PREDICTIVE VARIABLES FOR SEXUAL CONDUCT VIOLATIONS IN INDIANA STATE PRISONS

Megan Lalioff

Thesis

Submitted in partial fulfillment of the requirements for the degree

Master of Arts in Applied Sociology

Department of Sociology

University of Indianapolis

October 10th, 2019

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Colleen E. Wynn, Ph.D.

Assistant Professor, Advisor

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Mary Moore, Ph.D.

Professor of Sociology, Reader

# Abstract

Sexual violence in prisons is not a new concept, but rather has been a defining characteristic of the American Criminal Justice System. There have been many efforts to lower the risk of sexual violence in prison, such as the passing of the Prison Rape Elimination Act (PREA) in 2003 by President George W. Bush. However, almost all current programs geared towards lowering prison rape are retroactive and focus on helping the victim cope or helping to reduce the risk of recidivism for sexual violent victims. There fails to exist analysis on the characteristics of perpetrations of sexual violence in prison that might have predictive capability for reducing the number of sexual conduct violations committed. By using intersectionality and symbolic interactionism to identify characteristics that are common among offenders that commit sexual violence in Indiana State Prisons, a predictive, dynamic model can be used to flag potential perpetrators of sexual conduct violations.

# Reflexive Statement

The inception of this research topic was from my work experience as a data scientist with the Indiana Department of Correction. My knowledge on the data and possible applications of data shaped how I would structure this project. However, my background as a sociologist and researcher ultimately was the deciding factor on developing the project as a thesis research project rather than a computer science project. The impact of the variables chosen to be part of the model was important to this project as each variable selected had a purpose in relation to the theories used. I have had prior experience working with a predictive model, which acted as a rough skeleton for me to add the specific information for my project. This might have caused some bias as I did not spend a significant amount of time researching other methodologies as I believe my prior experience was sufficient in determining that the method used in this project was well suited to answer the problem statement. Admittedly, I did not fully consider all the implications of my research at the beginning of the project. Upon presentation at an academic conference, fellow sociology researchers presented possible negative consequences of the model. The most pressing consequence was pertaining to the future use of the model in the hands of ill-trained users who might use the model as a way to discriminate against offenders before any violation has occurred. One researcher brought up the issue of using information on past behavior of individuals to predict the current/present behavior of different individuals. This is directly contrary to the field of predictive analytics, as that is the function of the discipline. However, this comment had an impact on the project as the sample size of the population was limited to make sure all “participants” were comparable.

# Statement of the Problem

Sexual misconduct is a well-known characteristic of the prison system. More than 80,000 offenders are victims of sexual violence each year; however, only about 8% of offenders ever self-report being victimized while incarcerated (Kubiak 2017). The women’s liberation movement has been somewhat successful in altering the perceptions of rape and domestic violence since the early 1960s, but this movement was exclusive to life outside of incarceration. Due to the persistent overlook of prison rape and violence, the perception of prison rape has remained unaltered since the beginning of the modern American Prison System. Many politicians and officers of the law believe prison rape has an important effect on prison life, such as acting as “just deserts” to punish the offender for their crime or they see rape as a natural deterrent to committing crimes (Capers 2011). It is important to note that prison is primarily male, with male prisons being much larger than female prisons and men incarcerated for typically more violent crimes than women. Even though men make up a far majority of the prison population, male rape is still overlooked and studied far less than female offender rape or sexual violence. The sample includes females in the original model, but the analysis includes a comparison between male and female populations inside the Indiana Department of Correction. Throughout current research, different terminology is used to describe prison rape. The definition in this study for sexual misconduct refers to any sexual act that is deemed inappropriate in a specific social context. Sexual harassment is any unwanted sexual attention, while sexual violence, also called sexual assault, is any unwanted sexual act that causes physical harm: rape.

# Introduction

The first federal government action taken to decrease sexual assault in prisons was the Prison Rape Elimination Act (PREA) passed in 2003. Before PREA was passed, there were some state legislation aimed at reducing prison sexual assault, but no uniform standard across states. However, a major weakness of PREA is that the act relies on reported rapes, which are highly under-reported (McFarlane 2011). The Act tries to reduce the risk of recidivism upon release from prison as offender victims of sexual assault have been found to be more likely to commit violent crimes after release (Vega 2011). The Act also aims to provide resources to victims of sexual assault in prisons and to encourage offenders to report the sexual assault; this is completely retroactive. Each facility is required to have a PREA officer or someone designated to making sure the facility complies with the criteria outlined in PREA (Vega 2011). PREA requires that each offender be assessed at intake at the facility to determine risk of being sexual assaulted, but this only happens once while at this facility and is not updated during the offender’s stay (Vega 2011). While PREA aims to focus on the victim primary focus of this study is on the perpetrator to try to predict the likelihood of the violation before it occurs.

There have been many studies surrounding prison rape and sexual violence, but there fails to be a working dynamic model to predict sexual misconducts before they occur in the prison setting (Man 2001, Kubiak 2017, Worthen 2017). The most advanced predictive model related to sexual misconducts is geared towards sex offender recidivism. Almost all programming surrounding sexual misconducts in prison relates to reducing recidivism risk, rather than reducing the actual act of sexual misconduct itself. The Indiana Department of Correction uses a Static-99 form that is completed upon intake to assess risk of sexual violent recidivism upon release. The findings of this research will allow for further understanding what variables are correlated with sexual misconducts to help prisons prevent these sexual misconducts in the future. While geared towards prisons, this model could be adapted to other total institutions, such as the military.

# Theory

The theories used to create the model and influence the data collection are intersectionality and symbolic interactionism. Intersectionality helps break down the different levels as to what causes sexual violence. A study on the perceptions of sexual assault education and reporting on college campuses used an intersectional feminist framework to find that heterosexual men are less informed than others about sexual assault (Worthen 2017). The researchers also found that perceptions and reactions to sexual assault are often associated with their lived experiences (Worthen 2017). The key factors that informed their perception of sexual assault are gender, sexuality, race, and other structural forms of oppression and privilege, such as Greek life affiliation and years at higher education (Worthen 2017). Using this theory as a framework for variable selection, an offender’s gender, race, length of stay, and gang affiliation can impact an offender’s risk of sexual violence. One key area that is highlighted in this study that cannot be tracked is an offender’s sexuality. The relationship between sexuality and perception of sexual assault was high in the study so it is unfortunate that this cannot be included in the study. The relationship between these key variables will need to be noted in the model.

Symbolic interactionism allows for the exploration of the relationship between how perception and actions are used to interact in a setting that is related to rape. Javaid (2015) uses symbolic interactionism to compare the experiences of male victims of rape to female victims of rape, finding that men are more likely to retaliate and fight back against their attackers and that this response comes from hegemonic masculinity. The argument from this research is that males feel the need to appear as masculine among their peers by being aggressive and refusing to be submissive. In a hyper-masculine environment such as prisons, using conduct data to identify offenders acting out can have a relationship with sexual assault. These conduct violations could be a response to the offender trying to re-define himself as a man or trying to prevent rape by proving himself to be masculine and dominant. The physical attributes of an offender as well as how an offender is perceived will also determine what interactions the offender will have with other offenders. Looking at physical characteristics such as height, weight, and tattoos and how they relate to masculinity can also impact an offender’s likelihood of experiencing sexual violence.

# Literature Review

There is a drastic under-reporting of rape in prisons, which has caused prison rape to be difficult to address and reduce (Thompson 2009). One study found that in order to reduce prison rape, there needs to be a change in legislation. The legal definition of rape in multiple states only accounts for female rape and not male rape. The indifference of officials and administrators to look for signs of offender on offender rape has perpetuated a culture where rape is allowed (Thompson 2009).

Capers (2011) found that in a prison setting, men are less likely than women to report rape. This can be due to perceptions of rape and officers believing that the offenders deserve the rape. Prisons also do not allow for much movement for offenders to escape their rapists and protective custody or segregation would often place the offender in a location with highly violent offenders which could be more dangerous for them (Capers 2011). A significant finding of this study is that once an offender has been raped, it is very likely that they will continue to be raped for the duration of their sentence (Capers 2011).

When it comes to predicting what causes self-reporting of sexual assault in prison, a research study found that offenders are more likely to report the assault if they do not know the offender who assaulted them (Kubiak 2017). The location is also significant for if an offender will report the rape, if the setting of the rape was private would make the offender less likely to report the rape (Kubiak 2017). If an offender lives in a more public setting such as a dormitory style bunk rather than individual or two person cells, this would decrease the likelihood of an offender reporting a rape. This study also found that rape happens in higher level facilities even where there is an environment of total control. The lack of control an offender has over their environment increase the risk of rape, but decreases the likelihood that the offender will report the rape (Kubiak 2017). While it is difficult to determine the un-reported rapes, the researchers found that if offenders do not report the rape through official channels, they are more likely to report the rape to visitors (Kubiak 2017). If an offender has a dramatic increase in visitations, then this could be indicative of some kind of sexual conduct violation.

One study found that most sexual acts in prison are coerced actions driven by dominance, intimidation, and terror (Man 2001). This comes from the subculture of masculinity and power in prison that values dominant, aggressive traits. When sexual acts between two male offenders occur, it is not viewed as a homosexual encounter, but rather as a demonstration of power and masculinity for one offender and weakness for the other offender (Man 2001). Therefore, offenders who have more typical feminine traits, such as being smaller or acting more timid are more likely to be raped (Man 2001). A very critical time for new offenders is during the first 24 hours where offenders are trying to determine if the offender will be an easy target (Man 2001). Race is also one of the most polarizing characteristics and a dominant feature of inmate life (Man 2001). However, the effect of race in prison is not the same as the effect of race in regular society. Prison has an over-representation of people of color and gangs based off of race can often target white, younger men for sexual assault (Man 2001). The average age of a sexual assault victim is 21 in prison and offenders who are convicted of a non-violent crime are more likely to be victims. This study also found that offenders who commit sexual conduct violations are typically uneducated and lacked meaningful and financially rewarding employment before going to prison (Man 2001). These offenders are also more likely to internalize traditional, misogynistic notions of gender roles which perpetuates aggressive behavior (Man 2001). The longer an offender is sentenced, typically associated with violent offenses and a higher security level, the more likely an offender will be aggressive and commit a sexual conduct violation (Man 2001). The exception to this idea is for offenders who were committed of a child sex abuse. These offenders are more likely to be targeted by other offenders for sexual assault (Man 2001).

# Research Methods and Data

As discussed above, this project seeks to design and test a predictive model that can be used in order to prevent sexual conduct violations from happening in a prison setting. This model will be created using predictive analytics that can uncover variables with predictive capability to uncover situations that lead to sexual conduct violations. Predictive analytics is a wide field that uses data, statistical algorithms, as well as machine learning techniques to use historical data to identify the likelihood of future outcomes (SAS 2019). There are two types of predictive models: classification models and regression models. Classification models predict class membership while regression models predict a number. These models use one of the three predictive modeling techniques: decision trees, regression, and neural networks. Decision trees use classification models to partition data into subsets based on categories of input variables. The different branches represent a choice between a number of alternatives and uses the variable that is the most different to split the data. This model explains very well a path of decisions that explain a specific outcome. A regression estimates relationships among variables, but assumes that the variables follow a normal distribution. It can find key patterns in large sets of data but fails to show the different decision paths. Neural networks model highly complex relationships between all variables. This technique is often used to confirm findings from regression or decision trees. Since this technique tries to mimic the neurophysiology of the human brain by calculating the relation between each variable to each other variable, a very powerful computer and software is required to run this model.

This project uses a decision tree to find the different pathways that might increase an offender’s risk of sexual assault in prison. All variables used are defined in categorical groups. The model was run using Bayesian analysis that supports changing or updating the model based off of past results and prior belief regarding the probability distribution of an unknown parameter. This allows for the theories used to influence the model directly as the impact of certain variables can be altered to change the model to fit the theories. Each iteration of the model will be important in refining the final model that will show the different scenarios that lead to sexual conduct violations.

This project uses publicly available data for 2016-2018 from the Indiana Department of Corrections[[1]](#footnote-1). These data include individual level information for offenders including: sentencing information (charges, sentence length, intake date), commitment history, sex offender history, location history, length of stay at the current facility, bed history, demographic information (age, sex, race, county of commitment, mental health code, medical code, weight, tattoos), conduct history, current facility, gang affiliation, and job history.

The methodology of this project includes dividing the three years of collected data into testing periods. There are 12 testing period a year, for a total of 36. The datasets are already divided into monthly files making it the logical choice for dividing the data. It was important to create testing periods, as this model aims to predict the likelihood of the offender committing a sexual conduct violation (hereon referred to generally as the “target”) during a specific period of time. If the data were to stay aggregated to the three years, the model would only be able to say the likelihood of the offender committing a target within the three years and is not actionable for the department and does not allow for any sociological analysis on the targets to be conducted. Therefore, this model divides offenders into categories based off of the variables collected and determines the likelihood of the categories to have a committed a target. Therefore, 36 files are created with data pertaining to that testing period so as to get a snapshot of what the population looked like at that time and link that snapshot to any targets that were committed during that testing period.

The final file with all the testing periods is used for the decision tree model using SAS Enterprise Miner. A regression would not allow for the categories to be as clearly stated and greater sociological analysis can be conducted on the groupings of the categories rather than on the coefficients of each variable. Therefore the decision tree models are a better visual representation of the different categories. This project focused less on the statistics and computer science behind the model, but rather explaining why the variables would be predictive in relation to the theories used to select the variables.

It is important to note that predictive analytics are not to be used to flag individual behaviors and further stigmatize individuals before any actions have taken place. Predictive analytics are intended to be used to identify variables that are likely to contribute to a certain even occurring, and therefore provide a basis for evidence based changes in programming or other interventions that would decrease the chances of the event occurring in the future. Rather than focus just on the individual, projects using predictive analytics should be used in the research of undesirable events to find better interventions to provide a safer environment for every individual in the future.

## Data

The variables used in this study follow the variable naming scheme used by the Indiana Department of Correction and attached as Appendix B is the list of variables collected and their definitions for reference throughout this paper. The main key used to tie and offender to the data is the DOCNUM, which is the unique six digit number assigned to the offender at the time when the offender enters the prison system and is used in all data collection systems. The INTKDT is the date the offender first entered the prison system for their most recent commitment period. Along with the date at intake, there is an INTKSTCD which is a code that indicates the type of intake, such as a parole violation, new commitment, or first time felony offense. Basic demographic information was collected, such as gender, age, what county they came from, how long they were sentenced, weight, any tattoos, if the offender is in a security threat group (gang), if the offender is a sex offender or a violent sex offender, the offender’s current age, the offender’s age at intake, and the offender’s race. Information that pertains to the stay with the Indiana Department of Correction includes: how long the offender is sentenced to the Department (fixed term of incarceration), the medical code for the offender indicating if the offender needs certain medications or has any chronic medical issues, a code indicating if the offender has any disabilities, a code indicating if the offender has any mental health issues, a code indicating if the offender is in the general population, segregation, or protective custody, if the offender has a job, the number of times the offender has been sentenced to the Department, the number of times the offender has been released and returned to the Department, and the number of conduct violations an offender has had in the month prior to the reporting period.

There were six datasets used to create the final testing dataset. These datasets include a conduct file, population file, commitment file, length of stay file, and sexual conduct violation file. The conduct file is a count of the number of conducts and offender committed in the month prior to the testing period. There was no restrictions on the type of conduct violation. The population file included all of the demographic information for the offender at the first of the month (first day of the testing period). The commitment file was split into two files: the number of previous commitments the offender has with the Department, and the number of time the offender has returned to the Department. Returns include parole violations so this number would be higher than the number of commitments and is indicative of multiple movement while in custody and a revocation of control as the offender has been released outside of a facility to parole and then returned to a facility. The length of stay file is a count of the number of days the offender has been at the facility at the beginning of the testing period. This allowed for an indication of how integrated the offender might be in the facility. The sexual conduct violation file was simply a file of each instance of a sexual conduct violation over the testing years. This file was used as the target dataset.

The sample population had to meet the following criteria: physically in a facility at the beginning of the testing period, not in a work release facility at the beginning of the testing period, and in the general population during the beginning of the testing period. This would exclude offenders on temporary moves to court or other jurisdictions as well as removing offenders who are in segregation units or protective custody units. Work release facilities were excluded as all but two of these facilities are run by the counties. The level of control granted to offenders at work release facilities is also not comparable to the other facilities, as well as the fact that offenders must qualify to be sent to a work release facility and are less likely to act out in these facilities so they are not sent back to a higher level facility.

# Analysis

In total, 11 unique models were created. Appendix C includes the decision trees of all the models. The method for adjusting the models provides insight into the variables in the final model that have predictive capability. The first model was using all of the variables collected, but only allowed the modeling program to split the variables into two categories and only breakdown into six levels. The largest break was at the facility level, by breaking the facilities into two groups: IWP/RTC/NCN and all other facilities. The three facilities grouped together are two women’s facilities and one of the New Castle facilities. New Castle is the sex offender facility, with over 90% of the offenders at this facility having a sex offense conviction. This is an indication that the model is trying to split out women and the New Castle facilities from the other populations. Following Model 1 in Appendix C, node 6 splits based off of race. This node separates offenders with a race code of “B” (black) from the other race codes. In sociology, it is not a common practice to separate race in this fashion (black and non-black). The majority of the population is coded as either Black or White; however, it would make more sense to separate race into the binary categories of White and non-White. This model indicates that race might have predictive capability, but separating out White from the rest of the races might allow for further analysis. The variable with the most predictive capability in this model is node 33: intake age. This model is indicating that the older the offender at the time of intake, the more likely that they will commit a target.

For the second model, a new variable was added to the model file for race: RACE2. This variable breaks down the race of an offender into two groups: White and non-White. The original Race variable is now rejected from the model and will no longer be considered a viable variable for the model to use so it has no further impact on the model. This specifications for this model were the same as the first model, allowing the program to split variables into 2 categories and breakdown into 6 levels. Model 2 now includes disability as a variable with predictive capability, along with conduct count, and intake age. This model is still pulling out female facilities and New Castle from the other facilities.

Based off of Model 2, two new variables were added to the file: Facility Level (FACILITYLEVEL) and New Castle (NEWCASTLE). The new model file also excluded all offenders at work release facilities. Most work release facilities are under the control of the counties and provide a drastically different environment from the typical prison setting. This would have a significant impact on the model and by removing these offenders, the sample population is more consistent. The old facility variable is now a rejected variable, along with the original Race variable. The specifications of the model have not changed from Models 1 and 2. The new model now has SEX as the largest split, with female offenders having conduct count (CONDUCTCNT), facility level (FACILITYLEVEL), and length of stay (LOS) as variables with predictive capability. For men, being a sex offender (SEXOFFENDER\_FLAG) , not having a job (JOBFLAG), having been at their current facility for over 102 days (LOS), being housed at a New Castle facility (NEWCASTLE), and being over the age of 47 (CURAGE) is all predictive of a target. This model highlights the difference between the male and female population, while also indicating that the New Castle facility is also an important variable. The creation of the New Castle flag was originally created as a numeric value of 1 or 0, but was changed for Model 4 to be a yes or no. This would not allow for a split of 0.5 as seen in Model 3, node 42 and 43. The changing of this variable does not change the model as seen when you compare Model 4 and Model 3.

Model 5 was created by rejecting certain variables that have not been included on the model and are unlikely to have predictive capability: county of commitment (COUNTYCOMMIT), commitment count (COMMITCOUNT), intake date (INTKDT), intake status code (INTKSTCD), maximum release date (MRD), range category (RANGECAT), receive date (RECVDT), receive code (RECVCD), return count (RETURNCOUNT), test period date. County of commitment was rejected because it would have little to do with an offender once inside the prison. Also based off of the model, the commitment count, return count, and the date variables would not be predictive as the previous models have indicated that the longer someone is with the Department, the more likely they would be to commit a target and these variables all deal with the beginning of someone’s sentence. This model was allowed to split the variables into two groups and go down 6 levels, which is very similar to model 4 but not as predictive.

Model 6 returns to the variables used in Model 4 as the previous model was less predictive, but allowed the variables to be split a total of 50 times, rather than only twice; this model also was seven levels deep instead of six. The new model brought out two new variables as predictive: County of Commitment (COUNTYCOMMIT) and Weight (WEIGHT). While county of commitment was not a variable that was expected to be predictive, weight was expected to be predictive as the offenders weight would be a way of presenting himself in this environment. This model did find an expected variable, weight, to be predictive, but would only predict a handful of offenders. The model was allowed to split the variables into too many categories and the results are too spread out to be of any use.

Model 7 continues to use the same variables as Model 4, but the variables can be split into 10 categories and extend down 8 levels, allowing for fewer leaves but more relationships between variables. This model is the most predictive with Sex (SEX) and Current Age (CURAGE) having predictive capability at multiple levels. Length of stay (LOS), Intake Age (INTKAGE), and Weight (WEIGHT) also has predictive capability in this model. While Model 7 performs relatively well, there was an error in the calculation of the Conduct Count (CONDUCTCNT) variable that will need to be correct for future models.

Model 8 is a replica of the previous model, but with the corrected Conduct Count (CONDUCTCNT) variable. This model does not identify many variables with predictive capability, with New Castle (NEWCASTLE) being the most likely to be predictive. This model also highlights the need to remove women from the sample. Model 9 represents the model with the sample restricted to just men. This allows for the sample to be more consistent with the literature as most studies focus on men or women, but not the two sexes together. In this model, Current Age (CURAGE) has predictive capability as well as Zachary Violent Flag (ZACHARY\_VIOLENT\_FLAG). While the model has potential for predictive capability, the model is very wide, with allowing the model to split variables into 10 categories. Model 10 is a replica of Model 9 but only allows variables to split into 8 categories. This did not change anything in the model. Model 11 is another replica of Model 9, but rejects the following variables: County of Commitment (COUNTYCOMMIT), Commitment Count (COMMITCOUNT), Earliest Possible Release Date (EPRDT), Facility (FACILITY), Intake Date (INTKDT), Intake Status Code (INTKSTCD), Maximum Release Date (MRD), Race (RACE), Receive Date (RECVDT), Receive Code (RECVCD), Return Count (RETURNCOUNT), and Test Period Date. The resulting model is the same as Model 9, but is the model used for analysis. The largest break for the final model is at conduct count, with Current Age (CURAGE), Fixed Term of Incarceration (FTI), New Castle (NEWCASTLE), Sex Offender Flag (SEXOFFENDER\_FLAG), and Zachary Violent Flag (ZACHARY\_VIOLENT\_FLAG) being the variables with predictive capability.

## Discussion

The predictive variables for the final model (Model 11) all correlate to one of the two theories used for the original variable selection. By using symbolic interactionism and intersectionality we can draw assumptions as to why the variables would have predictive capability.

Through symbolic interactionism, we can analyze how offenders are presenting themselves and interacting with other people inside the prison environment. The higher the number of conduct violations within the past 30 days (CNDCTCNT) is an expression of control that might be used to assert dominance over their situation. The older an offender is (CURAGE), the longer they have been in their environment and the more likely they are to have a defined persona.

Intersectionality can help suggest relationships between different variables and how they are working together to create an environment that might lead an offender to commit a target. The length of time an offender is sentenced to serve with the Department (FTI) can impact how the offender will cope with the change in the environment, with a longer sentence leading to a sense of helplessness and the offender trying to find ways to regain a sense of control over their environment. Race is a common characteristic studied by researchers when looking at sex crimes, and it is unsurprising that race would have an impact on an offender’s experience in prison. An older age at the time of intake (INAGEYR) was found to increase the likelihood that an offender would commit a target, which follows the literature as the literature reports that the younger a person is, the more likely they are to be a victim.

The environment around the offender also has an impact on committing a target, as being at the New Castle Facility (NEWCASTLE) increased the likelihood of committing a target. This facility is known as the sex offender facility as over 90% of the population at this facility has been flagged as a sex offender. Not only does being around sex offenders increase the likelihood of committing a target, being flagged as a sex offender (SEXOFFENDER\_FLAG) or a violent sex offender (ZACHARY\_VIOLENT\_FLAG) increases the likelihood of committing a target as well. The interactions between being labeled a sex offender and classified as a sexually violent offender and being placed inside a facility that is labeled for sex offender would just reinforce the idea of sexual assault. The duration of time the offender has been at the facility also had predictive capability, which flips current research instead of focusing on the victim (which suggests the first 24 hours in prison are the most likely for a person to experience sexual assault) and instead focuses on the perpetrator and suggests that the longer the offender has been at the facility, the longer they have been able to assert their dominance and find where they fit in the social hierarchy of prison.

# Conclusions

There are many variables to look at when studying sexual conduct violations in prisons. Much of the existing research is based on looking at the victims by using data collected for PREA, however this analysis was conducted by looking at characteristics of perpetrators leading up to the offenders committing sexual conduct violations, what is called targets throughout this paper (Vega 2011). Two main theories were used during variable selection to create a basis for the predictive model which would attempt to predict the likelihood of an offender committing a target. These two models were symbolic interactionism and intersectionality. Symbolic interactionism was used to select variables that pertained to the environment of the offender and external characteristics that might impact the perception of the offender while in the prison system. Intersectionality was used to select variables directly related to the offender and how they choose to present themselves, as well as how other people might perceive the offender.

This project is less focused on the statistical significance of variables, but rather if the computer was able to find a relationship between variables which is why variables will be flagged as having predictive capability and not being flag as statistically significant. For the predictive model, a significance value of 0.2 was used in order for the variable to be placed in the model, and for this reason not every variable that is brought into the model is significant or even has predictive capability. A split validation test is used in the model, which means that half of the records will be hidden from the model and used as a validation set. The half of the dataset that is being used by the computer to create the model, called the train set, will then be tested against the validation set to determine if the model was able to accurately predict what occurred in the validation set. If the node average for the variable was over 50% and the train and validation averages were within 20% of each other, then the variable was considered to have predictive capability.

Based off of the current research in the field, variables that were anticipated to have predictive capability were the physical characteristics (weight, height, tattoos…), the gang status flag, job status, previous commitment history, and facility security level. Only a few of these were found to have any predictive capability in the final model, and with previous commitment history being excluded completely from the dataset used in the final model.

There were a total of nine unique models run, however throughout the process of finalizing the model there were 11 models run. Each model was adapted from the previous versions to create the most complete model by model 11. These changes included the recoding of race to be white and non-white instead of allowing for multiple denominations of race. The recoding of the facility, first splitting out female and male facilities before dropping female facilities altogether, as well as splitting out the New Castle Correctional Facility as it is the facility with the highest concentration of sex offenders (at over 90% of the population being flagged as a sex offender). The model criteria itself was also changed to keep the model from only splitting the groups into binary groups and instead allowing for more breakdown of the variables itself. This was significant with the conduct count variable being allowed to split into four groups instead of just two. The model was also altered to allow for more depth by allowing more variables to enter the model. This strengthened the model by allowing for more variety, which included more offenders.

In the model used for analysis, the first branch found that a majority of the population did not have any conduct violations in the month prior to the reporting period. The branch without any conduct violations did not lead to any variables with predictive capability. The branches over three conduct violations in the month prior to the reporting period also did not lead to any variables with predictive capability, which could be due to certain stipulations (consequences) being placed on the offender for those conduct violations that would keep them from committing a target. The one conduct violation in the month prior to the reporting period branched again at the NEWCASTLE variable, with offenders not at the New Castle Facility branching again into race, but no predictive capability. The offenders that were at the New Castle Facility branched by current age at the time of the reporting period and offenders over the age of 68 had predictive capability for committing a target. Offenders aged between 48 and 67 also had predictive capability and branched again into length of stay at their current facility. Offenders in this group that have been at their current facility for 181 days had predictive capability. Offenders under the age of 48 branched at the SEXOFFENDER\_FLAG variable, where there was no predictive capability for offenders who were not flagged as sex offenders. Offenders who were flagged as sex offenders branched again at the ZACHARY\_VIOLENT\_FLAG which indicates that the offender has been flagged for sexually violent crimes against minors. The branch for two violations branched first by RACE, with non-white not having predictive capability. The white group branched again at NEWCASTLE, with not being at New Castle Facility not having predictive capability. The offenders at New Castle branched again at CURAGE, with offenders being under the age of 47 not having predictive capability. Offenders over the age of 47 did have predictive capability for this branch. If the offender had 3 conduct violations then the first branch was at FTI (Fixed Term of Incarceration), with offenders being sentenced to over 24,289 days (roughly 66 years) had predictive capability. The offenders who were sentenced to less than 66 years branched at RACE, with the nonwhite group with no predictive capability. The white group branched again at INAGEYR, with offenders over the age of 45 at intake having predictive capability and offenders under the age of 45 at intake not having any predictive capability.

The final model suggests that having one to two conduct violations in the previous month, being at the New Castle Correctional Facility, being older in age, being older at the age of intake, staying at the same facility for longer, and having a history of being a sex offender or a violent sex offender increases the risk of committing a target.

Further research on the topic of sexual conduct violations in prison is necessary as not every aspect of the issue was able to be explored in this study. One area for further study is looking at the length of time an offender has been in prison, not only how long the offender has been at their current facility. If the offender has been released to parole and returned to prison, does this impact the likelihood of the offender committing a target? This would bring in the idea of how much institutional knowledge an offender has and if they are able to use that knowledge to their advantage. The return of a parolee would also suggest that they are not well equipped to return to life outside of prison. The type of sentence the offender is serving, such as serving concurrent sentences or consecutive sentences. This would get into the length of sentencing (FTI) that was already investigated in this study, but would also account for severity of offense and could also have a different impact on the outlook of the offender if they are looking at serving two sentences rather than one sentence. This study only looked at any conduct violation in the month prior to the reporting period, but looking for specific types of conduct violations that might lead to the offender committing a target would be another area of study. Labeling theory is a facet of symbolic interactionism and a further analysis looking at the behaviors of offenders attached to the label of sex offender or sexually violent offender would be interesting to see if they are treated differently in the system, considering that there is a specific prison facility to house sex offenders.

# References

Capers, Bennett. "Real Rape Too." *California Law Review* 99, no. 5 (2011): 1259-307.

Javaid, Aliraza. 2015. “The Dark Side of Men.” *Journal of Men’s Studies* 23 (3): 271–92.

Kubiak, Sheryl P., Hannah J. Brenner, Deborah Bybee, Rebecca Campbell, Cristy E. Cummings, Kathleen M. Darcy, Gina Fedock, and Rachael Goodman-Williams. 2017. “Sexual Misconduct in Prison: What Factors Affect Whether Incarcerated Women Will Report Abuses Committed by Prison Staff?” *Law & Human Behavior (American Psychological Association)* 41 (4): 361–74.

McFarlane, Linda, and Lovisa Stannow. 2011. “Corrections Officials + Advocates = Safer Prisons and Jails.” *Corrections Today* 73 (3): 52–58.

Man, Christopher D., and John P. Cronan. 2001. “Forecasting Sexual Abuse in Prison.” *Journal of Criminal Law & Criminology* 92 (1/2): 127.

Thompson, Anthony C. 2009. “What Happens Behind Locked Doors: The Difficulty of Addressing and Eliminating Rape in Prison.” *New England Journal on Criminal & Civil Confinement* 35 (1): 119–76.

Vega, Rhonda. 2011. “Understanding and Implementing Prea Programs.” *Corrections Today* 73 (3): 41–47.

Worthen, Meredith G. F., and Samantha A. Wallace. 2017. “Intersectionality and Perceptions About Sexual Assault Education and Reporting on College Campuses.” *Family Relations* 66 (1): 180–96.

# Appendix A

**ATTACHMENT A**

**APPLICATION FOR RESEARCH**

**Indiana Department of Correction**

INSTRUCTIONS:

* Replace the non-bolded test after each number with the answers to each question.
* Submit the application via email or by mail to:

Executive Director, Research and Technology Division

Indiana Department of Correction

302 W Washington St, E334

Indianapolis, IN 46204

* Include the Agreement of Security and Confidentiality (State Form 13251) with this application, if applicable.

1. **Title of Project** – Predicting Sexual Conduct Violations in Indiana State Prisons
2. **Name of Applicant** – Megan Lalioff
3. **Phone number of applicant** – 317-525-0253
4. **Address of Applicant** – 1400 E Hanna Ave, Indianapolis, IN 46227
5. **Organization or institutional affiliation pertaining to the project** – University of Indianapolis, Department of Sociology
6. **Name, title, department of faculty advisor** – Dr. Colleen Wynn, Professor of Sociology, Department of Sociology
7. **Project hypothesis** (es) – Using predictive analytics to create a model to predict sexual conduct violations can reduce the number of sexual conduct violations in prison.
8. **Project procedure** – State specifically the procedure involved to conduct the research. Include the following information:

* Facility(ies), parole district(s), division(s) to be involved;
* None
* Estimated start and completion dates of project;
* January through May
* What the Department is expected to provide (equipment, personnel, space, supplies etc);
* Offender level data for the past three years – list of variables to be provided on a separate list
* Number and type of subjects and method to select them;
* None
* Copy of each instrument to be used;
* None
* Name of all research personnel involved and their background as related to the project; and
* Megan Lalioff – Master’s student at the University of Indianapolis completing a final thesis project.
* Dr. Colleen Wynn – Thesis advisor for the project.
* Any other pertinent information necessary.

1. **Intended use and dissemination of findings** – State if the final report is to be used for educational credit, publication, in-house use, etc. Include a statement that two (2) copies will be submitted to the Executive Director, Research and TechnologyDivision.

Final thesis project to complete the Masters Degree program.

1. **Importance to Department of Correction** – State how the Department can benefit from the results of the project.

After completion of the project, the model will be given to the department to implement and manipulate as they see fit.

1. **Specific information required** – State all the information required from the Department in order to conduct the project. If information is to be obtained on individual subjects, specifically state the information required.

Offender level data for the past three years – list attached at the end of the applicaiton

1. **Justify need for subject’s information in identifiable form if applicable** – State why information is required in a manner which will identify the subjects and state the specific information associated with them that is needed. If criminal history information is required, a copy of AGREEMENT OF SECURITY AND CONFIDENTIALITY, State Form 13251, is to be completed. If information is not required in identifiable form enter “NA”.

Agreement of security and confidentiality form attached.

1. **If the project is being conducted in association with a college or university**, has it received approval from the Committee on the Protection of Human Subjects (or similar committee)?

If yes, please attach a copy of the committee approval. If no, please explain.

Yes, being conducted with the University of Indianapolis. No IRB approval necessary as not actual data collection is being done.

1. **Signature of Applicant** – please sign the application.

Megan Lalioff

1. **Title of Applicant** – Master’s student
2. **Date of signature** – 12/4/2018



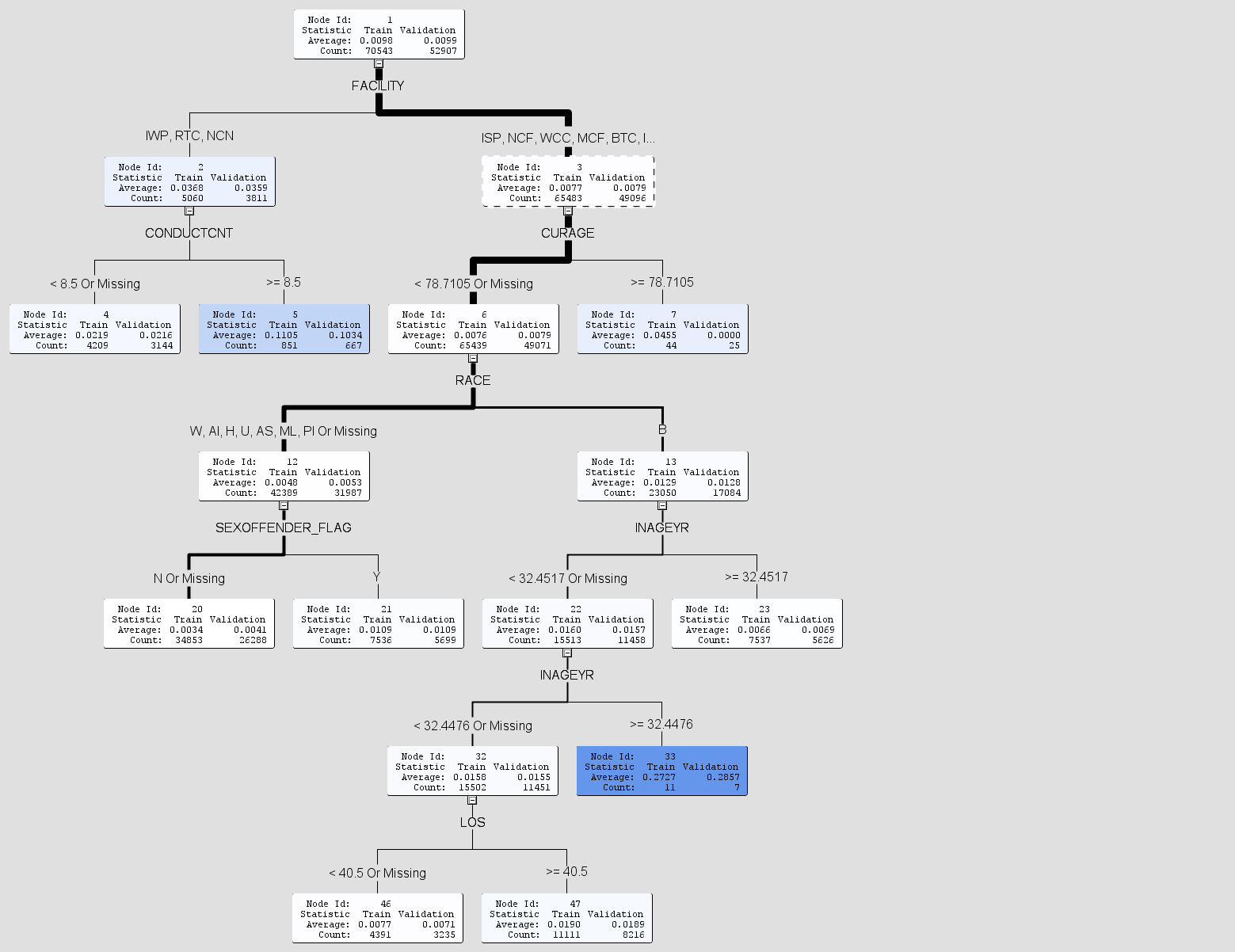
# Appendix B

List of variables and definitions:

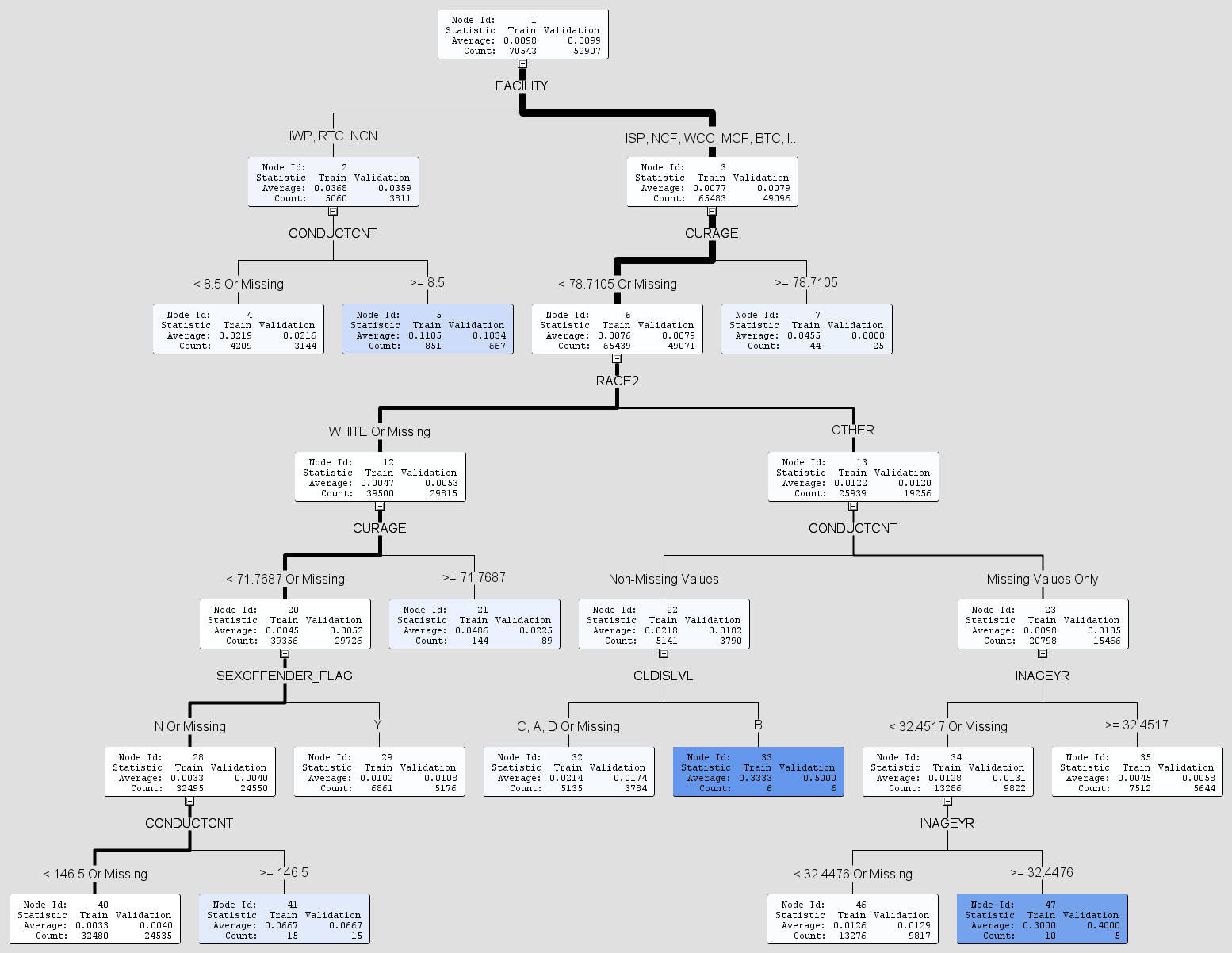
* DOCNUM – unique ID given to each offender, used to match datasets together
* INTKDT – the original intake date for the current commitment the offender is serving
* INTKSTCD – a code identifying the type of intake for the offender (new offense, etc….)
* RACE – the race of the offender
* SEX – the biological sex of the offender
* EPRDT – the earliest possible release date for the offender for their current commitment
* ACOUNTY – the county of commitment
* FTI – fixed term of incarceration, the number of days the offender must serve with the IDOC
* MRD – the maximum release date, the date that the offender must be discharged
* FACILITY – current facility the offender is housed at
* RECVDT – the receive date, the most recent date the offender has returned to the Department within the same commitment period
* CLMEDLVL – the medical level of the offender
* CLDISLVL – the disability code of the offender
* RANGECAT – the range category for the housing type of the offender, such as general population
* STGFLAG – security threat group flag, a security threat group is the terminology used in the department for groups like gangs
* SEXOFFENDER\_FLAG – flag for if the offender was convicted of a sex offender
* ZACHARY\_VIOLENT\_FLAG – flag for if the offender was convicted under the Zachary Law (sexually violent crimes against a minor)
* CLMHLVL – the mental health code of the offender
* RECVCD – the receive code of the offender for their most recent return to the Department within the same commitment period, such as a parole violation
* MSO – the most serious offense for the offender’s current commitment period
* INAGEYR – the offender’s age at intake
* WMDATE – the first date of the testing period
* CURAGE – the current age of the offender at the beginning of the testing period
* JOBFLAG – flag for if the offender was working a job or was idle at the beginning of the testing period
* WEIGHT – the weight of the offender at intake
* MARK\_TYPE\_CD – if the offender has any tattoos or body piercings
* COMMITCOUNT – the offender’s number of commitments
* RETURNCOUNT – the number of times the offender has been returned to the Department
* LOS – the number of days the offender has been at the facility at the beginning of each testing period
* TARGET2 – if the offender committed a sexual conduct violation during the testing period
* CONDUCTCNT – the number of conducts the offender committed in the month prior to the testing period
* COUNTYCOMMIT – then county of commitment for the offender
* FACILITYLEVEL – the security level of the facility the offender was housed at during the testing period
* NEWCASTLE – a flag for if the offender was housed at the New Castle Facility (a facility primarily for sex offenders)

# Appendix C

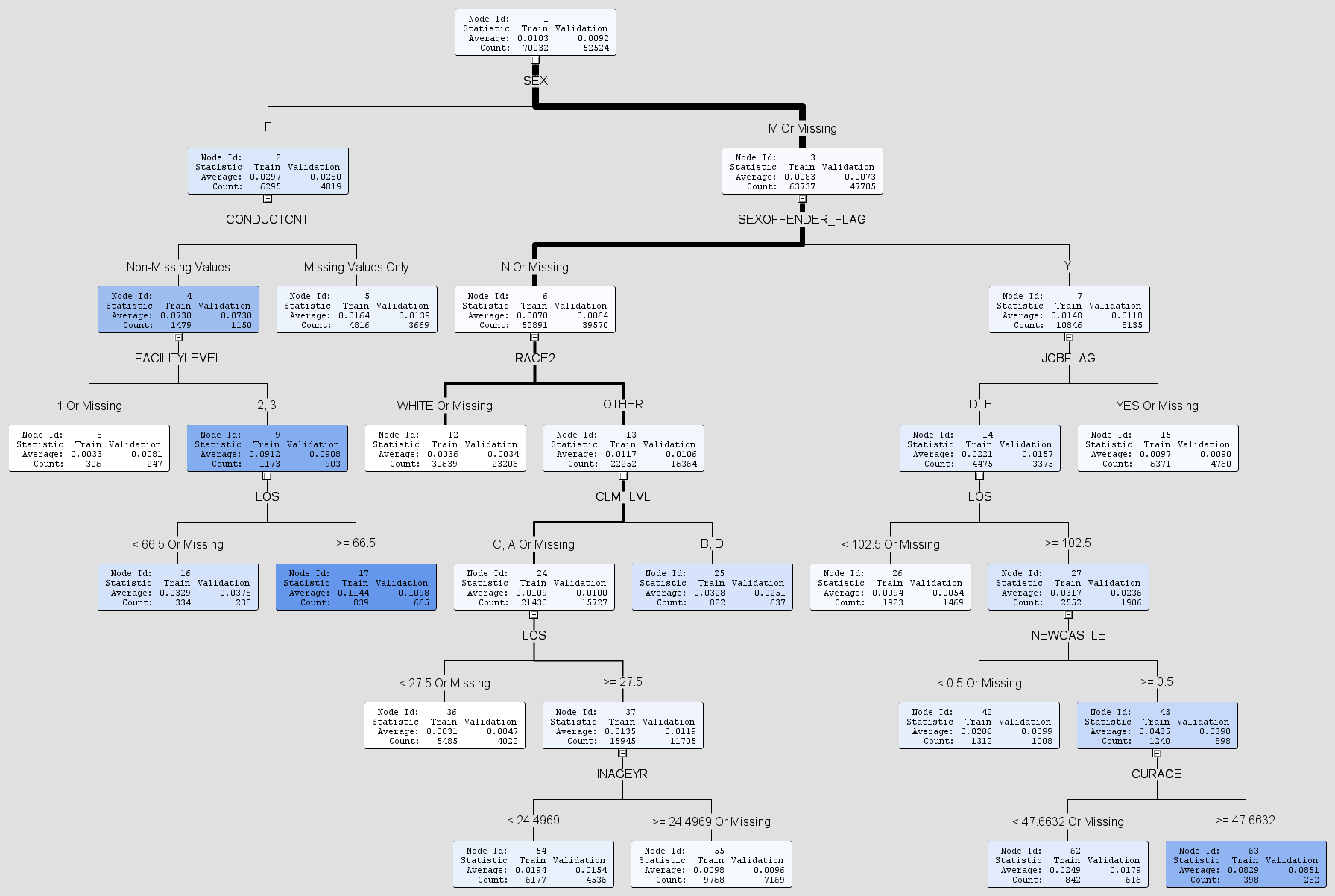
Model 1



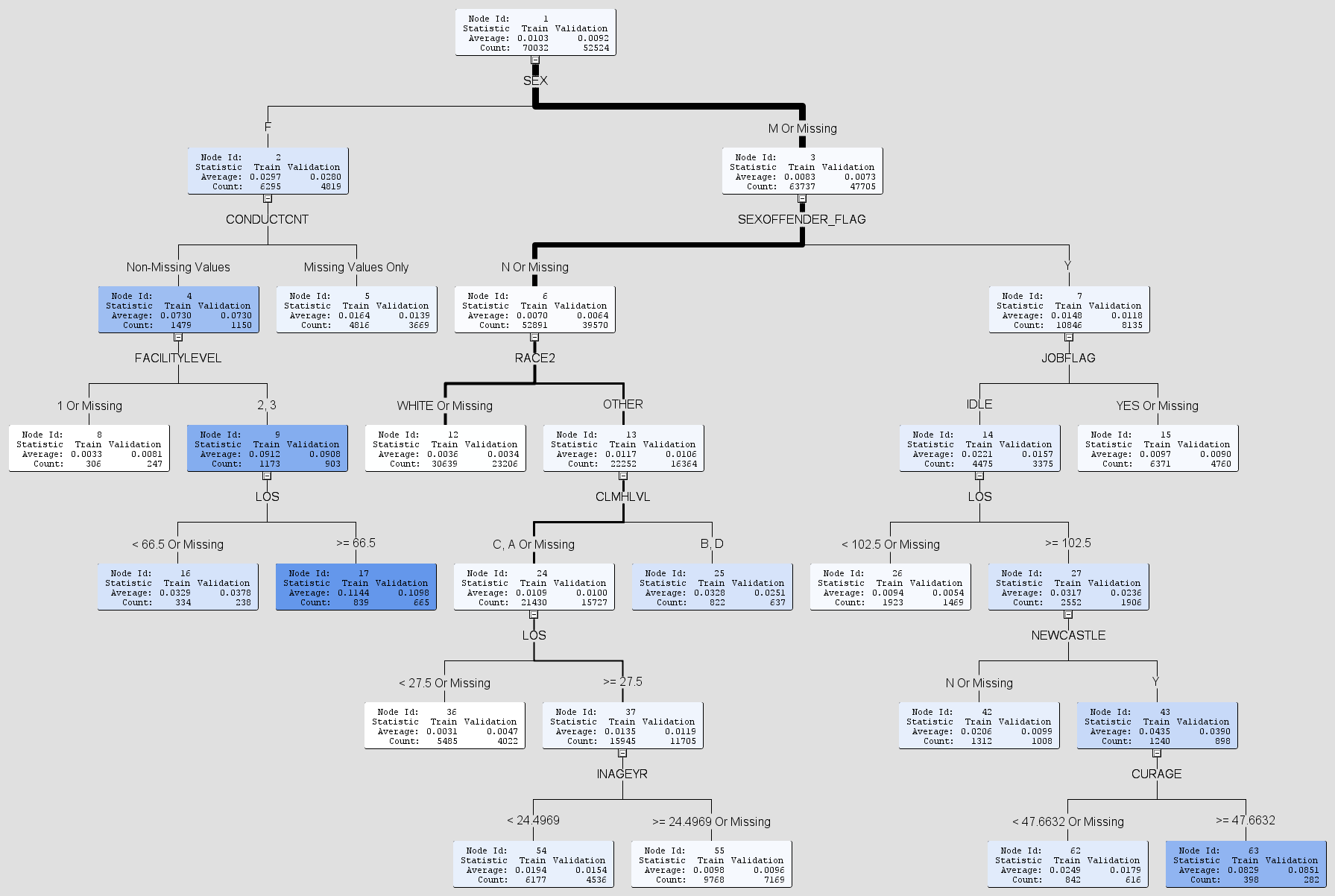
Model 2



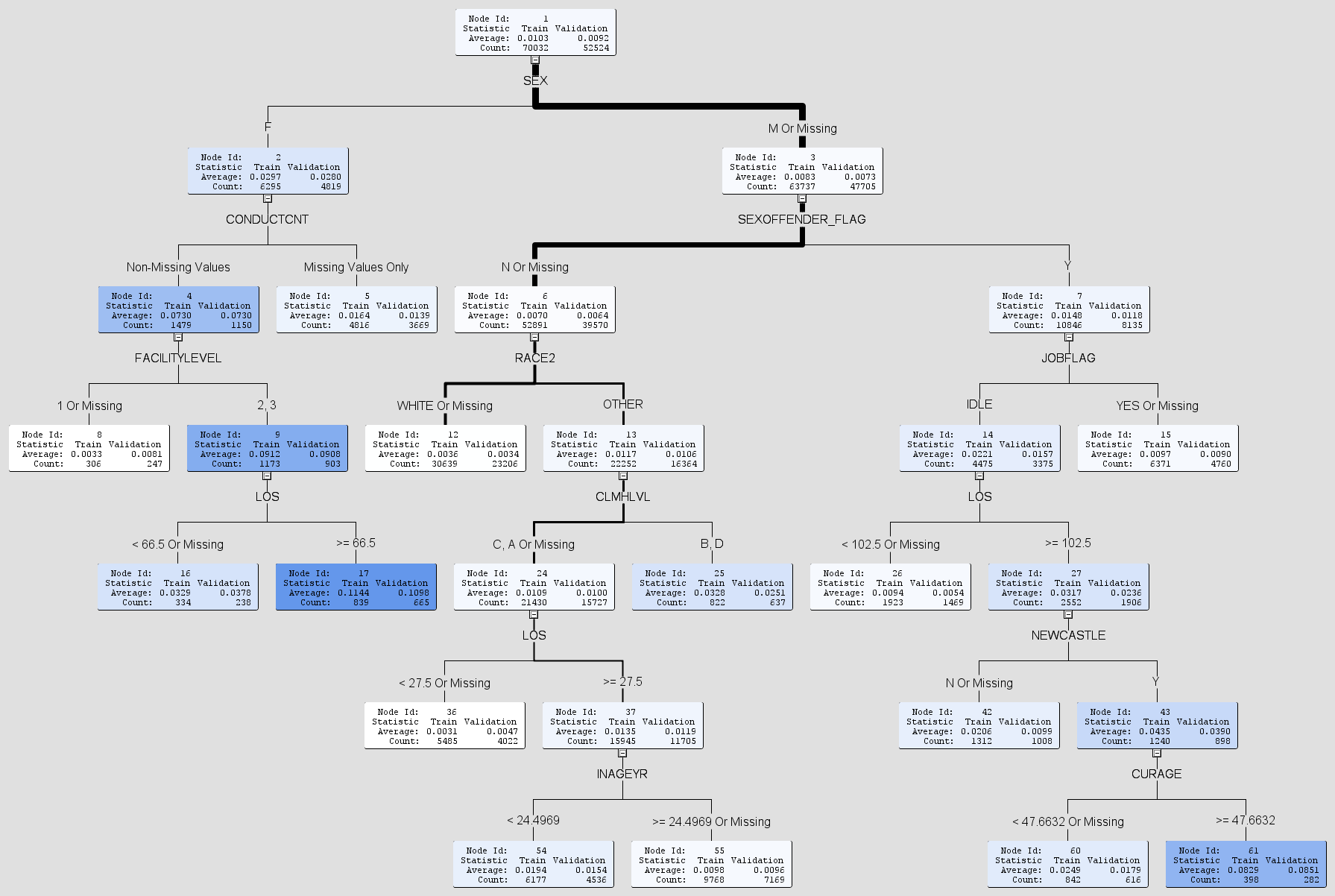
Model 3



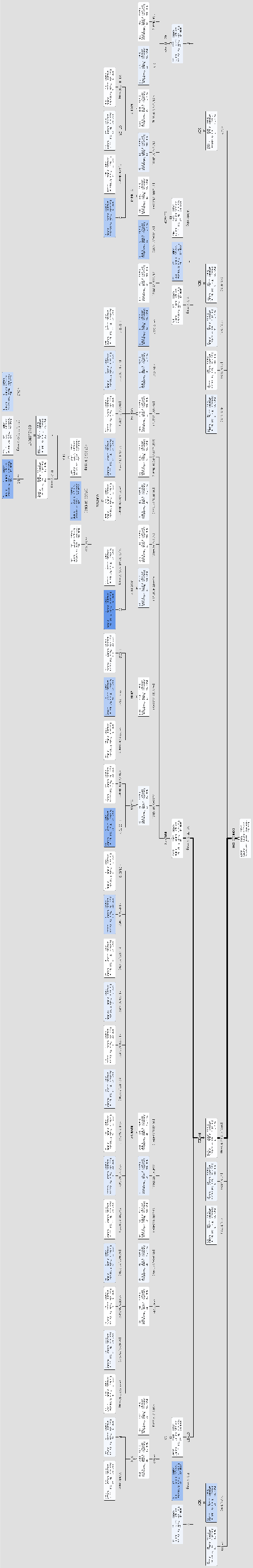
Model 4



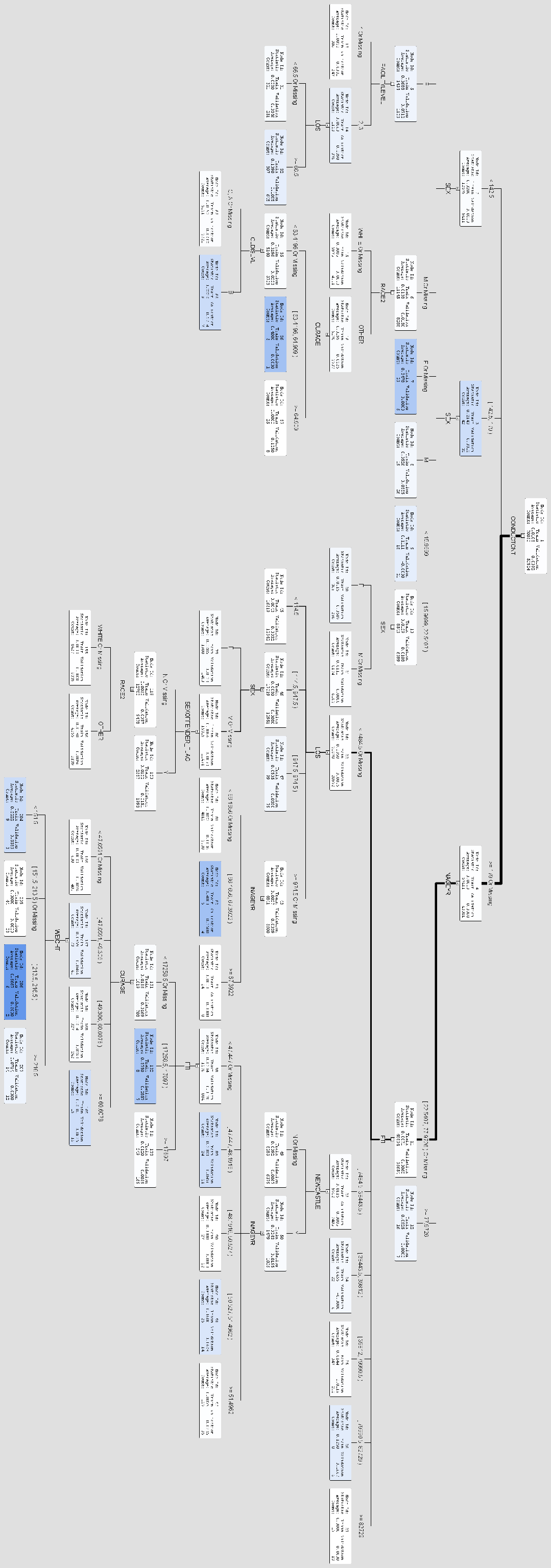
Model 5



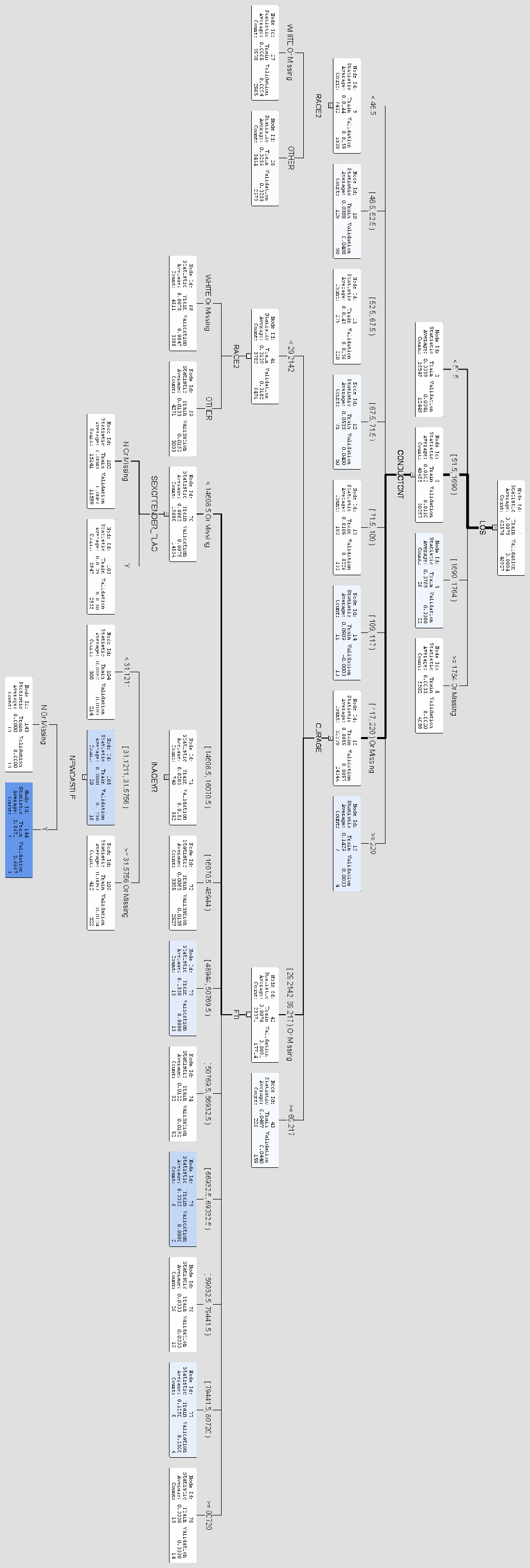
Model 6



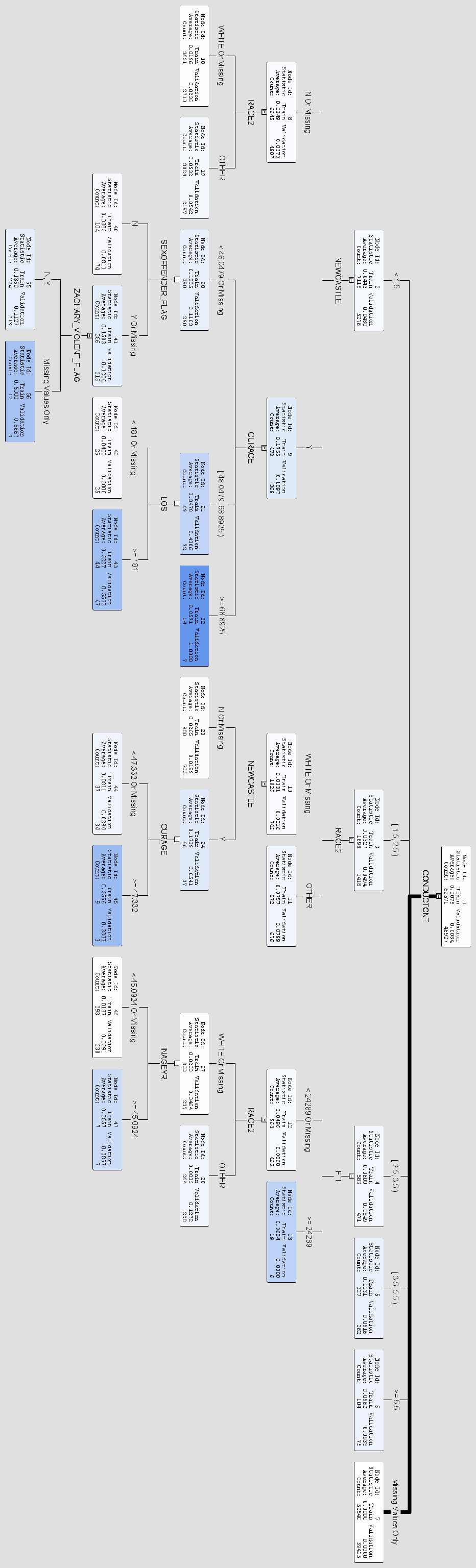
Model 7



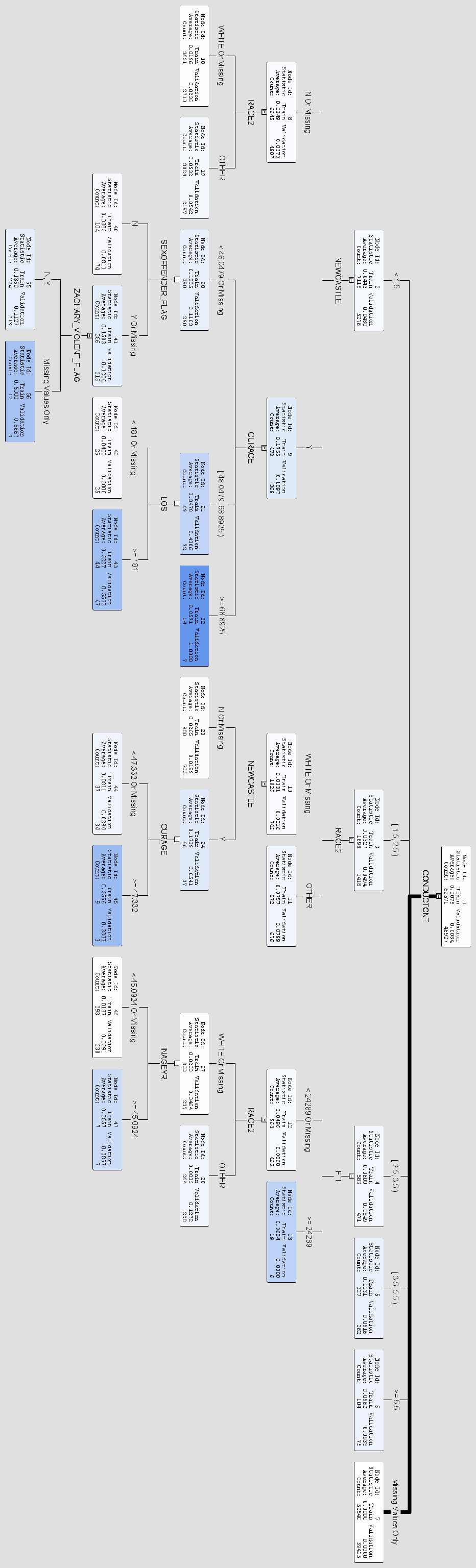
Model 8



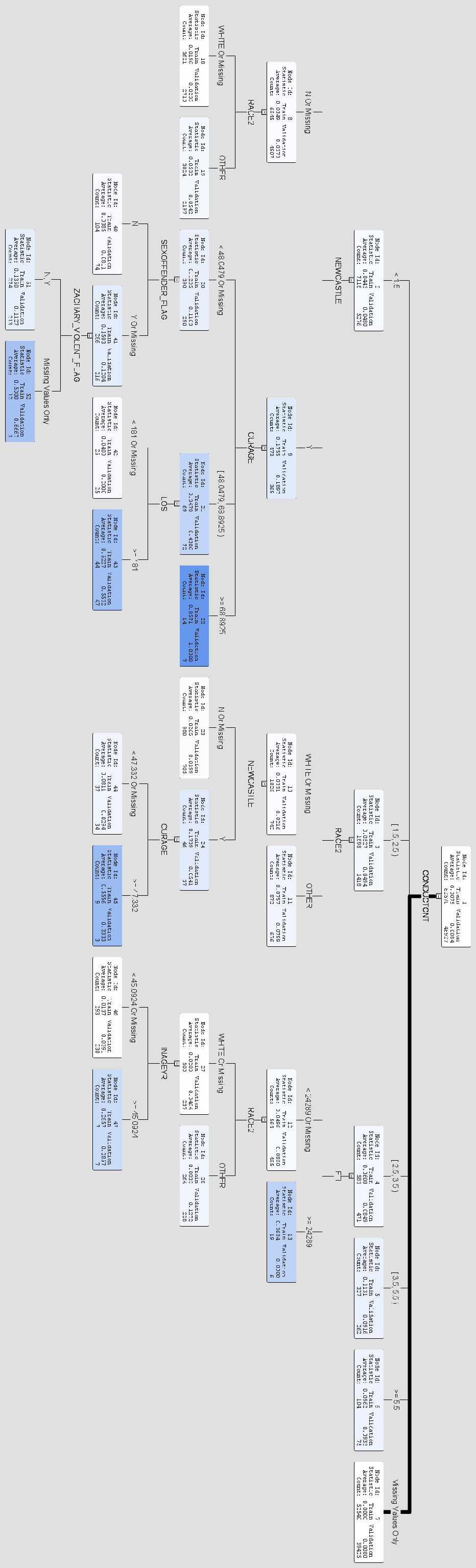
Model 9



Model 10



Model 11



1. The application used to obtain the data and is attached as Appendix A [↑](#footnote-ref-1)