NOT FOR CITATION WITHOUT PERMISSION

Data-Mining Approaches to Suicide and Suicidal Behavior

Final Report

March 2004

Submitted for Funding to Longer Life Foundation

Rumi Kato Price, Ph.D., M.P.E.[1]

Nathan K. Risk, M.A.[1]

Edward L. Spitznagel, Ph.D.[2]

1. Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, U.S.A.

2. Department of Mathematics, Washington University, St. Louis, Missouri, U.S.A.

Executive Summary

We proposed to improve prediction of suicide and suicidal behavior by employing computer-intensive "data-mining" techniques focusing on the genetic algorithms (GA), the artificial neural networks (ANN), and the tree-based regression (TBR) techniques. Although suicide is now considered preventable, predicting who will commit or attempt suicide and when such an act will occur, nonetheless, remains in the realm of intuition among clinicians or family members, thus making targeted and economical prevention difficult. So-called data-mining techniques can overcome limitations inherent in more traditional parametric-oriented approaches. The three techniques proposed in this application are complementary to each other with their combined use maximizing the strengths of each. We proposed to use three data-mining techniques to: **a)** select the most predictive measures of suicide and suicidal behavior using the GA; **b)** examine the patterns of interaction among the most predictive measures chosen by GA; **c)** maximize the predictive power of the selected measures using ANN and compare the results with those by other methods; and **d)** to examine the structure of associations among the most predictive measures using the information stored in the trained ANNs.

We utilized two large datasets available in the public domain: The National Comorbidity Survey, 1990-1992 (NCS, N=8,098) contains detailed timing information on suicidal behavior and environmental and vulnerability factors. The National Mortality Followback Survey, 1993 (NMFS93; N=22,957) includes 1% of all 1993 U.S. deaths ascertained, and 86% of the sampled deceased members followed back with informant interviews. These datasets were chosen for three reasons: large general-population samples to allow for generalization of findings; availability of sufficient numbers of suicide attempters or those who committed suicide; and availability of sufficient predictive measures.

From the NCS datafile, an “initial” dataset containing 57 variables were developed, which contained mostly dichotomous and psychiatric diagnostic variables. An “expanded” dataset containing 77 variables was developed after preliminary analyses indicated the need to refine measures. The latter dataset included many ordinal variables such as symptom counts. After variable selection analyses were completed, all 77 of the variables and the best 15 variables chosen by the GA were separately input to RPART in S-Plus, a relatively widely used tree-based regression (TBR) method. The 15 most predictive measures from each dataset were input to the Multi-Layer Perceptron (MLP) estimation, the most commonly used method in the ANN field, to assess the maximum predictive power of the selected variables. The Receiver Operating Characteristic (ROC) analysis was used to evaluate model performance of the MLP in comparison to the quadratic discriminant analysis (QDA) and logistic regression, where all analyses were cross-validated. Males and females were separately analyzed to obtain gender specific results. The analysis dataset from the NMFS93 was developed to replicate results obtained from NCS. Completed suicide was compared with the category of “deaths due to accident.”

<u>Finding Summary</u>:

a) Best predictors of past-year suicidal thought using the NCS: When applied to both the initial and expanded datasets derived from NCS, the GA-chosen “best” predictors included many variables that are non-significant, if the traditional statistical significance were required. However, those measures chosen by both the forward selection and GA yielded $p<.01$. Using the expanded dataset, the GA tended to pick more ordinal variables compared to the forward selection. The best predictors for males and females were very similar, although not identical.

b) Interaction among predictive measures using the NCS: When GA-QDA selected variables were input, major gender differences emerged. For males, fewer variables were sufficient in constructing the best generalizable tree: depression, impairment due to substance abuse, financial problems, and a loss of daily activities remained in the final model. For females, several more predictors intricately interacted with each other, including relational strengths with relatives. The final tree was more complex than that for males. When all 77 variables in the expanded dataset were input to RPART, the estimation yielded fewer variables after cross-validation.

c) Predictive power for past-year suicidal thought using the NCS: For both the initial and expanded datasets, predictive power was modestly improved with the GA-QDA, but the MLP further improved the predictive power by a considerable amount. Overall, MLP was able to improve the prediction by 8 to 18 % in the Area Under Curve (AUC) value when the GA-chosen variables were input. For example, the AUC=.98 was obtained for males using the NCS expanded dataset.

d) Replicating the results from the NCS using the NMFS93: With respect to the most predictive measures of suicide vs. accidental death, depression and depression-related variables were selected. Not unexpectedly, medically-related variables that may lead to reduced judgement, such as dementia symptom count, were selected to reflect their predictive power for accidental death. For this set of analyses, GA-selected variables did not vary greatly from the variables found with the forward selection. The MLP improved the prediction over QDA by a large margin (12% by the AUC value), and by a smaller margin (5%) over the logistic regression with forward selection. We found that most of the improvement from the logistic regression to MLP was due to MLP's estimation and not by GA variable selection.

e) Structure of ANN weights: Unlike regression coefficients of a parametric approach, weights on MLP paths do not have intuitive meanings. The ranking of the absolute values of weights obtained from the NCS ANN best models show that ranking of measures are more similar between males and females than would have been expected from results of TBR results. These rankings are not consistent with the importance of measures expected from logistic regression analyses run parallel to the GA variable selection runs.

<u>Significance to the Mission of the Foundation</u>:

Several concrete inferences can be drawn. The results of the GA analyses indicate that the current practice in medical research that emphasizes the use of diagnostic variables may undermine the ability to make accurate prediction. The patterns of interaction among predictive measures are rather different between males and females. The results suggest that clinical intervention should focus more on medication for depression and substance abuse intervention for males, but for females, more attention should be given to antisocial personality modification and improvement in social relations. Using a different numbers of predictors, we found that the TBR estimation yielded fewer variables to reach the minimum deviance, indicating that data-mining techniques such as TBR perform better when an initial effort is made to carefully select predictive measures. From the analyses attempting to show the predictive power of the epidemiologic measures, we ascertained high predictive power from our best measures (in particular, AUC=.98 for males in NCS). Such results indicate that existing detailed epidemiological measures contain sufficient information to predict past-year suicidal behavior, if appropriate variables are developed and more flexible estimation methods such as the GA and MLP are used with care.

Overall, the results from the NMFS data file were not as spectacular as those derived from the NCS. It is possible that the proxy measures by next of kin may contain a substantial amount of measurement errors. We also found that comparison of suicides with accidental death to be problematic for some analyses. A very large longitudinal dataset containing a sufficient number of suicides over many years would solve the problems we encountered with the NMFS93 dataset; however such a study would be prohibitively expensive and is very unlikely to be funded in the foreseeable future. The utility of ANN weight structure analysis is currently uncertain. Our analyses showed considerable inconsistencies with our results of linear models such as logistic regressions as well as non-linear modeling results such as the dissimilarities between the male and female TBR results.

Data-mining techniques are not a magic bullet. Researchers must work hard initially to select predictive measures, and create variables in such a way to be informative to particular techniques. Once this work is accomplished, the three methods together are likely to help improve predictive models of suicidal behavior and help us to better understand the patterns of interaction among predictive measures. Combined techniques would provide much improved prediction when sufficiently detailed predictive measures are available. With the arrival of a much faster processor, it may be possible to implement these data-mining techniques in a clinic and provide a rapid and up-to-date assessment of the patient's suicidal risk as new information is added on line to the patient chart.

Acknowledgment

Preparation of this final report was supported by the Independent Scientist Award (K02DA00221) and the Longer Life Foundation grant (LLF2001-003) to the first author. The authors thank Donald J Meyer, Tom Downey, William Shannon for technical assistance and Suzanne Culter and Angel Keller for clerical assistance.

Table of Contents

Chapter 1.

Prediction of Suicide and Suicidal Behavior

Suicide is the eighth leading cause of death in the United States (National Strategy for Suicide Prevention, 2001). Ever since the first life insurance policies were issued, the evaluation of the suicide risk has been one of the most important decisions made by medical doctors and underwriters (Ladd, 2001). Although suicide is now considered preventable, predicting who will commit or attempt suicide and when such an act will occur, nonetheless, remains in the realm of intuition among clinicians or family members, thus making targeted and economical prevention difficult.* The rationale for the study stems from our contention that analytical as well as assessment improvements are needed to increase the prediction of suicide as well as non-fatal suicidality.

Current Knowledge on Suicide and Suicidal Behavior

With consequences so pronounced and so counter-intuitive to the survival instinct universal across species, it is no coincidence that many brilliant philosophers (Camus, 1942), sociologists (Durkeim, 1951), and psychiatrists (Kraepelin, 1896; Freud, 1917) have searched for the meaning of suicide, its motives, the reasoning of those who commit suicide, and the underlying factors associated with suicide and suicidal behavior. Contemporary studies of suicide and suicidal behaviors, spanning genetics, biology, psychiatry, psychology, epidemiology, social sciences, social work, and theology (Maris, 1993), have advanced scientific knowledge in suicide research so much so that there is a wealth of information available (Clark and Fawcett, 1992; Moscicki, 1997; Hawton, 1987; Monk, 1987; Klerman, 1987; Blumenthal, 1996). There are numerous and replicated findings especially with risk

---

* Throughout this proposal, "suicide" is defined as "an act of voluntarily and intentionally taking one's own life" (Murphy, 1992), including physician-assisted suicide and euthanasia (American Association of Suicidology Report, 1996). The term is used interchangeably with "committed suicide." "Suicidal attempt" is a physical act to accomplish suicide, including parasuicide, which is often interpreted as the act with no real intention to die (O'Carroll et al., 1996; Soubrier, 1993). "Suicidal behavior," on the other hand, is more loosely and empirically defined, and covers such behaviors as suicide attempt, plans, and ideation.

factors for suicide and suicide behavior, such as psychiatric (especially affective) disorders (Roins, 1981; Allgulander et al., 1992; Allebeck and Allgulander, 1990; Marttunen et al., 1994; Harris and Barraclough, 1997), alcoholism (Murphy and Wetzel, 1990; Ohberg et al., 1996) and other drug abuse (James, 1967; Marzuk et al., 1992), physical illness (Harris and Barraclough, 1994; Fox et al., 1982; Murphy et al., 1992), personality, family history (Pitts and Winokur, 1964; Tsuang, 1983; Murphy and Wetzel, 1982), and psychosocial factors such as unemployment, hopelessness, living alone, interpersonal conflict, family violence, sexual and physical abuse (Marttunen et al., 1994; Krupinski et al., 1996; Shafii et al., 1985; Briere and Zaidi, 1989; de Wilde et al., 1991; Wagner, 1997; Resick, 1988), and war trauma (Kilpatrick, 1985).

Substantial knowledge is also available about potential genetic (Juel-Nielsen and Videbech, 1970; Roy et al., 1991; Statham et al., 1998; Schulsinger et al., 1979; Wender et al., 1986) and other biological mechanisms (e.g., lower levels of 5-hydroxyindoleacetic acid in the cerebrospinal fluid) (Asberg et al., 1976; Ninan et al., 1984; Winchel et al., 1990) which underlay vulnerability for suicide. It appears that with advances in medical technology, such as the use of functional brain imaging, findings of neurobiological markers could ultimately be incorporated in evaluations of patients' suicidal risk (Mann, 1998). However, at this time, predicting who will commit suicide and when it will occur still remains in the realm of intuition among clinicians or family members (Pokorny, 1983; Murphy, 1983), thus making targeted and economical prevention difficult.

Challenges in Suicidal Behavior Research

A unique challenge in suicide research relates to drastically different ascertainment and assessment methodologies when studying completed suicides, compared to studies of suicidal behaviors. The most common source of information about the deceased used in epidemiologic studies is death certificates (Permeger et al., 1993; Kircher and Anderson, 1987). A number of methodological issues are involved in the analyses of death certificates (Tardon et al., 1995; Mackenbach et al., 1995; Moyer et al., 1989) but from the viewpoint of epidemiologic follow-up research, major shortcomings facing the study of suicides are the rarity of suicides and the lack of information in death certificates that would allow linking causes of death to reasons that led to the ultimate ending of life. Hence, follow-up research often "throws away" the suicides as censored. Suicides are studied very rarely as part of follow-up research unless high-risk sample ascertainment was intentionally designed (Fawcett et al., 1990). Of those who did followup high risk groups, however, prior suicide attempt is a very significant predictor of both the next attempt and eventual suicide: The minimum of a 20% increase in risk is reported for a future suicidal attempt (Leon et al., 1990). Reports from foreign countries show that the risk for completed suicide is 40-to-60 fold up for five years after the previous attempt compared to the general population (Ryngestad, 1997; Soukas and Lonnqvist, 1991). The risk is decreased to a 10-to-25 fold range over 10 years after an attempt (Mehlum, 1994; Holley et al., 1998).

Suicide research has also made extensive use of followback investigation, often called psychological autopsy (Robins, 1981; Jacob and Klein, 1993; Schneidman, 1981). Its applications have been extended to childhood and adolescents (Gould et al., 1996; Brent, 1989), and to geriatric populations (Weisman and Kastenbaum, 1968). This approach has also been extended to the investigation of clinical, legal, and criminal cases (Jacob and Klein, 1993). Through the intensive work of psychological autopsies, many insightful findings have emerged (Gould et al., 1996; Brent et al., 1987; Heikkinen et al., 1995; Conwell et al., 1996). Substantial overlaps have been found between predictors of suicide and those of suicidal behaviors, such as age (Owens, 1990); marital status (Petronis et al., 1990; Bland et al., 1994); psychiatric illness (Rubin, 1993), particularly depression (Hintikka et al., 1998; Kovacs et al., 1993); alcoholism and drug abuse (Petronis et al., 1990; Saxon et al., 1978); physical illness (Owens, 1990; Perry et al., 1990); hopelessness (Wetzel, 1976); and living alone (Owens, 1990). There appear to be unique risk factors for completed suicide compared to non-fatal suicidality, such as leaving a suicide note and using a lethal method (Beck et al., 1985; Tuckman and Youngman, 1963). The field, nonetheless, is still divided on whether suicide completers, suicide attempters, and those who exhibit less severe suicidal behaviors represent the same or different populations (Hawton, 1987; Schneidman and Farberow, 1965).

In our previous work, we noted a dearth of information on the literature of protective factors in suicide research in contrast to the wealth of knowledge on risk factors. Substantial evidence supports the existence of several protective factors: being married, African-American (Kung et al., 1998), employed, having religious commitment (Stack, 1983), having a high level of social support or certain kind of social networks (Yang and Clum, 1994), and having increased family cohesion (McKeown et al., 1998; Hafner et al., 1988). However, a perplexing lack exists of documented mechanisms by which protective factors work. The present lack of research in this area is clearly echoed in the Surgeon General's Call to Action to Prevent Suicide (U.S. Department of Health and Human Services, 1999).

With the recent increased public interest in suicide prevention (Hyman, 2000), the need is even greater to understand the structure of protective factors and their interaction with risk factors, focusing on those at increased risk for suicide and suicidal behavior.

Advantages of Data-Mining Approaches

A data-mining approach relies primarily on empirical knowledge and minimizes the importance of generalization by statistical inference. It builds a model by "learning" as many details as possible about the raw data, through using computational power. In this approach, validity assessment depends more on the ability of the constructed model to predict "future" data. The importance of making statistical inferences is reduced if future data are indeed available to validate the empirically obtained constructs and rules. By their flexible modeling assumptions, data-mining approaches can overcome limitations inherent in linear models, which limit the ability to construct a potentially more predictive model (e.g., a substantively meaningful predictive measure with a large effect and a large standard error may be dropped in favor of another measure with a small effect and an even smaller standard error) (Price et al., 2000; Achen, 1982). In our previous study (Price et al, 2000), we found that data-mining techniques can improve the prediction when the associations of predictors are highly collinear, as in the case with many of the risk and protective factors of suicide and suicidality. Estimation by data-mining techniques, in some instances, provide better results than do conventional linear techniques when the prevalence of the outcome measure is very low, as in suicide and suicide attempt. Furthermore, when the patterns of association or causality are not well known, as in the patterns of interaction among risk and protective factors, and among protective factors themselves, data-mining techniques can offer flexible and exploratory analyses, which can then guide linear modeling.

Research Questions

Predicting suicide and suicidality is difficult. However, it is not clear whether the lack of precision is due to the randomness of such behavior, a lack of measures conducive to quantification, or to inadequate analytical techniques, all of which reduce the potential utility of predictive measures. In the subsequent chapters, we will address three questions: **a)** what are the best available quantitative measures that maximize the prediction of suicidal behavior? **b)** what are the patterns of interactions among many risk and protective factors which predict suicide and suicidality? and, **c)** how good is the predictive utility of the available quantitative measures? Using three data-mining techniques, we will attempt to answer these questions. We will also address limitations of linear methods by comparing results of the three data-mining techniques with those of selected linear models.

In the subsequent chapters, the two datasets and three data-mining techniques are introduced (Chapter 2), followed by results of the applications of the data-mining techniques to selected measures from the two datasets (Chapter 3). This report concludes with summaries of the study and a description of future directions.

Chapter 2. Methodology

Two existing public-domain datasets were utilized because large datasets were needed: National Comorbidity Survey, 1990-1992 (NCS, N=8,098) and National Mortality Followback Survey, 1993 (NMFS93; N=22,957). The NCS was used because the data file contained general-population living respondents; the NMFS93 datafile was chosen because it was a general-population study of the deceased. The three data-mining techniques utilized are complementary to each other with their combined use maximizing the strengths of each. For example: **a)** the genetic algorithms (GA) is most suited for selecting the optimal information for a non-linear target function; **b)** the tree-based regressions (TBR) is most suited for detecting the patterns of interaction among input measures, and identifying subgroups whose outcome behavior is most predictable; and **c)** the artificial neural networks (ANN) is best suited for arriving at the maximum potential predictive utility of the given input information, and we can learn how it reaches its maximum predictive utility.

Two Datasets

The two datafiles, the National Comorbidity Survey, 1990-1992 (NCS, N=8,098) and National Mortality Followback Survey, 1993 (NMFS93; N=22,957) were chosen for three reasons: **a)** both surveys were conducted on epidemiologically ascertained samples, thus the results of our study will be generalizable to general populations; **b)** sufficient numbers of suicidal positives or completed suicides were available so that an adequate number of pattern variations will be captured by the data-mining techniques; and **c)** a sufficient number of measures of interest were available, including suicide measures, psychiatric disorders, substance abuse, physical illness, environmental risk and protective factors, and health care utilization. The investigative team's familiarity with the datasets was an additional factor.

National Comorbidity Survey, 1990-1992 (NCS):

The NCS (Warner et al., 1995; Kessler et al., 1994; Kessler et al., 1997) was designed to study the comorbidity of substance use disorders and psychiatric disorders, using a representative sample of persons aged 15 to 54 in the United States. The survey work was carried out by the Survey Research Center at the University of Michigan, Ann Arbor, from 1990 to 1992. The study was designed to assess major substance abuse and psychiatric disorders according to DSM-III-R, based on a modified version of the Composite International Diagnostic Interview.

The NCS was based on a stratified, multistage area probability sample of the civilian non-institutionalized population with ages between 15 and 54, who were residing in the contiguous 48 states. The study also included a supplemental sample of students living on-campus group housing. The response rate was 82.6%. The total sample size was 8,098. Sampling weights are available to adjust for prevalence estimates; however, weights were not used in our study because our primary interest was application of data-mining techniques.

National Mortality Followback Survey, 1993 (NMFS-93):

The 1993 National Mortality Followback Survey (National Center for Health Statistics, 1998) is the sixth in a series of surveys first initiated by the National Center for Health Statistics (NCHS) in the 1960's to provide additional information related to the mortality experience of the United States beyond that which could be obtained through the vital registration of deaths. The survey was a collaborative project between NCHS and various federal agencies, state and local governments, colleges and universities, and private associations and organizations. This survey was arguably the largest national general-population "incidence" study of deaths available. However, as it used a cross-sectional data collection design, information about the deceased prior to death relied heavily on the information provided by their informants. The NMFS Medical Examiner/Coroner Abstract drew upon information from the death certificates, investigative reports, and case summaries. The NMFS93 Respondent Questionnaire and Medical Examiner or Coroner Abstract included the assessment of the living situations of the deceased during the last year of their lives, including: demographics, living condition, circumstances of death, violence, medical and psychiatric conditions, alcohol and drug use, services utilization, and access to care.

The population was defined as total deaths of those aged 15 years or over in the U.S. in 1993 (over 2.2 million), except deaths registered in North Dakota. Roughly 1% of the deaths was ascertained from a 10% drawing of death certificates from 50 sampling strata, with an oversampling of those with external causes, very young or old, and blacks; and informant sampling of 10% of the deceased identified in death certificates. Informant interviews were not obtained for 14% of the deceased sample.The total sample size was 22,957. The results can be weighted and standard errors can be adjusted, however, again, we did not employ these adjustments since most data analyses were done using data-mining techniques.

Measures

Table 2.1. presents a detailed cross-reference of measures from the two datasets. The measures cover demographic/SES, life experience, mental health, access to care, violence, and suicidality measures. The NCS has more measures relating to violence, while the NMFS93 has more measures on the circumstances of death, including suicide-related measures. Information on parasuicide is unique to NCS. Most measures in NCS include information on onset and recency. Measures included in NMFS93 relate to experiences for the last year of the deceased, mostly reported by informants for the deceased.

Basics of the Three Data-Mining Techniques

<u>Genetic algorithms as an optimization technique</u>:

Although entering a large number of input measures by eschewing the minimum sample size considerations is commonly done (Zou et al., 1996), it is important to elminate measures that contribute only to noise learning when applying data-mining techniques. As an alternative to linear and logistic regression variable selection based on the significance levels, we propose genetic algorithms (GA) as a search technique. GA was originally devised as a special case of evolutionary computation to select the best population members, using the principles of fitness, reproduction, recombination and mutation (Koza, 1992). Our application departs from the original GA algorithms (Holland, 1975; Goldberg, 1989), so that a string is composed of a set of variables (**Figure 2.1.**). GA selects variable sets (mates) as follows: **a)** it starts out with a randomly generated initial population consisting of N sets of variables, with a constraint of unique elements within the string; **b)** each subset of variables (string) is rank-ordered according to a fitness function; **c)** a "mate" of a variable set is chosen with a monotonically decreasing probability corresponding to the fitness function ranking, with a constraint that two subsets must contain more than one different variable; and, **d)** the process is repeated until N variable sets have attempted to find "mates." A crossover is achieved, resulting in two new sets of variables which contain common variables of the "parent" variable sets, with remaining variables randomly assigned to the paired elements of the new sets from the paired elements of "parent" variable sets. Mutation is applied after crossover to each variable set of the population. Each variable is randomly replaced by a variable unique to the set with a defined (usually low) probability, resulting in a new variable set (Downey and Meyer, 1994).

GA is a highly flexible technique because the fitness function can be specified to be tailored to the optimization of interest. Examples include mean square errors (Vivarelli et al., 1995), minimization of weights and maximization of frequencies (Dhingra and Lee, 1994), optimization of profit (Koza, 1992), and least square minimization (May and Johnson, 1994).

For variable selection, we use the GA in combination with the quadratic discriminant analysis (QDA) as the evaluating model of the GA selected measures, because the QDA provides a non-linear discriminant function with use of quadratic curves. In comparison, we use logistic regression models and the forward selection of individual measures, which uses the Wald χ^2 as the criterion for variable selection (**Figure 2.2.**).

Tree-based regression (TBR) methods:

We use the tree-based regression (TBR) method to capture the patterns of interactions among predictive measures and to identify subgroups who are most at risk for suicide and suicidal behavior. The TBR results are at once informative, in contrast to results of the ANNs which is often considered a “black-box” operation (since ANN weights do not have real meanings) (Duh et al., 1998).

Known most commonly as the classification and regression tree (CART®) (Breiman et al., 1984), the tree-based regression method is a flexible non-parametric alternative to linear regression models. The most common TBR, which follows the methods used in CART®, operates by using a procedure known as recursive partitioning. The tree function is given an output variable y and a set of regressors x_i. If the output y is a scaled variable, the tree is called a regression tree; if the output is a binary variable, the tree is called a classification tree. As shown in **Figure 2.3.**, the tree function starts at the root node (the complete set of all data) and then divides the data into two child nodes, by splitting the data into two sets based on the regressors x_i. The choice of variable to use in the split and where to make the split is determined by the “impurity” which measures how homogeneous a node is; the tree function chooses the split that lowers the impurity of the two child nodes by the greatest amount, subject to minimum necessary split size. The tree function continues to split the data recursively until either minimum impurity or minimum node size is met. The minimum node size and the minimum split size are used to prevent the tree from over-fitting. Without the constraints, any tree with enough variables could exactly specify the outcome variable. Several measures of impurity are available, such as deviance, and the Gini Index (Mingers, 1989; Therneau and Atkinson, 1997). A simple example of how to compute the deviance for a three node analysis is shown in **Figure 2.4.**

The artificial neural network (ANN) models :

The center piece of the analytic plan is modeling ANN architectures. Artificial neural network (ANN) modeling has been used for various tasks, including speech recognition, stock market prediction, mine detection, cancerous cell identification, and handwriting recognition. We use the ANN as a computational modeling tool, the center piece of the data-mining approaches taken in this application. In this context, ANN can be characterized as a highly flexible nonlinear regression technique. ANN's ability to transform nonlinear input covariances to linearly separable ones via activation functions supposedly gives ANN an edge over the linear modeling technique in some situations. In practice, however, reports of ANN's ability to surpass linear models for classification and prediction problems have been inconsistent (Doig et al., 1993; Baxt, 1995; Dybowski et al., 1996; Anderer et al., 1994; Duh et al., 1998), depending on data complexity, evaluation measures, and the user's familiarity with the data as well as modeling.

Multi-layer perceptron (MLP) modeling has been used most often in clinical medicine and psychological literature. MLP models are considered to be parametric, quasi-parametric or nonparametric (infinitely parametric to be more precise), depending on the complexity of the model. As a reference point, the generalized additive model (Hastie and Tibshirani, 1990) can be placed somewhere between parametric linear regression and fully nonparametric nonlinear regression such as MLP (Cheng and Titterington, 1994; Sarle, 1994; White, 1992). An MLP model consists of multiple layers of inter-connected neurons.

Figure 2.5. illustrates the iterative process of MLP training for single hidden-layer models. A hidden layer is used for ANN's internal transformation: hidden neurons transform the "signals" from the input layer and generate signals to be sent to the next layer, so that MLP is often called the "feed-forward" neural network. The input layer consists of input variables plus an input bias, which adjusts the weighted mean of the input variables. All input variables are regressed on the next layer, in the same fashion as multiple regression estimation. With a high number of hidden neurons with nonlinear activation functions and multiple layers ANN is highly flexible. Any function can be approximated with increasing numbers of hidden neurons using the sigmoid activation function (Bishop, 1995). Once the predicted values are forwarded from the hidden layer to the output layer, the error amount is

computed based on the target value of outputs. Weights are recalculated after each observation until all observations are processed. More details of the MLP models are found in neural network textbooks (Bishop, 1995; Ripley, 1996; Rumelhart and McClelland, 1986; Anderson, 1995; Price et al, 2000).

In our applications, sigmoid and Gaussian activation functions are used for transformation in hidden neurons (f(Vpj) in Figure 2.5) (**Figure 2.6.**). The sigmoid activation function is most commonly used in MLP and it is essentially a logistic function. It turns the signal at a certain value in the monotonically increasing curve. Another activation function is called the generalized Gaussian activation function. Using this function, the neuron produces the signal when the input is near a certain value of the hill. ANNs sometimes reach conversion faster using the Gaussian function; however, efficiency really depends on the nature of the data and the models applied.

<u>ANN model evaluation</u>:

Prediction needs to be based on the minimization of errors but weighted by substantive importance. We employed the Receiver Operating Characteristic (ROC) Analysis as our primary method of evaluating ANN models. Originally invented as an engineering technique for radar detection, the ROC analysis has been subsequently developed in the medical decision theory and signal detection literature (Weinstein and Fineberg, 1980; Swets and Pickett, 1982). It combines information on both true positives and false negatives by plotting sensitivity and (1 - specificity); therefore, ROC is free from problems of cut-offs (Dwyer, 1996). To assess the overall predictive utility of measures, the area under the ROC curve (AUC) is commonly used, which ranges between .5 and 1.0, where .5 is random prediction (see an example in **Figure 2.7.**). The AUC is usually computed using the trapezoid method, and is believed to be asymptotically equivalent to the value of the “c”statistic; (Bamber, 1975) thus, the predictive utility of ANN can be compared with the standard techniques such as logistic regression. Empirically, when prediction is made from self-reported behaviors, an AUC value rarely extends beyond .8 even with established risk factors (Robins and Price, 1991).

The MLP which is evaluated by the Receiver Operating Characteristics (ROC) analysis at each iteration, is extremely CPU intensive when combined with genetic algorithms (GA), and is not advantageous until GA has narrowed the variable selection to a manageable number. Alternatively, we obtained initial estimations by the quadratic discriminant analysis (QDA) with GA (**Figure 2.8.**). Theoretically, this strategy should not reduce predictive power in the subsequent stage of MLP estimation after a manageable number is selected by the QDA with GA.

Cross-Validation:*

A main weakness of data-mining techniques is danger of over-fitting to the data. Lacking real future data in practice, cross-validation is an essential part of the ANN methodology. For example, ten-fold validation, a typical choice in ANN modeling (Bishop, 1995), means that 90% of the data is used to train ANN, and the remaining 10% of the data is used to compute the numbers of positive and negative errors (**Figure 2.9**). Results of each validation are obtained from the predicted outcome for each observation based on estimates obtained from the testing phase. The process is repeated 10 times and results are summed over 10 validation results to draw ROC's. Choice for the number of validation sets depends on sample size and time constraints.

In TBR, the problem of over-fitting is also addressed by cross-validation, but in a different way.

---

* Cross-validation addresses the issue of over-fitting, it does not directly address sample size considerations. Two considerations are important: **a)** the sample size needs to be large enough to avoid over-fitting, i.e., to avoid drawing conclusions about the population based on the sample being used to train an ANN; and **b)** the sample size needs to be large enough to enable ANN to formulate a "true" model, which is analogous to a "power" computation in more traditional statistics. The proposed ANN analyses will avoid making inferences based on overfitted models, since cross-validation will automatically accompany most analyses. Since a prediction is based on data that was not used in creating the model, the final fit functions will reveal over-fitting among models that are too complex for the amount of data provided. The second issue is not easily addressed in the ANN literature. While cross-validation can assure generality of the results, it does not address whether over-fitting is in fact ANN's uncovering fine reality. Many factors affect the necessary sample size to discern a small effect, including the total number of weights, the degree of precision needed, the degree of dependence among covariates, and the degree to which the modeled distribution fits the actual distribution; therefore computing an explicit necessary sample size is considered an intractable problem in general (White, 1989). Using the total number of weights, *s*, "rules of thumb" state that anywhere from *s* x 2 to *s* x 30 is an acceptable sample size (Sarle, 1999). However, any rule of thumb should be considered only as an educated guess.

We will focus on the cross-validation method implemented in RPART (Therneau and Atkinson, 1997; “MathSoft,” 1997). First, in the model building stage, the deviance measure is used to build a standard classification tree. Then the original tree is subjected to cross-validation. If a tenfold cross-validation is chosen, 10 trees will be constructed from 90% of each of the ten partitions that are evenly divided randomly. The 10 trees are then tested for predictive accuracy with the remaining “held out” sample. Since the test trees are evaluated with data that were not used to construct them they are not subject to over-fitting. The proper depth or size of the tree itself is determined by looking for the point at which the new test trees fail to improve the misclassification rate. The original tree is then pruned to this point. This process is diagramed in **Figure 2.10.**

Infrastructure

We found 44 freeware and 40 commercial software (Sarle, 1999) at a few ANN websites. For a majority of GA and ANN analyses, we used Partek’s software (PARTEK 1993) because it included ANN and GA, and ROC was implemented during our pilot data collaboration a few years ago. For TBR modeling, the TREE and RPART procedures in S-Plus (Therneau and Atkinson, 1997; MathSoft, 1997) were utilized. Comparison statistical analyses were conducted using standard packages such as SAS and STATA (Stata Statistical Software, 1999). Until a few years ago, a practical problem with most data mining techniques was CPU’s. With explosive expansion in computer technology, fast machines are much more affordable these days. A PC configured for high memory handled most of the ANN estimations as well as the TBR models. Because many iterative and repetitive analyses were conducted, it was necessary to organize the runs in a way to minimize manpower while estimation continued.

Summary

In this chapter, the two datasets used for this study and the basics of the three data-mining techniques are introduced. **Table 2.2.** summarizes the four purposes of the data analyses (variable selection, identifying subgroups at risk, overall prediction and corresponding data-mining techniques used (GA, TBR and ANN), and comparison of the biostatistical methods utilized (forward selection, logistic regression, discriminant analysis). To assure generality of the data-mining techniques, cross-validation was employed to TBR and ANN models.

Table 2.1

Measures from NCS and NMFS

Domains/Measures	NCS	NMFS
Demographic/SES		
Sex	*	*
Age	*	*
Country born		*
Race	*	*
Income/assets	*	*
Education	*	*
Life Experience		
Marital status	*	*
Living alone	*	*
Nursing home		*
Home assistance		*
Mobility outside of home		*
Level of daily activities	*	*
Recreational activities		*
Family connectiveness	*	*
Support from friends	*	

Employment	*	*
Religion	*	*
Money problems	*	
Trauma	*	*
Physical characteristics		*
Physical health	*	*
Mental health		
Disease: Alz., Dementia		*
Cognitive impairment		*
Resolve/hope	*	*
Depression	*	*
Paranoia	*	*
Post traumatic stress disorder	*	
Anti-social personality	*	*
Alcohol dependence/use	*	*
Drug dependence/use	*	*
Tobacco dependence/use		*
Access to care		
Insurance	*	
Desire to seek help	*	*
Belief in efficacy	*	
Received treatment	*	
Any medical	*	*

Any psych/mental	*	*
Any alc/drug	*	*
Psych/mental facility	*	*
Alc/drug facility	*	*
Other barriers to care	*	*
Violence		
Combat	*	
Rape	*	
Physical attack	*	
Held captive/kidnapped	*	*
Murdered		
Suicidality		
Suicide specific		*
Thought	*	*
Attempt	*	
Completion		*
Availability of method		*

Table 2.2

Data-Mining Approaches to Suicide and Suicidal Behavior: Purposes and Techniques

Purpose	Data Mining Techniques	Comparison Biostatistical Method
Variable Selection:		
Select the most predictive predictive measures of risk and protective factors	Genetic Algorithms (GA)	Forward Selection (FS)
Identify Subgroups at Risk:		
Delineate patterns of interaction among predictive measures	Tree based Regression (TBR)	Not Examined*
Overall Prediction:		
Maximize the predictive power of the selected measures	Artificial Neural Networks (ANN)	Logistic Regression, Quadratic Discriminant Analysis (QDA)
Generalization:		Validation method

Figure 2.1. Principles of GA

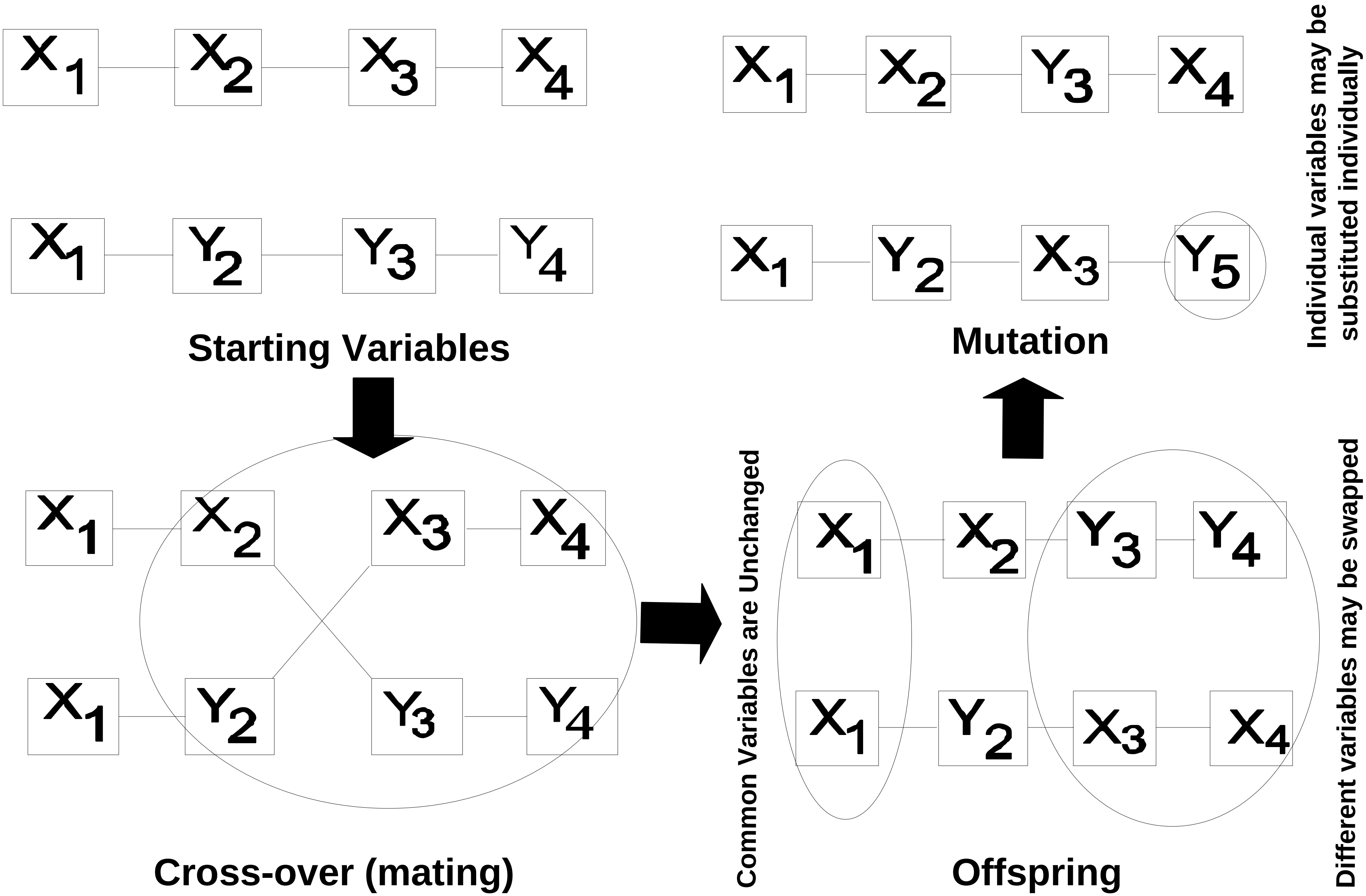


Figure 2.2. Analysis Steps Comparing Traditional Statistical vs. Data-Mining Techniques

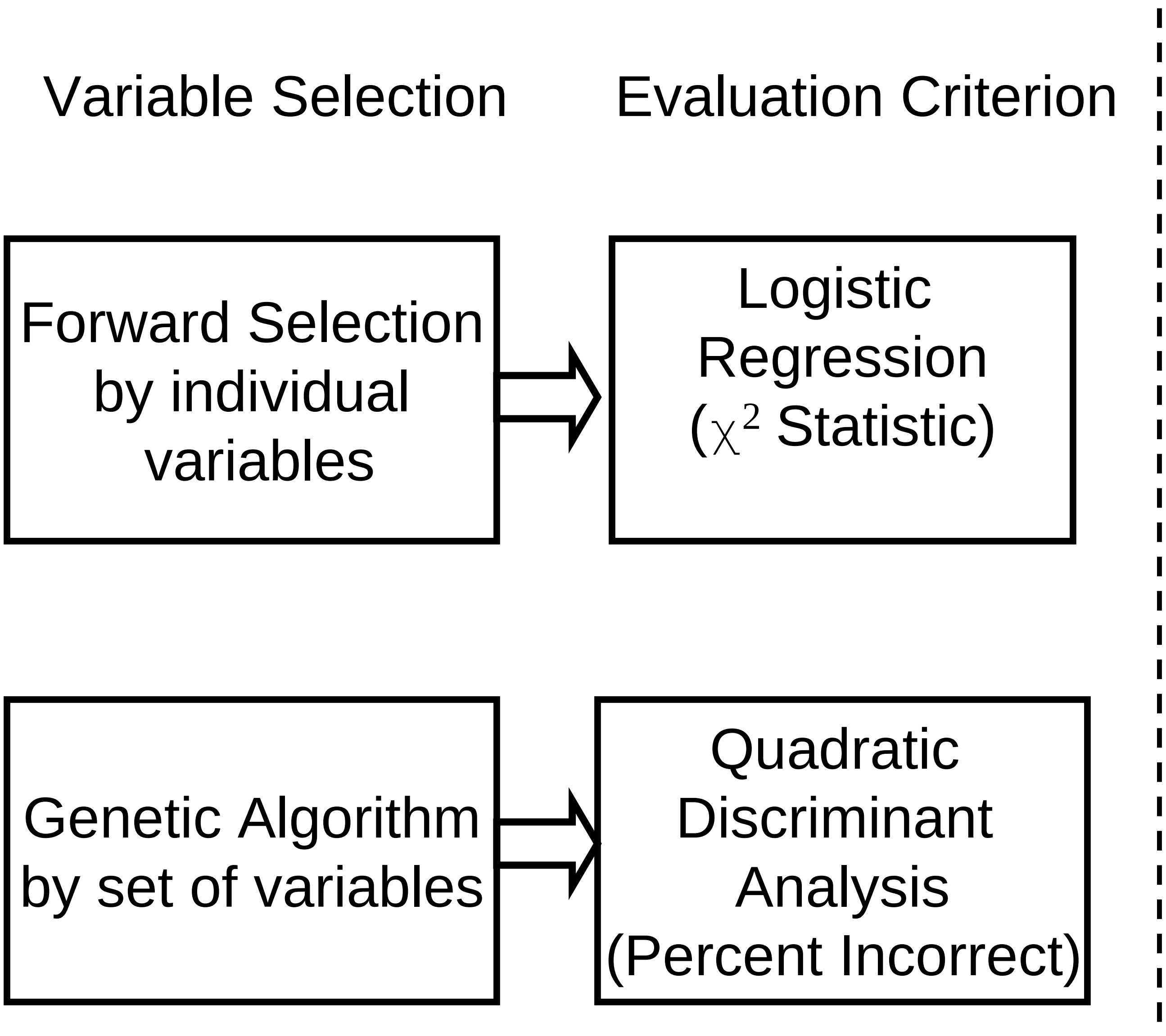
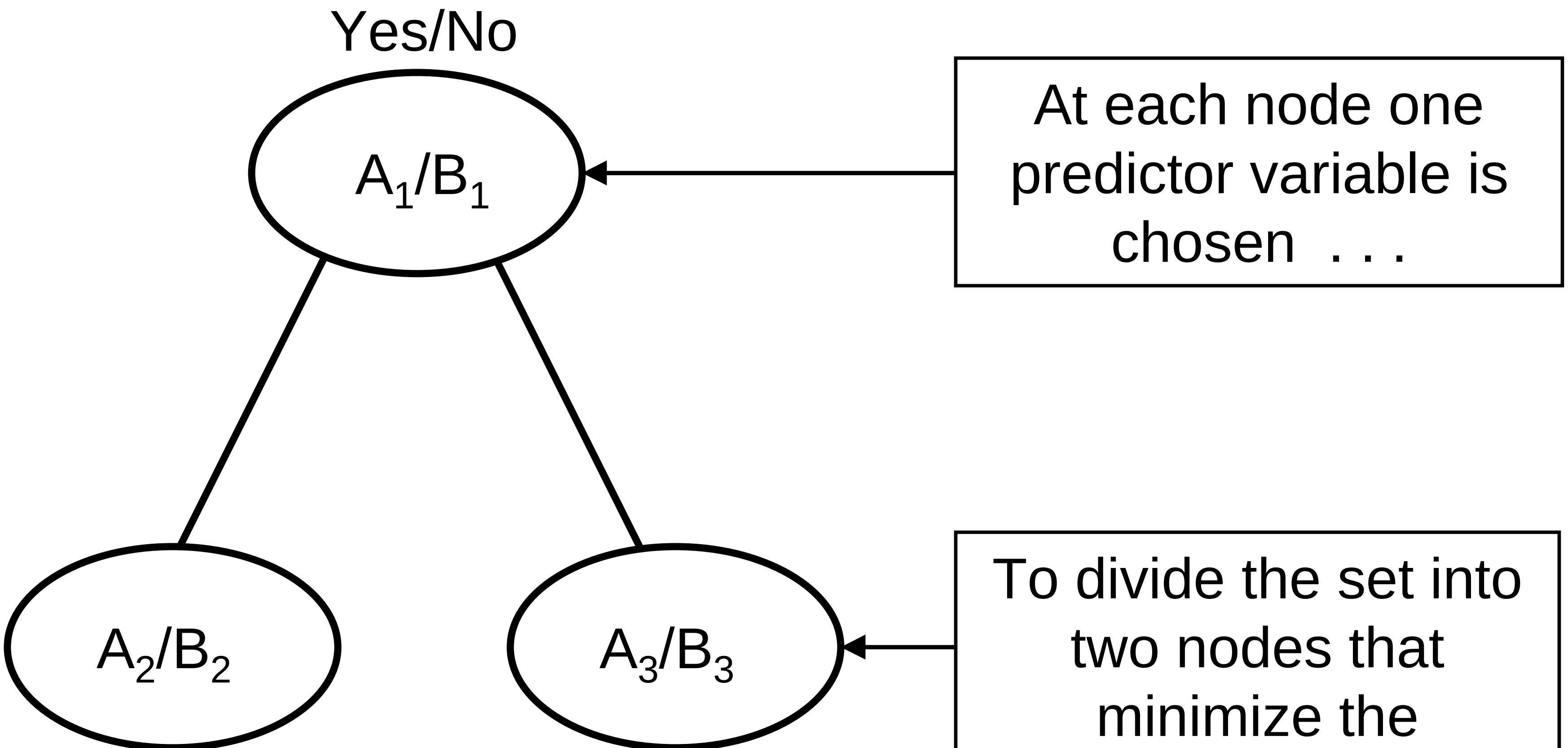
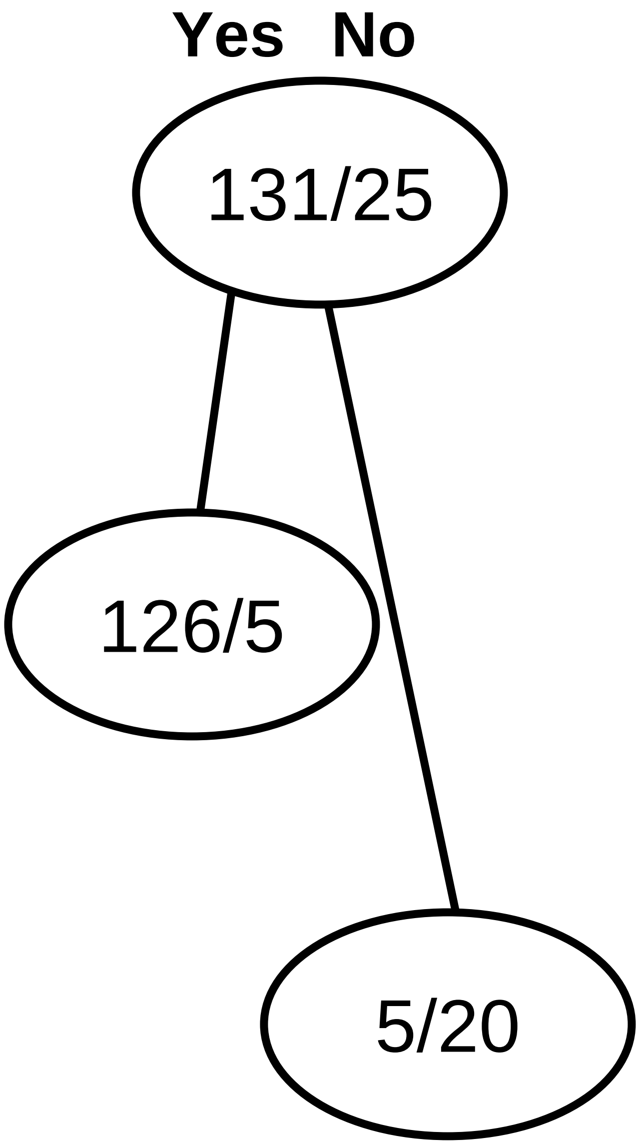


Figure 2.3. Basics of Tree-Based Regression (TBR)



$$D_i = -2\left(A_i \log \frac{A_i}{A_i + B_i} + B_i \log \frac{B_i}{A_i + B_i} \right) = \text{Deviance}$$

Figure 2.4. An Example of TBR



$$D = -2\left(25\,\mathbf{log}\frac{25}{156} + 131\,\mathbf{log}\frac{131}{156}\right) = 85.52$$

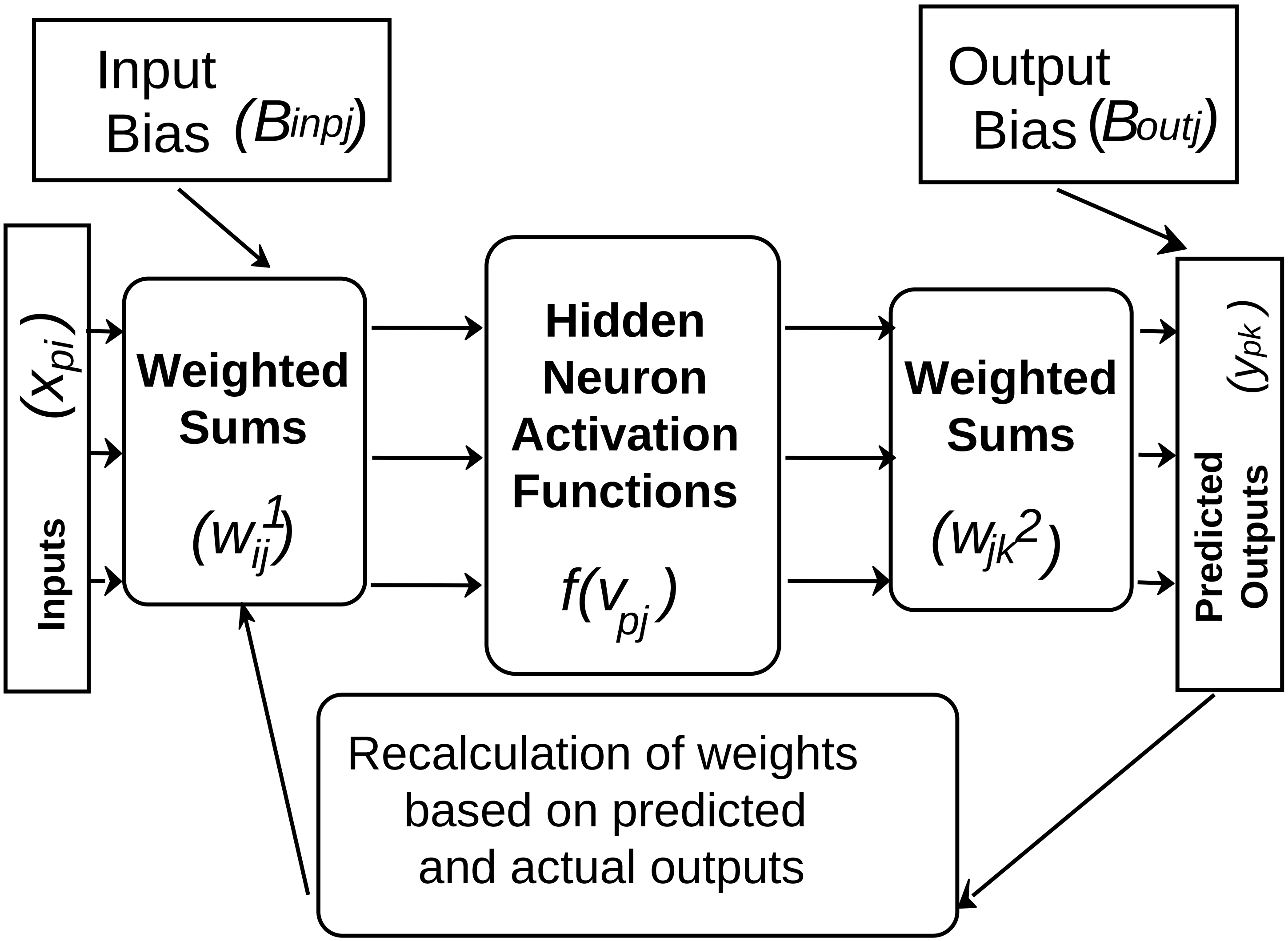
$$D = -2\left(5\,\mathbf{log}\frac{5}{131} + 126\,\mathbf{log}\frac{126}{131}\right) = 18.44$$

+

$$D = -2\left(5\,\mathbf{log}\frac{5}{25} + 20\,\mathbf{log}\frac{20}{25}\right) = 10.86$$

= 29.30

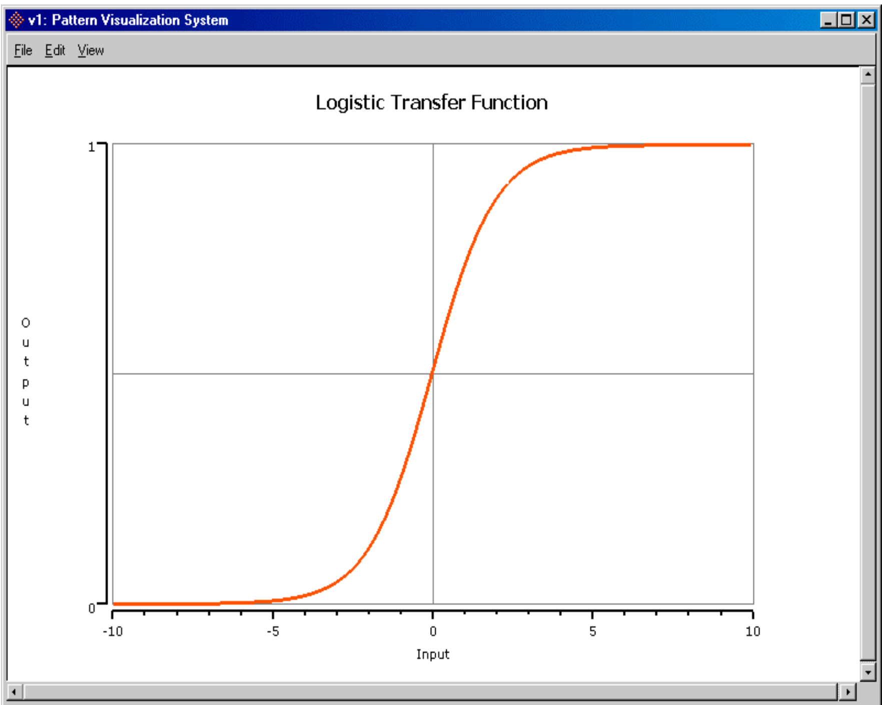
Figure 2.5. Flow of the ANN a Multilayer Perceptron (MLP) Model*



*** Source: Price et al, 2000**

Figure 2.6. Two Common Activation Functions

Sigmoid (Logistic, Squashing)
(turns on at about certain value)



Gaussian
(produces signal when input is near certain value)

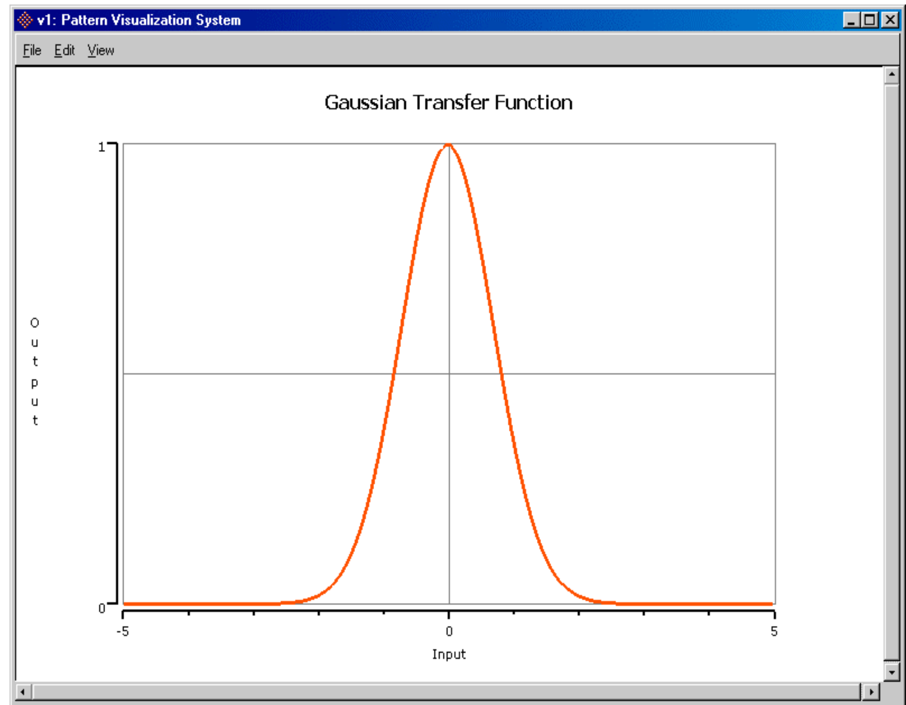


Figure 2.7 Example of ROC

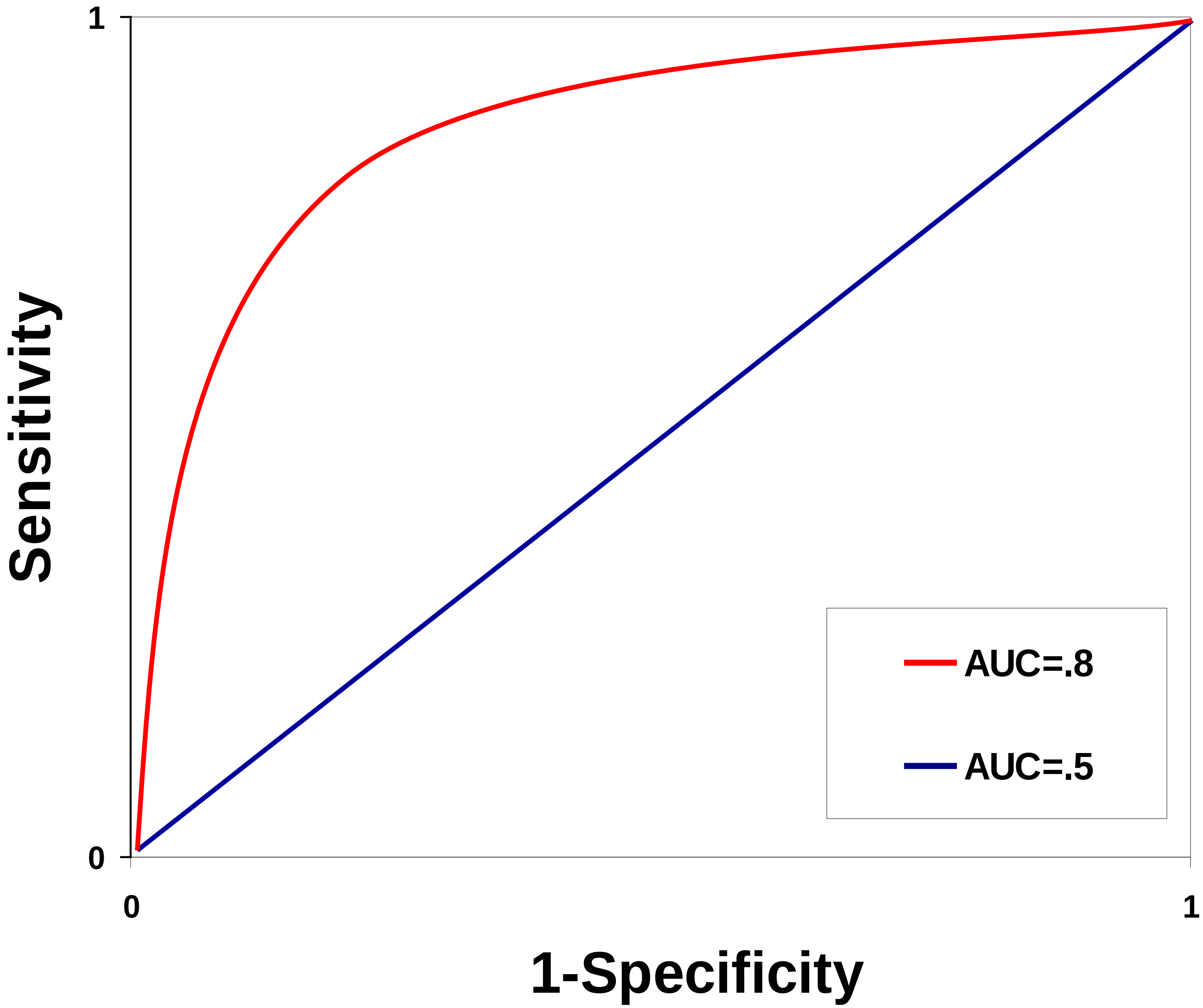


Figure 2.8. Analysis Steps Comparing Traditional Statistical vs. Data-Mining Techniques

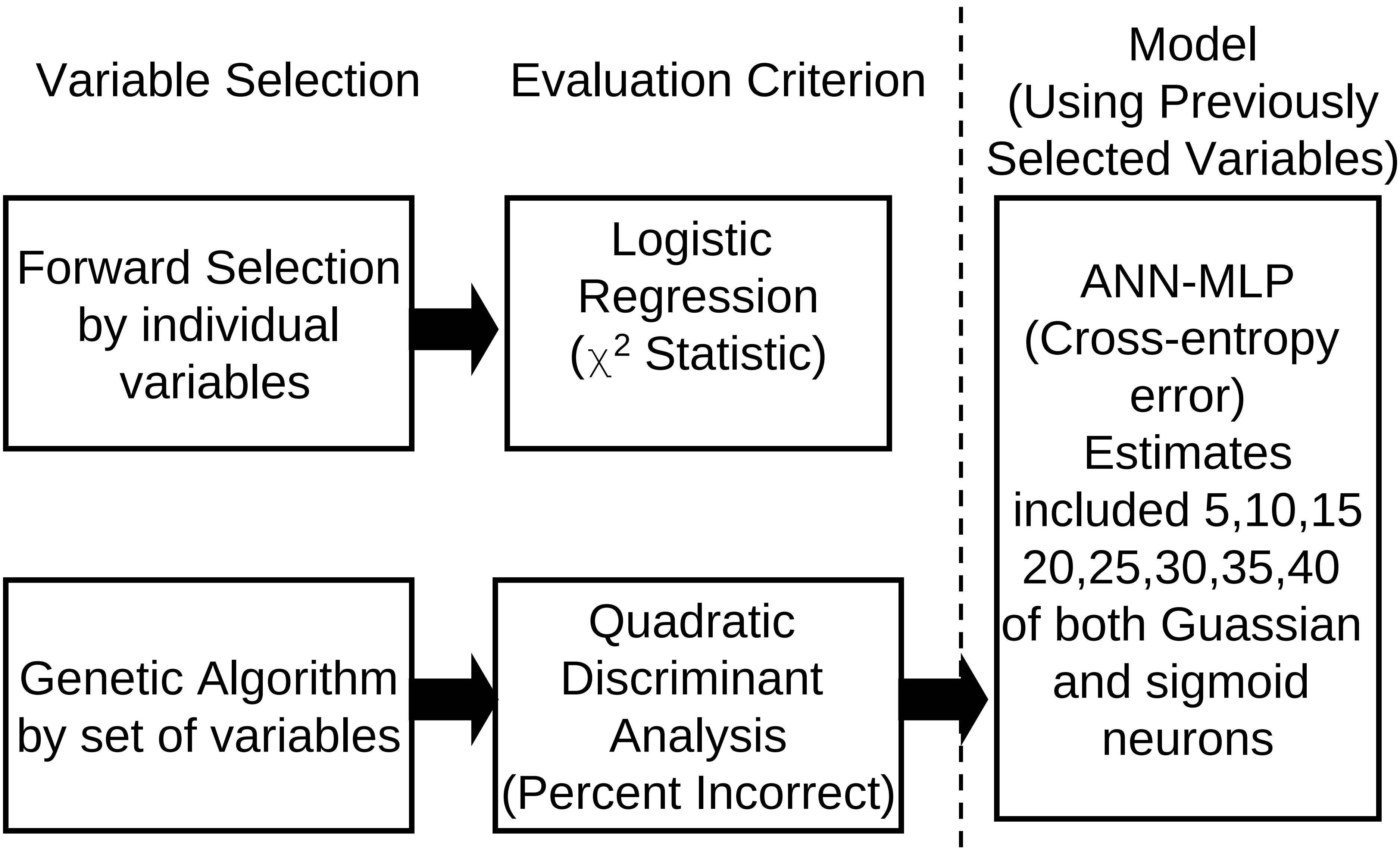


Figure 2.9 Cross-Validation Mimicking Prediction of “Future” Data

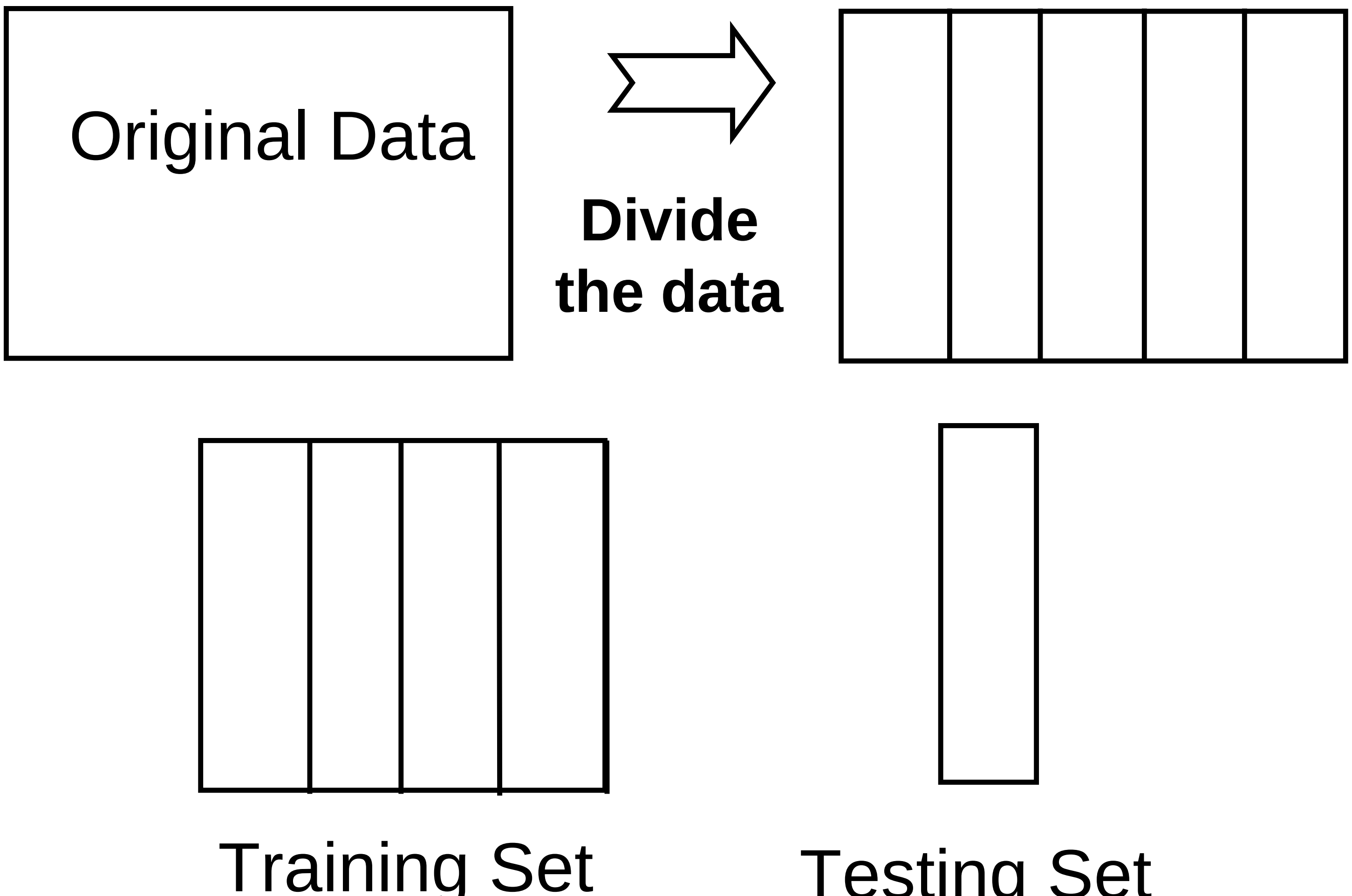
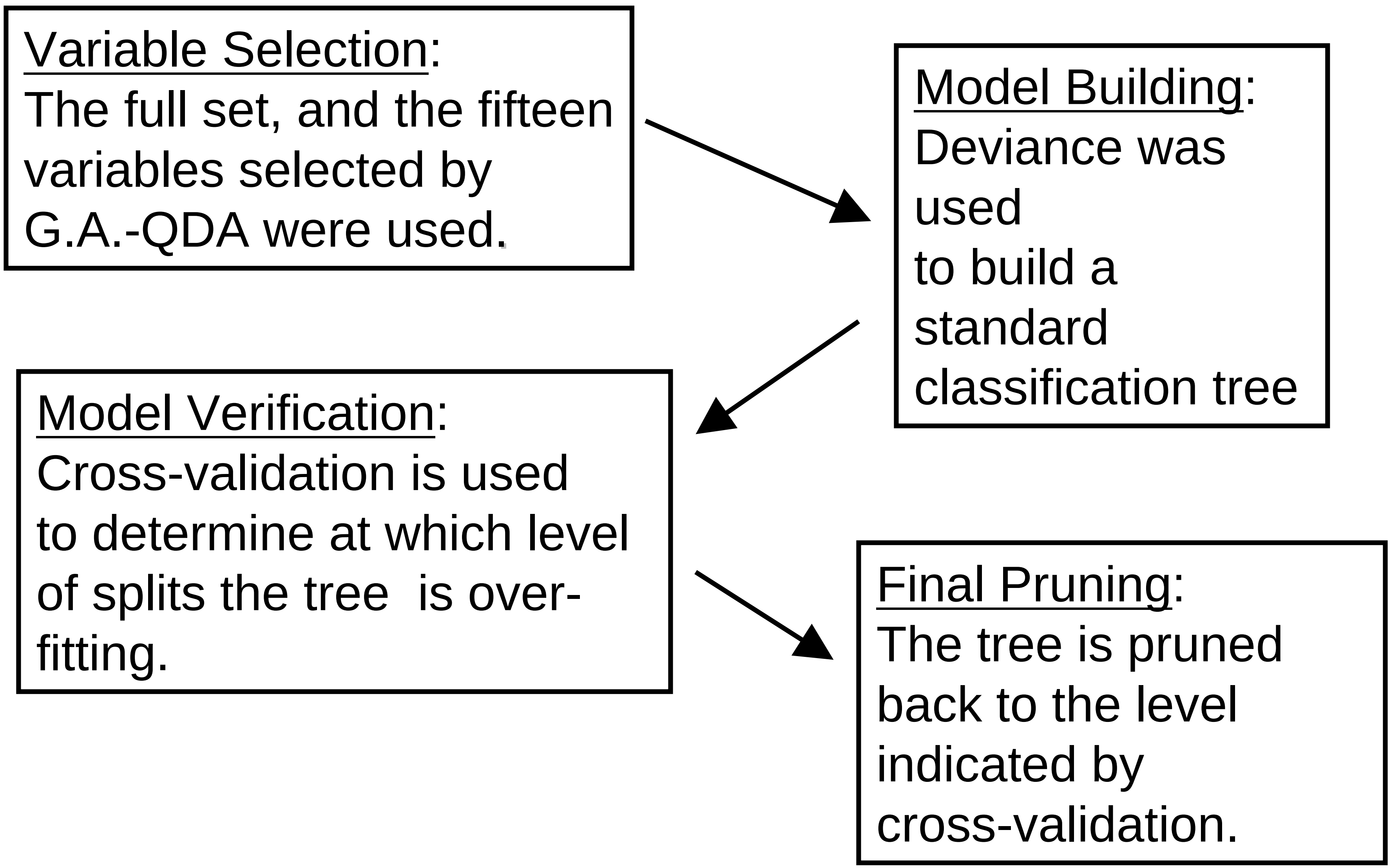


Figure 2.10. Analysis Scheme for TBR Using RPART



Chapter 3.

Analysis Results of the National Comorbidity Study, 1992, and the National Mortality Follow-Back Survey, 1993

After a series of preliminary analyses, we chose "past-year suicidal thought" as the main outcome measure for the NCS dataset to assess performance of the three data-mining techniques. Measures for suicide attempts were considered, but were judged less desirable because of a limited number of positives. For prediction of suicide, we chose accidental death as a comparison (referent) group, after a series of preliminary analyses, because natural deaths overwhelm suicide deaths by their sheer number, that is, predictors tended to be those which predict natural death, vis-a-vis suicide. Choice of accidental death was considered appropriate because their sudden death implied that the individuals who died by accident would have led normal lives up to the point of death. All data analyses were conducted for males and females separately, since findings on gender differences are considerable in the suicide literature. Gender specific analysis sample sizes are as follows: NCS males n=4,251, NCS females n=3,047; and NMFS males n=2,682, NMFS females n=1,074.

Predicting Past-Year Suicidal Ideation Using the NCS Dataset

The NCS dataset was explored first, as this dataset contained detailed timing information on suicidal behavior and environmental and vulnerability factors, including demographic, socio-economic status, life experience, mental health, substance abuse, exposure to violence, and access to care. Some of these measures were known to be predictive risk and protective factors of suicidal risk. From this pool of apparent-predictive measures, an "initial" dataset containing 57 variables were developed, which contained mostly dichotomous and psychiatric diagnostic variables according to the Diagnostic and Statistical Manual, Revision IV (DSM-IV). An "expanded" dataset containing 77 variables was developed after preliminary analyses indicated the need to refine measures. The latter dataset included many ordinal variables such as symptom counts. After GA variable selection analyses were completed, and the 15 most predictive measures selected, all 77 variables and the best 15 variables chosen by the

QDA with GA were separately input to RPART in S-Plus to run tree-regression (TBR) models. The 15 best predictors were also input to the Multi-Layer Perceptron (MLP) estimation in ANN to assess the maximum predictive power of the selected variables.

Best predictors of past-year suicidal thought:

The GA runs were set to select 15 best measures with which to maximize the goodness function, in this case, % incorrect. The selection was evaluated by QDA. The initial population was set as 200 and the mutation rate was set as .05. The GA continued to evolve over 100 generations, at which point it would output the results. When applied to both the initial and expanded datasets, the GA-chosen "best" predictors included many variables that were non-significant, if the traditional statistical significance were required. However, those measures chosen by both the forward selection and GA yielded $p<.01$ when GA measures were rerun with logistic regressions.

Using the expanded dataset, the GA tended to pick more ordinal variables compared to the forward selection implemented in logistic regressions. **Table 3.1.** and **Table 3.2.** show results for males and females, respectively. The "FS" columns show the measures that were significant at $p<=.05$; the GA columns show the 15 best measures chosen by GA according to the above parameter specifications; and the "P" columns show the probability level of logistic regressions applied to the GA-selected 15 "best" measures. The "D" indicates dichotomous measures, and "O" indicates ordinal measures.

The forward selection tended to select only a subset of the GA chosen 15 measures. Those chosen by both the forward selection and GA tended to be of highly significant measures when logistic was applied to the 15 measures without the variable selection option (e.g., depression, PTSD, supportive spouse, major expenses). This indicates when the measure was highly predictive and significant, both linear and GA selection yield similar results. GA tended to select more ordinal measures; presumably they provide finer cut-off options. Those measures that were not selected by the forward selection but selected by GA tended to yield higher probability level when logistic regression was applied. These would have been considered unimportant measures if a linear model was the only choice of variable selection method.

The best predictors for males and females were very similar, although not identical. As in results for males, depression and supportive spouse were chosen both by logistic forward selection and GA. However, some measures were chosen both by forward selection and GA only for males, including education level and antisocial personality symptom count. In addition, bipolar disorder, and alcohol abuse and dependence were chosen either by the forward selection or GA for females, while drug abuse and dependence, marijuana abuse, and impairment due to drug abuse were chosen by either selection method. These differences indicate varying psychiatric and substance abuse predictors of suicidality between males and females.

Interaction among predictive measures:

For NCS, the specification in RPART used the default minimum split size of 20 and the default minimum node size of 7. This means that no node under the size of 20 was considered for a split and no node of size less than seven was created. The specifications were considered adequate on the basis of sample size and the number of predictors. When all 77 variables in the expanded dataset were input to RPART which provided cross-validated trees, the estimation yielded fewer variables to reach the minimum deviance. This result indicated that TBR performs better when an initial effort is made to carefully select predictive measures.

Figure 3.1. and **Figure 3.2.** show TBR cross-validated results for males and females, repectively, when GA-QDA selected 15 variables were input. Major gender differences emerged. For males, fewer variables were sufficient in constructing the best generalizable tree: depression, impairment due to substance abuse, financial problems, and a loss of daily activities remained in the final model. For females, several more predictors intricately interacted with each other, including relational strengths with relatives. The final tree was more complex than that for males and more evenly spread. For both males and females, psychiatric syndromes, most notably depression, was the first and most important node, consistent with the suicide and suicidality literature. Other psychiatric syndromes (antisocial personality and bipolar) were more prominent predictors for females, whereas impairment by drug abuse was the second important measure for males.

Predictive power for past-year suicidal thought:

For this set of estimations, both sigmoid and Gaussian activation functions were used with the number of hidden neurons varying from 5 to 40 in an increment of 5 neurons. Two different learning rates that effect the size of change from one iteration to the next were also varied. For the male sample the best prediction was achieved with 40 Gaussian neurons. For females, the best prediction was achieved with 35 Gaussian neurons, using model performance evaluation of the Receiver Operating Characteristic (ROC) analysis. Model performance of the MLP was compared to those by QDA and logistic regression, where all analyses were validated using a 10-fold cross-validation.

For both the initial and expanded datasets, predictive power was modestly improved with the GA-QDA. But the MLP further improved the predictive power by a considerable amount. **Figure 3.3.** and **Figure 3.4.** show the results of MLP performance as compared to logistic regression and QDA for males and females, respectively. The value of the Area Under the Curve (AUC) is shown for each estimation. Overall, MLP was able to improve the prediction by 8 to 18 % in the Area Under Curve (AUC) value when the GA-chosen variables were input. In particular, AUC=.98 for males with 15 measures as input from the expanded dataset is an impressive prediction, not usually seen. Our method of past-year suicidal ideation prediction for males was able to select measures that contain sufficient information to reach near perfect prediction. In Figure 3.3., the differences of predictive power across three estimation methods are clearly shown. While MLP achieved AUC=.98, the comparison logistic runs reached AUC=.87, still very good but not as good as MLP's performance. MLP's performance for females (AUC=.93) (Figure 3.4.) was not as good as that for males, although the value indicates prediction was still in an excellent range. The other estimation methods performed equally well compared to their prediction level for males; resulting in smaller differences in prediction among the three methods for females.

Replicating the NCS Results Using the NMFS93 Dataset

With initial results indicating some limitations of the NMFS93 dataset, we attempted to focus on replication of the results produced from the NCS dataset. Parameter specification for each of the three data-mining techniques follow those used for the NCS dataset as much as possible.

<u>Predictors of suicide compared to accident death</u>:

The basic GA parameters used in NCS were also used for the NMFS data set as well, except for the number of variables selected. The forward selection for the logistic regression selected only 9 and 10 variables for the males and females, respectively, for the NCS dataset. However, it selected 20 and 22 variables for the NMFS dataset. Therefore, we decided to use a larger number of predictive variables for GA to provide a good contrast with the logistic regression. GA runs were performed to select 20 and 25 variables for the NMFS data set. Both runs produced similar AUCs (.79), so the 20 variables selected by GA were used for the TBR and ANN analysis to maintain parsimony.

Table 3.3. and **Table 3.4.** show the results of measures selected by forward selection and GA for males and females, respectively. The predictive measures of suicide when accidental death was a

referent group, included depression and depression-related variables. Not unexpectedly, medically-related variables that may lead to reduced judgement, such as dementia symptom count, were selected to reflect their predictive power for accidental death. For this set of analyses, GA-selected variables did not vary greatly from the variables found with the forward selection.

Interaction of predictive measures of suicide compared to accident death:

As with the NCS, we used the RPART default minimum split size of 20 and the default minimum node size of 7. **Figure 3.5**. and **Figure 3.6.** show the results of TBR models for males and females, respectively. Unlike the TBR results from the NCS dataset, the cross-validated tree for males was much "fuller," indicating that many important predictors of suicide were selected which differentiated suicide from accident death. As expected, predictors of suicidal ideation shown in the results from the NCS dataset were shown to be important. However, "being withdrawn" was the first split, not "depression," indicating the predictors of suicide may be considerably different from those of suicidal ideation. Unlike the predictors of suicidal behavior, "gun around" and "feeling of worthlessness" were shown to interact with other measures (Figure 3.5).

The cross-validated TBR result for females, on the other hand, was a lot simpler, in part reflecting a smaller sample size (n=1,074). Although "feelings of worthlessness" was shown to be an important factor, as in the male tree, a different set of predictors were chosen, such as seeing "psychiatric doctor," "not being medicaid," and being "older than 44."

Predictive power of best measures of suicide compared to accident death:

For NMFS, the number of neurons in the ANN models varied from 5 to 35 in an increment of 5, and both sigmoid and Gaussian neurons were used. The learning rate was varied following the specifications used for NCS analyses. A momentum term that effects the direction of change from one iteration to the next was also varied, because preliminary runs indicated that the ANN was not converging properly. For the male sample, the best prediction was achieved with 25 sigmoid neurons. For the female sample the best prediction was achieved with 15 sigmoid neurons The inclusion of a momentum term doubled the number of runs to be tested. Furthermore, each cycle takes considerable time to complete with 10-fold cross-validations. Therefore, the models using 40 hidden neurons were not attempted. As the predictive power of the ANN models dropped or held steady from 25 to 35 neurons, it was deemed unnecessary to include the 40 neuron runs.

For the male runs, the MLP improved the prediction over QDA by a good margin (12% by the

AUC value), and by a smaller margin (5%) over the logistic regression with forward selection. We found that most of the improvement from the logistic regression to MLP was due to MLP's estimation and not by GA variable selection. Overall, the results were not as spectacular as those derived from the NCS. The results of MLP for females indicated inadequate sample size to apply complex ANN architecture (data not shown). The model with 5 sigmoid neurons achieved a cross-validated ROC AUC of .791. Models with additional neurons did not improve the AUC much. The highest AUC of .809 was achieved with 15 sigmoid neurons. The improvement of MLP over QDA was less than 3% compared with the improvements of 8%, 12% and 18% showed earlier. While the uncross-validated models could achieve high ROC AUC values (e.g., .905 with 25 sigmoid neurons), increasing the number of neurons did not improve the cross-validated model fit and generally hurt the cross-validated model as more and more noise was fitted. It is also possible that the proxy measures by next of kin may contain a substantial amount of measurement error, which could have contributed to MLP fitting too much noise.

Summary

Through the application of data-mining techniques to the NCS dataset, we learned that GA is more suitable than linear forward selection when many additional variables are available. GA-selected variables would often be thrown out as uninformative if a linear selection method were to be used. Cross-validated trees were more informative when GA-selected variables were input compared to inputting all variables in the dataset. TBR results showed major gender differences. For males, depression was a very powerful and recurring predictor. But for females, more intricate interactions with a support environment, as well as other psychopathology, were shown. The TBR results clearly showed the patterns of interaction among strong predictors. For the NCS male dataset, the ROC evaluation showed a 22% increase in the predictive power compared to logistic. The improvement appears primarily to have resulted from better prediction of suicide ideation positives, which is more helpful (than predicting who do not become suicidal) for precise prediction of past-year suicidal thought. Such an improvement was achieved "gradually," involving both GA and ANN models and careful human evaluation of measures and results.

With respect to prediction of suicide compared to accident death, variable selection using the GA resulted in the choice of measures such as depression, and depression related variables ("taking antidepressants in the last year"). Some variables such as "high blood pressure" and "dementia symptom count” were selected, but they were related to prediction of accidents rather than suicide. GA selected variables did not vary greatly from the variables found with forward selection. The male tree was well balanced and showed a number of interesting interaction patterns (e.g., gun & worthlessness). Measures not available in the NCS dataset were shown to be predictive (e.g., gun, antidepressant). However, some measures appeared to be the predictors of accident death, rather than of suicide (e.g., blood pressure). The ANN-MLP estimation improved prediction over QDA by a good margin (12%). ANN-MLP improved prediction over logistic with forward selection by a smaller amount (5%). Most of the improvement from logistic regression to ANN was due to ANN-MLP itself and did not seem to depend on the method of variable selection. MLP’s less spectacular performance in the NMFS analysis may be due to a number of reasons. For example, not as many ordinal variables could be created, which affected the MLP’s ability to take advantage of learning non-linear associations. Proxy measures by next of kin are not as good as probands’ self-reports. Unfortunately, this is a fundamental limitation in the study of suicide. Limited by the nature of the NMFS dataset, the accident comparison group was chosen, but this choice may still be problematic.

Table 3.1

Variable Selection by Logistic with Forward Selection and GA-QDA, NCS-Males

Variables	FS	GA	P
Major expenses	D	D	<.01
Income problems		D	0.25
Education level		O	0.07
Spouse supportive	O	O	<.01
Closeness to relatives		O	0.09
Support from friends		O	0.75
Daily activities stopped			
Daily activities stopped by drug use	D		
Daily act. stopped/reduced		O	0.36
Daily act. stopped/reduced by drugs		O	0.09
Drug abuse, past year	D		
Drug dependence, past year	D		
Marijuana abuse lifetime symptom cnt.		O	0.07
Depression, past year	D	D	<.01
Depression, lifetime symptom count	O	O	<.01
PTSD, past year	D	D	<.01
Has insurance		D	0.20

Comfortable with treatment		D	0.68
Belief in efficacy of treatment		D	0.31
Sexual molestation, lifetime	D		

<u>Note.</u>

FS: Forward selection with logistic regression.

GS: Genetic algorithms with quadratic discriminant analysis.

D: Dichotomous variable.

O: Ordinal variable

P: P-value of the variable using a logistic regression for the variables chosen by GA with QDA

Table 3.2

Variable Selection by Logistic with Forward Selection and GA-QDA, NCS-Females

Variables	FS	GA	P
Income		O	0.52
Income problems		D	0.08
Education level	O	O	0.01
Spouse supportive	O	O	<.01
Closeness to relatives		O	0.17
Relatives supportive		O	0.17
Daily activities reduced		D	0.41
Daily act. stopped/reduced by drugs	O		
Adult anti-social symptom count	O	O	<.01
Alcohol abuse, past year	D		
Alcohol dependence, past year		D	0.63
Childhood anti-social, symptom count		O	0.25
Depression, past year	D	D	<.01
Depression, lifetime symptom count	O	O	<.01
PTSD, past year	D		
Non-affective psychosis, lifetime	D		
Bi-polar, past year		D	0.03
Desire to seek help		D	0.57

Belief in efficacy of treatment		D	0.28
Rape, past year	D		

<u>Note.</u>

FS: Forward selection with logistic regression.

GS: Genetic algorithms with quadratic discriminant analysis.

D: Dichotomous variable.

O: Ordinal variable

P: P-value of the variable using a logistic regression for the variables chosen by GA with QDA

Table 3.3

Variable Selection by Logistic with Forward Selection and GA-QDA, NMFS-Males

Variables	FS	GA	P-value
Married		D	.02
Lived alone	D		
Religious	D	D	<.01
No care received in the past year	D		
No method for paying for care available		D	.23
Method for payment of care utilized		D	.99
Difficulty doing chores	D	D	<.01
On workers comp in the past year		D	.17
Total number of health problems	O		
Heart problems		D	.45
Heart attack		D	.03
High blood pressure	D	D	<.01
Has had stroke[1]			
Dementia symptom count		O	.35
Trouble remembering in the past year	D	D	.03
Total number of depressive symptoms	O	O	.08
Withdrawn in the last month	D	D	

Trouble sleeping in the last month	D	D	<.01
Feelings of worthlessness in the last month	D	D	<.01
Drowsy in the last month	D		
Worried in the last month	D		
Destroyed property in the last year	D	D	<.01
Saw a psychiatrist in the last year	D		
Took antidepressants in the last year	D	D	<.01
Took cocaine in the last year	D	D	.18
Took stimulants in the last year	D		
Took heroin in the last year	D	D	<.01
Took tranquilizers in the last year	D		
Alcohol abuse/dependence symptom count	O		
Guns around in the last year	D	D	<.01

<u>Note.</u>

1. "Stroke" status was selected by forward selection, but later dropped due to co-linearity with other variables

FS: Forward selection with logistic regression.

GA: Genetic algorithm with quadratic discriminant analysis.

D: Dichotomous variable.

O: Ordinal variable.

P-value: P-value of the variable using a logistic regression for the variables chosen by GA with QDA.

Table 3.4

Variable Selection by Logistic with Forward Selection and GA-QDA, NMFS-Females

Variables	FS	GA	P-value
Age		O	.35
Education level		O	.63
High school education		D	.18
Lived alone	D	D	<.01
Foreign born	D		
Religious	D	D	<.01
Employed in past year	D	D	<.01
Promoted in past year	D		
Spouse ill in past year	D		
Number of friends/family with illness in past year		O	.56
Medicaid paid for care	D	D	<.01
Medicare paid for care		D	.54
Has health insurance	D		
Total number of health problems	O	O	.09
Asthma in the past year	D		
Lung condition in the past year	D		
Heart attack		D	.71

Chest pain	D	D	.13
Dementia symptom count	O		
Trouble remembering in the past year		D	<.01
Trouble recognizing people in the past year	D		
Withdrawn in last month	D	D	<.01
Feelings of worthlessness in last month	D	D	<.01
Worried in last month	D		
Caused complaints among family/friends in past year	D		
Saw psychiatrist in past year	D	D	<.01
Took antidepressants in the past year	D	D	<.01
Took blood pressure medicine in past year		D	.09
Took heroin in last year	D		
Took cocaine in past year		D	.36
Guns around in last year	D		
Amount of firearms available in last year		O	.02

Note.

FS: Forward selection with logistic regression.

GA: Genetic algorithm with quadratic discriminant analysis.

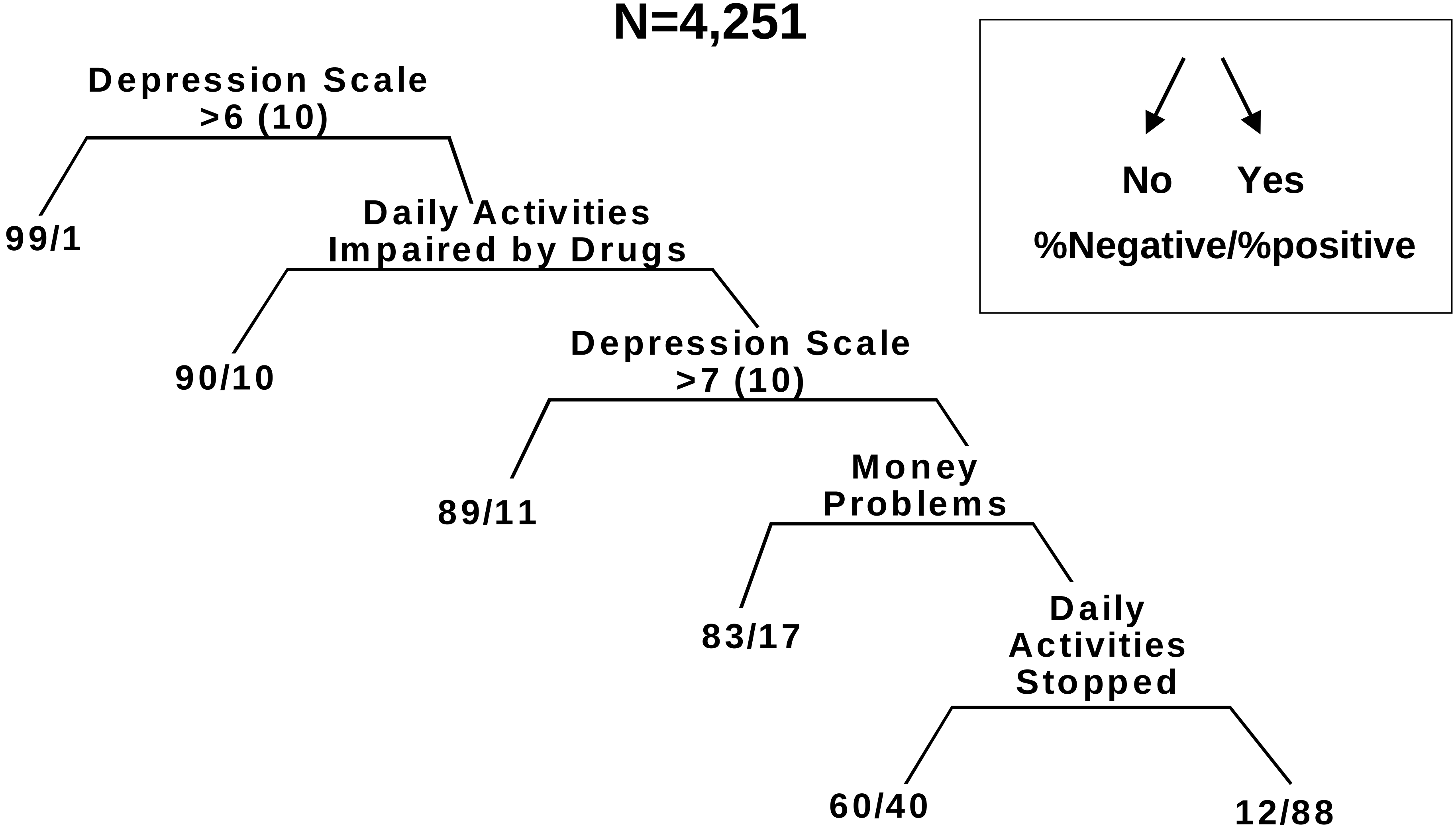
D: Dichotomous variable.

O: Ordinal variable.

P-value: P-value of the variable using a logistic regression for the variables chosen by GA with QDA.

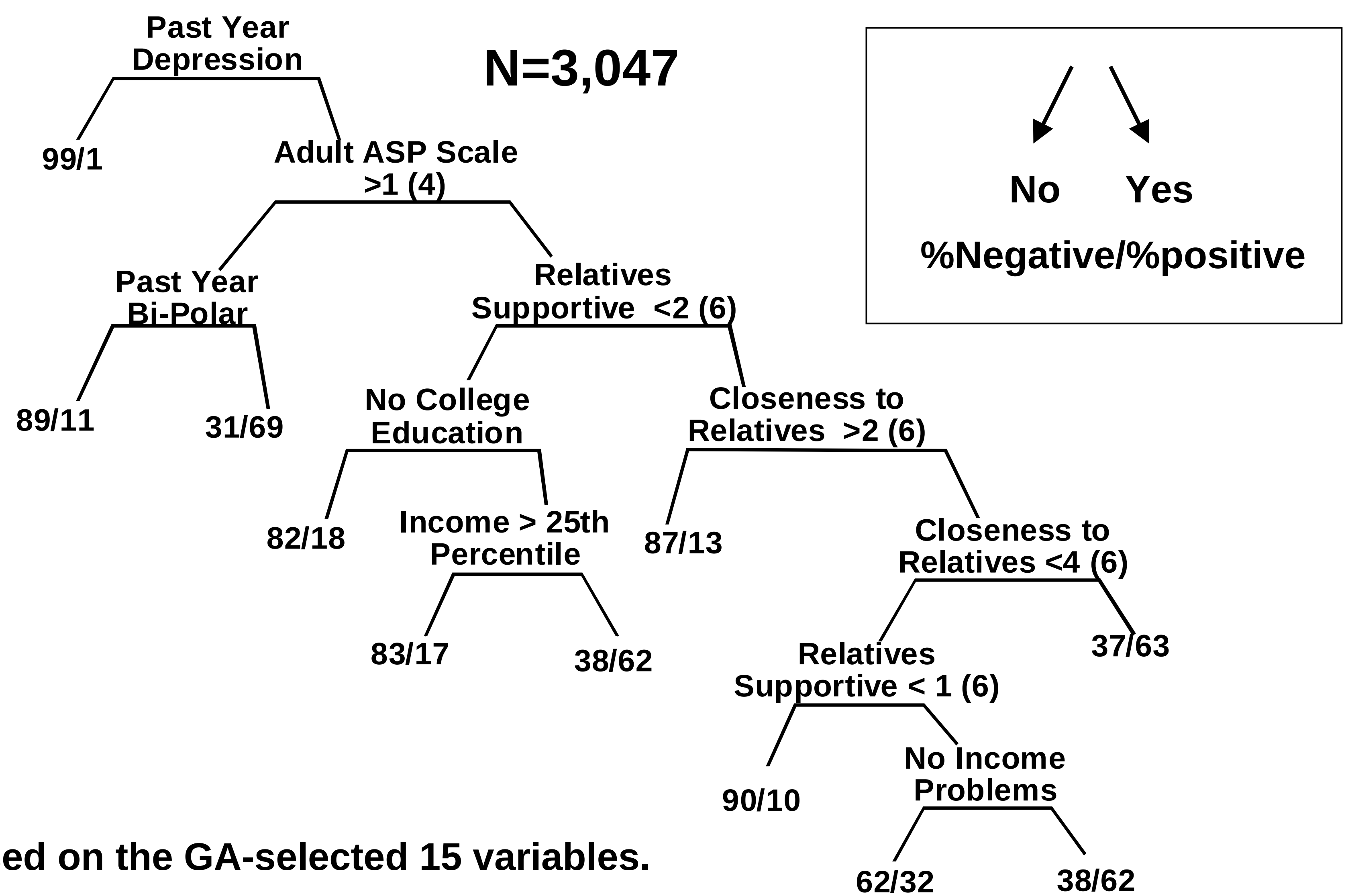
Figure 3.1. TBR Predicting Past-Year Suicidal Thought. NCS Males*

N=4,251



* Based on the GA-selected 15 variables.

Figure 3.2. TBR Predicting Past-Year Suicidal Thought. NCS Females*



* Based on the GA-selected 15 variables.

Figure 3.3. Predictive Power for Past Year Suicidal Thought Evaluated by ROC: NCS-Males

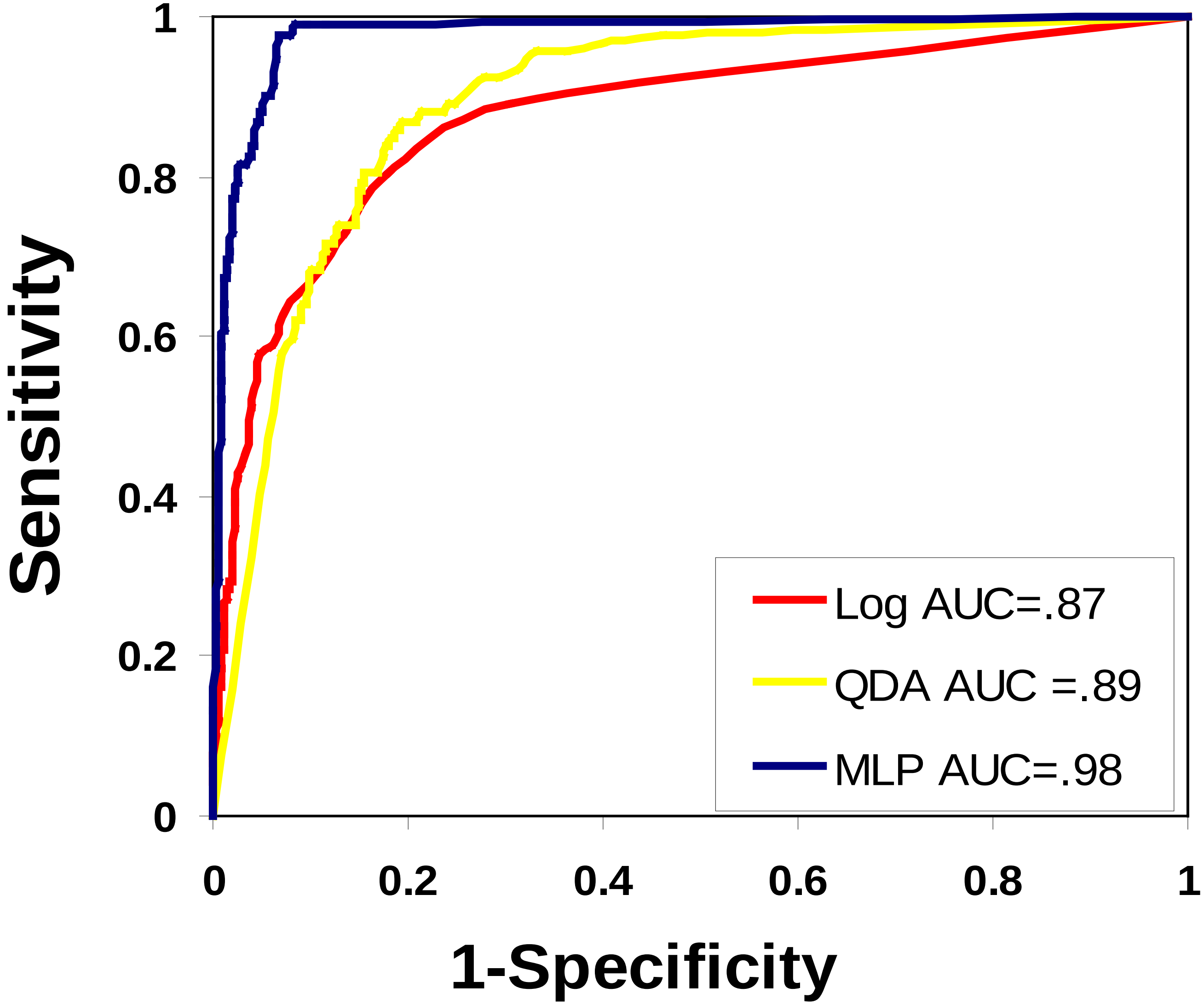


Figure 3.4.Predictive Power for Past Year Suicidal Thought Evaluated by ROC: NCS-Females

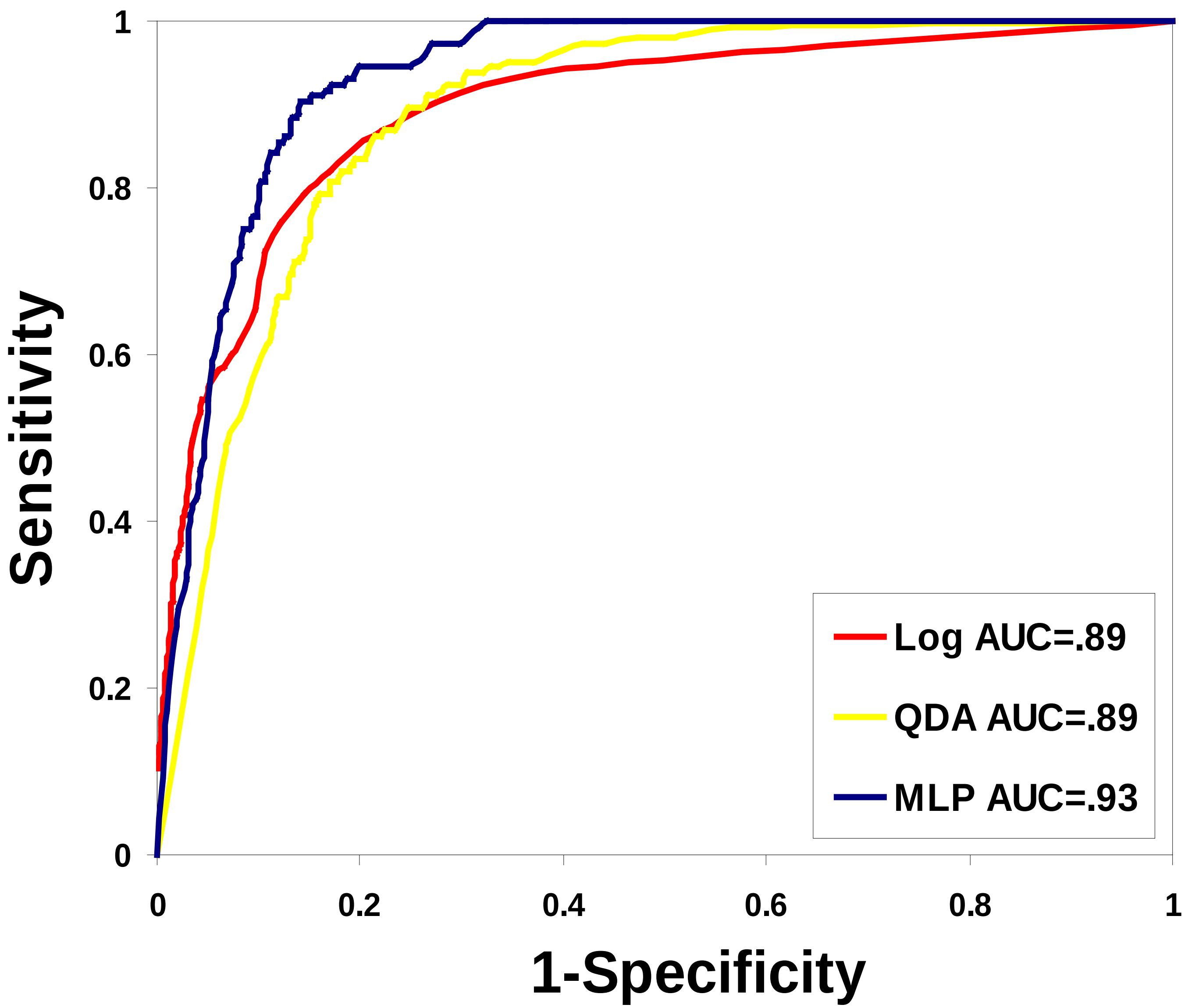


Figure 3.5 TBR Predicting Past-Year Suicide vs. Accident. NMFS Males

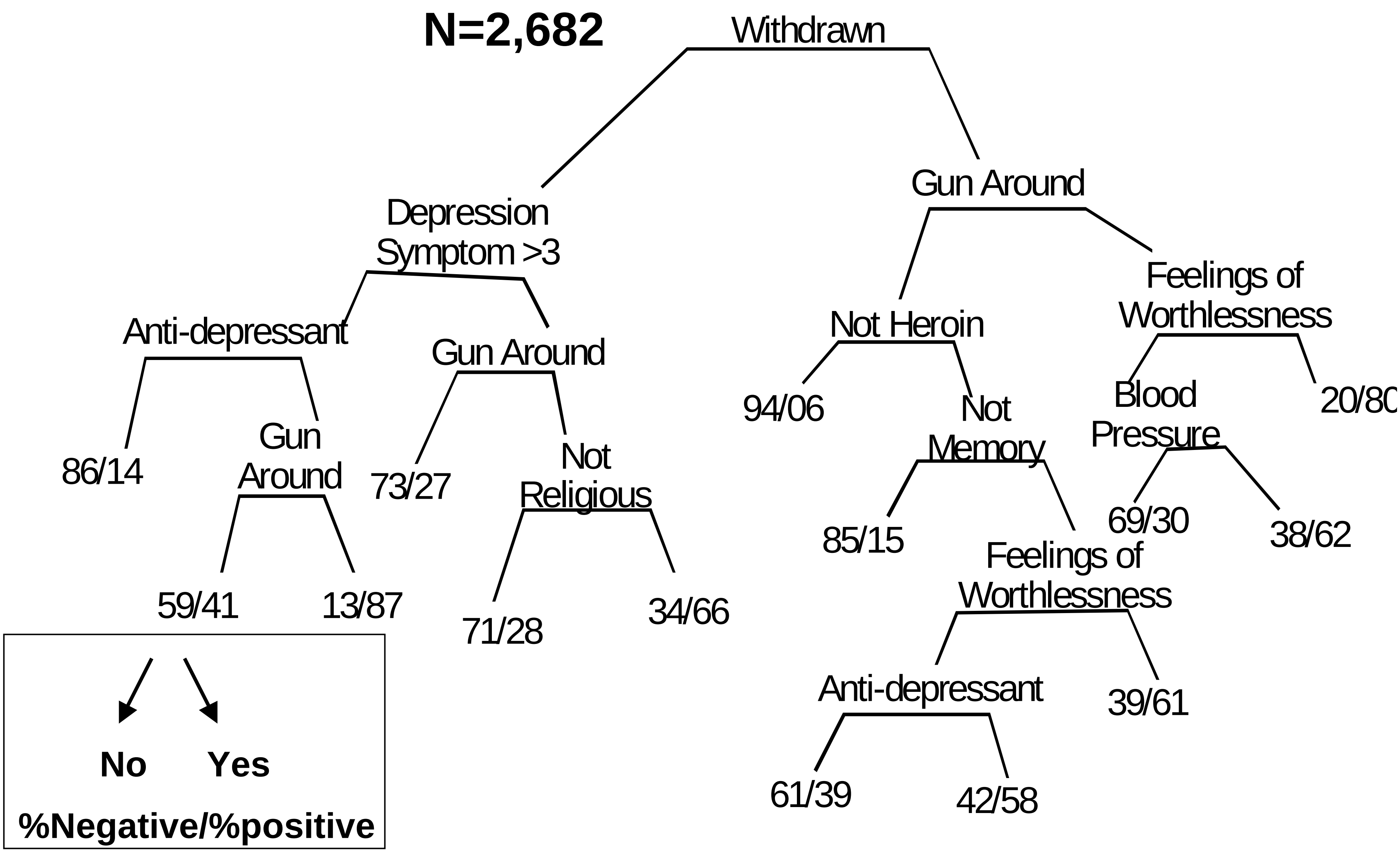


Figure 3.6 TBR Predicting Past-Year Suicide vs. Accident. NMFS Females

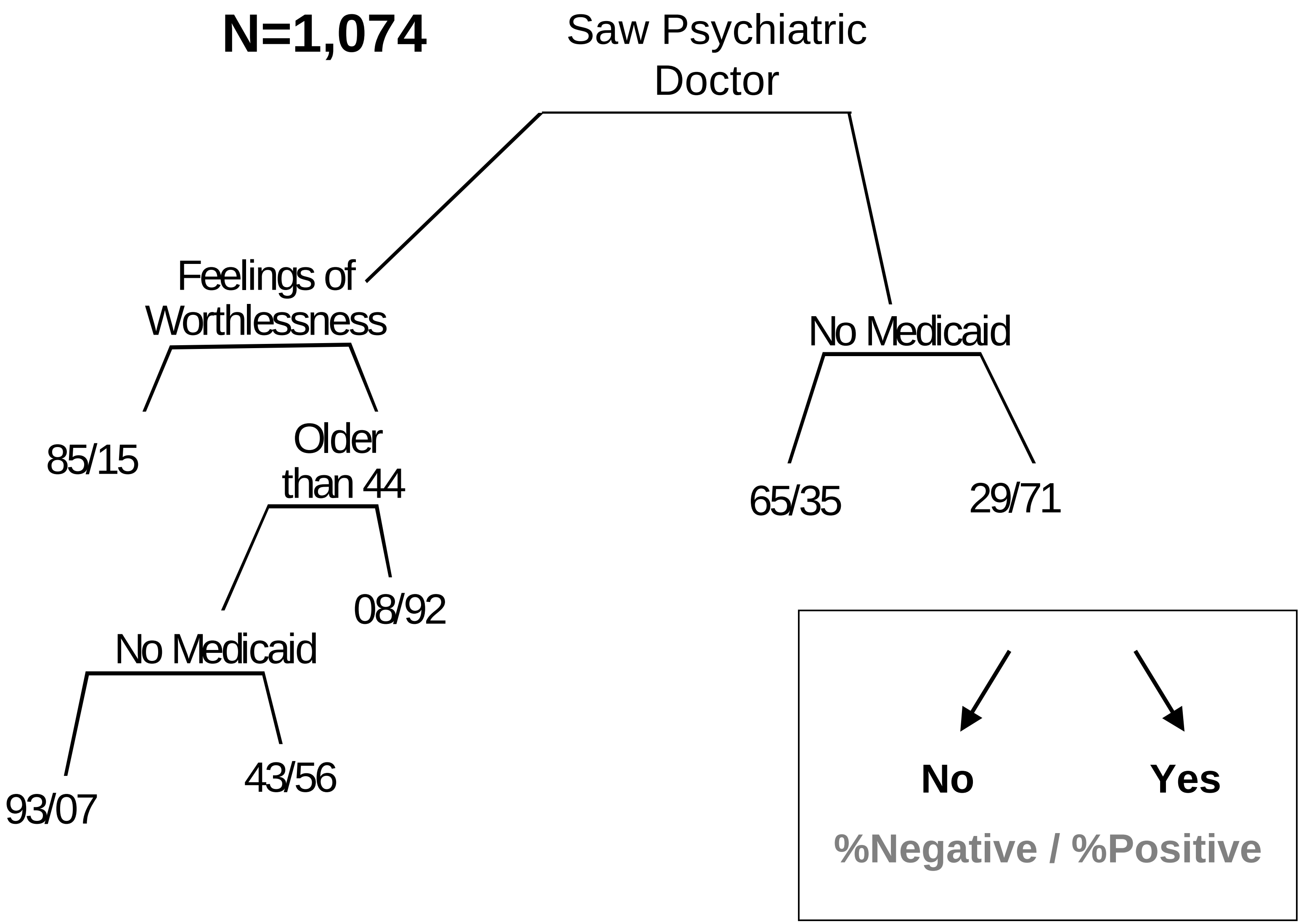
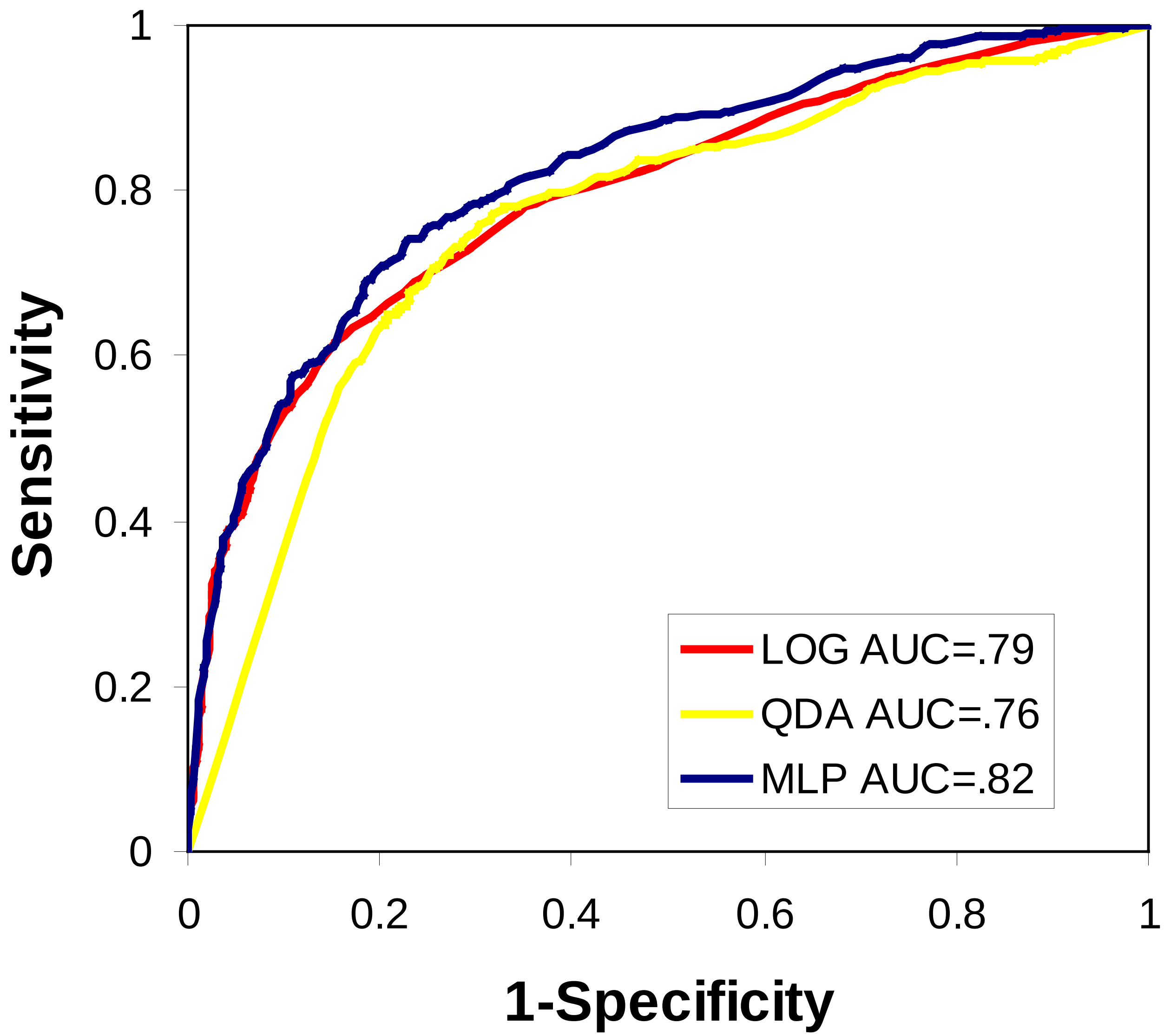


Figure 3.7.Predictive Power for Past Year Suicide vs. Accident Evaluated by ROC: NMFS-Males



Chapter 4.
Future Directions and Summary

In this chapter, we describe our effort to understand the learning process of data-mining techniques, in particular ANN learning process. The work here is preliminary at this point. We conclude the report by providing an overall summary that is relevant for the mission of the Longer Life Foundation.

Opening the ANN Black Box

The center piece of the analytic plan of this study is modeling ANN architectures. Among ANN practitioners, at least two types of information have been analyzed: **a)** weights, and **b)** hidden neuron activation vectors. Both can be seen as traces of ANN's best learning during its training phase. It is important, therefore, that weights and neuron vectors to be examined be derived from a model which uses only predictive input measures.

<u>Examining ANN weights</u>:

Unlike regression coefficients of a parametric approach, weights on MLP paths do not have intuitive meanings except in simple models under certain assumptions. In a very simple ANN model, weights can be examined visually. In a very restricted single-layer model, weights become identical or close to logistic unweighted beta parameters, thus weights are perfectly interpretable (Duh et al., 1998). If a model is more complex, but still relatively simple, the total contribution of an input measure can be computed by tracing the paths, analogous to the path analytic method. Genetic applications use a summation of the product of the weights along each pathway from an input to the output measure, or its variant (Lucek and Ott, 1997; Saccone et al., 1999). These methods are useful when the number of paths are relatively small and, therefore, the meaning from one stage to another is interpretable. Learning about the contributions of weights provides an understanding of which input measures have large impacts on which neurons or collection of neurons and which layers, and overall contributions to the outcome.

In **Table 4.1.** and **Table 4.2.**, the ranking of products of the weights (w_{ji}^1 x w_{jk}^2) and the ranking of the absolute values of the weights ($|w_{ji}^1$ x $w_{jk}^2|$) (see Figure 2.5. for specific paths where weights are placed) are shown in the “P” column and “PA” columns, respectively. Weights here were obtained separately for male and female MLP runs using the NCS datasets, and were derived from the learning phase of MLP models (best test models), prior to the start of cross-validation (cross-validated results of prediction which are subsequently plotted in ROC are shown in Figure 3.3. and Figure 3.4.) Not surprisingly, the lifetime depression measure was loaded with the largest weights if absolute values are taken (bottom of each table), assuring that weight ranking by absolute values is a reasonable method to examine weights for relatively simple MLP models. In these figures, ranking of measures are more similar between males and females than would have been expected from results of TBR results (Figure 3.1. and Figure 3.2.). For example in Table 4.1., relational measures (e.g., support from friends, closeness to relatives) are shown to have loaded with large absolute values of weights, which are not expected from the TBR results for males. However, results of females are largely consistent with the TBR results for females. These rankings are also inconsistent with the importance of measures expected from logistic regression analyses run parallel to the GA variable selection runs (See Table 3.1. and Table 3.2.). For example, for male results, support from friends, closeness of relatives and marijuana abuse/dependence were chosen only by GA, but ANN weights show they were more important than other measures. Likewise for female results, childhood antisocial symptoms and income are among those that were selected only by GA, which are loaded with large weights in the ANN best learning model.

Examining hidden neurons:

Examining hidden neurons themselves may provide a better understanding of how input information is learned by ANN, especially for a complex model with a large number of hidden neurons. In discussing the insights of the connectionist modeling for the development of learning words, McClelland (McClelland, 1991; McClelland, 1988; McClelland, 1981) envisioned that visual recognition of a “word” takes three steps: **a)** raw inputs are processed to elements of features; **b)** features are process to letters, “w,” “o,” “r,” and “d”; and **c)** individual letters are processed to form “word” which has a meaning. Though not numerous, ANNs with multiple hidden layers have been applied to empirical data with explicit conceptual reasons as in the fuzzy neural networks (e.g., “and” and “or” constitute different hidden layers) (D’Alche-Buc et al., 1994; Fu and Shann, 1994), or empirically derived (Carbone and Piras, 1998). The complexity within a hidden layer depends in part on the complexity of data. In general, it is believed that a higher nonlinear function can be approximated with fewer numbers of hidden neurons when multiple-hidden layers are used rather than when a single-hidden layer is used (Sarle,1994; Bishop, 1995). We had envisioned that pathways to different expressions of suicidal behavior or types of suicide would be shown in the complexity of multiple-hidden layers, where pathways would be observable in weights passed on to different neurons from one layer to another. In practice, we reverted to single hidden-layer models after initial trials of multi-hidden layer models using simpler datasets did not show obvious advantage for modeling multi-hidden layer models.

Regardless of the complexity of ANN architecture, the second approach of examining hidden neurons has taken an approach of data reduction, by factor analysis for example, but the summarization is done at a more abstract level. To the extent that ANN correctly learned the structure, the interpretation of reduced data should be clearer. Hidden neuron vectors have been analyzed by various standard methods such as principal component analysis, canonical discriminant analysis (Wiles and Bloesch, 1992), cluster analysis (Elman, 1990), and multidimensional scaling (Cottrell and Metcalfe, 1991).

However, this approach requires considerable experience in organizing the ANN architecture, before data reduction can be attempted. In Elman's (1990) experiment, the author took 29 words and randomly generated 10,000 sentences following certain pre-entered rules of the sentence structure. He did this by creating a 31 digit binary code that was different for each word. Then he concatenated the random sentences together with no delimiters or punctuation, yielding 27,534 words. The next step was to run these words (observations) through an ANN with a hidden layer of 150 neurons, which then fed forward to "context units" to loop back to the hidden layer (see **Figure 4.1.**). The ANN then was given the "task" to predict the next word. In the learning phase, the sentences were run through six times with the ANN updating after the addition of each word. At the 7th run, the connections of the ANN network was "frozen." The information on the number of times the word activated a given neuron was extracted from neurons in the hidden layer, yielding a construction of a 29X150 matrix. The matrix consisted of words vs. average number of times the neuron was activated by the word. Hierarchical cluster analysis was applied as a data reduction method, which recovered the original grouping on the words that drove sentence structure.

The neurons "learned" the pattern of clusters even though the individual network has about an 80% error rate, due to the fact that the neural net can learn the "grammar" of the sentence structure but not the random words that would fit into the "grammar." The cluster analysis results of words are shown in **Figure 4.2**. In the final stage, a fake word "zog" was substituted for the originally used noun "man" to see if the ANN learned the structure of the words so that it will place "zog" to where it should be, even though the word is not "man."

Neurons are what connect the input variables to the output. Furthermore, examination of weights can become too complex as the complexity of the ANN architecture increases. For a complex multi-hidden layer model, it is not clear if products of absolute values of weights yield straight-forward interpretation. For these reasons, examining the ANN's neurons directly has a certain intuitive appeal. For our purposes, we may be able to use the predictors and an output to create the neural network with the observations being people. When every observation goes through the network, we would discard the results to see what neuron fired for each observation, then averaging out over observations with a certain value of a variable in the input. A cluster analysis can be derived from a matrix of the number of predictors and the number of neurons. This is a direction of our future research.

Findings Summary

<u>Best predictors of suicidal ideation</u>:

Given more detailed measures, the GA tended to pick more ordinal variables compared to the forward selection implemented in logistic regressions. Those chosen by both the forward selection and GA tended to be of highly significant measures when logistic was applied to the 15 measures without the variable selection option. This indicates that when the measure was highly predictive and significant, both linear and GA selection yield similar results. Those measures that were not selected by the forward selection, but selected by GA, tended to yield higher probability levels when logistic regression was applied. These would have been considered unimportant measures if a linear model was the only choice of variable selection method. The best predictors for males and females were very similar, although not identical. Some measures were chosen both by forward selection and GA only for males, including education level and antisocial personality symptom count. In addition, bipolar disorder, and alcohol abuse and dependence were chosen either by the forward selection or GA for females, while drug abuse and dependence, marijuana abuse and impairment due to drug abuse were chosen by either selection method. These differences indicate different psychiatric and substance abuse predictors of suicidality between males and females.

Interaction among predictive measures of suicidal ideation:

When predictive measures were not sorted out prior to TBR input, cross-validated estimation yielded fewer variables to reach the best model. This result indicated that TBR performs better when an initial effort is made to carefully select predictive measures. When the fewer predictors selected by GA were input to TBR, major gender differences emerged. For males, fewer variables were sufficient in constructing the best generalizable tree; for females, several more predictors including relational measures intricately interacted with each other. The final tree is more complex than that for males and more evenly spread. For both males and females, psychiatric syndromes, most notably depression, was the first and most important node, consistent with the suicide and suicidality literature. Other psychiatric syndromes are more prominent predictors for females, while drug abuse appears to be a good discriminator of suicidal males.

Predictive power for past-year suicidal thought:

Predictive power was modestly improved with the variable selection, but the ANN further improved the predictive power by a considerable amount. Overall, ANN models were able to improve the prediction by 8 to 18 % in a measure of prediction that takes into account both sensitivity and specificity. For males, the ANN achieved the level of prediction that can be considered near perfect. ANN's performance for females was not as good as that for males, although the value indicates prediction was still in an excellent range. The other estimation methods performed equally well compared to their prediction level for males; resulting in smaller differences in prediction between ANN and traditional biostatistical methods.

Replicating results predicting suicidal ideation for suicide:

The predictive measures of suicide when accidental death was a referent group included depression and depression-related variables. Not unexpectedly, medically-related variables that may lead to reduced judgement, such as dementia symptom count, were selected to reflect their predictive power for accidental death. Unlike the TBR results from the NCS dataset, the cross-validated tree for males is much "fuller" indicating many important predictors of suicide were selected, that differentiate suicide from accident death. As expected, predictors of suicidal ideation shown in the results from the NCS dataset were shown to be important. However, several predictors of suicide were different from those of suicidal ideation. The cross-validated TBR result for females was a lot simpler which reflected a smaller sample size for deceased females. The results indicated a different set of predictors for females as compared to males. Most of the improvement achieved by ANN's did not result from the improvement obtained by GA variable selection. Overall, the results were not as spectacular as those derived from the NCS. The results of MLP for females indicated inadequate sample size to apply complex ANN architecture.

Examining ANN weights:

Learning about the contributions of weights provides an understanding of which input measures have large impacts on which neurons or collection of neurons and which layers, and overall contributions to the outcome. When weights were computed from the ANN's best-learning models for males and females using the NCS dataset, the lifetime depression measure was loaded with the largest weights, consistent with the suicide literature. The ranking of measures by weight values are more similar between males and females than would have been expected from TBR results. The rankings are also inconsistent with the importance of measures expected from logistic regression analyses run parallel to the GA variable selection runs. These inconsistencies will need to be resolved in the future. Potential sources of the inconsistencies are mis-specification of TBR models, over-fitting by ANNs, which potentially resulted from small sample sizes available for some of the analyses conducted with use of the three data-mining techniques.

Significance to the Mission of the Foundation

The current study is highly significant to the Longer Life Foundation's mission, because our effort in suicide and suicidality research fits to the Foundation's mission to study factors that assist in predicting mortality and morbidity of selected populations and to research methods to promote improvements in longevity and health by analyzing the effects of changes in medicine and advances in public health practice. Suicide is the eighth leading cause of death in Americans; everyday 86 Americans take their own lives and over 1,500 attempt suicide. There are now twice as many deaths due to suicide than due to HIV/AIDS. Yet, given the knowledge and tools that have become available recently, suicide can be approached as a public health problem that is preventable.

Data-mining techniques are not a magic bullet. Successful implementation of analyses using data-mining techniques depends heavily on the knowledge of the measures, knowing how to take advantage of the strengths of these techniques. The researcher's ability to organize the data analysis progression efficiently is critical because affording extensive CPU time is a major factor for obtaining better results. Researchers must work hard initially to select predictive measures, and create variables in such a way as to be informative to particular techniques. Once this work is accomplished, the three methods together are likely to help improve predictive models of suicidal behavior and help us to better understand the patterns of interaction among predictive measures. Although our variable selection method still did not provide perfect prediction even with the use of flexible data-mining techniques, we showed much improved prediction when sufficiently detailed predictive measures were available. With the arrival of a much faster processor, it may be possible to implement these data-mining techniques in a clinic and provide a rapid and up-to-date assessment of the patient’s suicidal risk as new information is added in the patient chart.

Table 4.1

Ranked weights of predictors input to ANN, NCS - Males

	PA	P
Activities stopped or impaired	62.1	19.0
Comfortable with treatment	62.7	12.1
Major expenses	63.9	3.9
Has insurance	68.6	2.4
PTSD, past year	69.3	-3.2
Depression, past year	69.5	-14.5
Spouse supportive	70.6	-24.5
Act. stopped or imp. by drugs	76.6	-15.0
Income problems	79.3	0.9
Education level	87.8	-18.1
Marijuana abuse/dep lifetime	91.8	10.4
Belief in efficacy of treatment	97.0	1.7
Support from friends	103.1	18.4
Closeness to relatives	110.5	-18.4
Depression, lifetime symptoms	110.5	-51.7

PA: Product of the absolute value of the weights on the “Yes” neuron.

P: Product of the weights on the “Yes” neuron.

Table 4.2

Ranked weights of predictors in put to ANN, NCS - Females

	PA	P
Desire to seek help	55.9	-4.3
Bi-polar, past year	57.7	2.7
Alcohol dependence, past year	60.5	-21.8
Depression, past year	62.7	24.8
Income problems	66.5	4.3
Daily activities reduced	69.0	-8.9
Belief in efficacy of treatment	76.2	29.8
Closeness to relatives	79.1	-4.1
Education level	79.2	0.2
Income	81.7	-7.0
Relatives supportive	82.9	-33.1
Spouse supportive	88.3	-3.5
Adult ASP, symptom count	96.2	-47.2
Child. ASP, symptom count	97.8	21.1
Depression, lifetime symptoms	116.3	9.9

PA: Product of the absolute value of the weights on the “Yes” neuron.

P: Product of the weights on the “Yes” neuron.

Figure 4.1. Elman's ANN Architecture

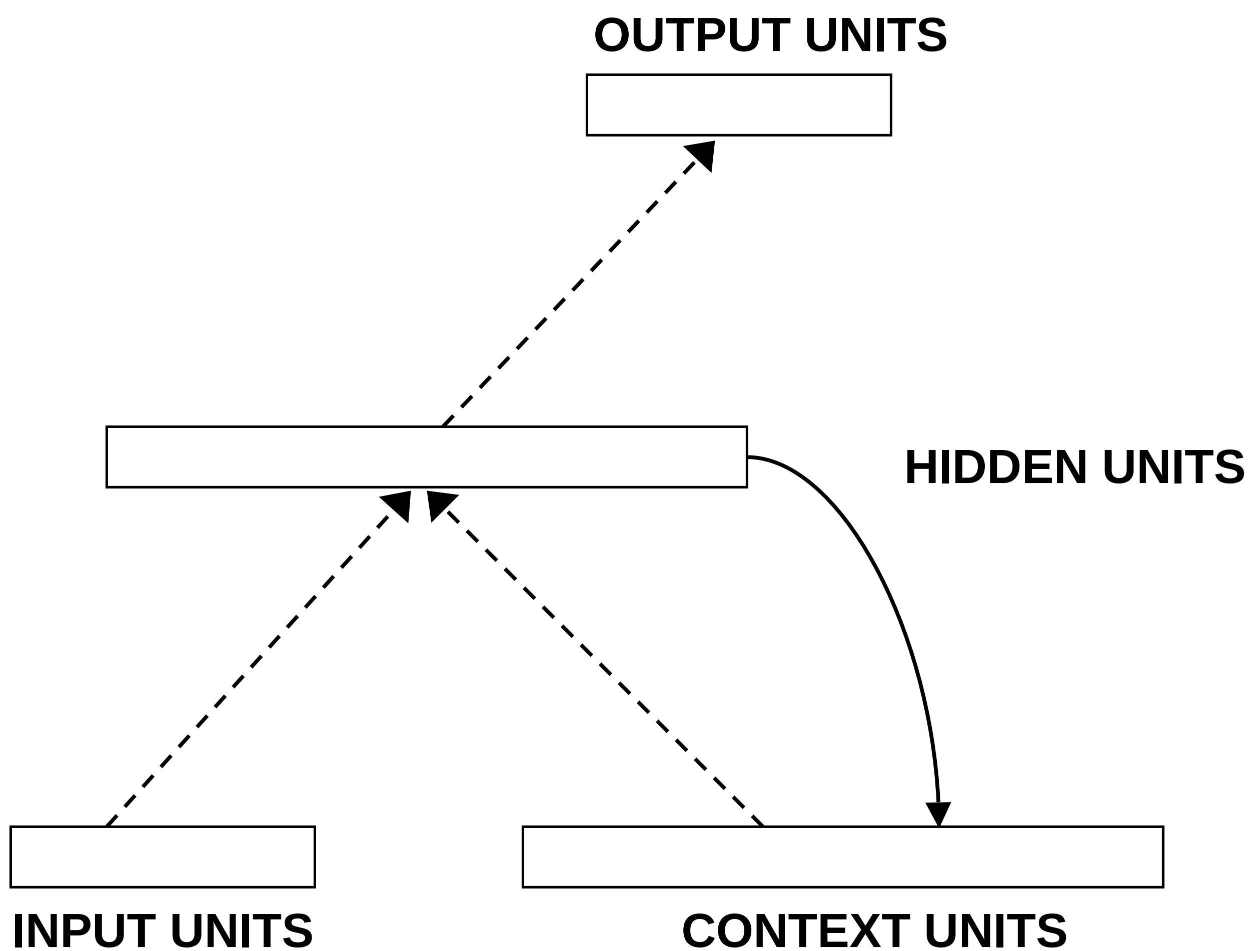
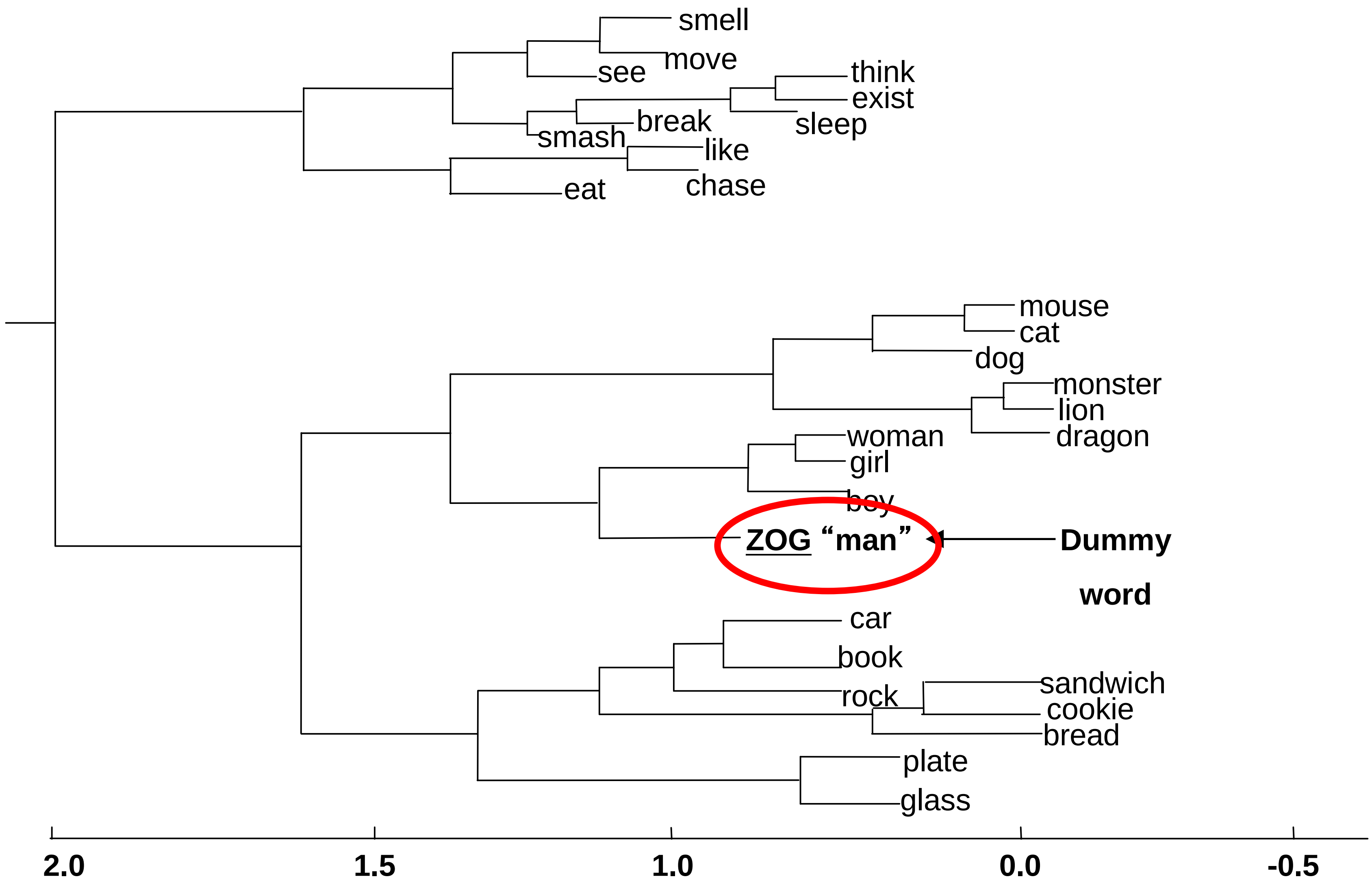


Figure 4.2 Cluster Analysis Results of Hidden Context Layers*



Source: Elman, 1990

References

Achen, C. Interpreting and using regression. In Sullivan, JL and Niemi, RG (eds.), Sage University Paper Series on Quantitative Applications in the Social Sciences (07-001). Beverly Hills, CA: Sage, 1982.

Allebeck, P and Allgulander, C. Suicide among young men: psychiatric illness, deviant behavior and substance abuse. Acta Psychiatr Scand 1990;81:565-570.

Allgulander, C, Allebeck, P, Przybeck, TR, and Rice, JP. Risk of suicide by psychiatric diagnosis in Stockholm County. European Archives of Psychiatry and Clinical Neuroscience 1992;241:323-326.

American association of suicidology report of the committee on physician assisted suicide and euthanasia. Suic Life Threat Behav 1996;26:2-19.

Anderer, P, Saletu, B, Kloppel, B, Semlitsch, HV, and Werner, H. Discrimination between demented patients and normals based on topographic EEG slow wave activity: Comparison between z statistics, discriminant analysis and artificial neural networks. Electroencephalography Clin Neurophysiol 1994;91:108-117.

Anderson, JA. An Introduction to Neural Networks. Cambridge, MA: MIT Press, 1995.

Asberg, M, Traskman, L, and Sjostrand, L. Monoamine metabolites in CFS and suicidal behavior. Arch Gen Psychiat 1976;33:1193-1197.

Bamber, D. The Area above the Ordinal Dominance Graph and the Area below the Receiver Operating Characteristic Graph. J Math Psychol 1975;12:387-415.

Baxt, WG. Application of artificial neural networks to clinical medicine. Lancet 1995;346:1135-1138.

Beck, A, Steer, R, Kovacs, M, and Garrison, B. Hopelessness and Eventual Suicide: a 10-year prospective study of patients hospitalized with suicidal ideation. Am J Psychiat 1985;142:559-563.

Bishop, CM. Neural networks for pattern recognition. Oxford: Oxford University Press, 1995.

Bland, RC, Newman, SC, and Dyck, RJ. The epidemiology of parasuicide in Edmonton. Can J Psychiat 1994;39:391-396.

Blumenthal, SJ. Suicide and gender. American Foundation for Suicide Prevention, Internet, www.afsp.org/research, 1996.

Brent, DA. The psychological autopsy: methodological considerations for the study of adolescent suicide. Suic Life Threat Behav 1989;19:43-57.

Brent, DA, Perper, JA, and Allman, CJ. Alcohol, firearms and suicide among youth: temporal trends in Allegheny County. J Am Med Assoc 1987;257:3369-3372.

Breiman, L, Friedman, JH, Olsen, RA, and Stone, CJ. Classification and Regression Trees. Boca Raton, LA: Chapman & Hall/CRC, 1984.

Briere, J and Zaidi, LY. Sexual abuse histories and sequella in female psychiatric emergency room patients. Am J Psychiat 1989;146:1602-1606.

Camus, A. Le Mythe De Sisyphe. Paris: Gallimard, 1942.

Carbone, V and Piras, G. Palomar Project: Predicting school renouncing dropouts, using the Artificial Neural Networks as a support for educational policy decisions. Subst Use Misuse 1998;33:717-750.

Cheng, B and Titterington, DM. Neural networks: a review from a statistical perspective. Statistical Sci 1994;9:2-54.

Clark, DC and Fawcett, J. Review of empirical risk factors for evaluation of the suicidal patient. In Bongar, B (eds.), Suicide. Guidelines for assessment, management and treatment. New York: Oxford University Press, 1992, pp. 16-48.

Conwell, Y, Duberstein, PR, Cox, C, Hermann, JH, Forbes, NT, and Caine, ED. Relationships of age and Axis I diagnoses in victims of completed suicide: a psychological autopsy study. Am J Psychiat 1996;153:1001-1008.

Cottrell, GW and Metcalfe, J. EMPATH: Face, emotion, and gender recognition using holons. NIPS 1991;3:564-571.

D'Alche-Buc, F, Andres, V, and Nadal, JP. Rule extraction with fuzzy neural network. Int J Neural Systems 1994;5:1-11.

de Wilde, E, Keinhorst, ICWM, Diekstra, RFW, and et al. The relationship between adolescent suicidal behavior and life events in childhood and adolescence. Am J Psychiat 1991;149:45-51.

DHHS. The Surgeon General's call to action to prevent suicide 1999. U.S.Department of Health and Human Services. Internet, www.surgeongeneral.org/osg/calltoaction/calltoaction.htm, 1999.

Dhingra, AK and Lee, BH. A genetic algorithm approach to single and multiobjective structural optimization with discrete-continuous variables. Int J for Numerical Methods in Engineering 1994;37:4059-4080.

Doig, GS, Inman, KJ, Sibbald, WJ, Martin, CM, and Robertson, JM. Modeling mortality in the intensive care unit: comparing the performance of a back-propagation, associative-learning neural network with multivariate logistic regression. The Annual Symposium on Computer Applications in Medical Care 1993;361-365.

Downey, TJ and Meyer, DJ. A genetic algorithm for feature selection. In Dagli, CH, Fernandes, BR, Ghosh, J, and Kumura, RTS (eds.), Intelligent Engineering Systems through Artificial Neural Networks. New York, NY: ASME Press, 1994, pp. 363-369.

Duh, M-S, Walker, AM, and Ayanian, JZ. Epidemiologic interpretation of artificial neural networks. Am J Epidemiol 1998;147:1112-1122.

Duh, M-S, Walker, AM, Pagano, M, and Kronlund, K. Prediction and cross-validation of neural networks versus logistic regression: Using hepatic disorders as an example. Am J Epidemiol 1998;147:407-413.

Durkeim, E. Suicide. A study in sociology, 1987. New York: Free Press, 1951.

Dwyer, CA. Cut scores and testing: Statistics, judgment, truth and error. Psychol Assess 1996;8:360-362.

Dybowski, R, Weller, P, Chang, R, and Gant, V. Prediction of outcome in critically ill patients using artificial neural network synthesized by genetic algorithm. Lancet 1996;347:1146-1150.

Elman, JL. Finding structure in time. Cognit Sci 1990;14:179-211.

Fawcett, J, Scheftner, WA, Fogg, L, Clark, DC, Young, MA, Hedeker, D, and Gibbons, R. Time-related predictors of suicide in major affective disorder. Am J Psychiat 1990;147:1189-1194.

Fox, BH, Stanek, EG, Boyd, SC, and Flannery, JT. Suicide rates among cancer patients in Connecticut. J Chron Dis 1982;35:85-100.

Freud, S. Mourning and melancholia. In Strachey, J (eds.), London: Hogarth Press, 1917, pp. 237-258.

Fu, HC and Shann, JJ. A fuzzy neural network for knowledge learning. Int J Neural Systems 1994;5:13-22.

Goldberg, D. Genetic Algorithms in Search, Optimization and Machine Learning. Reading, MA: Addition-Wesley, 1989.

Gould, MS, Fisher, P, Parides, M, Flory, M, and Shaffer, D. Psychosocial risk factors of child and adolescent completed suicide. Arch Gen Psychiat 1996;53:1155-1162.

Hafner, H, Veiel, HOF, and Welz, R. Epidemiology of suicide and attempted suicide. In Chute, K (eds.), Psychiatria Fennica: Year Book. 1988, pp. 103-123.

Harris, EC and Barraclough, B. Suicide as an outcome for medical disorders. Suicide as an outcome for medical disorders 1994;73:281-296.

Harris, EC and Barraclough, B. Suicide as an outcome for mental disorders. Br J Psychiatr 1997;170:205-228.

Hastie, TJ and Tibshirani, R. Generalized Additive Models. London, UK: Chapman & Hall, 1990.

Hawton, K. Assessment of suicide risk. Br J Psychiatry 1987;150:145-153.

Heikkinen, ME, Isometsa, ET, Aro, HM, Sarna, SJ, and Lonnqvist, JK. Age-related variation in recent life events preceding suicide. Journal of Nervous and Mental Disease 1995;183:325-331.

Hintikka, J, Viinamaki, H, Koivumaa-Honkanen, H-T, Saarinen, P, Tanskanen, A, and Lehtonen, J. Risk factors for suicide ideation in psychiatric patients. Soc Psychiat Psychiatric Epidemiol 1998;33:235-240.

Holland, JH. Adaption in natural and artificial systems. Arbor, MI: University of Michigan Press, 1975.

Holley, HL, Fick, G, and Love, EJ. Suicide following an inpatient hospitalization for a suicide attempt: A Canadian follow-up study. Soc Psychiat Psychiatric Epidemiol 1998;33:543-551.

Hyman, SE. National Institute of Mental Health. Hearing on suicide awareness and prevention, 2000.

Jacob, D and Klein, ME. The expanding role of psychological autopsies. In Leenaars, AA (eds.), Suicidology. Essays in honor of Edwin Schneidman. Northvale, NJ: Jason Aronson, 1993.

James, IP. Suicide and mortality among heroin addicts in Britain. Br J Addictions 1967;62:391-398.

Juel-Nielsen, N and Videbech, T. A twin study of suicide. Acta Genetic Med Gemellol 1970;19:307-310.

Kessler, RC, McGonagle, KA, Zhao, S, Nelson, CB, Hughes, M, Eshleman, S, Hans-Ulrich, W, and Kendler, KS. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Arch Gen Psychiat 1994;51:8-19.

Kessler, RC, Crum, RM, Warner, LA, Nelson, CB, Schulenberg, J, and Anthony, JC. Lifetime co-occurrence of DSM-III-R alcohol abuse and dependence with other psychiatric disorders in the National Comorbidity Survey. Arch Gen Psychiat 1997;54:313-321

Kilpatrick, DG, Best, CL, Veronen, LJ, Amick, AE, Villeponteaux, LA, and Ruff, GA. Mental health correlates of criminal victimization: a random community survey. J Consult Clin Psychol 1985;53:866-873.

Kircher, T and Anderson, RE. Causes of death. Proper completion of the death certificates. J Am Med Assoc 1987;258:349-352.

Klerman, GL. Clinical epidemiology of suicide. Journal of Clinical Psychiatry 1987;48:33-38.

Kovacs, M, Goldston, D, and Gatsonis, C. Suicidal behaviors and childhood-onset depressive disorders: a longitudinal investigation. J Am Acad Child Adolesc Psychiatry 1993;32:8-20.

Koza, JR. Introduction to genetic algorithms. In Genetic Programming on the Programming of Computers by Means of Natural Selection. Cambridge, MA: MIT Press, 1992, pp. 17-61.

Kraepelin, E. Psychiatry: Ein Lehrbuch. Leipzip: Brazil, 1896.

Krupinski, J, Tiller, JWG, Burrows, GD, Mackenzie, A. Predicting suicide risk among young suicide attempters, 1996.

Kung, HC, Liu, X, and Juon, H-S. Risk factors for suicide in Caucasians and African-Americans: a matched case-control study. Soc Psychiat Psychiatric Epidemiol 1998;33:155-161.

Ladd, CB. Medical Director's Notebook: Suicide. Internet, http://www.rgare.com/index_main.htm, 2001.

Leon, AC, Friedman, RA, Sweeney, JA, Brown, RP, and Mann, JJ. Statistical issues in the identification of risk factors for suicidal behavior: The application of survival analysis. Psychiatry Res 1990;31:99-108.

Lucek, PR and Ott, J. Neural network analysis of complex traits. Gen Epidemiol 1997;14:1101-1106.

Mackenbach, JP, Kunst, AE, Lautenbach, H, Bijlsma, F, and Oei, YB. Competing causes of death: An analysis using multiple-causes-of-death from the Netherlands. Am J Epidemiol 1995;141:466-475.

Mann, JJ. The neurobiology of suicide. American Foundation for Suicide Prevention 1998.

Maris, RW. The evolution of suicidology. In Leenaars, AA (eds.), Suicidology: Essays in Honor of Edwin Schneidman. Northvale, NJ: Jason Aronson, 1993.

Marttunen, MJ, Aro, HM, Henriksson, MM, and Lönnqvist, JK. Antisocial behavior in adolescent suicide. Acta Psychiatr Scand 1994;89:167-173.

Marzuk, PM, Tardiff, KT, Leon, AC, Stajic, M, Morgan, EB, and Mann, JJ. Prevalence of cocaine use among residents of New York City. Am J Psychiat 1992;149:371-375.

MathSoft. S-Plus 4 Guide to Statistics. 1997.

May, ACW and Johnson, MS. Protein structure comparisons using a combination of a genetic algorithm, dynamic programming and least-squares minimization. Protein Eng 1994;7:475-485.

McClelland, JL. Connectionist models and psychological evidence. Ann Mem Lang 1988;27:107-123.

McClelland, JL. Stochastic interactive processes and the effect of context on perception. Cog Psychol 1991;23:1-44.

McClelland, JL and Rumelhart, DE. An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. Psychol Rev 1981;88:375-407.

McKeown, RE, Garrison, CZ, Cuffe, SP, Waller, JL, Jackson, KL, and Addy, CL. Incidence and predictors of suicidal behaviors in a longitudinal sample of young adolescents. J Am Acad Child Adolesc Psychiatry 1998;37:612-619.

Mehlum, MAJL. Young male suicide attempters 20 years later: the suicide mortality rate. Mil Med 1994;159:138-141.

Mingers, J. An empirical comparison of selection measures for decision-tree induction. Machine Learning 1989;3:319-342.

Monk, M. Epidemiology of suicide. Epidemiologic Review 1987;9:51-69.

Moscicki, EK. Identification of suicide risk factors using epidemiologic studies. In Mann, JJ (eds.), Psychiatric Clinics of North America. Washington, DC: American Psychiatric Press, 1997, pp. 499-517.

Moyer, LA, Boyle, CA, and Pollock, DA. Validity of death certificates for injury-related causes of death. Am J Epidemiol 1989;130:1024-1032.

Murphy, GE. Problems in studying suicide. Psychiatric Development 1983;4:339-350.

Murphy, GE. Suicide in Alcoholism. New York: Oxford University Press, 1992.

Murphy, GE and Wetzel, RD. Family history of suicidal behavior among suicide attempters. J Nerv Ment Dis 1982;170:86-90.

Murphy, GE and Wetzel, RD. The lifetime risk of suicide in Alcoholism. Arch Gen Psychiat 1990;47:383-392.

Murphy, GE, Wetzel, RD, Robins, E, and McEvoy, L. Multiple risk factors predict suicide in alcoholism. Arch Gen Psychiat 1992;49:459-463.

Ninan, PT, van Kammen, D, Scheinin, M, Linnoila, M, Bunny, W, and Goodwin, F. 5-hydroxindoleacetic acid levels in suicidal schizophrenic patients. Am J Psychiat 1984;141:566-569.

National Center for Health Statistics. The description of the National Mortality Followback Data, December 1998. National Center for Health Statistics. Internet, www.cdc.gov.nchswww/about/major/nmfs , 1998.

NSSP. National Strategy for Suicide Prevention: Goals and objectives for action. Internet, http://www. mentalhealth.org/suicideprevention/summarytoc.htm, 2001.

O'Carroll, PW, Berman, AL, Maris, RW, Mosicki, EK, Tanney, BL, and Silvermann, MM. Beyond the tower of babel: a nomenclature for suicidology. Suic Life Threat Behav 1996;26:237-252.

Ohberg, A, Vouri, E, Ojanpera, I, and Lonnqvist, J. Alcohol and drugs in suicides. Br J Psychia 1996;169:75-80.

Owens, MJ. Age and attempted suicide. Acta Psychiatr Scand 1990;82:385-388.

PARTEK. Analytical/Visual Base Reference Manual. Version 1. St. Charles, MO: PARTEK,, 1993.

Permeger, TV, Klag, MJ, and Whelton, PK. Cause of death in patients with end-stage renal disease: death certificate vs. registry. Am J Public Health 1993;83:1738.

Perry, S, Jacobsberg, L, and Fishman, B. Suicidal ideation and HIV testing. J Am Med Assoc 1990;263:679-682.

Petronis, KR, Samuels, JF, Moscicki, EK, and Anthony, JC. An epidemiologic investigation of potential risk factors for suicide attempts. Soc Psychiat Psychiatric Epidemiol 1990;25:193-199.

Pitts, F and Winokur, F. Affective disorder. III. Diagnostic correlates and incidence of suicide. J Nerv Ment Dis 1964;139:176-181.

Pokorny, AD. Prediction of suicide in psychiatric patients: report of a prospective study. Arch Gen Psychiat 1983;40:249-257.

Price, RK, Spitznagel, EL, Downey, TJ, Meyer, DJ, Risk, NK, and El-Ghazzawy, OG. Applying artificial neural network models to clinical decision making. Psychol Assess 2000;12:52-60.

Resick, PA, Jordan, CG, Girelli, SA, Hutter, CK, and Marphefer-Dvorak, S. A comparative outcome study of behavior group therapy for sexual assault victims. Behavr Ther 1988;19:385-401.

Ripley, BD. Pattern recognition and neural networks. Cambridge: Cambridge University Press, 1996.

Robins, LN and Price, RK. Adult disorders predicted by childhood conduct problems: Results from the NIMH Epidemiologic Catchment Area Project. Psychiatry 1991;54:116-132.

Roins, E. The final months. A study of the lives of 134 persons who committed suicide. New York: Oxford University Press, 1981.

Roy, A, Segal, N, Centerwall, D, and Robinette, D. Suicide in twins. Arch Gen Psychiat 1991;48:29-32.

Rubin, MN. Mental illness and social integration among suicide attempters in Copenhagen. Acta Psychiatr Scand 1993;88:278-285.

Rumelhart, DE and McClelland, JL. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations. Massachusetts: MIT Press, 1986.

Ryngestad, T. Mortality after deliberate self-poisoning. A prospective follow-up study of 587 persons observed for 5279 person years: risk factors and causes of death. Soc Psychiatr Psychiatric Epidemiol 1997;32:443-450.

Saccone, NL, Downey, TJ, Meyer, DJ, Neuman, RJ, and Rice, JP. Mapping genotype to phenotype for linkage analysis. Genetic Epi 1999;17:S703-S708.

Sarle, WS. Neural networks and statistical models. Proceedings of the Nineteenth Annual SAS Users Group International Conference, April, 1994.

Sarle, WS. Neural Network FAQ, parts 1 - 7. Internet, ftp://ftp.sas.com/pub/neural/FAQ.html, 1999.

Saxon, S, Aldrich, SY, and Kuncel, EE. Suicide and drug abuse. In Lettieri, DJ (eds.), Drugs and Suicide When Other Coping Strategies Fail. Beverly Hills: Sage, 1978.

Schneidman, E. The psychological autopsy. Suic Life Threat Behav 1981;11:325-340.

Schneidman, ES and Farberow, NL. Statistical comparisons between attempted and committed suicides. In Schneidman, ES and Farberow, NL (eds.), The Cry for Help. London: Academic Press, 1965, pp. 19-47.

Schulsinger, F, Kety, SS, Rosenthal, D, and Wender, PH. A family study of suicide. In Origin, prevention and treatment of affective disorders. Orlando, Fl: Academic Press, 1979, pp. 277-287.

Shafii, M, Carrigan, S, Whittinghill, JR, and et al. Psychological autopsy of completed suicide in children and adolescents. Am J Psychiat 1985;142:1061-1064.

Shah, BV, Barnwell, BG, and Bieler, GS. SUDAAN User's Manual, Release 7.0. Research Triangle Park, NC. Research Triangle Institute, 1996.

Soubrier, J-P. Definitions of suicide. In Leenaars, AA (eds.), Suicidology. Essays in honor of Edwin Schneidman. Northvale, NJ: Jason Aronson, 1993.

Soukas, J and Lonnqvist, J. Outcome of attempted suicide and psychiatric consultation: risk factors and suicide mortality during a five-year follow-up. Acta Psychiatr Scand 1991;84:545-549.

Stack, S. The effect of religious commitment on suicide: a cross-national analysis. J Health Soc Behav 1983;24:362-374.

Statham, DJ, Heath, AC, Madden, PAF, Bucholz, KK, Bierut, LJ, Dinwiddie, SH, Slutske, WS, Dunne, MP, and Martin, NG. Suicidal behaviour: an epidemiological and genetic study. Psychol Med 1998;28:839-855.

Swets, JA and Pickett, RM. Evaluation of Diagnostic Systems: Methods from Signal Detection Theory. New York, NY: Academic Press, 1982.

Tardon, AG, Zaplana, J, Hernandez, R, and Cueto, A. Usefulness of the codification of multiple causes of death in mortality statistics. Int J Epidemiol 1995;24:1132-1137.

Therneau, TM, Atkinson, EJ. An Introduction to Recursive Partitioning Using the RPART Routines. Presented at the Mayo Foundation, September 3, 1997. 1997.

Tsuang, MT. Risk of suicide in the relatives of schizophrenics, manics, depressives and controls. J Clin Psychiatr 1983;44:396-400.

Tuckman, J and Youngman, WF. Identifying suicide risk groups among attempted suicide. Pub Health Rep 1963;78:763-766.

Vivarelli, F, Giusti, G, Villani, M, Campanini, R, Fariselli, P, Compiani, M, and Casadio, R. LGANN: A parallel system combining a local genetic algorithm and neural networks for the prediction of secondary structure of proteins. CABIOS 1995;11:253-260.

Wagner, BM. Family risk factors for child and adolescent suicidal behavior. Psychol Bull 1997;121:246-298.

Warner, LA, Kessler, RC, Hughes, M, Anthony, JC, and Nelson, CB. Prevalence and correlates of drug use and dependence in the United States. Arch Gen Psychiat 1995;52:219-229.

Weinstein, MC and Fineberg, HV. Clinical Decision Analysis. Philadelphia, PA: W.B. Saunders, 1980.

Weisman, AD and Kastenbaum, R. The psychological autopsy. A study of terminal phase of life. New York: Behavioral Publications, 1968.

Wender, PH, Kety, SS, Rosenthal, D, Schulsinger, F, Ortmann, J, and Lunde, I. Psychiatric disorders in the biologic and adoptive families of adopted individuals with affective disorders. Arch Gen Psychiat 1986;43:923-929.

Wetzel, RD. Hopelessness, depression and suicide intent. Arch Gen Psychiat 1976;33:1069-1073.

White, H. Learning in artificial neural networks: A statistical perspective. Neural Comp 1989;1:425-464.

White, H. Artificial Neural Networks: Approximation and Learning Theory. Oxford, UK: Blackwell, 1992.

Wiles, J and Bloesch, A. Operators and curried functions: Training and analysis of simple recurrent networks. In Moddy, JE, Hansen, SJ, and Lippman, RP (eds.), Advances in Neural Information Processing Systems 4. San Mateo, CA: Morgan Kaufmann, 1992, pp. 325-332.

Winchel, RM, Stanley, B, and Stanley, M. Biochemical aspects of suicide. In Blumenthal, SJ (eds.), Suicide over the life cycle. Washington, DC: American Psychiatric Press, 1990, pp. 97-126.

Yang, B and Clum, GA. Life stress, social support, and problem-solving skills predictive of depressive symptoms, hopelessness, and suicide ideation in an Asian student population. A test of a model. Suic Life Threat Behav 1994;24:127-139.

Zou, Y, Shen, Y, Shu, L, Wang, Y, Feng, F, Xu, K, Qu, Y, Song, Y, Zhong, Y, Wang, M, and Liu, W. Artificial neural network to assist psychiatric diagnosis. Br J Psychiat 1996;169:64-67.