Title
Developing a Practical Forecasting Screener for Domestic Violence Incidents for the Los Angeles County Sheriff's Department

Authors
Berk, Richard
He, Yan
Sorenson, Susan B.

Publication Date
2004-06-03

Developing a Practical Forecasting Screener for Domestic Violence Incidents for the Los Angeles County Sheriff's Department*

Richard Berk and Yan He
Department of Statistics, UCLA

Susan B. Sorenson
School of Public Health, UCLA

June 3, 2004

*This project would have been impossible to undertake without the hard work of Sergeant Robert Jonson, Deputy Cecilia Ramirez, Lieutenant Charles Stringham, and Sergeant Christopher Cale, all of the Los Angeles Sheriff's Department. Thanks also go to the deputies who helped in the data collection.

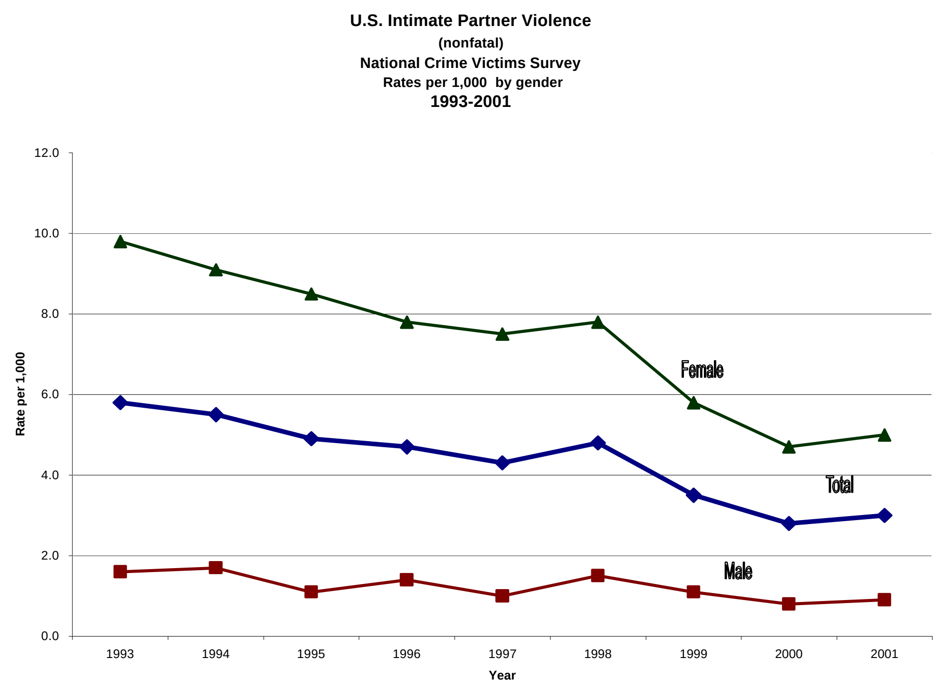
Summary

The goal of the research was to develop a short screener that deputies could use in the field to help forecast domestic violence incidents in particular households. Based on information collected from over 600 households, forecasting procedures were developed. They are based on a few simple questions that can be asked at the scene, selected from a larger set of over thirty questions. From these questions alone, correct forecasts of future calls for service can be made about 60% of the time. Future calls involving domestic violence misdemeanors and felonies can be correctly forecast about 50% of the time. The 50% figure is especially important because such calls require a law enforcement response and yet are a relatively small fraction of all domestic violence calls for service.

1 Introduction

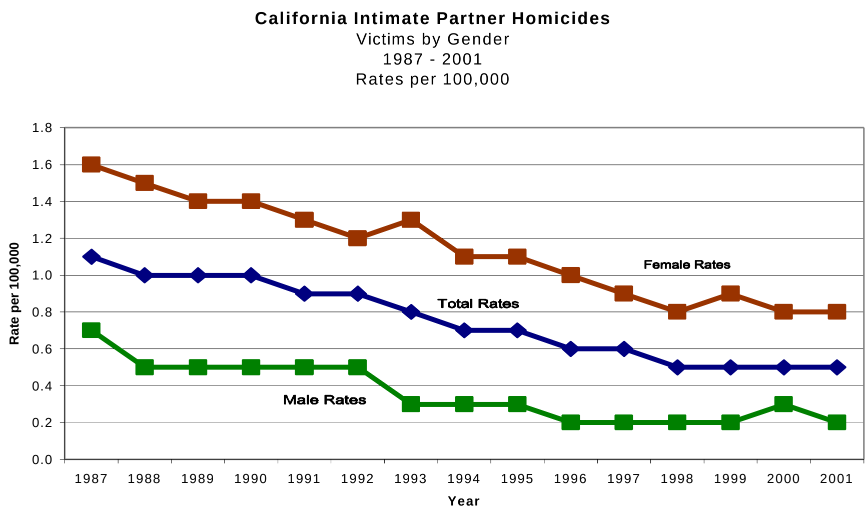
Domestic violence remains a very serious public health and law enforcement problem, especially for women. For the country as a whole, Figure 1 shows that the general declines in nonfatal domestic violence events since the early 1990's have recently leveled off and that women are far more likely than men to be victimized. Figure 2 shows much the same pattern for partner homicides in California. Finally, Figure 3 indicates that in California in 2002, there were 196,569 domestic violence calls to law enforcement with 119,850 involving weapons, which also represents a leveling off since the declines of the 1990s. It follows that the allocation of law enforcement resources for domestic violence incidents needs careful scrutiny, and these resources should be allocated in an efficient manner.

In response, the overall goal of the project was to develop a short screening instrument that would help predict future domestic violence incidents and their seriousness as well. Ambitions were modest. The intent was to help sheriff's deputies elicit information at the scene that may help determine the future services required. No effort was made to develop a state-of-the-art forecasting tool to predict such things as "lethality." An enterprise of that sort would have required far more resources and would have been very difficult. Moreover, the instrument that would likely have resulted would almost certainly have been impractical to employ in the field. We hoped to find a set of approximately five questions deputies might ask, not a full battery of items that would take considerably longer to administer.



	1993	1994	1995	1996	1997	1998	1999	2000	2001
Total	5.8	5.5	4.9	4.7	4.3	4.8	3.5	2.8	3.0
Male	1.6	1.7	1.1	1.4	1.0	1.5	1.1	0.8	0.9
Female	9.8	9.1	8.5	7.8	7.5	7.8	5.8	4.7	5.0

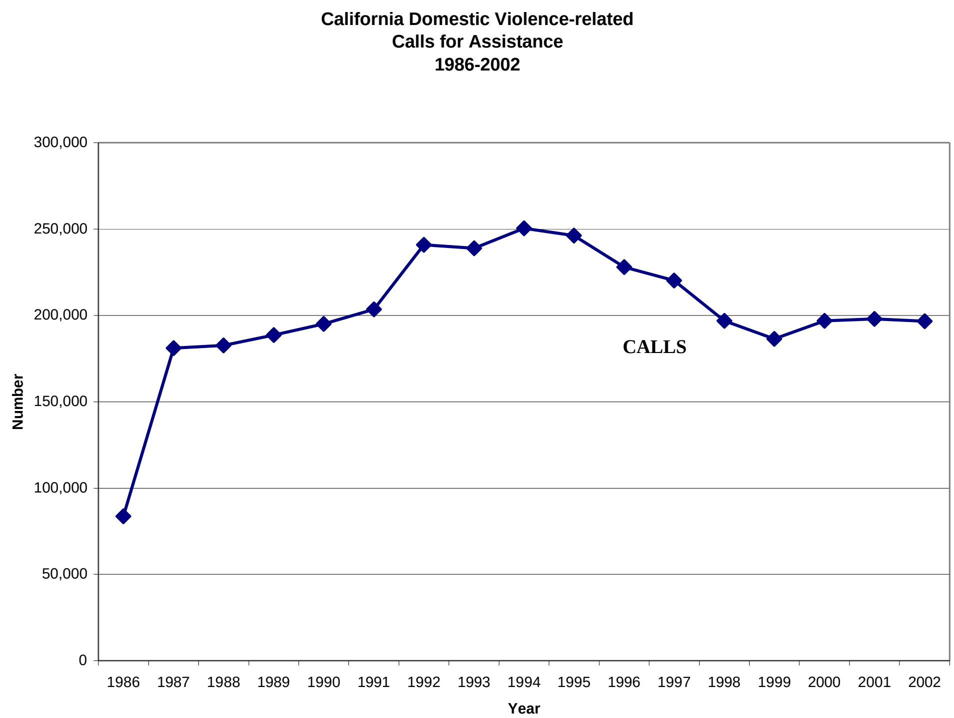
Figure 1: Intimate Partner Violence in the US



	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
Total Rates	1.1	1.0	1.0	1.0	0.9	0.9	0.8	0.7	0.7	0.6	0.6	0.5	0.5	0.5	0.5
Male victims	0.7	0.5	0.5	0.5	0.5	0.5	0.3	0.3	0.3	0.2	0.2	0.2	0.2	0.3	0.2
Female victims	1.6	1.5	1.4	1.4	1.3	1.2	1.3	1.1	1.1	1.0	0.9	0.8	0.9	0.8	0.8

Source: California Dept. of Justice, Criminal Justice Statistics Center, Homicides 1987-2001, ages 20-44, July 2003.

Figure 2: Intimate Partner Homicides in California



Source: *Crime and Delinquency in California*,. State of California, Office of the Attorney General, California Department of Justice, Bureau of Criminal Information and Analysis. Data collection began in July 1986, only six months of data are available. Access www.ag.ca.gov/cjsc/index.htm for more information.

Figure 3: Domestic Violence Calls for Assistance in California

2 Research Design

The original research design specified a representative sample of 1500 households. These were to be households to which sheriff's deputies had been dispatched for incidents that were likely to involve domestic violence. The deputies were to employ a screener of about 30 questions (see Appendix A) as part of their usual duties at the scene. The questions to be asked were selected because of their potential to predict domestic violence in the future. We anticipated that much of the time the victim would be the primary respondent and that information would be obtained from others.

In a three-month follow-up period, all new dispatches to the 1500 household were to be recorded. The answers to the screener would then, with the help of new data mining procedures, be used to predict which households had subsequent calls and how serious those incidents were. It was anticipated that a few of the items would prove to be far better predictors than others. Those better predictors would then comprise the short screener recommend to the Los Angeles Sheriff's department.

Three features of the data need to be emphasized. First, the outcome of interest was a call to the same household during the three-month follow-up period. While these calls were likely to be for a domestic dispute, the calls could involve other law enforcement concerns. The boundaries between domestic violence and other household incidents are sometimes unclear, and we wanted initially to caste a wide net.

Second, the information collected with the screener, based largely on perceptions of victims and others at the scene, was to be used solely for forecasting. Whether these perceptions were always fully accurate does not matter as long as useful forecasts follow. In the same spirit, the screener items were not collected to unravel the many causes of domestic violence. That would entail a different study.

Third, the study was mounted in six substations selected, as a matter of efficiency, because they accounted for largest numbers of domestic violence calls.[1] It was assumed that all deputies in these substations would cooperate. However, it took far longer than expected for the project staff within the Sheriff's Department to obtain permission to field the screener. Subsequent cooperation from deputies was spotty. As a result, deputies collected on-

[1]The substations are Century City, Compton, East Los Angeles, City of Industry, Lakewood, and Lancaster.

the-scene data from fewer than half the households specified by the research design. It is clear that deputies were able to exercise wide discretion in when to use the screeners, but we do not know what consequences there are, if any, for how representative the sample of households is. We turn to some tabulations now that may help to address this issue.

3 Data from the Long Screener

The tabulation from the long screener provides important background information on the kinds of domestic violence cases to which sheriff's deputies are dispatched. Appendix B contains all of these tabulations.

Nearly three-quarters of the households in this study had experienced domestic violence in the past. For these households, the most recent earlier occurrence was within the preceding 6 months. For about half of these households, the police had been called twice before or more. According to what the deputies could learn at the scene, about a quarter of the time an earlier incident had led to an arrest and about 15% of the time, a conviction for domestic violence followed. For nearly three-quarters of these households, the violence is reported to be getting worse with nearly a quarter of the victims seeking medical attention from a previous assault. About 16% said they were treated in an emergency room.

About half of the perpetrators are reported to engage in various forms of domination and intimidation. In addition, a majority have problems with jealousy, drugs and drinking. About 60% destroy property in the household when angry and about a third have threatened to kill someone in the victim's family. In about 10% of the households with children, one or more of the children are reported to have been injured by the perpetrator when he or she is angry. About 10% of the perpetrators are said to have a handgun and a bit more than 40% of those were said to have threatened the victim with it in the past. About 4% of the perpetrators are reported to have a rifle and a bit less than 40% are said to have threatened the victim with it in the past.

Restraining orders were reported to be in place in approximately 10% of the households. A little more than half of the victims reported that they left the perpetrator in the past. Among those who had left in the past, about half had left 2 or more times. A bit less than half of the perpetrators have regular jobs.

These tabulations are about what one would expect from a sample of households to which deputies are called to resolve a domestic dispute. The figures imply that generalizations can be usefully made from the sample to all such households served by the Los Angeles Sheriff's Department. But, there are no guarantees, and the results to follow must be treated with a bit more caution than had the research design been implemented fully as intended.

4 Developing Forecasting Procedures

4.1 Setting the Stage

We had data from the screening instrument for 671 households. However, some of the items on that instrument were not filled out, and once households with missing data were eliminated, there were 516 complete observations.[2] For the 516 households with complete data, there was a least one return call for 109. Thus, about 21% had a return call within three months after the screener information was collected. While there is no guarantee that all of these calls were for domestic violence incidents, no doubt most were.

We began the analysis using simple cross-tabulations to see which screener items were related to whether there were subsequent calls in the households studied. It was apparent that some items had considerable promise. However, which items would be most effective would necessarily depend on the consequences associated with forecasting errors. In this instance, there were two kinds: 1) failing to predict high risk for households that really were and 2) predicting high risk for households that really were not. The former can be viewed as "false negatives" and the latter can be viewed as "false positives." Thus, a predictor that produced few false positives but many false negatives might be discarded if the undesirable consequences from the false negatives were larger than the undesirable consequences from the false positives. We needed, therefore, information from the Los Angeles Sheriff's Department on the consequences of false positives and false negatives.

Efforts to elicit this information from the Los Angles Sheriff's Department led to a general conclusion that false negatives were substantially more

[2]We also considered approaches that would impute the missing data as part of the analysis. But imputation for data mining is quite new and not yet fully accepted, especially when there are substantial amounts of missing data. Many of the screeners had some fields not filled in when they should have been. In the end, we could not use about 15% of the cases.

problematic than false positives. One false negative would produce about the same potential harm as several false positives, but the precise figures for these "costs" could not be determined. As a technical matter, all we needed for our statistical procedures was the ratio of false negative costs to false positive costs, but this was also too demanding. We proceeded, therefore, with four reasonable, but different, ratios of the costs of false negatives to the costs of false positives that would cover the range of likely values: 1 to 1, 2 to 1, 5 to 1, and 10 to 1. Consistent with the information provided by the Sheriff's Department, for none of the ratios was the failure to accurately forecast a new call for service more costly than incorrectly forecasting a new call for service.

One can gain further understanding about the key role of costs using the obtained 21% return call figure. If for every household, one predicted another call within three months, one would be correct about 20% of the time. And, one would also be wrong about 80% of the time. Conversely, if for every household, one predicted no calls within three months, one would be correct about 80% of the time. And one would also be wrong about 20% of the time. Which is a better strategy: always predicting a future call or not? The answer depends on the costs of false negatives compared to the costs of false positives.

If both were equally costly, the best strategy would clearly be to never predict a subsequent call. But now suppose that failing to anticipate future calls was very costly; suppose that these false negatives were 10 times more costly than false positives. Then, the best strategy would clearly be to always predict a subsequent call.[3] In short, the relative costs of false negatives compared to the relative costs of false positives can affect how forecasting is done. And it also affects, therefore, which predictors are likely to be important.

4.2 Forecasting Calls for Service

Credible forecasting of calls for service requires two steps. First, strong associations are required between information contained in the screener and whether there were subsequent calls during the three-month follow-up period. Finding these associations is a task for multivariate statistics. Second, once

[3] This is because $(10 \times .2) > (1 \times .8)$. When the costs are the same $(1 \times .2) < (1 \times .8)$. One can think of these calculations as providing comparisons between the expected costs of different forecasting strategies.

some strong associations are found, these can be used in the future to link screener information from new households to the chances of return calls to these new households. The assumption is that the associations found in the data on hand apply to new households in the future.

Focusing initially on the first step, it is common for criminologists to apply logistic regression when the goal is to determine which predictors are associated with outcome such as ours. For our enterprise, however, there are three serious problems. First, there is no way to effectively introduce costs directly into logistic regression, despite the fact that they are essential. Second, if one ignores costs and proceeds with logistic regression anyway, the findings can be very misleading. One has implicitly assumed that the costs of false negatives and false positives are the same. Third, for these data, logistic regression does not help much. In fact, only 10 true subsequent calls out of 109 were identified correctly as such, even when all of the most promising predictors are used.

Figure 4 shows a histogram of the probabilities from the logistic regression used to identify households with one or more calls for service during the follow-up period. One can see that only a few of these probabilities are larger than .50. The .50 threshold is important because households with probabilities greater than .50 would be classified as having subsequent calls. For these households the chances are better than 50-50. The key message is that only for these very few households does the statistical model imply that the chances are better than 50-50 of a future call for service. Thus, about 91% (99/109) of the true calls are incorrectly determined to have not occurred. Clearly, this is unsatisfactory.

We turned then to data mining techniques and found that Classification and Regression Trees (CART) performed far better for the range of costs ratios elicited from the Sheriff's Department. The 5 to 1 ratio of the costs of false negatives to false positives produced results that were based on a sensible set of predictors that had useful associations with the calls for service during the follow-up period. The 1 to 1 and 2 to 1 cost ratios generated far too many false negatives, while the 10 to 1 cost ratio generated far too many false positives. Figure 5 shows that, compared to the earlier results for logistic regression, there are now a substantial number of probabilities greater than .50.

But how well do we sort households into those that had new calls and those that did not? Table 1 shows the relevant "classification table." From the first row we learn that 56% of the time households with no subsequent

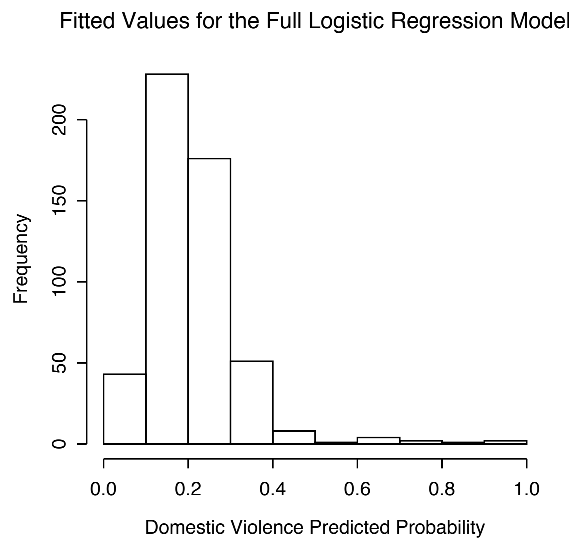


Figure 4: Probability Distribution of A New Call Produced by Logistic Regression

calls are correctly identified. From the second row we learn that 66% of the time households that have a subsequent call are correctly identified. These results are certainly a dramatic improvement.

Another way to look at the table is to consider the relationship between the false negatives and the false positives. There are about 4.9 false positives for every false negative (181/37), which is virtually the same as the 5 to 1 costs introduced into the model. Because false negatives are taken to be 5 times more costly than false positives, the *total* costs for the two kinds of errors balance. This implies that CART is performing as intended.

	Identified as No Call	Identified as Call	Proportion Correct
No Call	226	181	0.56
Call	37	72	0.66

Table 1: Classification Table for the 5 to 1 CART Model

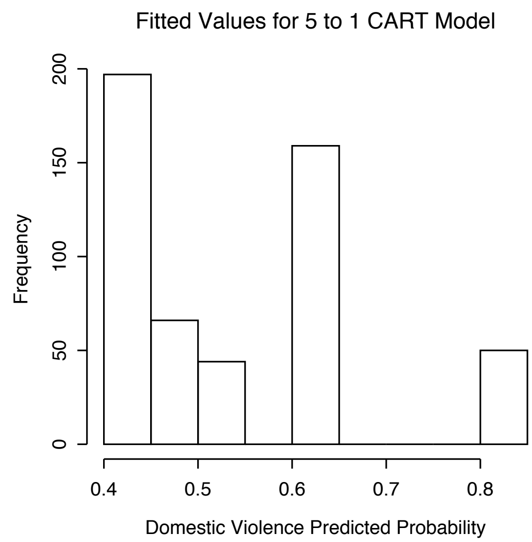


Figure 5: Probability Distribution for New Call Produced by the 5 to 1 CART Model

4.3 Which Screener Items Work?

Figure 6 shows the classification tree produced by CART. The tree is "read" from top to bottom because that reflects the order in which screener items were selected. The ovals represent intermediate subsets of the data while the rectangles represent "terminal" subsets of the data. The letter "b" means that all of the households in that group were identified by CART as "call households" while the letter "a" means that all households in that group were identified by CART as "no call households." The figures in each oval or rectangle show from left to right the actual number of households without a subsequent call and the actual number of households with a subsequent call. At the top, for example, which shows what happens when no predictors are used, there are 410 households without a subsequent call and 109 household with a subsequent call. Given the 5 to 1 cost differential, that "node" conveys that if no predictors are used, one's best guess is to identify all households as "call households." But one can do better moving down the tree to the terminal nodes represented by the rectangles.

From Figure 6, one can see that four screener items were selected because

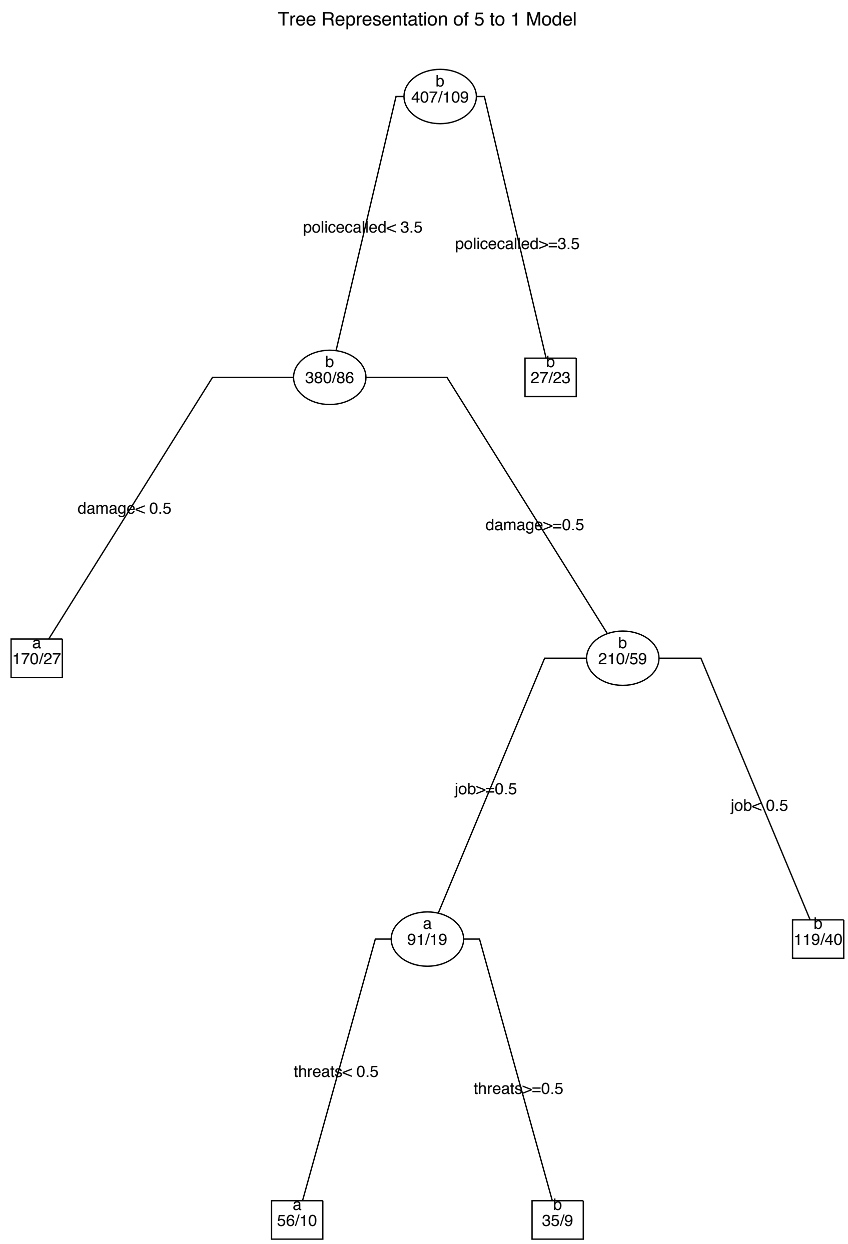


Figure 6: Classification Tree for 5 to 1 CART Model

they substantially help sorting households into "call" and "no call" groups. These four items, in the order selected were: the reported number of previous calls to that household, whether the perpetrator was reported to destroy household property when angry, whether the perpetrator was reported to be unemployed, and whether the perpetrator was reported to have threatened to kill the victim or someone else in the family in the past. The tree structure conveys how these items can be used to forecast calls for service from new households in the future. A prediction of a new call would be made under three scenarios as follows:

I. if the police have been called 4 times or more to a household;

II. if the police have been called 3 times or less, and

   1. the perpetrator is reported to destroy property when angry, and
   2. the perpetrator also is unemployed;[4]

III. if the police have been called 3 times or less, and

   1. the perpetrator is reported to destroy property when angry, and
   2. is employed, and
   3. is reported to have threatened to kill the victim or someone else in the family in the past.

The role of prior police calls is certainly not surprising. It makes sense that past police calls is a good predictor of future police calls. Using that predictor alone with the cutoff of 4 past calls or more, forecasts of future calls will be correct nearly half the time. However, the real power of the analysis is that even when there have been *fewer* police calls in the past, there are other predictors that can be very useful. The first such predictor is whether the perpetrator is reported to do property damage around the home when angry. And the stakes are raised even higher if such individuals are unemployed. Finally, if such a person is employed, the fallback predictor is whether the perpetrator is reported to have threatened in the past to kill the victim or someone else in the family. Note that while there were approximately 30

[4]Binary predictors are coded 1 for the presence of the attribute or action and 0 otherwise. Splits at .5 or greater imply a value of 1 while splits less than .5 imply a value of 0.

items in the long screener, only 4 are needed. No other items meaningfully improve the results.

Unfortunately, it is well known that CART is vulnerable to overfitting. The gist of the problem is that the data used to build the classification tree are also used to determine how well it forecasts. This kind of double dipping, all too common in statistical tools used to predict the future, is likely to produce an overly optimistic picture of how accurate the forecasts really are. A far more realistic assessment can be obtained from the the data mining procedures called "random forests" (see Appendix C). In effect, data used to evaluate how well the model forecasts are not used to build the model. Thus, true forecasts are being undertaken. We emphasize this in the classification tables by distinguishing between CART *identifying* calls and random forest *forecasting* calls.

Given the 5 to 1 cost ratio, the random forest results can be seen in Table 2. Because of the sampling process, the total number of cases is slightly different, but that does not affect the interpretation. As expected, forecasting skill declines a bit. Instead of being able to forecast future calls correctly nearly two-thirds of the time, a more accurate expectation is to be correct about 60% of the time. Likewise, the 56% figure for accurately predicting an absence of calls is more reasonably pegged at about 47%. Still this is far better than one would do ignoring the four predictors.

	No Call Forecasted	Call Forecasted	Proportion Correct
No call	190	217	0.47
Call	45	65	0.59

Table 2: Classification Table for the 5 to 1 Random Forest Model

5 Forecasting New Domestic Violence Offenses

It also may be important to forecast not just any calls to the same household, but new calls for misdemeanor or felony domestic violence. There are 29 such events during the follow-up period, representing 5.6% of the households. We will proceed in essentially the same manner as we did for predicting any new calls.

Figure 7 shows the probability distribution for a logistic regression that used all available screener items. As before, logisitic regression does not perform well. There are only 4 households for which the chances of a subsequent call for misdemeanor or felony domestic violence are greater than .5.

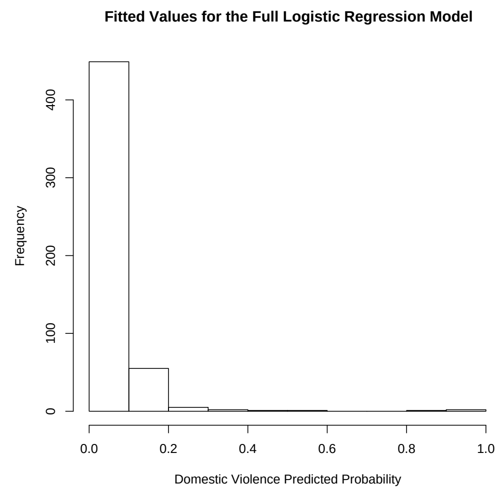


Figure 7: Probability Distribution for a New DV Offense Produced by Logistic Regression

We again turn to CART. But this time we use a 10 to 1 cost ratio of false negatives to false positives. This too is consistent with information elicited from the Sheriff's department and leads to useful results. The harm of failing to forecast accurately DV crimes is considerably larger than the harm of failing to forecast accurately any new calls. Figure 8 shows that CART generates a substantial number of households for which the chance of a subsequent domestic violence misdemeanor or felony are larger than .5.

Table 3 shows that despite the small number of new domestic violence crimes in the 3 month follow-up period, CART does a good job of identifying them. Over 80% of the households with no new DV crimes are properly identified and 50% of the households with a new DV crimes are properly identified. However, because the ratio of false positives to false negatives is only a little over 6 to 1 (88/14), our results do not place quite as much weight on false negatives as the Sheriff's Department would prefer.[5]

[5]This results from constructing a tree (see Figure 9) for which the results were stable

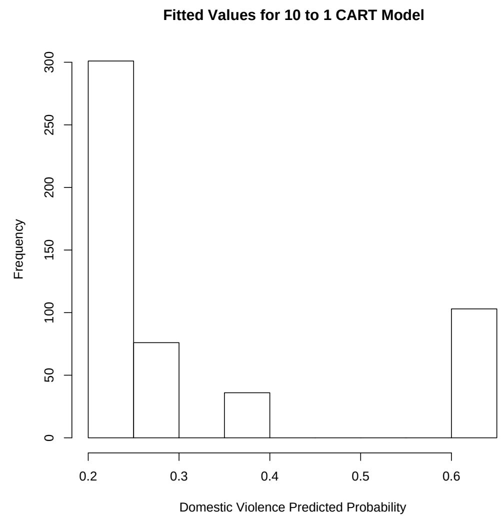


Figure 8: Probability Distribution for a New DV Offense Produced by 10 to 1 CART Model

Table 4 shows that as before, CART overfits the data a bit. The forecasting skill for true positives drops from 52% to 49% while the forecasting skill for true negatives dropsfrom 82% to 70%. But the price paid is small. Note that we now have approximately the 10 to 1 ratio of false negatives to false positives needed.

	Identified as No DV	Identified as DV	Proportion Correct
No DV	399	88	0.82
DV	14	15	0.52

Table 3: Classification Table for The 10 to 1 CART Model

Figure 9 shows the classification tree. Three useful predictors are represented: 1) whether the police are reported to have been called in the past, 2) whether the perpetrator is reported to be unemployed, and 3) whether the

and easily interpreted. Larger trees would have produced the desired 10 to 1 ratio but would not have provided a useful forecasting tool. These are the kinds of statistical tradeoffs one often faces when the event to be identified is relatively rare.

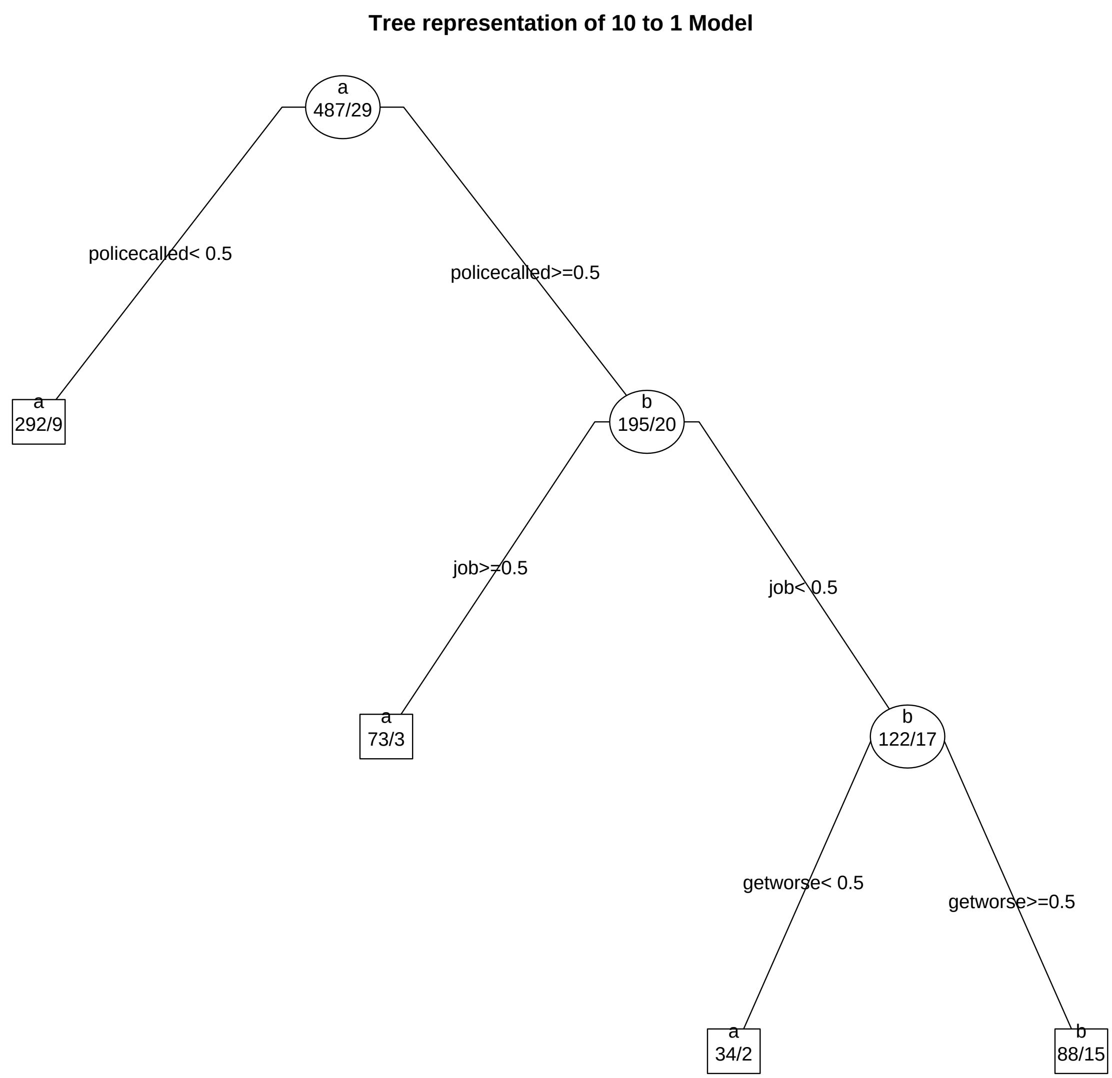


Figure 9: Classification Tree for 10 to 1 CART Model

	No DV Forecasted	DV Forecasted	Proportion Correct
No DV	341	146	0.70
DV	15	14	0.49

Table 4: Classification Table for The 10 to 1 Random Forest Model

violence is reported to be getting worse.

A prediction of a new domestic violence crime would be made under three scenarios as follows:

I. the police have been called to the household in the past;

II. if in addition the perpetrator is unemployed;

III. if in addition the violence is getting worse.

In other words, there are three situations in which a future domestic violence crime should be forecast in order of increasing likelihood. But even the lowest level (i.e. level I) shows useful forecasting skill despite the fact that new calls involving domestic violence misdemeanors and felonies are relatively rare.

6 Conclusions

The conclusions are straightforward. First, it is possible to forecast with useful skill future calls for service. Using a cost ratio of 5 to 1 for false negatives (incorrectly forecasting no future calls) to false positives (incorrectly forecasting future calls), one can accurately forecast future calls about 60% of the time and accurately forecast the absence of domestic violence calls nearly 50% of the time.

Second, one can accomplish this with just four predictors from our larger screening instrument. One cannot do meaningfully better adding more predictors. These four predictors are: 1) whether the victim reports that there have been more than 3 police calls to the household before, 2) whether the perpetrator is reported to damage household property when angry, 3) whether the perpetrator is reported to be unemployed, and 4) whether the perpetrator is reported to in the past have threatened the life of the victim or victim's family.

Third, using a cost ratio of 10 to 1 for false negatives to false positives, one can accurately forecast future *domestic violence* calls about 50% of the time and accurately forecast the absence of future *domestic violence* calls nearly 70% of the time. These are calls involving a domestic violence misdemeanor or felony.

Fourth, for these kinds of calls, just three predictors are required. These three predictors are: 1) whether the police are reported to have been called in the past, 2) if the perpetrator is reported to be unemployed, and 3) if the violence is reported to be getting worse. One cannot do meaningfully better including more predictors.

However, one must keep in mind several important caveats. New calls for service do not necessarily mean that a domestic violence incident has occurred, although in most cases it probably has. Even for calls in which deputies report a domestic violence offense, the offense may later prove to be "unfounded." And there are surely a large number domestic violence incidents for which no call to the police is made. The calls for service forecasted in this analysis are significantly related to domestic violence incidents, but are not the same thing.

In addition, although it may be tempting to infer that our best predictors are also important *causes* of domestic violence, we counsel great caution. For example, unemployment may contribute to domestic violence or, alternatively, the kinds of individuals who have trouble finding and holding jobs may tend to be the same kinds of individuals who can be violent at home. A proper causal analysis would require a different kind of study.

Finally, forecasting is necessarily data dependent. Here, it is likely that forecasting skill would change a bit if the forecasts were for either shorter or longer term outcomes. We suspect that one would forecast a bit better in the shorter term and a bit worse in the longer term. More important, only about half of the specified number of households were included in the study, and it is unclear how these were chosen. The deputies who participated in the data collection had wide discretion in when the screener was employed. We do not know how this discretion was exercised. It is possible, therefore, that the group of households on which the analysis rests is somehow atypical. We find no evidence of this in the data, but it remains a possibility. The best solution would be to replicate the study with a sample of households known to be representative.

Appendix A— The Long Screening Instrument

1. Is this the first time he/she has tried to hurt you? (*Circle one*)
   (a) No [1]
   (b) Yes [2] — **Skip to #12**
2. When was the last time? (*Circle the one for the most recent event*)
   (a) Earlier today [1]
   (b) Within the past week [2]
   (c) Within the past month [3]
   (d) Within the past 6 months [4]
   (e) Within the past year [5]
   (f) Longer than a year ago [6]
3. How many times before has he/she tried to hurt you?
   (a) . . . . . . Times
4. How many times before have the police been called?
   (a) . . . . . . Times — **If 0 skip to #7**
5. Was he/she ever arrested for domestic violence as a result? (*Circle one*)
   (a) No [1]
   (b) Yes [2]
   (c) Don't know [3]
6. Was he/she ever convicted of domestic violence as a result? (*Circle one*)
   (a) No [1]
   (b) Yes [2]
   (c) Don't know [3]
7. Is the violence getting worse as time goes on? (*Circle one*)

(a) No [1]
(b) Yes [2]
(c) Don't know [3]

8. How long ago did the violence start? (*Circle the most recent time that is appropriate*)

    (a) Within the past week [1]
    (b) Within the past month [2]
    (c) Within the past 6 months [3]
    (d) Within the past year [4]
    (e) Longer than a year ago [5]

9. Has he/she ever hurt you so that you needed to see a medical doctor? (*Circle one*)

    (a) No [1] — **Skip to #12**
    (b) Yes [2]

10. How many times?

    (a) . . . . . . Times

11. Were you ever treated for those injuries in a hospital emergency room? (*Circle one*)

    (a) No [1]
    (b) Yes [2]

12. Does he/she have a problem with jealousy? (*Circle one*)

    (a) No [1]
    (b) Yes [2]

13. Does he/she keep track of whom you talk to on the phone? (*Circle one*)

    (a) No [1]

(b) Yes [2]

14. Does he/she try to determine which of your friends you can see? (*Circle one*)

    (a) No [1]
    (b) Yes [2]

15. Does he/she try to put you down in front of your friends or family? (*Circle one*)

    (a) No [1]
    (b) Yes [2]

16. Does he/she have a drinking problem or a problem with drugs? (*Circle one*)

    (a) No [1]
    (b) Yes [2]

17. When he/she is angry with you, does he/she ever try to destroy things around the house? (*Circle one*)

    (a) No [1]
    (b) Yes [2]

18. Has he ever threatened to kill you or someone in your family?

    (a) No [1]
    (b) Yes [2]

19. Are there any children in the home? (*Circle one*)

    (a) No [1] — **Skip to #23**
    (b) Yes [2]

20. How many?

    (a) . . . . . . Children

21. What are their ages, starting with the youngest? (*List the ages in chronological order*)
    (a) (. . .) (. . .) (. . .) (. . .) (. . .) (. . .) (. . .) (. . .) (. . .) (. . .)
22. Has he/she ever intentionally hurt him/her/any of them just because he/she was angry? (*Circle one*)
    (a) No [1]
    (b) Yes [2]
23. Does he/she have a handgun he/she can get to? (*Circle one*)
    (a) No [1] — **Skip to #26**
    (b) Yes [2]
24. Did he/she purchase it himself/herself? (*Circle one*)
    (a) No [1]
    (b) Yes [2]
25. When he/she is angry, has he/she ever threatened you with it? (*Circle one*)
    (a) No [1]
    (b) Yes [2]
26. Does he/she have a rifle he/she can get to? (*Circle one*)
    (a) No [1] — **Skip to #29**
    (b) Yes [2]
27. Did he/she purchase it himself/herself? (*Circle one*)
    (a) No [1]
    (b) Yes [2]
28. When he/she is angry, has he/she ever threatened you with it? (*Circle one*)
    (a) No [1]

(b) Yes [2]

29. Has he/she ever threatened you with some other weapon like a knife? (*Circle one*)

    (a) No [1]
    (b) Yes [2]

30. Is there a restraining order against him/her right now? (*Circle one*)

    (a) No [1]
    (b) Yes [2]
    (c) Don't know [3]

31. Have you ever left him/her (*Circle one*)

    (a) No [1] — **Skip to # 33**
    (b) Yes [2]

32. How many times?

    (a) . . . . . . Times

33. Does he/she have a regular job? (*Circle one*)

    (a) No [1]
    (b) Yes [2]

Appendix B — Tabulations for the Screener (using the 516 complete cases — or less depending on skip patterns)

1. First time he tried to hurt you? — **31% said yes**
    2. When was the last time? — **median is within the past 6 months**
    3. How many times has he tried to hurt you? — **median is 4**
    4. How many times police been called before? — **median is 2**
    5. An arrest for domestic violence as a result? — **28% said yes**
    6. A conviction for domestic violence as a result? — **17% said yes**
    7. Is the violence getting worse as time goes on? — **71% said yes**
    8. How long ago did it start? — **median is within the past year**
    9. Need to see a medical doctor? — **22% said yes**
        10. How many times? — **median is 1**
        11. Treated in a hospital emergency room? — **16% said yes**
12. A problem with jealousy? — **66% said yes**
13. Keep track of whom you talk to on the phone — **48% said yes**
14. Determine which of your friends you can see? — **48% said yes**
15. Try to put you down? — **52% said yes**
16. A drinking or drug problem? — **53% said yes**

17. Destroy things around the house? — **61% said yes**

18. Threatened to kill you or someone in your family? — **34% said yes**

19. Any children in the home? — **74% said yes**

    20. Ever intentionally hurt any of them? — **10% said yes**

21. Have handgun? — **10% said yes**

    22. He/she purchase it? — **50% said yes**

    23. Ever threatened you with it? — **42% said yes**

24. Does he/she have a rifle he/she can get to? —**4% said yes**

    25. He/she purchase it? — **63% said yes**

    26. Ever threatened you with it? — **37% said yes**

27. Ever threatened you with other weapons? — **15% said yes**

28. Retraining order? — **10% said yes**

29. Have you ever left him? — **52% said yes**

    30. How many times? — **median is 2**

31. Regular job? — **44% said yes**

Appendix C: A Primer on Classification and Regression Trees and Random Forests

Classification and Regression Trees (CART) uses a set of predictors to partition the data. Within each partition, the values of the response variable are as homogeneous as possible. The space defined by the data is partitioned one partition at a time in stagewise fashion. Once a partition is defined, it is unaltered by later partitions. The partitioning is accomplished with a series of straight-line boundaries, as in Figure 10, which define a break point for each selected predictor.

Figure 10 illustrates the CART partitioning. There is a binary outcome coded "A" or "B" and this simple illustration, just two predictors x and z. The red vertical line defines the first partitioning. The green horizontal line defines the second partitioning. The yellow horizontal line defines the third partitioning.

The data are first segmented left from right and then for the two resulting partitions, the data are further segmented separately into an upper and lower part. In this simple example, the upper left partition and the lower right partition are perfectly homogeneous. There remains considerable heterogeneity in the other two partitions and in principle, their partitioning could continue. Thus, cases that are high on z and low on x are always "B." Cases that are low on z and high on x are always "A." In a real analysis, the terms "high" and "low" would be precisely defined by where the boundaries cross the x and z axes.

Usually, CART output is displayed as an inverted tree. Figure 11 is a simple illustration. The full data set is contained in the root node. The final partitions are subsets of the data placed in the terminal nodes. The internal nodes contain subsets of data for intermediate steps.

The CART algorithm employs a fitting criterion that determines which binary split for each predictor is best and which predictor should be selected for each split. When the response variable is equal interval, that criterion is usually the error sum of squares. When the response variable is categorical, there are two popular criteria, the deviance or the Gini Index.

CART is vulnerable to overfitting because the fitting function is so flexible. One response has been to construct many trees and then average over these trees. Random samples of the data are drawn, as in the bootstrap, and then at each split a random subset of predictors is selected. This results in

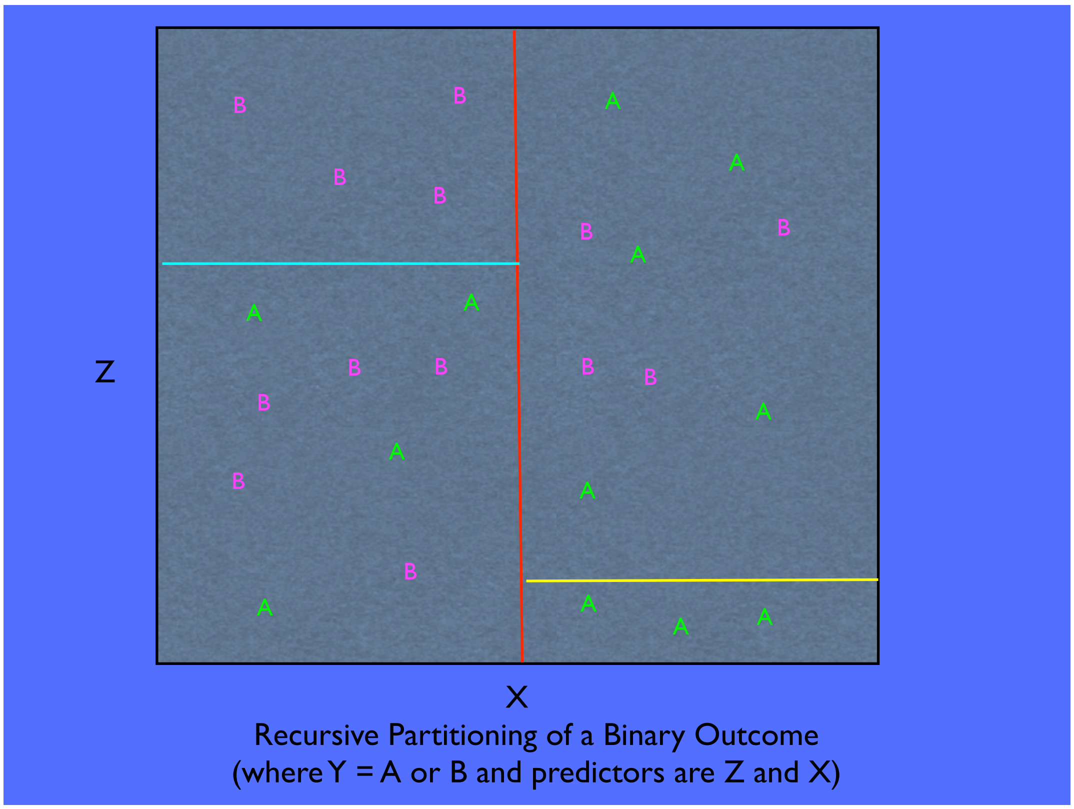


Figure 10: Recursive Partitioning Logic in CART

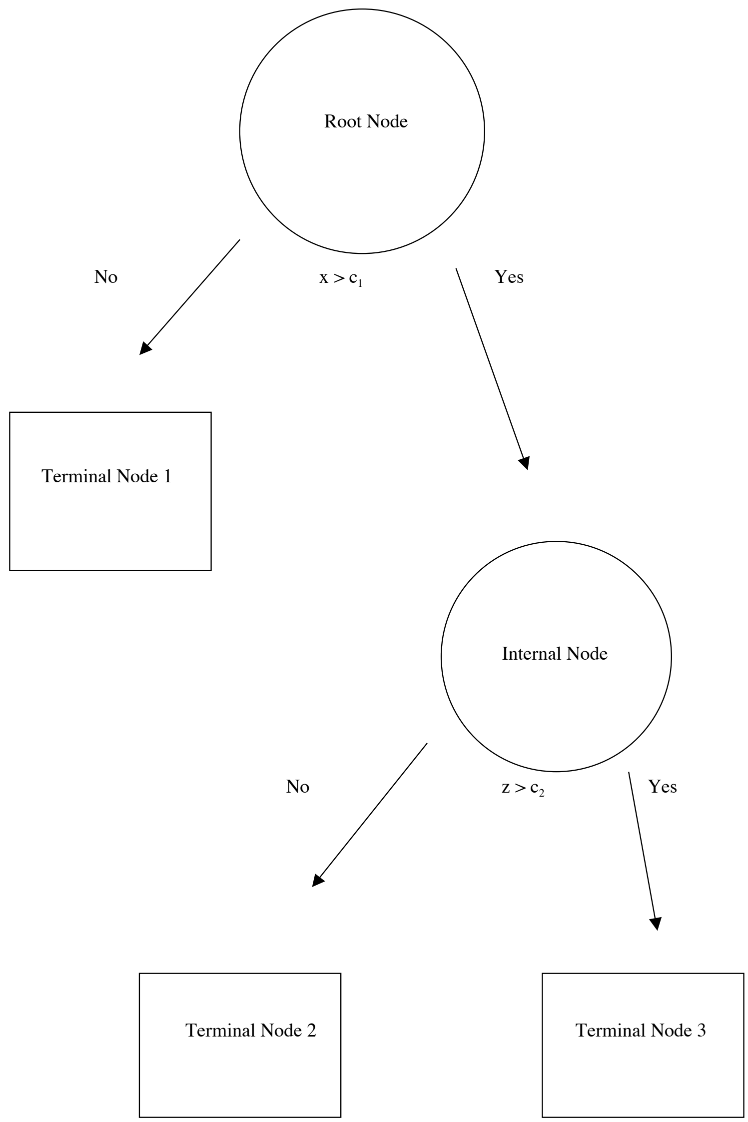


Figure 11: CART Tree Structure

large random sample of trees, the results of which are then averaged. The use of multiple trees (often as many as 1000) makes the fitting function even more flexible, while the averaging over trees compensates for the overfitting. This procedure is called "random forests." CART will generally fit the data better than standard regression models, and random forests will generally fit the data better than CART. Some key references are provided below.

References

Berk, R.A. (2004) “Data Mining with a Regression Framework,” UCLA Department of Statistics Preprint Series # 371.

Breiman, L. (2001). “Random Forests.” *Machine Learning* 45, 5-32.

Breiman, L., Friedman, J.H., Olshen, R.A., and C.J. Stone (1984) *Classification and Regression Trees.* Monterey, Ca: Wadsworth

Hastie, R., Tibshirani, R. and J. Friedman. (2001) *The Elements of Statistical Learning.* New York: Springer-Verlag.