Simulation Study Protocol

Simulation(s) to assess the impact of class imbalance corrections on calibration of prediction models.

Alex Carriero

November 14, 2022

1 ADEMP

1.1 Aim

We aim to assess the impact of common class imbalance corrections on the performance of clinical prediction models. In this study, we will systematically compare the effect of imbalance corrections on the performance of several clinical prediction models developed using various machine learning methods. All models are developed for dichotomous risk prediction. In particular, we aim to determine if class imbalance corrections can lead to improved model performance without compromising model calibration.

1.2 Data-Generating Mechanisms

~ insert data generating mechanism details ~

- discuss how control of auc is maintained, present mean (sd) auc for generated data sets
- how are the interaction effects achieved

Imbalanced data will be simulated to reflect 27 (3 x 3 x 3) unique scenarios. This is achieved by varying the following three properties of the data: number of predictors, event fraction and sample size. The number of predictors will vary through the set {8,16,32} and event fraction, through the set {0.5, 0.2, 0.02}. The minimum sample size for the prediction model (N) will be computed according to formulae presented in Riley et al. (2022). Sample size will then vary through the set $\{\frac{1}{2}$N, N and 2N$\}$.

Table 1: Summary of factors to be varied in data simulation.

Factor	**Levels**
No. of predictors	8, 16, 32
Event fraction	0.5, 0.2, 0.02
Sample Size	$\frac{1}{2}N$, N, $2N$

* N represents the minimum sample size for the prediction model.

Under each scenario, 2000 data sets will be generated. Each data set will be comprised of training and test data with a ratio of 10:1. For each data set, five imbalance corrections will be applied to the training set. Subsequently, six prediction models will be developed for each of the imbalance corrected training sets. In other words, for each data set, the training data will be used to train 5 x 6 = 30 imbalance correction - prediction model combinations. Finally, out-of sample predictive performance will be assessed for each imbalance correction - prediction model combination using the test data.

1.3 Estimands

The focus of this study is the out-of-sample predictive performance of clinical prediction models for dichotomous risk prediction.

1.4 Methods

To investigate the effect of common class imbalance corrections on model performance, a full-factorial simulation design will be implemented. Five imbalance corrections will be implemented for each of six classification algorithms. The classification algorithms and imbalance corrections we will include in our simulation are detailed in Tables 2 and 3 respectively.

All models will be trained using training data sets. Out-of-sample performance will be then be assessed using the test data.

Table 2: Summary of class imbalance corrections to be implemented.

Imbalance Correction	**R Package**	**Python Library**
Random Under Sampling	ROSE	imblearn
Random Over Sampling	ROSE	imblearn
SMOTE	smotefamily	imblearn
SMOTE-ENN	*IRIC	imblearn
None	—	—

* IRIC package not available on CRAN

Table 3: Summary of classification algorithms to be implemented.

Method	**R Package**	**Python Library**
Logistic Regression	glmnet	scikit-learn
Support Vector Machine	e1701	scikit-learn
Random Forest	randomForest	scikit-learn
XG Boost	xgboost	xgboost
RUSBoost	ebmc	imblearn
EasyEnsemble	*IRIC	imblearn

* IRIC package not available on CRAN

1.5 Performance Measures

Out-of-sample model performance will be assessed using measures of discrimination, accuracy and calibration. Discrimination will be measured by area under the receiver operator curve (ΔC-statistic). Four measures of accuracy will be reported: overall accuracy, Matthew's correlation coefficient, sensitivity and specificity. Finally, calibration will be measured in terms of calibration intercept and slope.

2 Error Handling

References

Riley, Richard D., Gary S. Collins, Joie Ensor, Lucinda Archer, Sarah Booth, Sarwar I. Mozumder, Mark J. Rutherford, Maarten van Smeden, Paul C. Lambert, and Kym I. E. Snell. 2022. "Minimum Sample Size Calculations for External Validation of a Clinical Prediction Model with a Time-to-Event Outcome." *Statistics in Medicine* 41 (7): 1280–95.