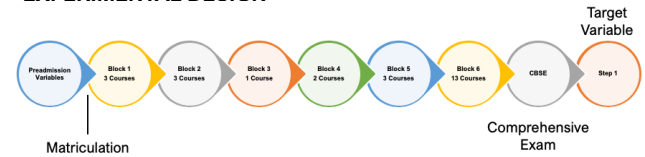
BCM medical students' Step 1 outcomes can be predicted as early as the end of the first course, Foundations Basic to the Science of Medicine.

Students with a final course grade lower than 85.35 are predicted to fail Step 1.



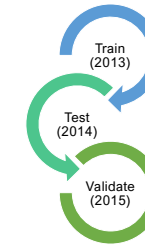
Take a picture to download the full paper

## EXPERIMENTAL DESIGN



1. Preadmission Variables (9 Predictors)
2. Experiment 1 + 1<sup>st</sup> Block Course Grades (12 Predictors)
3. Experiment 2 + 2<sup>nd</sup> Block Course Grades (15 Predictors)
4. Experiment 3 + 3<sup>rd</sup> Block Course Grades (16 Predictors)
5. Experiment 4 + 4<sup>th</sup> Block Course Grades (18 Predictors)
6. Experiment 5 + 5<sup>th</sup> Block Course Grades (21 Predictors)
7. Experiment 6 + 6<sup>th</sup> Block Course Grades (34 Predictors)
8. Experiment 7 + CBSE (35 Predictors)

## VALIDATION PLAN



- 3-fold cross validation
- Train with students matriculating in 2013
  - Unbalanced
  - Random Under Sampling (RUS)
  - Random Over Sampling (ROS)
  - Synthetic Minority Over Sampling (SMOTE)
- Test with students matriculating in 2014
- Validate with students matriculating in 2015

## CRITERIA FOR SUCCESS

	Predicted Step 1 Failure	Predicted Step 1 Passing
Actual Step 1 Failure	Actual failure predicted to fail (True Positive)	Actual failure predicted to pass (False Positive)
Actual Step 1 Passing	Actual passing predicted to fail (False Negative)	Actual passing predicted to pass (True Negative)

- Accuracy – percentage of correctly predicted outcomes
- Precision – ability of the model to predict failed outcomes
- Recall – strength of the model to predict failed outcomes
- F-measure (F1) – indicator of model effectiveness

## EXPERIMENT RESULTS

Experiment	TP	TN	FP	FN	Accuracy	Precision	Recall	F1
1: ROS	2	130	6	8	90.4%	0.25	0.20	0.22
2: SMOTE	4	136	4	2	95.9%	0.50	0.67	0.57
3: SMOTE	8	123	0	15	89.7%	1.00	0.35	0.52
4: RUS	6	81	2	57	59.6%	0.75	0.10	0.17
5: RUS	6	81	2	57	59.6%	0.75	0.10	0.17
6: SMOTE	1	136	7	2	93.8%	0.13	0.33	0.18
7: SMOTE	1	136	7	2	93.8%	0.13	0.33	0.18
8: SMOTE	1	136	7	2	93.8%	0.13	0.33	0.18

Results using 2015 matriculating year as the validation dataset (n=146, 8 failed Step 1 observations). RUS = Random Under Sampling, ROS = Random Over Sampling, SMOTE = Synthetic Minority Oversampling Technique, F1 = indicator of model effectiveness.