Evaluation Decisions for CST and MSO: The Impact of Evaluator and Defendant Characteristics

Kevin Nguyen

October 6, 2022

1 Executive Summary

The court will seek an evaluation if the defendant's mental health is in doubt during a trial. Competency to Stand Trial (CST) and Mental State at the Time of the Offense (MSO) evaluations are both used in Virginia. The CST assesses if the defendant is able to comprehend how the court process works. The MSO assesses the individual's mental state and sanity at the time of the offense. Trial verdicts and the lives of defendants are significantly impacted by these evaluations. It is critical to assess the evaluators' tendencies because these could influence their evaluation opinion. Both of the client's questions were properly addressed in this report using logistic regression. It is unlikely an evaluator's propensity for evaluating defendants as competent in the CST evaluations have any correlation with his or her propensity for evaluating defendants as sane in the MSO evaluations. Moreover, the evaluator's discipline and defendant's age are likely to be significant in predicting an evaluation decision for CST, and the evaluator's employer and evaluation setting are likely to be significant in predicting an evaluation decision for MSO.

2 Introduction

2.1 General Background

When the defendant's mental health is in question during a trial, the court will require an evaluation. In Virginia, the Competency to Stand Trial (CST) and Mental State at the Time of the Offense (MSO) evaluations are both utilized. Using a combination of interviews and testimonies, forensic psychologists primarily conduct these evaluations. The CST evaluates whether the defendant is capable of understanding how the court proceedings operate. The

MSO evaluates the individual's mental state and sanity at the time of the offense. These evaluations have a significant impact on the results of trials and the lives of defendants. Therefore, it is important to take into consideration the evaluators' tendencies because of how they may influence their evaluation opinion.

2.2 Objectives

This report's primary objective is to provide answers to the following two questions:

1. Does an evaluator's propensity for evaluating defendants as competent in the CST evaluations have any correlation with his or her propensity for evaluating defendants as sane in the MSO evaluations?

2. Which characteristics of evaluators and defendants are significant in predicting an evaluation decision for CST and MSO?

2.3 Data Description

The client's raw data consists of two distinct data sets: one pertaining to CST and the other to MSO. The CST data set contains 1,126 observations and 17 columns, whereas the MSO data set contains 926 observations and 20 columns.

The main variables of interest in the CST and MSO data sets are as follows:

- **Discipline**: The evaluator's discipline.
- **Employer**: The evaluator's employer.
- **Training**: The evaluator's training year.
- **Postdoc**: Whether the evaluator completed a postdoctoral fellowship.
- **Boarded**: Whether the evaluator has received board certification.
  - If yes, values will contain the organization that granted the evaluator his or her board certification.
- **EvalGender**: The evaluator's gender.
- **ReportDate**: The report's publication date.
- **Setting**: The evaluation setting.

- **Opinion**: The evaluator's opinion.
- **DefendantSex**: The defendant's gender.
- **DefendantAge**: The defendant's age.
- **DefendantRace**: The defendant's race.

3 Project Approach

3.1 Data Cleaning

Processing the raw data involved the following seven actions:

1. Extraneous columns were removed from the CST and MSO data sets.
2. Trailing spaces were removed from character strings in the CST and MSO data sets.
3. ReportDate was converted to a date format.
4. Opinion was converted to a binary variable to produce a new column.
5. For EvalGender and DefendantRace, NA values were imputed with the value Unknown.
6. For DefendantAge, NA values were imputed with the median.
7. Several columns in the CST and MSO data sets were converted to a factor data type.

3.1.1 Removal of Extraneous Columns

Evaluator, ForensicProgram, ReportID, BasicEvalType, and EvaluationType were removed from the CST data set. Furthermore, Evaluator, ForensicProgram, ReportID, BasicEvalType, EvaluationType, ThresholdMIatOffense, NatureCharacterConseq, Wrongfulness, and IrrImpulse were removed from the MSO data set.

3.1.2 Removal of Trailing Spaces

Trailing space is all the whitespace at the end of a character string that is not followed by any other characters. Many character strings had trailing spaces when we used the `read.spss()` function to import the CST and MSO data sets. Although we are unaware of the reason behind this result, we suspect it may be related to how SPSS stores data.

3.1.3 Conversion to Date Format for ReportDate

Dates are stored in SPSS as the number of seconds since October 14, 1582. For easier interpretability, we would prefer the values in ReportDate be in the Year-Month-Day format. Furthermore, the values in ReportDate are in the tens of billions, which might cause modeling issues due to their magnitude. As there are 86,400 seconds in a day, we first divided each value in ReportDate by 86,400. The column was then converted into a date format using the `as.Date()` function with the argument `origin = "1582-10-14"`.

3.1.4 Conversion to Binary Variable for Opinion

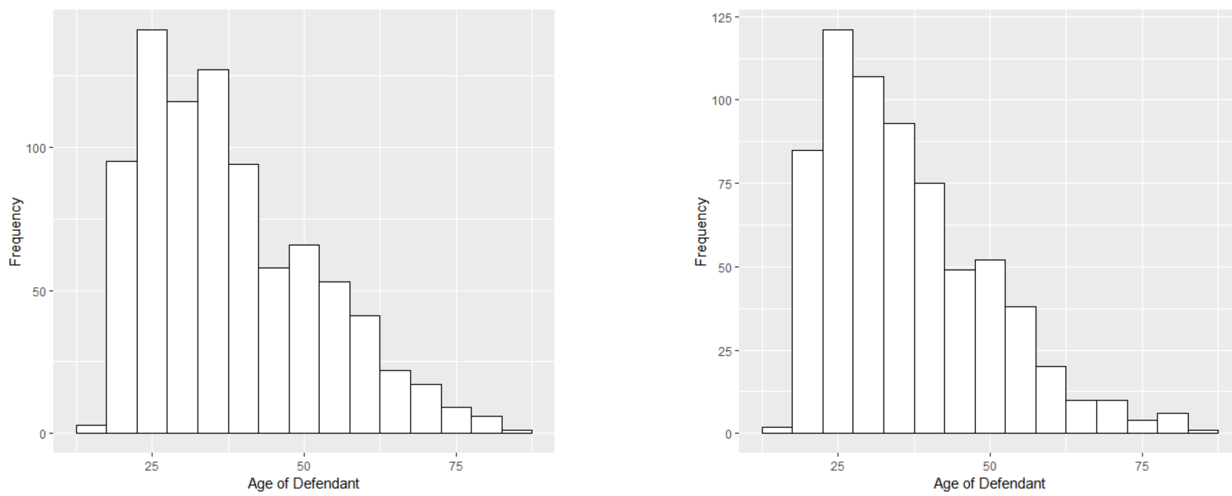
To apply logistic regression, which we will elaborate on later in the report, the response variable should be binary. Therefore, by using the `ifelse()` function, we created a new column called BinOpinion that converts Opinion to a binary variable. For the CST data set, a value is labeled as 1 if Opinion = "CST" and 0 otherwise. For the MSO data set, a value is labeled as 1 if Opinion = "Sane" and 0 otherwise.

3.1.5 Imputation of Unknown in EvalGender and DefendantRace

The CST data set contains 540 NA values for DefendantRace and 701 NA values for EvalGender. Additionally, the MSO data set contains 440 NA values for DefendantRace and 551 NA values for EvalGender. We believe we cannot ethically impute a race or gender for an individual. However, we do not want to omit the NA values either because doing so would drastically reduce the sample size for both data sets, which would most likely be undesirable for a multitude of reasons. Consequently, we determined the best course of action would be to impute the value "Unknown" for the NA values and treat it as an additional level for EvalGender and DefendantRace.

3.1.6 Imputation of Median in DefendantAge

The CST data set contains 227 NA values for DefendantRace. Additionally, the MSO data set contains 253 NA values for DefendantRace. Even though there are less NA values in DefendantRace than in EvalGender and DefendantRace, omitting them will still substantially reduce the sample size. We decided imputing the mean or median would most likely be the best option. Further examination of DefendantAge reveals the variable has a right-skewed distribution in both data sets, as shown in Figure 1. Consequently, we opted to impute the median rather than the mean.



(a) CST (b) MSO

Figure 1: Distributions of DefendantAge

3.1.7 Conversion to Factor Data Type for Several Columns

EvalGender, Boarded, Setting, DefendantSex, BinOpinion, and DefendantRace were converted to a factor data type using the `as.factor()` function in both data sets because we wish to use them as categorical variables.

3.2 Methodology

We implemented logistic regression, a supervised machine learning model, to answer the client's questions. Based on a given data set of independent variables, logistic regression calculates the likelihood that an event will occur. A logit transformation is applied on the odds, the probability of success divided by the probability of failure, in logistic regression. The transformed odds are often referred to as log odds. In this model, the beta parameters, or coefficients, are often estimated via maximum likelihood estimation. This algorithm evaluates various beta values over a number of iterations to get the best match for the log odds. To determine the most accurate parameter estimate, logistic regression aims to maximize the log likelihood function, which is produced by all of these iterations. Following the identification of the best coefficients, the conditional probabilities for each observation can be computed, logged, and summed to provide a predicted probability. There are five assumptions of logistic regression:

1. The response variable is binary.

2. The observations are independent.

3. The explanatory variables do not exhibit significant multicollinearity.

4. There are no extreme outliers.

5. There is an adequate sample size.

$$Ln\left(\frac{P}{1-P}\right)=\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_k X_k$$

Figure 2: Logit Function

4 Results

First, we checked to see if there were any NA values left in each data set after the data cleaning. We found there were 63 NA values remaining in the CST data set and 48 NA values remaining in the MSO data set. Due to the potential for modeling difficulties, we opted to remove observations containing NA values in any column.

We then fitted a full logistic regression model for each data set, with BinOpinion as the response, using the `glm()` function with the argument `family = binomial`:

1. $\hat{y}_{CST} = \hat{\beta}_0 + \hat{\beta}_1 Discipline + \hat{\beta}_2 Employer + \hat{\beta}_3 Training + \hat{\beta}_4 Postdoc + \hat{\beta}_5 Boarded + \hat{\beta}_6 EvalGender + \hat{\beta}_7 ReportDate + \hat{\beta}_8 Setting + \hat{\beta}_9 DefendantSex + \hat{\beta}_{10} DefendantRace + \hat{\beta}_{11} DefendantRace$

2. $\hat{y}_{MSO} = \hat{\beta}_0 + \hat{\beta}_1 Discipline + \hat{\beta}_2 Employer + \hat{\beta}_3 Training + \hat{\beta}_4 Postdoc + \hat{\beta}_5 Boarded + \hat{\beta}_6 EvalGender + \hat{\beta}_7 ReportDate + \hat{\beta}_8 Setting + \hat{\beta}_9 DefendantSex + \hat{\beta}_{10} DefendantRace + \hat{\beta}_{11} DefendantRace$

As previously stated in the Methodology section, the explanatory variables should not exhibit significant multicollinearity. We tested for multicollinearity in the two full models using the `vif()` function. The GVIF indicates the increase in the variance of a regression coefficient due to its collinearity with the other variables. A VIF threshold of 10 is the standard practice. However, we opted for a more conservative VIF threshold of 5 due to the multicollinearity assumption of logistic regression. As shown in Figure 3, Employer, Boarded, EvalGender, and Setting in the CST data set as well as Discipline, Boarded, and EvalGender in the MSO data set have a GVIF value greater than 5.

	GVIF	Df	GVIF^(1/(2*Df))
Discipline	2.201	1	1.484
Employer	5.629	2	1.540
Training	1.445	1	1.202
Postdoc	1.854	1	1.362
Boarded	13.057	3	1.535
EvalGender	6.483	2	1.596
ReportDate	1.067	1	1.033
Setting	5.200	2	1.510
DefendantSex	1.022	1	1.011
DefendantAge	1.102	1	1.050
DefendantRace	1.797	5	1.060

(a) $\hat{y}_{CST}$

	GVIF	Df	GVIF^(1/(2*Df))
Discipline	8.172	1	2.859
Employer	2.942	2	1.310
Training	1.291	1	1.136
Postdoc	1.375	1	1.173
Boarded	53.269	4	1.644
EvalGender	6.551	2	1.600
ReportDate	1.082	1	1.040
Setting	2.616	2	1.272
DefendantSex	1.025	2	1.006
DefendantAge	1.079	1	1.039
DefendantRace	2.042	5	1.074

(b) $\hat{y}_{MSO}$

Figure 3: VIF Output for Full Models

Consequently, we dropped these explanatory variables from the full models and fitted a reduced logistic regression model for each data set:

1. $\hat{y}^*_{CST} = \hat{\beta}_0 + \hat{\beta}_1 Discipline + \hat{\beta}_2 Training + \hat{\beta}_3 Postdoc + \hat{\beta}_4 ReportDate + \hat{\beta}_5 DefendantSex + \hat{\beta}_6 DefendantRace + \hat{\beta}_7 DefendantRace$

2. $\hat{y}^*_{MSO} = \hat{\beta}_0 + \hat{\beta}_1 Employer + \hat{\beta}_2 Training + \hat{\beta}_3 Postdoc + \hat{\beta}_4 ReportDate + \hat{\beta}_5 Setting + \hat{\beta}_6 DefendantSex + \hat{\beta}_7 DefendantRace + \hat{\beta}_8 DefendantRace$

We tested for multicollinearity in the two reduced models using the `vif()` function. None of the explanatory variables have GVIF values above the VIF threshold of 5, as depicted in Figure 4. Thus, there is no need to further reduce the models.

	GVIF	Df	GVIF^(1/(2*Df))
Discipline	1.018	1	1.009
Training	1.379	1	1.175
Postdoc	1.391	1	1.179
ReportDate	1.014	1	1.007
DefendantSex	1.011	1	1.006
DefendantAge	1.078	1	1.038
DefendantRace	1.117	5	1.011

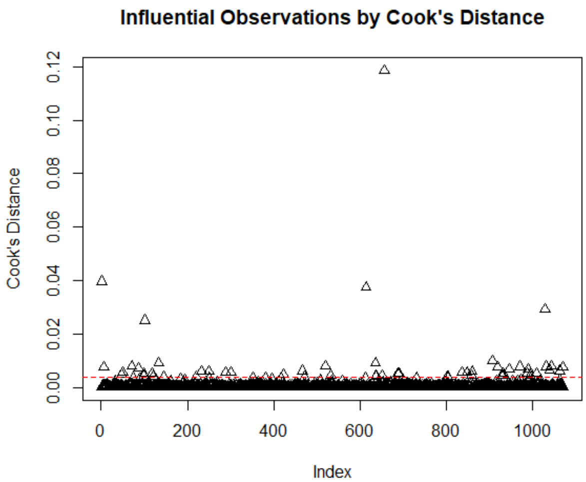
(a) $\hat{y}^*_{CST}$

	GVIF	Df	GVIF^(1/(2*Df))
Employer	2.450	2	1.251
Training	1.220	1	1.105
Postdoc	1.300	1	1.140
ReportDate	1.038	1	1.019
Setting	2.322	2	1.234
DefendantSex	1.021	2	1.005
DefendantAge	1.071	1	1.035
DefendantRace	1.200	5	1.018

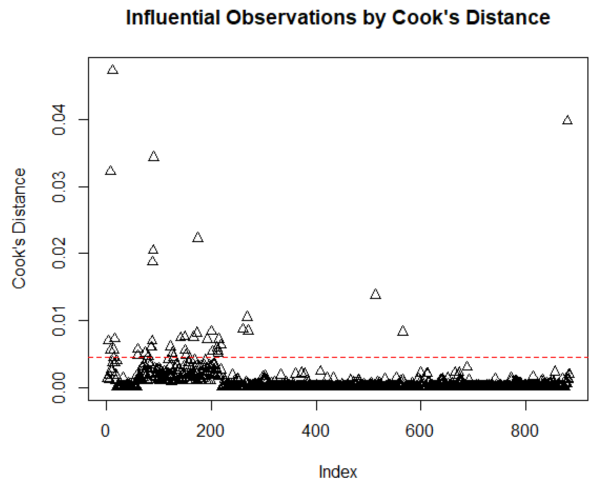
(b) $\hat{y}^*_{MSO}$

Figure 4: VIF Output for Reduced Models

Next, we checked for extreme outliers in both data sets using the `cooks.distance()` function because there should be no extreme outliers, as aforementioned in the Methodology section. In regression analysis, Cook's distance is used to detect influential observations that may have a detrimental impact on a regression model. It measures how the removal of a single observation affects all of the fitted values in the model. An observation with a large Cook's Distance value suggests that it has a significant influence on the fitted values. Any observation with a Cook's Distance exceeding 4/n, where n is the sample size, is generally regarded as an outlier. In Figure 5, a triangle represents the Cook's Distance for an observation. Additionally, the red horizontal line indicates the Cook's Distance threshold of 4/n. Any observations above this line are considered to be extreme outliers. Figure 5 illustrates both reduced models have several observations above the line. Therefore, we identified these observations and created a subset without them for the CST and MSO data sets.



(a) $\hat{y}^*_{CST}$



(b) $\hat{y}^*_{MSO}$

Figure 5: Cook's Distance Plots for Reduced Models

Using the subset rather than the entire data set, we once more fitted a reduced logistic regression model for each data set. Figure 6a reveals that DisciplinePsychiatrist and DefendantAge are statistically significant at a significance level of 0.05.

- For every one unit change in DefendantAge, the log odds of evaluating defendants as competant decreases by 0.017.
- When the evaluator is a psychiatrist, as opposed to when the evaluator is a psychologist, increases the log odds of evaluating defendants as competent by 0.693.

Furthermore, Figure 6b reveals that EmployerDBHDS, SettingOutpatient, and SettingUnknown are statistically significant at a significance level of 0.05.

- When the evaluator's employer is the DBHDS (Department of Behavioral Health and Developmental Services), as opposed to when the evaluator's employer is private, the log odds of evaluating defendants as sane increases by 0.672.

- When the evaluation setting is outpatient, as opposed to when the evaluation's setting is inpatient, the log odds of evaluating defendants as sane increases by 0.946.

- When the evaluation setting is unknown, as opposed to when the evaluation's setting is inpatient, the log odds of evaluating defendants as sane increases by 0.94.

```
Call:
glm(formula = BinOpinion ~ . - Employer - Boarded - EvalGender -
    Setting, family = "binomial", data = CST_noout)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7422  -1.3654   0.8787   0.9685   1.4425

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)              -6.927      9.613  -0.721    0.471
DisciplinePsychiatrist    0.693      0.417   1.662    0.096 .
Training                  0.000      0.000   0.558    0.577
PostdocYes                0.362      0.275   1.316    0.188
ReportDate                0.000      0.001   0.917    0.359
DefendantSexMale          0.033      0.152   0.221    0.825
DefendantAge             -0.017      0.005  -3.214    0.001 ***
DefendantRaceBlack       -1.077      1.126  -0.956    0.339
DefendantRaceHispanic    -1.615      1.219  -1.325    0.185
DefendantRaceOther      -15.037    378.387  -0.040    0.968
DefendantRaceUnknown     -0.935      1.125  -0.831    0.406
DefendantRaceWhite       -0.683      1.132  -0.604    0.546
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1348.0  on 1019  degrees of freedom
Residual deviance: 1324.7  on 1008  degrees of freedom
AIC: 1348.7

Number of Fisher Scoring iterations: 12
```

(a) $\hat{y}^*_{CST}$

```
Call:
glm(formula = BinOpinion ~ . - Discipline - Boarded - EvalGender,
    family = "binomial", data = MSO_noout)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1359   0.5185   0.6141   0.6610   1.0423

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)              3.012     13.680   0.220    0.826
EmployerDBHDS            0.672      0.399   1.683    0.092 .
EmployerCSB             -0.224      0.482  -0.465    0.642
Training                 0.000      0.000  -0.837    0.402
PostdocYes              -0.138      0.453  -0.306    0.760
ReportDate               0.000      0.001  -0.100    0.920
SettingOutpatient        0.946      0.386   2.452    0.014 *
SettingUnknown           0.940      0.525   1.790    0.073 .
DefendantSexMale        -0.239      0.214  -1.114    0.265
DefendantSexTrans       12.756    882.743   0.014    0.988
DefendantAge            -0.012      0.008  -1.580    0.114
DefendantRaceBlack       0.069      1.142   0.061    0.952
DefendantRaceHispanic    0.752      1.565   0.481    0.631
DefendantRaceOther      13.265    623.962   0.021    0.983
DefendantRaceUnknown     0.301      1.139   0.264    0.792
DefendantRaceWhite       0.261      1.149   0.227    0.821
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 808.36  on 844  degrees of freedom
Residual deviance: 795.36  on 829  degrees of freedom
AIC: 827.36

Number of Fisher Scoring iterations: 13
```

(b) $\hat{y}^*_{MSO}$

Figure 6: Summary Output for Reduced Models

5 Conclusions

In this report, we are able to adequately answer both of the client's questions. In response to the first question, our analysis suggests there is unlikely to be any correlation between an evaluator's propensity for evaluating defendants as competent in the CST evaluations and his or her propensity for evaluating defendants as sane in the MSO evaluations because there is not a single explanatory variable that is statistically significant in both the CST and MSO reduced logistic regression models. In response to the second question, the evaluator's discipline and defendant's age are likely to be significant in predicting an evaluation decision for CST. Furthermore, the evaluator's employer and evaluation setting are likely to be significant in predicting an evaluation decision for MSO.