

Paper 073-31

Predicting Child Support Payment Delinquency Using SAS[®] Enterprise Miner[™] 5.1

Jodi Blomberg, SAS Institute Inc., Denver, CO
Jacqueline K Long, SAS Institute Inc., Chicago, IL

ABSTRACT

To ensure the health and welfare of children and to reduce welfare costs, child support agencies around the country are tasked with successfully collecting child support payments for the children in their respective states. This paper describes two case studies that illustrate how a data mining approach has helped agencies collect more support payments and use their collection resources more effectively. Using data mining to identify payers who are likely to become delinquent helps child support agencies build effective intervention strategies for collecting payments. Predictive modeling is also used to determine which intervention strategies are effective for various types of payers. Sequence analysis and binary response modeling are emphasized in this approach, using SAS Enterprise Miner 5.1 as the modeling environment.

INTRODUCTION

There are currently 15.9 million child support cases in the United States. State agencies are responsible for the collection of these cases and are reimbursed by the federal government for some of the costs associated with collection. Of cases in which the court has ordered the non-custodial parent (NCP) to pay specific child support, and no support has been collected, about 30% remain uncollected. Cumulative unpaid child support at the end of fiscal year 2003 had reached \$96 billion and continues to grow each year (Office of Child Support Enforcement (a) 2005).

State child support enforcement agencies have many tools at their disposal to help them enforce child support such as professional license revocation, liens against property, and passport revocation, but the single most effective means of collecting child support is by automatic withholding of wages (Office of Child Support Enforcement (b) 2005). States employ numerous resources and energy toward enforcing child support, and are constantly seeking ways to become more efficient and effective.

There are many ways in which data mining could be used to improve enforcement efforts. One possibility is to determine which cases are most likely to be collected and focus the field offices on those cases to increase cost effectiveness. A second possibility is to determine effectiveness of various legal strategies for different types of cases. In the case studies below, we examine two current approaches to improve enforcement efforts and suggest further possibilities.

Data mining is always an iterative process. The two case studies presented here show the progressive nature of the data mining life cycle, using data from two states with differing levels of data availability and resources. These studies provided insight into the collection process and subsequently resulted in increased revenue for the states. To avoid revealing specifics of child support collection practices and outcomes until their data mining models are more mature and thoroughly tested, the states in our case studies requested to remain anonymous for this paper. We therefore refer to them as State A and State B to avoid confusion. Our first case study is from a state that is in the initial stages of using data mining for collection efforts. They used exploratory techniques to make some conclusions about their collection of efforts. Our second case study is from a state that is farther along in their data mining efforts. They completed a predictive model that has been tested in their field enforcement offices.

In general, predictive models try to find good rules (models) for predicting the values of one or more variables in a data set from the values of other variables in the data set. The variable being predicted is known as the *target* variable. After a good rule is determined, the rule can be applied to new data sets (scoring) that do not contain the target variable being predicted. This is known as *supervised learning*. In the absence of a good target variable, we can use unsupervised learning. *Unsupervised learning* involves using exploratory techniques to learn as much as possible about the independent variables and interactions between them.

In the following case studies, we outline both our failures and successes with respect to the process of building a good predictive model using supervised and unsupervised techniques. By outlining where our efforts fell short, we hope to provide insight on some of the challenges of finding a good predictive model.

CHILD SUPPORT ENFORCEMENT OVERVIEW

A child support case may be initiated for a single child or multiple children. An NCP may have multiple cases for separate families or separate children. Cases do not become obligated until a court or administrative order requires the NCP to provide financial support. A case cannot be delinquent until it is obligated because no money is officially owed. Therefore, for the purposes of predictive modeling of child support delinquency, only obligated cases are considered.

Child support delinquency can be defined any number of ways, and determining an appropriate definition for data mining is a significant hurdle to getting a good predictive model. Determining how to define delinquency as a target variable involves making decisions on two issues: time and money. A target definition must be defined with respect to payments or delinquency over time. Should an NCP be considered delinquent after missing one payment or two payments? Conversely, should a payer be considered non-delinquent if they have paid a certain number of months in a row, even if they were delinquent in previous years? Second, a target definition must consider the amount paid. Should an NCP be considered delinquent if they make regular payments, but not of the full court-ordered amount? Should they be considered delinquent if they have made payments every month, but not always for the full amount?

Determining the appropriate target variable definition must take into consideration the policies of the state and the desired outcome of the data mining model. Decisions regarding the data mining target definition of delinquency will depend on the policies and legal actions available to the state child support enforcement agency. In our case studies, we will see two ways of defining delinquency for modeling and some consequences of those choices.

CASE STUDY 1: STATE A

There are many reasons to undertake the task of data mining. In the case of State A, there was a genuine desire to improve the quality of life for as many children as possible. They felt that by identifying those who were likely to become delinquent, they could work with them towards a successful payment. In our initial attempts at mining, we focused on identifying common characteristics of delinquency. Armed with this data, domain experts can define and develop programs to work with these individuals. The domain experts might also make policy adjustments that increase the number of children receiving the aid they legally owed.

DATA MINING PROBLEM DEFINITION

Our overall objective was to reduce the delinquency in child care payments through the design of intervention policies and programs. Our analytic objective in support of this effort was two-fold:

1. Use historical data with actual payment information to train and fine tune a predictive model.
2. Apply this model to data with an unknown target in order to assign a probability of payment.

The intent is to identify what factors contribute to an NCP becoming delinquent in making child support payments. Domain experts subsequently review and analyze significant parameters to identify non-payment trends and develop intervention policies and programs to address the issues uncovered. As new data becomes available, we score the new data in an effort to identify those individuals with a high probability of non-payment. Where possible, child support counselors work with the high-risk NCPs to enroll them in a program, or take other action to avoid non-payment.

DATA PREPARATION

Table 1 lists the fields (variables) that are common in a child support database.

FIELD	DESCRIPTION
ID	Unique member identification
Demographic Information	Gender, race
Address Information	City, state, county
Payment Information	Case dollars obligated; how often the payment is required; whether there were arrears assigned
Case Type	Case type
Number of Dependents	Dependents covered in this child support case
Education Level	Education level
Bankruptcy Flag	Whether the NCP ever declared bankruptcy

Table 1. Common Variables in a Child Support Database

The transaction data contained multiple records for each NCP (member). Each row in the original data represented a unique child support order and Member ID combination. A child support order is also referred to as a "case" and is defined by a single Custodial Parent (CP) but may represent multiple dependents. As part of the data preparation, these transactional records were combined at the member level. The final table structure for the data mining data base contained a unique row for each NCP and represented all associated case, dependent, and payment history information.

Data issues to be considered included the following:

- How do we combine multiple records?
- Which records do we include?
- Do we need additional variables?
- What is the relevant date range?

We used a three-year cutoff date to establish a valid sample, and collected four years of history for each NCP. Any case that was "open" at any point during the defined time period was included in the sample.

An NCP was considered delinquent the first time they had a case in which arrears were greater than 100% of their monthly child support obligation; in other words, the point at which the member went 30 days or more without paying the full obligation. When DELINQUENT = YES, the only data included were actions which occurred prior to the first delinquency. The four years of history was determined using the date of first delinquency. In the case of DELINQUENT = NO, the historical time period for the data collection was determined based on the four year period preceding the latest payment date.

These time restrictions are necessary to allow for good prediction and applicability of results. In order to make proactive decisions about cases that are likely to become delinquent, we must be able to describe what was going on before the delinquency occurred. For example, the current job is not of interest if the NCP was delinquent two years ago and is now paying. The job two years ago, when the NCP was delinquent, is the data point of interest.

Additional metrics were created from the transaction (case) level data:

- Frequent Mover flag
- Job Jumper flag
- Various counters (number of employers, number of children born out-of-wedlock)
- Indicator flags (home phone, work phone, e-mail address, employment status)

The Frequent Mover and Job Jumper flags enabled us to establish a sense of stability for the NCP. It is generally accepted that those with higher stability are more likely to pay consistently. An NCP received a value of 1 for the frequent mover flag if their address changed more than once in a 12 month period. Similarly, the Job Jumper flag was assigned a value of 1 when an NCP changed jobs more than once in a 12 month period.

Various counters were created to account for the multiple cases per member. We tallied things such as the number of children born out-of-wedlock, total number of employers during the four-year period, and a count of cases in which there were Initial Arrears assigned. Variables were also created to represent metrics associated with the maximum number of consecutive months ever paid by the NCP and total number of monthly payments ever made by the NCP.

Indicator Flags were created to recognize records in which the NCP had, for example, a home phone, an e-mail address, or was currently employed.

Once the data mining database was created at the member level and target correlations examined, additional data exploration and data preparation were required to address the following:

- Missing Values
- Filter Outliers, establish reasonable caps
- Transformation into grouped variables
- Skewed variable distribution

While no upper limit was placed on the Member age, a lower limit was established at 14. All month counts were capped at 48 because we were working with four years of historical data. Based on input from domain experts, the maximum monthly payment considered was \$15,000. For cases in which values fell outside the valid range, the 'outlier' value was replaced with the maximum/minimum established.

Due to the high cardinality of the Employer Source and Professional License variables, new group variables were defined describing which employer source group and which professional license group an NCP was in. The groups were determined using a combination of domain expertise and results from the Variable Selection Node in Enterprise Miner. The Variable Selection Node in Enterprise Miner can automatically group categorical data values according to their relationship with the target variable. These groups may be better predictors of the target variable than the original values. In addition, for the purposes of techniques such as regression in which a dummy variable is created for each value of a categorical variable, these grouped variables can reduce the number of input variables and model complexity.

INITIAL MODELING EFFORTS

One of the most fundamental criteria for determining a successful model is verifying that the model generalizes well. It is imperative that we choose appropriate inputs for the model because extraneous inputs can impact model performance. The *Curse of Dimensionality* in its simplest form refers to having variables in your model that don't *improve* your model, and may, in fact, reduce the accuracy of the model. As the number of model dimensions increases within a model, so does the complexity. This combination of increased dimension and complexity results in decreased generalization-- which is counter to our goal. In training the model, it is important to look out for three types of variables that do not provide value-add.

- **Proxy variable** is a variable whose effect *mirrors* that of another variable already included in the model
- **Extraneous input** is a variable that provides no additional predictive value to the model
- Variables with **known data quality issues** are omitted from the modeling process.

For example, data exploration revealed that the Commercial Drivers License flag is represented within the Professional License variable, so the Commercial Drivers License flag field was removed from consideration in the model. In some cases, both a count and a flag existed for a variable, therefore one or the other was selected, not both. In addition, the maximum consecutive months paid prior to the reference date was removed because it is highly correlated with maximum consecutive months EVER paid.

Because we had a target variable, we initially began applying supervised modeling techniques such as decision trees and regression. We soon discovered that the data collected did not support these. As previously mentioned, this client was in the initial stages of integrating data mining processes into their decision support systems. As such, they were also in the initial stages of both collecting and cleansing data. One of the main benefits of mining at this stage is recognizing what data is available, necessary, and useful. Because most of the data available to us was demographic in nature, we speculated that if we clustered the data prior to any predictive modeling attempts, we might increase the robustness of our model. As a result, we decided to explore unsupervised modeling methods, which do not use the target variable. We obtained mild improvements by clustering the data prior to any predictive modeling attempts.

INITIAL MODELING EFFORTS: UNSUPERVISED LEARNING

After we completed the data preparation, we used cluster analysis to analyze our data. Cluster analysis attempts to group the NCPs into clusters (that is, groups) where NCPs within the group are very similar to each other but the groups themselves are very dissimilar to each other. In this case, we were interested to see what percentage of each cluster was delinquent.

The population norm for delinquency was 31%. We were able to define seven clusters within the data. These clusters were deemed valuable because five of the clusters exhibited delinquency rates that differed significantly from the population norm. In other words, we were able to isolate NCPs into distinct segments with similar behavior tendencies toward or away from delinquency (Table 2).

	Cluster ID						
	1	2	3	4	5	6	7
NOT Delinquent NCPs	30%	25%	80%	84%	71%	66%	89%
Delinquent NCPs	70%	75%	20%	16%	29%	34%	11%
Size of Cluster as % of Population	10%	8%	21%	12%	20%	14%	16%

Darkest gray indicates a higher than normal probability of delinquency; lighter gray indicates a close-to-normal probability of delinquency; white indicates lower than expected probability of delinquency.

Table 2. Distribution of Delinquency across Clusters

We assign names to clusters based on the most important demographic characteristics that determine the assignment of an NCP to that cluster. Cluster 1 and Cluster 2 contain “Unstable Non-Payers.” Members who fall into one of these two clusters are considered unstable because they are nearly twice as likely to be a Job Jumper and a Frequent Mover. An interesting apparent contradiction to this is that they are 20% more likely to have a home phone number. There are two reasons an NCP does not have a recorded phone number in the data: either they lack a phone number for work or home or the agency has not been able to determine the information. Cluster 1 and Cluster 2 also isolate those members who are twice as likely to become delinquent payers.

In addition, characteristics common to those in Cluster 1 include:

- Higher than average number for Dependents Born Out-of-Wedlock
- Lower than average value for Maximum Consecutive Months Paid
- Twice as likely to have ever been on public assistance
- Twice as likely to be hunter and fishing license holders

Characteristics common to those in Cluster 2 include:

- Maximum Consecutive Months Paid just under half the norm
- Fewer work phones than average
- Monthly obligation (amount legally owed) is 90% lower than the average

Cluster 3 and Cluster 4 contain “Stable Payers.” Members who fall into one of these two clusters are 30–50% less likely to become delinquent, and are considered stable because they rarely meet the definition of Job Jumper or Frequent Mover. On average, they have paid two more consecutive months than the population overall.

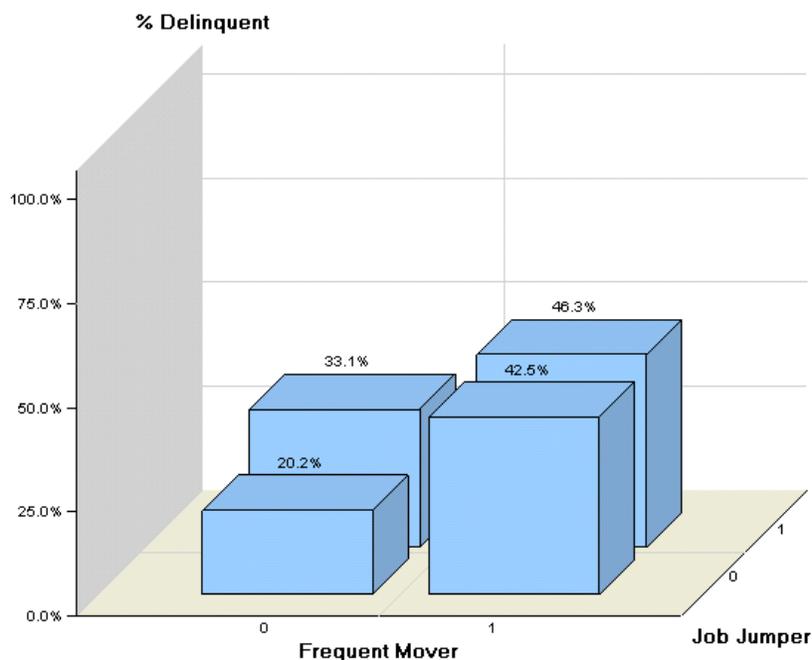


Figure 1. Delinquency Rates for Frequent Movers and Job Jumpers. Here is an example of an “expected” result. Stable NCPs who do not switch jobs or addresses have the lowest delinquency rates. Job Jumping seems to be slightly more important than frequent moving.

It is interesting to note a few things about Cluster 3. Although these NCPs are more likely to pay, the monthly obligation for this group is relatively high—18% higher than the norm. And although we considered them stable based on the definition of Job Jumper and Frequent Mover, they are half as likely to have a known work phone number and also less likely to have a known home phone number than the norm. This may be because the state has a withholding on their wages, which means they cannot choose to not pay as long as they continue to work for the listed employer. As such, the state may be less motivated to track down home and work phone numbers. Cluster 4 continues to behave in a stable manner with respect to phones: members are twice as likely to have a known work phone number and 50% more likely to have a home phone number.

While members in both Cluster 3 and Cluster 4 are more likely to have children born out-of-wedlock, those in Cluster 4 are twice as likely to be on public assistance, while those in Cluster 3 are less likely to have been on public assistance.

Cluster 7 is heavily weighted with reliable payers—these NCPs are 70% less likely to default than the overall population. This may be partly related to the fact that their total monthly obligation is 80% lower than average. Though we can usually rely on their payment, these members are less likely to have a known work phone number. We find this unusual because we observed in other populations that a known work phone number is usually associated with a higher likelihood of payment. They also share similarities with respect to licenses; they are three times as likely to have a hunting and/or fishing license and half as likely to have a commercial driver’s license.

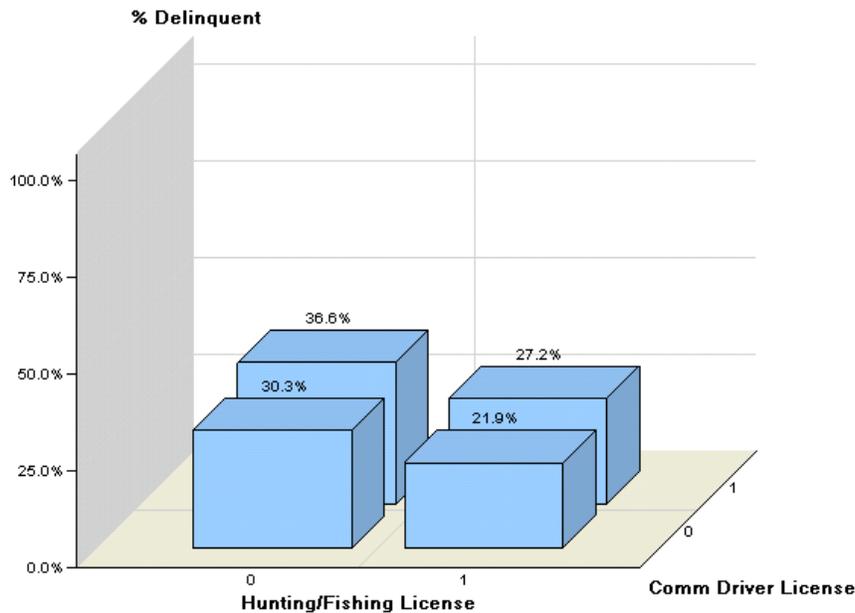


Figure 2. Delinquency Rates for Hunting, Fishing, and Commercial Drivers License Holders

Anecdotal evidence from this state and others we work with suggests that a commercial driver's license seems to be positively correlated with delinquency, so we do not consider this a surprising result. However, the negative correlation between hunting/fishing licenses and delinquency is somewhat unexpected. Perhaps even more unexpected is that hunters with no commercial driver's licenses have the lowest delinquency rates.

REVISED MODELING EFFORTS

Once we were able to isolate those with a high probability of non-payment into Cluster 1 and Cluster 2, we returned to our supervised modeling techniques of regression and decision trees. We created a distinct supervised model for each cluster. Using lift charts, misclassification rates, and averaged squared error as our assessment criteria, we determined that decision trees provided the best fit for this data with respect to performance (Figure 3). Child support agencies generally have limited funds to invest toward identifying potential delinquent payers. The fact that a decision tree will identify more non-payers up front is desirable. Trees also consistently provided slightly lower misclassification rates.

The following parameters were deemed significant in our final models:

- Maximum number of consecutive months paid
- Total monthly obligation
- Job Jumper flag
- Total number of months ever paid
- Count of cases for each NCP where initial arrears existed

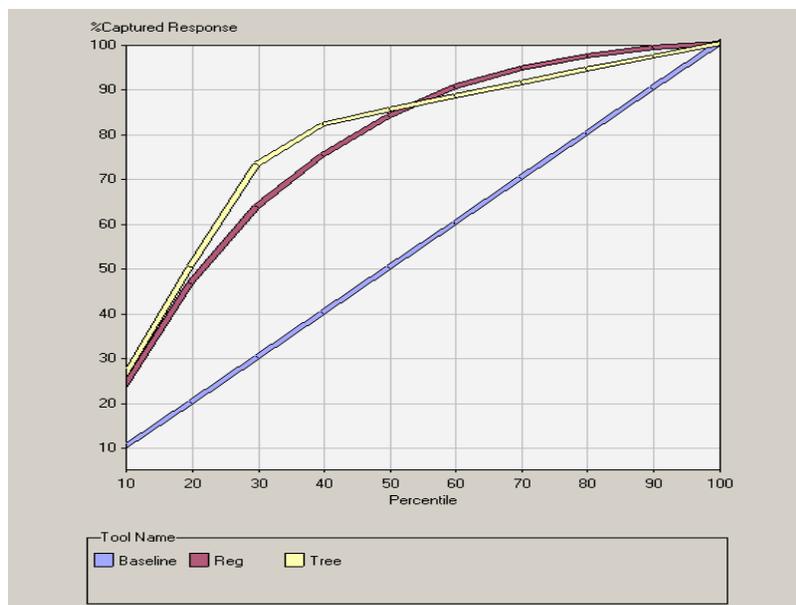


Figure 3. Lift Chart Comparing a Regression Model to a Decision Tree

MODEL DEPLOYMENT

Over the long term, this state will move towards deployment as follows:

- Apply the model to new data as it becomes available
- Determine a threshold probability for potential delinquency
- Identify NCPs who are above the threshold
- Determine which policy/program will be most effective
- Engage the NCP in a delinquency prevention program

This will be a long-term process because they have just begun to apply data mining. Even though we were able to find supervised models for this population, the state moved forward with their initial meetings with the domain experts using only the cluster results. Because the cluster results were based on demographic data, the state felt there was enough information in the clusters alone to start identifying what types of intervention strategies are currently used with these groups. In the future they plan to further explore the use of predictive modeling techniques, especially as they are able to add data about the agency's interactions with the NCPs. There is the promise of cleaner data and collection of new fields. All this, in addition to a deepening understanding of the data itself, will contribute to producing a final model with as much predictive power as possible.

CASE STUDY 2: STATE B

In our second case study of State B, we had many years of demographic data on the caseload and the collection efforts. The state wanted to use data mining to resolve issues including:

- What cases have the highest probability of paying continuously?
- For cases in which the children are supported by welfare, how can they increase the probability of finding and requiring the NCPs to pay?
- What child support enforcement activities are the most productive?

Before narrowing the initial data mining problem, we had to resolve several data issues.

DATA MINING PROBLEM DEFINITION

Initially, the most important data issue for the problem definition for State B was whether to model at the case level or at the NCP level. This state can only take legal actions on an NCP by case. However, because we were ultimately trying to model the behavior of an individual and not a case, we decided to compile the data at the NCP level.

The data mining problem was initially defined as follows: For a set of NCPs that includes delinquent and non-delinquent payers, develop a predictive model to determine who is likely to be delinquent.

We quickly determined that defining who was delinquent was not a simple task. As stated earlier, NCPs have many different patterns of paying, ranging from those who pay the full amount every month, to those who pay something less than the full amount but do so regularly, to those who pay sporadically, or not at all.

After much discussion, the state determined that an NCP could be classified as “not delinquent” only if they had paid the full amount for six months. Due to our intent to sample cases from a point in time six months prior to the current date, we constructed our target variable by designating a “non-delinquent” payer to be any NCP who paid in full for the previous six months. We used T_7 to indicate the current month, where T_i represented month i . A “non delinquent” payer was characterized as someone who made a payment for the full obligation for the time period T_1 to T_6 . All others were designated “delinquent”. This definition made for a somewhat unsatisfactory target. For example, an NCP who paid the previous five months (that is, no payment at T_1 but full payment for the period T_2 to T_6) and who would have paid the following month (T_7) would be designated “delinquent” by our definition. While this NCP was delinquent at T_1 , they became “non delinquent” from the time period T_2 to T_7 . In addition, NCPs who made regular payments for the period T_1 to T_6 but for an amount less than the full obligation would also be designated as “delinquent”. This strict definition resulted in the “non-delinquency” classification being quite a rare event and could possibly confound model effects looking for differences between delinquent and non-delinquent payers. For this reason, we constructed an additional target variable in our model for testing. Our second target variable was an interval variable that contained the number of months paid in full by the NCP in the time period T_1 to T_6 . This interval target variable enabled us to account for more of the variation in payment habits. While we were restricted to the binary target definition for creating the final predictive model, we did use the interval variable during model development for comparison purposes.

DATA PREPARATION

Data for the independent variables were assembled from the time of the NCPs entry into the system until time T_0 . The following types of data were available for this modeling effort (Tables 2, 3, and 4).

FIELD	DESCRIPTION
Date of Birth	Availability of date of birth (Yes/No)
SSN Flag	Availability of SSN (Yes/No)
Address Flag	Availability of address (Yes/No)
TANF	CP receiving welfare (Yes/No)
Employer Source	Source of employer information (for example, self report, automated system, etc)
Phone Number Flags	Availability of phone numbers – cell, work, home, other (Yes/No)
Drivers License Type	Commercial, regular, motorcycle, etc

Table 2. Common Variables for Demographic Characteristics

FIELD	DESCRIPTION
Ever Paid	Whether ever paid on any case
AOP	Whether paternity has been acknowledged
Kid Count	Number of children
BOW Count	Number of children born out of wedlock
Case Number	Number of cases for the NCP
Marital Status	Marital status at this time
Arrears	Current arrears owed
Employers	Employer count
CP to NCP Miles	Miles between NCP and CP -available when both parties lived in State B

Table 3. Common Variables for NCP Behavior

FIELD	DESCRIPTION
Cusp Disposed Order	Last disposed court order
Cusp File Order	Last filed court order
Region	Region/Office (where case is being treated)
Out Phone and In Phone	Counts of incoming and outgoing phone calls between NCP and field office
Capias	Whether an arrest warrant had been filed for nonpayment

Table 4. Common Variables for Agency Behavior

For many variables, we were able to calculate not only the current value, but also whether any changes had occurred in the month previous to T_7 . In contrast to State A, we had more information on the behavior of the NCP and their interactions with the child support collection agency and legal system.

As outlined in Case Study 1, data cleansing was also necessary. For example, NCPs with more than three employers on file were given a count value = 33 in the system, which would unnecessarily skew results. In addition, many of the attempts to narrow the list of candidate independent variables described in Case Study 1 were employed here.

INITIAL MODELING EFFORTS

As stated in the data mining problem definition section, non-delinquent payers as defined by the binary target were very rare, so we over-sampled our data to create a modeling data set with approximately 1/3 non-delinquent payers and 2/3 delinquent payers. We corrected for bias in the model by entering the prior probabilities within SAS Enterprise Miner, where they would be automatically accounted for in our assessment statistics.

In our initial modeling efforts, we found some very simple but telling results. We were able to easily identify three groups of NCPs: one group of consistent delinquent payers, one group of never delinquent payers, and a final group with a mix of delinquent and non-delinquent NCPs.

We discovered the first group by discovering that NCPs who have never paid on any case will almost certainly never pay. They are almost always “forever delinquent.” We discovered the second group of non delinquent payers by looking at those NCPs who had a known employer from a reliable source. Because automatic withholding is nearly always instituted in cases in which the agency has employment information on the NCP, there were very few NCPs that had employment information but were delinquent. The final group had no reliable recorded employment but had made a payment at some point on a case. In this third group, 83% of NCPs with no current recorded employment was delinquent. While these results were very statistically significant, on their own they were neither interesting nor actionable.

Other variables with weak but significant effects on the target were the miles between the non-custodial and custodial parent, marital status, whether an arrest warrant had been issued for non-payment, the number of children born out-of-wedlock, the source of employer information, most recent court order, most recent disposed legal action, possession of a commercial drivers license, and which region was handling the case.

Because this was in a state of larger geographic size, we suspect it has some unique characteristics that may not apply to other states. First, the miles between the non-custodial and custodial parent was not something we found significant in our other case study, but we speculate that distance is not a factor in states of smaller geographic size. State B had hypothesized before the modeling effort began that the region handling the case would likely be significant and this proved to be correct. The offices that handle cases range from rural to urban, and delinquent NCPs are less likely to avoid authorities in small rural communities. Smaller states may not be large enough to realize significant regional differences.

As a result of the initial modeling efforts, we made two conclusions. The first was that we needed to reconsider our data mining problem and data sample. Cases with automatic wage withholding require no action by the state’s offices, and therefore are not really of interest for a modeling effort. Rather, they wanted to find and focus their resources on cases that do not already have withholding but show some other likelihood of paying. Delinquent NCPs are often under-employed or unemployed and as a result, the existence of employment information could be considered a measure of the ability to pay. The real value in data mining would be the ability to find NCPs with an ability to pay that was as yet undiscovered.

Our second conclusion was that we were missing data on the greatest indicator of delinquency--the ability to pay. While we could account for actions by the state enforcement agency and demographic characteristics of the non-custodial and custodial parent, we felt that without a credit score or similar measure, we could not accurately judge ability to pay. The state enforcement agency believed a certain number of NCPs had avoided reporting their employment to avoid automatic withholding, but without some financial measure, we could not accurately locate those NCPs that had an ability to pay but had no employment record with the state enforcement agency.

Based on our model development using the binary target versus the interval target, we found significant improvement in model accuracy using the target variable that describes the number of months paid consecutively in the six-month period. We speculate from this that the definition of delinquency was too broad and we would get better results by relaxing the definition or using a target definition of delinquency based on the total number of months the full amount was paid.

REVISED MODELING EFFORTS

After our initial modeling efforts, State B sought to gain additional data points. The state enforcement agency was able to purchase a financial score from a major credit reporting agency for a number of their cases. Also, they were able to create an indicator variable on whether the NCP is receiving any benefits at the federal level (for example, Social Security). This last information is derived from The Federal Parent Locator Service, a national location system for NCPs provided to state agencies by the federal government.

Based on our previous conclusions, a new sample of 2,000 cases was drawn of NCPs that made at least one payment, but did not have employment information. In addition, we relaxed the delinquency definition. The target variable for this revised modeling effort was based on four months of data, and an NCP was considered "not delinquent" if they had paid full support in three or four of the previous four months. A new model was created using the same independent variables described in the "Data Preparation" section and the two new independent variables.

Not surprisingly, the new variables proved to be significant and improved the model predictions considerably. In addition to the new variables, the following variables remained significant: most recent court order, existence of phone and address information, miles between CP and NCP, and region handling the case. For this final modeling effort, we used a logistic regression model.

MODEL DEPLOYMENT

Beyond using validation and test data sets to validate the model, State B wanted to implement a field test of the results. This first model deployment was a little unusual but highly effective. Rather than scoring a new set of cases and sending top scoring cases to the field, they tested the model on the delinquent cases in the model. They took the cases with the highest predicted probability to NOT be delinquent, but who were actually delinquent as defined by the target variable. These are usually known as the *false negatives*. These cases were sent to their field offices to investigate. This tested why a group of delinquent payers seemed to have so much in common with some known non-delinquent payers. In all, they sent 218 cases to three offices in the field to investigate. In 60% of those cases, the field offices were able to work the case resulting in some next action such as filing new legal actions or issuing an automatic withholding. Because automatic withholding is the single most effective means of collecting child support (Office of Child Support Enforcement 2005), any data mining efforts that result in more withholdings can be considered very successful. The results of this effort were successful enough that the agency felt confident in two things: purchasing financial scores was worth the monetary investment and the model was appropriate for continued use.

FUTURE MODELING DIRECTIONS

There are several other data mining problems and approaches that could be explored to help the agency improve its child support recovery efforts. We discuss two possibilities here.

Legal action paths are not standardized across the state in State B. Office and regions tend to develop their own best practices with respect to handling cases. In order to determine what path of actions give the best results, the agency could first use the Sequence Analysis Node in SAS Enterprise Miner to determine all implemented legal action paths. From just this analysis, they could discover the most and least common legal paths. They could then use this analysis as part of a data mining effort to determine if certain legal paths are much more effective than others. They can also examine this data against each region to determine how different regions use their available legal options and whether this is instrumental to their success or failure. This exercise will have to be undertaken with some data preparation regarding the automatic withholding legal action in order to be useful, because we know automatic withholding is a 100% effective way to recover funds.

Another promising direction for the predictive modeling effort is to examine the value of available textual data. Auditors in the field keep notes of conversations with custodial parents and non-custodial parents regarding cases. These notes may contain information gained during phone calls or other correspondence the non-custodial parent has with the agency. Valuable information regarding an NCP's willingness or ability to pay may be contained in this text and using text mining to analyze this data could prove fruitful.

CONCLUSION

In comparing and contrasting these case studies, our conclusions are:

1. While we can shorten the learning curve for states using our experience, each state has different policies, data collection methods, and demographic characteristics and will have a unique model for child support delinquency. For example, distances seemed important in one state, not so in another.
2. The definition of delinquency should be made with some care. The expected usage of the predictive model, the policies of the state, and the time period for modeling should be considered. This definition has a large effect on the predictive power of the model.
3. There is more work to be completed regarding predicting what types of intervention strategies and enforcement efforts result in converting delinquent payers to non-delinquent payers.

The two states in our case studies are in very different stages of the data mining life cycle. For State A, many variables that proved helpful in State B simply weren't available. In addition, many socio-economic characteristics are unique to regions around the country and individual states. Whether a state is in the initial stages of collecting data for data mining or already testing the results of their model, we conclude that each state will have its own unique combination of factors that contribute to child support delinquency and data mining holds the key to understanding that combination.

REFERENCES

Office of Child Support Enforcement (a). June 2005. "Child Support Enforcement, FY 2004, Preliminary Report." Available acf.hhs.gov/programs/cse/pubs/2005/reports/preliminary_report/.

Office of Child Support Enforcement (b). June 2005. "Child Support Enforcement, FY 2004, Preliminary Report." www.acf.dhhs.gov/programs/cse/pubs/2005/reports/preliminary_report/.

SAS Institute. 2004. Data Preparation for Data Mining Using SAS Software. Course Notes.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors:

Jodi Blomberg
SAS Institute Inc.
6501 S Fiddlers Green Cir Ste 600
Greenwood Village, CO
Work Phone: (303) 723-8503
E-mail: Jodi.Blomberg@sas.com

Jacqueline Long
SAS Institute Inc.
Chicago, IL
Work Phone: (312) 664-4211
Email: Jacky.Long@sas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.