**EdX Data Science Capstone Project**

**Predictive Models to Identify Risk of In-Hospital Mortality after an Abnormal In-patient Stress Test-Analysis of a Nationwide Sample**

**Author: Saurav Chatterjee MD**

**Contents:**

1. **Introduction 2**
2. **Methods 3**
3. **Results 4**
4. **Limitations 5**
5. **Conclusion 5**
6. **Figures 6**
7. **Table 8**
8. **References 9**

**Introduction:**

The goals of treating patients with stable coronary disease are to reduce their risk of death and ischemic events and to improve their quality of life[1]. All patients with coronary disease should be treated with guideline-based medical therapy (hereafter, medical therapy) to achieve these objectives [2]. But it remains unknown if a strategy of performing revascularization with either percutaneous coronary interventions (PCI) or coronary artery bypass surgery (CABG) decreases risk of dying from coronary artery disease.

Several theories have been advanced to explain why previous strategy trials involving patients with stable coronary disease have not shown a decrease in death or myocardial infarction with revascularization [3-6]. In trials requiring angiographic evidence of obstructive coronary disease, patients with high-risk anatomical features may have been excluded and knowledge of the anatomy may have led to revascularization in patients who were randomly assigned to a conservative strategy. Previous studies allowed the enrollment of patients with any level of ischemia, which resulted in a minority of patients with moderate or severe ischemia for whom an invasive strategy might have been most beneficial. In a single-center observational study involving 10,627 patients, the incidence of death from cardiac causes was lower among those with at least 10% ischemia on myocardial perfusion imaging who underwent early revascularization than among those who did not undergo revascularization [7]. The recent International Study of Comparative Health Effectiveness with Medical and Invasive Approaches (ISCHEMIA) trial was done to determine the effect of adding cardiac catheterization (hereafter, angiography) and revascularization when feasible to medical therapy in patients with stable coronary disease and moderate or severe ischemia [8]. However, the trial failed to show a decrease in risk of mortality with adoption of this strategy [9]. Most of the stress tests done for the trial were out-patient stress tests-ie, performed in physician’s offices and hospitals on out-patient basis with patient having been discharged the same day. It remains unknown if the results of the same trial are relevant for patients admitted to a hospital and receive a stress test as in-patient. Nor are factors associated with mortality in this population well known.

Methods:

HCUP’s National Inpatient Sample and Nationwide Inpatient Sample data between January 2008 and December 2017 was used. These data set uses 2 coding systems, International Classification of Diseases-Ninth revision (ICD-9), and -Tenth revision (ICD-10) and the Clinical Classifications Software (CCS) developed at Agency for Healthcare Research and Quality (AHRQ) [https://www.hcup-us.ahrq.gov/nisoverview.jsp] . Our target population was adults admitted to the hospital, and undergoing a stress test as a primary procedure. Patients younger than 18 years of age or missing age information were excluded. Patients with ‘abnormal stress test’ were identified by the following ICD-9 codes: 794.39, and ICD-10 codes: R94.39. This population comprised the dataset for analysis. The outcome variable for this analysis was death in-hospital, or discharged from the hospital (variable DIED). All visits with missing information on the mortality status of the patient were removed. R and R-studio was used to write scripts and perform all analyses. The dataset and code for reproducing the results is available at [https://github.com/sauravsphs1981/DataScience-ML-Project/upload].

### 2.2. Variables selection and pre-processing

All variables included in all 2008–2017 NIS databases were considered, then filtered out unrelated and confounding variables. The dataset was then split randomly and evenly into a training and validation dataset. Briefly, the variables selected were grouped according to: 1) patient-level variables, including age, gender, race, etc; and 2) hospital-related information, such as hospital bed size, teaching status, hospital region and hospital ownership, etc; and 3) procedures undergone by the patients including PCI/CABG/ pacemaker-implantation and defibrillator implants. All nominal/categorical variables were recoded as binary. Codebooks for variables included in github repositiory [https://github.com/sauravsphs1981/DataScience-ML-Project/upload]

### 2.3. Machine learning models

We used and compared both linear and non-linear algorithms to identify an optimal model associated with the outcome of interest.

**Linear models**: Logistic regression is commonly used in clinical research, it converts linear regression to a binary classifier with sigmoid function, we here used both OLS(Ordinary Least Squares) and ridge logistic regression, which employs a L2 norm to mitigate overfitting. Additionally, Least Absolute Shrinkage and Selection Operator (lasso) was used to hypertune parameters and identify/select variables with optimal penalization (lambda). Elastic net is a penalized linear regression model that includes both the L1 and L2 penalties during training-elastic net

**Non-linear model**: Random Forest tree is an ensemble of decision trees, which categorizes data by setting up decision rules, with boosting algorithm that extracts important information from inputs, and then does classification using the extracted features. Model parameters were tuned by three-fold cross-validation and grid-search.

### 2.4. Model evaluation

The loss function used in this study is the area under the receiver operating characteristic (AUROC) due to its robustness on imbalanced datasets. Models were trained on 50 % training data and validations were executed through 50 % test set created by stratified split. Finally, variable importance was evaluated by bootstrapped average decrease in impurity from random forest and by means of coefficients from ridge logistic regression.

## 3. Results

The final sample size for the analysis in this study was N=26,076, among which 23 were declared as deceased (0.088%)-all patients had a resulted ‘abnormal cardiac stress test’ . In a multivariable logistic regression model in the training dataset, mortality was significantly (p < 0.01) associated with a variety of characteristics including older age, presence of diabetes, hypothyroid, abnormal electrolytes, history of a stroke, having treatment at a teaching hospital, heart failure and those transferred in from a different institution. Conversely, we found that patients that had insurance and a history of peripheral vascular disease had a lower risk of in-hospital mortality. The model using lasso, additionally identified obesity as a variable significantly associated with lower risk of in-hospital mortality. The elastic net regression additionally identified implantation of pacemakers, and history of chronic obstructive pulmonary disease (COPD) as associative variables.

Figure 1 shows the receiver operating curve for all models, which were based on 50 % test set (N=13,078). AUROC are reported by Table 1, the greatest value of each column was highlighted with bold. We found that traditional logistic regression (OLS) achieved the highest value of AUROC (0.875), while the other models had smaller AUROC. The derived models had good consistency with similar AUROC achieved in the testing dataset (0.918). Interpretation of linear and non-linear models (Figure 2) shows the most influential predictors of mortality using random forest, and ridge logistic regression models included obesity, teaching-status and PCI etc. The models performed with similar AUROCs when applied on the validation dataset.

Limitations:

The study has several limitations. One of the limitations of the current work is that discharge-level data was used, which might lead to multiple observations (admissions) per patient. Furthermore, the data contained no subjective content (e.g. clinical notes) which is needed to identify inpatients with high mortality risk. Another limitation is the lack of clinical and laboratory test variables. In terms of machine learning model development, only four of them were evaluated, more advanced learning algorithms (e.g., ensemble) have potential to improve the performance. Some imbalanced data processing techniques (e.g., under sampling, over sampling) which not used in the current analysis might also enhance the model predictive power.

Conclusion: Traditional OLS logistic regression appears to derive the best fitted model for identifying variables associated with in-hospital death in patients with a positive in-patient stress test, in comparison to linear and non-linear models derived with machine learning techniques.

Figure 1. AUROC for Models: Panels A: OLS Logistic Regression B: LASSO C: Ridge Logistic Regression D: Elastic Net

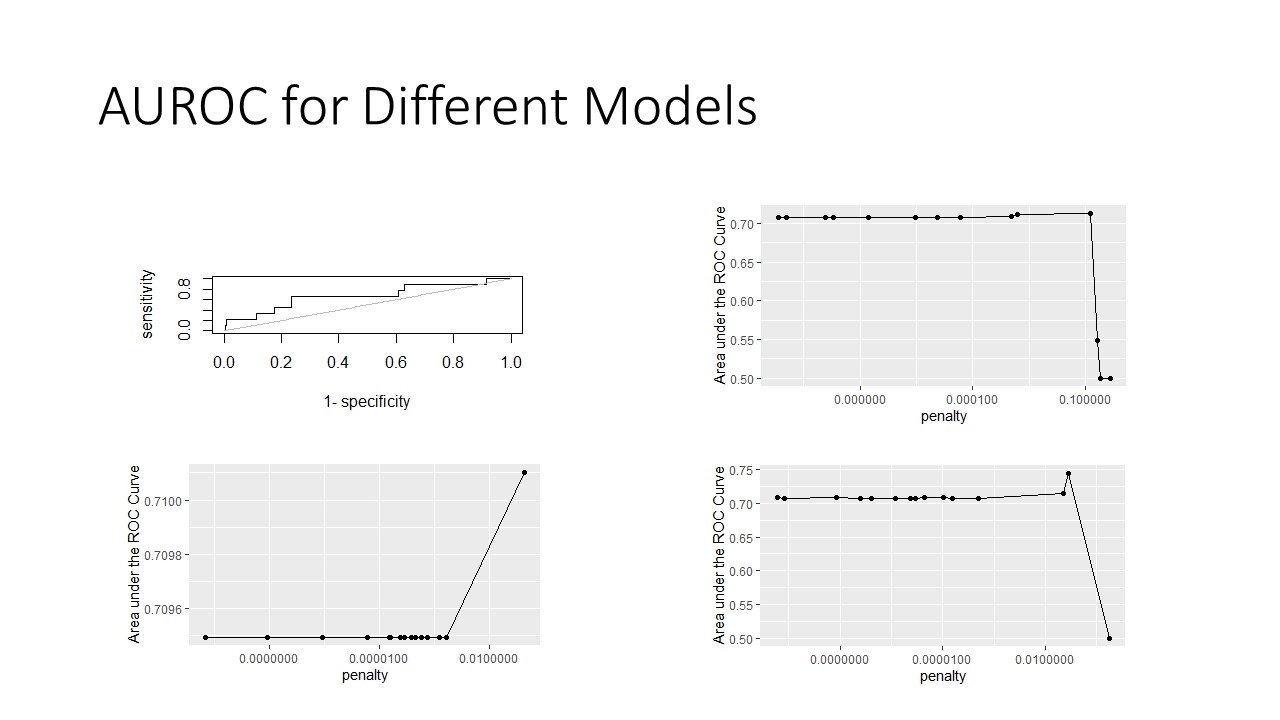


Figure 2. Variable Importance Plot

Chart, scatter chart

Description automatically generated

Table 1. AUROC(c-statistic) for Different Models

|  |  |  |
| --- | --- | --- |
| Group | Model | AUROC |
| Linear | Traditional OLS Logistic Regression | 0.897 |
|  | Lasso | 0.721 |
|  | Ridge Logistic Regression | 0.712 |
|  | Elastic Net | 0.75 |
| Non-Linear | Random Forest | 0.64 |

References:

1. Fihn SD, Gardin JM, Abrams J, et al. 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease: executive summary: a report of the American College of Cardiology Foundation/American Heart Association task force on practice guidelines, and the American College of Physicians, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. J Am Coll Cardiol 2012;60:2564-2603.
   * [Crossref. opens in new tab](https://www-nejm-org.libraryproxy.quinnipiac.edu/servlet/linkout?suffix=r1&dbid=16&doi=10.1056%2FNEJMoa1915922&key=10.1016%2Fj.jacc.2012.07.012)
   * [Medline. opens in new tab](https://www-nejm-org.libraryproxy.quinnipiac.edu/servlet/linkout?suffix=r1&dbid=8&doi=10.1056%2FNEJMoa1915922&key=23182124)

[Google Scholar. opens in new tab](http://scholar.google.com.libraryproxy.quinnipiac.edu/scholar_lookup?hl=en&publication_year=2012&pages=2564-2603&author=SD+Fihnauthor=JM+Gardinauthor=J+Abrams&title=2012+ACCF%2FAHA%2FACP%2FAATS%2FPCNA%2FSCAI%2FSTS+guideline+for+the+diagnosis+and+management+of+patients+with+stable+ischemic+heart+disease%3A+executive+summary%3A+a+report+of+the+American+College+of+Cardiology+Foundation%2FAmerican+Heart+Association+task+force+on+practice+guidelines%2C+and+the+American+College+of+Physicians%2C+American+Association+for+Thoracic+Surgery%2C+Preventive+Cardiovascular+Nurses+Association%2C+Society+for+Cardiovascular+Angiography+and+Interventions%2C+and+Society+of+Thoracic+Surgeons.)

1. **2.**Knuuti J, Wijns W, Saraste A, et al. ESC guidelines for the diagnosis and management of chronic coronary syndromes. Eur Heart J 2020;41:407-477.
   * [Crossref. opens in new tab](https://www-nejm-org.libraryproxy.quinnipiac.edu/servlet/linkout?suffix=r2&dbid=16&doi=10.1056%2FNEJMoa1915922&key=10.1093%2Feurheartj%2Fehz425)
   * [Web of Science. opens in new tab](https://www-nejm-org.libraryproxy.quinnipiac.edu/servlet/linkout?suffix=r2&dbid=128&doi=10.1056%2FNEJMoa1915922&key=000521321000017)
   * [Medline. opens in new tab](https://www-nejm-org.libraryproxy.quinnipiac.edu/servlet/linkout?suffix=r2&dbid=8&doi=10.1056%2FNEJMoa1915922&key=31504439)

[Google Scholar. opens in new tab](http://scholar.google.com.libraryproxy.quinnipiac.edu/scholar_lookup?hl=en&publication_year=2020&pages=407-477&author=J+Knuutiauthor=W+Wijnsauthor=A+Saraste&title=ESC+guidelines+for+the+diagnosis+and+management+of+chronic+coronary+syndromes.)

1. **3.**Boden WE, O’Rourke RA, Teo KK, et al. Optimal medical therapy with or without PCI for stable coronary disease. N Engl J Med 2007;356:1503-1516.
   * [Free Full Text](https://www-nejm-org.libraryproxy.quinnipiac.edu/doi/full/10.1056/NEJMoa070829)
   * [Web of Science. opens in new tab](https://www-nejm-org.libraryproxy.quinnipiac.edu/servlet/linkout?suffix=r3&dbid=128&doi=10.1056%2FNEJMoa1915922&key=000245624400004)
   * [Medline. opens in new tab](https://www-nejm-org.libraryproxy.quinnipiac.edu/servlet/linkout?suffix=r3&dbid=8&doi=10.1056%2FNEJMoa1915922&key=17387127)

[Google Scholar. opens in new tab](http://scholar.google.com.libraryproxy.quinnipiac.edu/scholar_lookup?hl=en&publication_year=2007&pages=1503-1516&author=WE+Bodenauthor=RA+O%E2%80%99Rourkeauthor=KK+Teo&title=Optimal+medical+therapy+with+or+without+PCI+for+stable+coronary+disease.)

1. **4.**The BARI 2D Study Group. A randomized trial of therapies for type 2 diabetes and coronary artery disease. N Engl J Med 2009;360:2503-2515.
   * [Free Full Text](https://www-nejm-org.libraryproxy.quinnipiac.edu/doi/full/10.1056/NEJMoa0805796)
   * [Web of Science. opens in new tab](https://www-nejm-org.libraryproxy.quinnipiac.edu/servlet/linkout?suffix=r4&dbid=128&doi=10.1056%2FNEJMoa1915922&key=000266813800005)
   * [Medline. opens in new tab](https://www-nejm-org.libraryproxy.quinnipiac.edu/servlet/linkout?suffix=r4&dbid=8&doi=10.1056%2FNEJMoa1915922&key=19502645)

[Google Scholar. opens in new tab](http://scholar.google.com.libraryproxy.quinnipiac.edu/scholar_lookup?hl=en&publication_year=2009&pages=2503-2515&author=The+BARI+2D+Study+Group&title=A+randomized+trial+of+therapies+for+type+2+diabetes+and+coronary+artery+disease.)

1. **5.**De Bruyne B, Pijls NHJ, Kalesan B, et al. Fractional flow reserve–guided PCI versus medical therapy in stable coronary disease. N Engl J Med 2012;367:991-1001.
   * [Free Full Text](https://www-nejm-org.libraryproxy.quinnipiac.edu/doi/full/10.1056/NEJMoa1205361)
   * [Web of Science. opens in new tab](https://www-nejm-org.libraryproxy.quinnipiac.edu/servlet/linkout?suffix=r5&dbid=128&doi=10.1056%2FNEJMoa1915922&key=000308649100005)
   * [Medline. opens in new tab](https://www-nejm-org.libraryproxy.quinnipiac.edu/servlet/linkout?suffix=r5&dbid=8&doi=10.1056%2FNEJMoa1915922&key=22924638)

[Google Scholar. opens in new tab](http://scholar.google.com.libraryproxy.quinnipiac.edu/scholar_lookup?hl=en&publication_year=2012&pages=991-1001&author=B+De+Bruyneauthor=NHJ+Pijlsauthor=B+Kalesan&title=Fractional+flow+reserve%E2%80%93guided+PCI+versus+medical+therapy+in+stable+coronary+disease.)

1. **6.**Xaplanteris P, Fournier S, Pijls NHJ, et al. Five-year outcomes with PCI guided by fractional flow reserve. N Engl J Med 2018;379:250-259.
   * [Free Full Text](https://www-nejm-org.libraryproxy.quinnipiac.edu/doi/full/10.1056/NEJMoa1803538)
   * [Web of Science. opens in new tab](https://www-nejm-org.libraryproxy.quinnipiac.edu/servlet/linkout?suffix=r6&dbid=128&doi=10.1056%2FNEJMoa1915922&key=000439063900008)
   * [Medline. opens in new tab](https://www-nejm-org.libraryproxy.quinnipiac.edu/servlet/linkout?suffix=r6&dbid=8&doi=10.1056%2FNEJMoa1915922&key=29785878)

[Google Scholar. opens in new tab](http://scholar.google.com.libraryproxy.quinnipiac.edu/scholar_lookup?hl=en&publication_year=2018&pages=250-259&author=P+Xaplanterisauthor=S+Fournierauthor=NHJ+Pijls&title=Five-year+outcomes+with+PCI+guided+by+fractional+flow+reserve.)

1. **7.**Hachamovitch R, Hayes SW, Friedman JD, Cohen I, Berman DS. Comparison of the short-term survival benefit associated with revascularization compared with medical therapy in patients with no prior coronary artery disease undergoing stress myocardial perfusion single photon emission computed tomography. Circulation 2003;107:2900-2907.
   * [Crossref. opens in new tab](https://www-nejm-org.libraryproxy.quinnipiac.edu/servlet/linkout?suffix=r7&dbid=16&doi=10.1056%2FNEJMoa1915922&key=10.1161%2F01.CIR.0000072790.23090.41)
   * [Web of Science. opens in new tab](https://www-nejm-org.libraryproxy.quinnipiac.edu/servlet/linkout?suffix=r7&dbid=128&doi=10.1056%2FNEJMoa1915922&key=000183612700007)
   * [Medline. opens in new tab](https://www-nejm-org.libraryproxy.quinnipiac.edu/servlet/linkout?suffix=r7&dbid=8&doi=10.1056%2FNEJMoa1915922&key=12771008)

[Google Scholar. opens in new tab](http://scholar.google.com.libraryproxy.quinnipiac.edu/scholar_lookup?hl=en&publication_year=2003&pages=2900-2907&author=R+Hachamovitchauthor=SW+Hayesauthor=JD+Friedmanauthor=I+Cohenauthor=DS+Berman&title=Comparison+of+the+short-term+survival+benefit+associated+with+revascularization+compared+with+medical+therapy+in+patients+with+no+prior+coronary+artery+disease+undergoing+stress+myocardial+perfusion+single+photon+emission+computed+tomography.)

1. **8.**ISCHEMIA Trial Research Group, Maron DJ, Hochman JS, et al. International Study of Comparative Health Effectiveness with Medical and Invasive Approaches (ISCHEMIA) trial: rationale and design. Am Heart J 2018;201:124-135.
   * [Crossref. opens in new tab](https://www-nejm-org.libraryproxy.quinnipiac.edu/servlet/linkout?suffix=r8&dbid=16&doi=10.1056%2FNEJMoa1915922&key=10.1016%2Fj.ahj.2018.04.011)
   * [Web of Science. opens in new tab](https://www-nejm-org.libraryproxy.quinnipiac.edu/servlet/linkout?suffix=r8&dbid=128&doi=10.1056%2FNEJMoa1915922&key=000436562100017)
   * [Medline. opens in new tab](https://www-nejm-org.libraryproxy.quinnipiac.edu/servlet/linkout?suffix=r8&dbid=8&doi=10.1056%2FNEJMoa1915922&key=29778671)

[Google Scholar. opens in new tab](http://scholar.google.com.libraryproxy.quinnipiac.edu/scholar_lookup?hl=en&publication_year=2018&pages=124-135&author=ISCHEMIA+Trial+Research+Groupauthor=DJ+Maronauthor=JS+Hochman&title=International+Study+of+Comparative+Health+Effectiveness+with+Medical+and+Invasive+Approaches+%28ISCHEMIA%29+trial%3A+rationale+and+design.)

1. **9.**Stone GW, Hochman JS, Williams DO, et al. Medical therapy with versus without revascularization in stable patients with moderate and severe ischemia: the case for community equipoise. J Am Coll Cardiol 2016;67:81-99.