# NAVAL POSTGRADUATE SCHOOL
## Monterey, California



# THESIS

**ANALYSIS OF NAVY DELAYED ENTRY PROGRAM AND RECRUIT TRAINING CENTER ATTRITION**

by

Bryant W. Knox

June 1998

Thesis Advisor: Samuel E. Buttrey
Second Reader: Harold J. Larson

# REPORT DOCUMENTATION PAGE

*Form Approved OMB No. 0704-0188*



| 1. AGENCY USE ONLY *(Leave blank)* | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | June 1998 | Master's Thesis |

**4. TITLE AND SUBTITLE**
**ANALYSIS OF NAVY DELAYED ENTRY PROGRAM AND RECRUIT TRAINING CENTER ATTRITION**

**5. FUNDING NUMBERS**

**6. AUTHOR(S)** Knox, Bryant W.

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Naval Postgraduate School
Monterey, CA 93943-5000

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**1. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
Commander, Navy Recruiting Command, Arlington, Va.

**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES**
The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

**12a. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for public release; distribution is unlimited.

**12b. DISTRIBUTION CODE**

**ABSTRACT *(maximum 200 words)***

Attrition from the Navy's Delayed Entry Program (DEP) and attrition from Bootcamp are costly phenomena. The Commander of Naval Recruiting (CNRC) and Center for Naval Analysis (CNA) have periodically modeled both DEP and Bootcamp attrition with logistic regression. This thesis analyzes current data provided by CNRC and CNA. Both DEP and Bootcamp attrition are modeled using logistic regression and tree-structured classification. For DEP, the logistic model indicates that individuals who accept incentives prior to enlistment (i.e., Navy College Fund or Enlisted Bonus Program) and individuals who change enlistment programs (while in DEP) have a significantly lower propensity to attrite from DEP than others. The DEP tree model indicates that an individual with a low Armed Forces Qualification Test (AFQT) score, no high school diploma and a long scheduled DEP duration has a 97% probability of attriting. For Bootcamp, the logistic model indicates that individuals who use tobacco products, individuals who do not exercise, and individuals that have criminal waivers have a significantly higher propensity to attrite than others. The Bootcamp tree model shows that smokers and individuals with low AFQT scores have higher propensities to attrite than others. The models are tested using random partitions and this analysis shows that all of the models predict poorly at the individual level, despite strong statistical significance.

**13. SUBJECT TERMS**
DEP, Bootcamp, tree-structured classification, logistic regression, AFQT

**15. NUMBER OF PAGES**
74

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFI- CATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | UL |

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

**Approved for public release; distribution is unlimited**

# ANALYSIS OF NAVY DELAYED ENTRY PROGRAM AND RECRUIT TRAINING CENTER ATTRITION


**Bryant W. Knox - Lieutenant, United States Navy**
**B.A., Economics, Davidson College, 1985**
**M.S., Business Administration, Boston University, 1998**


Submitted in partial fulfillment of the
requirements for the degree of

**MASTER OF SCIENCE IN OPERATIONS RESEARCH**

from the

**NAVAL POSTGRADUATE SCHOOL**

**June 1998**

Author: ______________________
Bryant W. Knox

Approved By: ______________________
Samuel E. Buttrey, Thesis Advisor

______________________
Harold J. Larson, Second Reader

______________________
Richard E. Rosenthal, Chairman
Department of Operations Research

# ABSTRACT

Attrition from the Navy's Delayed Entry Program (DEP) and attrition from Bootcamp are costly phenomena. The Commander of Naval Recruiting (CNRC) and Center for Naval Analysis (CNA) have periodically modeled both DEP and Bootcamp attrition with logistic regression. This thesis analyzes current data provided by CNRC and CNA. Both DEP and Bootcamp attrition are modeled using logistic regression and tree-structured classification. For DEP, the logistic model indicates that individuals who accept incentives prior to enlistment (i.e., Navy College Fund or Enlisted Bonus Program) and individuals who change enlistment programs (while in DEP) have a significantly lower propensity to attrite from DEP than others. The DEP tree model indicates that an individual with a low Armed Forces Qualification Test (AFQT) score, no high school diploma and a long scheduled DEP duration has a 97% probability of attriting. For Bootcamp, the logistic model indicates that individuals who use tobacco products, individuals who do not exercise, and individuals that have criminal waivers have a significantly higher propensity to attrite than others. The Bootcamp tree model shows that smokers and individuals with low AFQT scores have higher propensities to attrite than others. The models are tested using random partitions and this analysis shows that all of the models predict poorly at the individual level, despite strong statistical significance.

## TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# EXECUTIVE SUMMARY

The accession of quality personnel continues to be a challenge for the Navy. Strong economic growth and low unemployment have decreased the pool of potential recruits and the Navy is having difficulty meeting its recruiting goals. The situation is exacerbated by a dwindling budget. The Navy is confronted with the challenge of doing more with less and must constantly find areas where financial savings are possible.

Attrition, in both the Delayed Entry Program (DEP) and in Bootcamp, is one such area. It costs the Navy an average of $6500.00 per person to recruit an individual and an average of $1200.00 to begin their training in Bootcamp. These cost estimates aggregate the costs of testing, physical examinations, recruiter effort, DEP maintenance, shipping to Bootcamp and initial Bootcamp screening. An average of 19% of the individuals who enter DEP attrite, while an average of 13% of the individuals who enter Bootcamp attrite. DEP and Bootcamp attrition cost the Navy upwards of $139,000,000.00 per year (based on a shipping goal of 55,000 new recruits).

Attrition has been the focus of numerous studies, most of which predicted the probability of attrition as the dependent variable in a multivariate logistic regression model. This thesis analyzes attrition as a dependent variable using logistic regression and also models the probability of attrition using tree-structured classification. Tree-structured classification is an effective alternative to logistic regression and often provides insight into the data which is not discernible with the logistic models.

The data used for this thesis were provided by CNRC, Code 20, and represented every individual scheduled to report to Bootcamp between October 1995 and December 1997. There were 130,486 records in the data set. For the analysis, the data are

randomly partitioned into sets for building DEP and Bootcamp models and sets for testing the models. Further, since the ouput of both the logistic regression models and the classification tree models is a "probability of attrition", an optimal decision criterion (for scoring a fitted value as an attrite) is developed. This threshold is used to test the predictive power of each model.

Several significant factors are found with the logistic models. For DEP attrition, the factors that increase the probability of attrition with an increase in their value are age, race (white or black), Government Equivalency (GED) high school diplomas and scheduled DEP duration. The factors that decrease the likelihood of attrition with an increase in their value are Armed Forces Qualification Test (AFQT) score, sex (male), accepting incentive programs (Navy College Fund or Enlisted Bonus), enlisting as a senior in high school and changing programs while in DEP.

For Bootcamp attrition, the logistic models indicate that the probability of attrition increases with increases in age, race (white and black), GED high school diplomas, waivers (crime and other), tobacco use and program changes. The factors that decrease the probability of attrition with an increase in their value are AFQT score, long DEP duration, and exercise (running or jogging at least three times a week).

The tree models identify several interesting relationships. First, the DEP tree shows that individuals who enlist as seniors but do not graduate from high school or graduate with a GED have a 98% chance of attriting. Second, individuals with no high school degree and an AFQT score below 49.5 who do not enlist as seniors in high school have a 76% chance of attriting. Third, individuals who do not graduate from high school, have an AFQT score below 49.5 and are scheduled for long DEP durations have a 97%

chance of attriting. The Bootcamp tree identifies smoking and low AFQT scores as increasing the probability of attrition. The trees reveal structure within the data which is not identified through logistic regression.

Once the models are constructed, they are tested using the random partitions mentioned earlier. The DEP tree node, with a 98% attrite probability (mentioned above), correctly predicts 3954 attrites, while the DEP logistic model predicts only 71. Both of the Bootcamp models predict poorly. Further analysis of the DEP tree node with 3954 correct predictions reveals that the educational codes of individuals who quit from the DEP are suspect and the tree's predictive power should be scrutinized.

Many of the predictive factors found in this analysis have been identified in previous research, but the classification methodology identifies several interesting relationships not previously documented. All of the models have strong statistical significance and weak predictive performance. Policies that exclude individuals, based on these results, are not recommended.

# I. INTRODUCTION

## A. BACKGROUND

Technological advancements in both modern warfare and its strategies have enabled the Navy to reduce its force structure while maintaining operational readiness. Despite all of the new hardware and software, the key asset remains people; it is naval personnel who man the high-tech workstations and the ships at sea.

Naval personnel needs are met with an all-volunteer force that is either actively recruited by representatives of the Naval Recruiting Command (CNRC) or accessed through one of the Officer Programs. Since the Navy is all-volunteer, it is competing in the domestic job market with both the other branches of the military (Army, Air Force, Marine Corps) and the civilian sector. This dependence upon the job market for personnel subjects the Navy to the same economic forces as corporate America. For example, when unemployment is high, it is much easier for the Navy to recruit than when it is low. Currently, the United States is experiencing a 20-year low with respect to unemployment while the Navy is having difficulty meeting its recruiting goals and many fleet units are undermanned.

There is more to both manning and recruiting difficulties than the unemployment rate. The fiscal constraints that accompany the mandated reduction in forces and the changing roles of the military have been mentioned as possible causes of the difficulties (CNRC, Code 20, 1997). Given the changing environment, the Navy must continuously review its manpower policies and find areas with potential for improvement.

One of these areas is attrition, the unplanned loss of individuals who have promised to join or are already in the Naval Service. Thirty two percent of the individuals who initially sign contracts attrite before their fleet service begins. These attrition losses inflate goals and quotas and waste assets because the Navy expends resources when recruiting and conducting initial skill training. This thesis analyzes the attrition phenomenon.

## B. THE RECRUITING PROCESS

For the purposes of this paper, the recruiting process is defined as "Enlisted Recruiting." "Officer Recruiting" will not be included in this analysis.

### 1. Setting Goals and Quotas

The recruiting process is driven by congressional mandates and fleet needs. Congress, after reviewing budgetary and strategic considerations, sets the force size in terms of numbers of personnel required to fill each pay-grade within the naval force structure. This set of numbers is a target, which must be maintained within 1% (CNRC, Code 20,1997). Given the congressional requirements, the Bureau of Naval Personnel (BUPERS) is charged with continuously analyzing the status of forces to determine accession requirements. Figure 1 summarizes the goal/requirements process.

BUPERS answers fleet needs generated by the various Operational, Administrative and Training Commanders (represented in Figure 1 as fleet units). Each of these Commanders has actual billets (or jobs) authorized within the force structure. For example, an aviation squadron with sea-going detachments may be authorized eight aviation electricians below the pay-grade of E-5 (Petty Officer, Second Class); if the

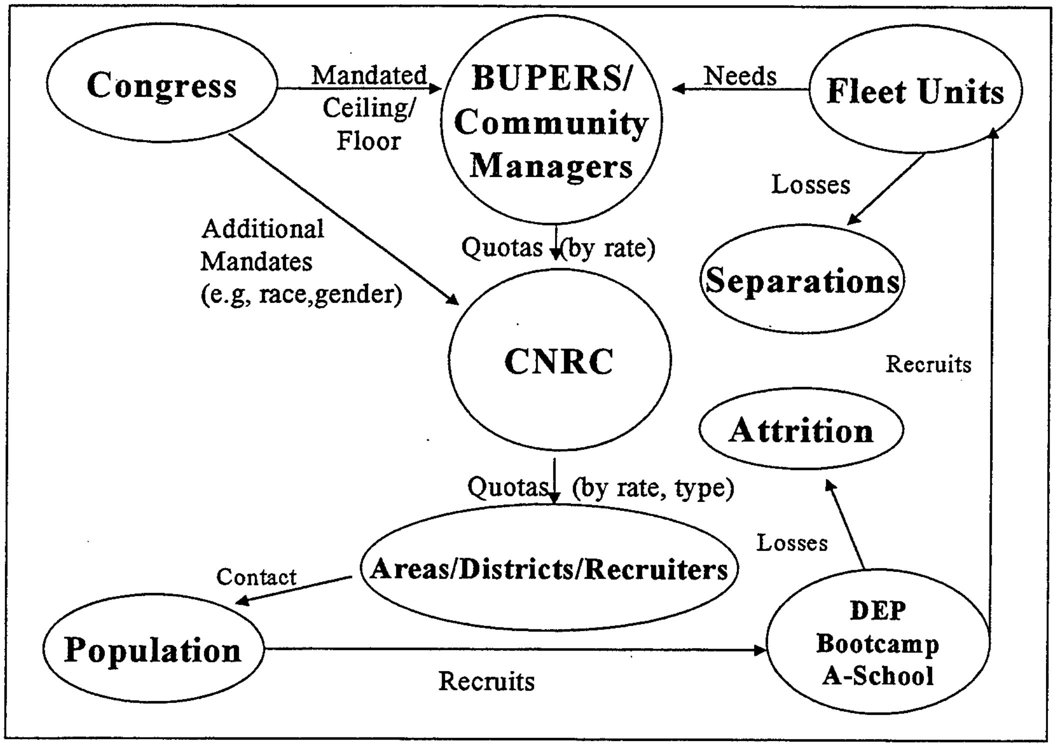


**Figure 1. Process Overview**

billets are not completely filled, the Commanding Officer will request additional personnel via BUPERS. BUPERS will weigh this request with the requests of other Commanders and with the overall status of forces. BUPERS will then either fill or "gap" the billet (gapping a billet implies that the billet will remain vacant until a suitable replacement is identified). Not every fleet need is planned for; sailors may separate from service for disciplinary reasons or new operational requirements may arise. In any case, if BUPERS elects to fill the billet, it has several choices.

First, an individual already in service may fill the billet. Depending upon the nature of the vacancy, this may warrant gapping another Commander's unit. For example, if a sea-going detachment from the aviation squadron needs an electrician for a

detachment departing for a regional conflict, BUPERS might transfer an electrician from a non-deploying aviation unit.

Second, BUPERS may identify an individual who is currently in the training pipeline to fill the billet. This transfer will occur at the completion of training. In this instance, a member of a training class with an appropriate graduation date is selected rather than a specific individual. The third method is to recruit a new individual. This method transfers the requirement to the recruiting command. These three methods require increasingly longer periods of time to fill the billet.

In each case, it takes some unspecified period of time before the Commander has his billet filled. If the need is not planned for, and the only way to fill the billet is with a new recruit, it will take at least three months (in the case of a non-rated sailor) and may take as long as two years (in the case of a nuclear power plant technician) to fill the billet. For an aviation electrician, the process would take approximately eight months. Planning for these needs is critical in maintaining fleet manning levels.

BUPERS employs an array of planning models that forecast these fleet requirements. The specific models are beyond the scope of this paper but it suffices to say that they help the community managers within BUPERS balance the fleet needs and congressional mandates by using historical data. The end result is that the community managers generate quotas for new accessions. The quotas are rating, month, and gender specific (e.g., the Navy may need 460 male aviation electricians to enter bootcamp in April). These quotas are designed to get individuals into the training pipeline to meet fleet requirements in the future. Filling these quotas is the responsibility of CNRC. CNRC analyzes the quotas and incorporates additional congressional mandates. For

example, quotas are often sub-categorized by CNRC to include race and educational background. CNRC divides the quotas into goals for each of its recruiting areas.

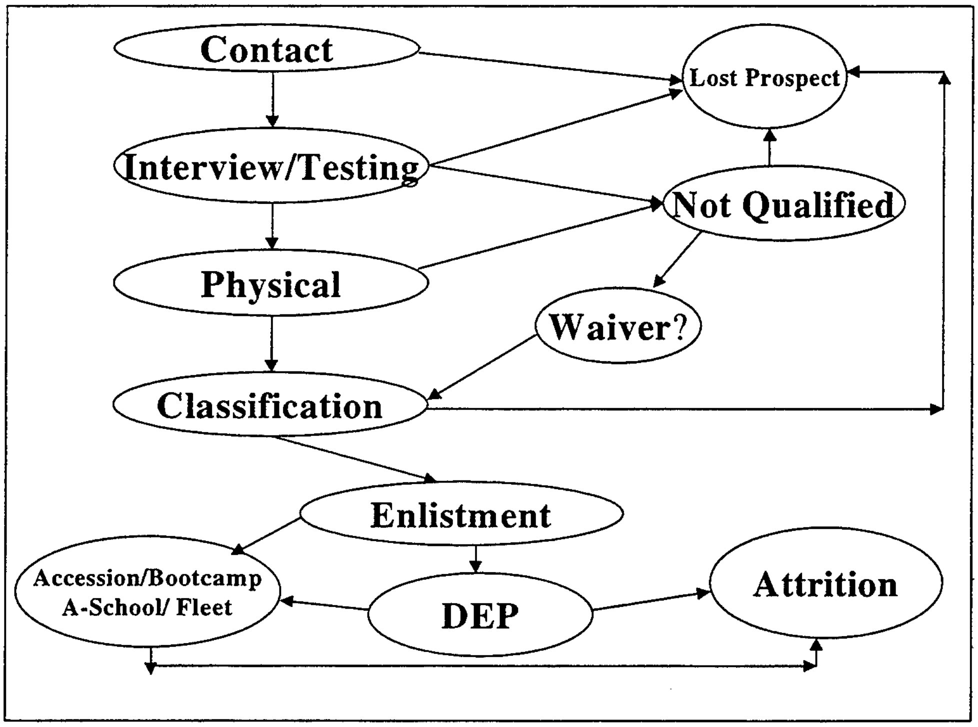


**Figure 2. The Recruiting Process**

The goals are ultimately transferred to the recruiting districts and the individual recruiters. There are approximately 3500 active recruiters in the Navy, with an aggregate goal of approximately 55,000 new recruits (for FY 1998). Simple analysis shows each recruiter should send an average of 1.3 new recruits to bootcamp per month. At the recruiter level, the quotas are specified with respect to race, educational background and gender and individual recruiter goals reflect the demographics of the recruiting region. For example, at a given instant in the Seattle recruiting district there may be only two slots for female aviation electricians for the month of May. Such restrictions, combined

with the management practices of the districts, may yield individual recruiter goals as low as one new recruit per month or as high as five new recruits per month.

### 2. Recruiting

The transition from civilian life to naval service is a complex process for the majority of accessions. This process is summarized in Figure 2. Armed with quotas, field recruiters seek to contact as many potential recruits as possible. Some interested individuals simply walk into a recruiter's office; others may fill out the information page on the Navy's website and be directly called by a recruiter. Many initial contacts come from recruiter presentations to local high schools and community colleges. The goal of the contact phase is to generate interviews.

The interview is where the prospective recruit (prospect) sits down with the recruiter to get the sales pitch. This pitch describes all of the possible opportunities (within the Navy) available to a new recruit. This is also the first opportunity for the recruiter to query the individual. The recruiter may directly ask the individual about past drug use, legal problems, or other barriers to recruitment.

If a qualified recruit remains interested, he or she may then be scheduled for the Armed Forces Qualification Test (AFQT). The AFQT is a standardized test designed to evaluate an individual's cognitive abilities and to determine the military tasks in which he or she might excel (if any). It is scored on a percentile scale from 1 to 99, with 99 being considered outstanding (CNRC, Code 20, 1997). After the interview and AFQT, a recruiter may do an initial classification of the individual by using the CNRC recruit quality matrix (RQM), which is depicted in Figure 3.

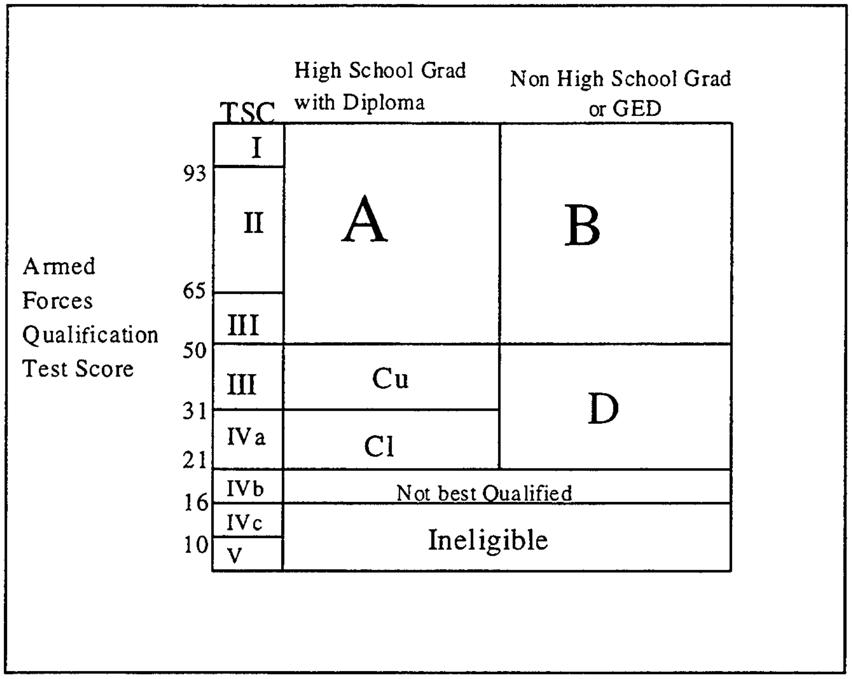


**Figure 3. Recruit Quality Matrix (Courtesy of CNRC, Code 20)**

The left side of the matrix shows AFQT scores; breakpoints are indicated in the picture. These scores are used to categorize prospects according to Test Score Category (TSC). Each prospect falls into a cell based upon TSC and his or her educational level. An individual who falls in cell A is highly desirable while an individual who falls in cell D is accepted only when severe recruiting shortages occur. There are mandated percentage limits on the maximum number of individuals from certain cells who may be recruited during normal operations. 95% of the total accessions must be high school graduates (this is more stringent than the congressional mandate of 90%), with 65% from category III-A or above (BUPERS LTR, 15 Jul 1997).

If the prospect is found to be qualified he or she will then be scheduled for a physical examination. Physicals are conducted at the Military Entrance Processing Stations (MEPS) located throughout the country. If something wrong is apparent during

the physical, the individual may be disqualified or a waiver package may be submitted by the recruiter. Upon completion of the physical, the qualified prospects proceed to classification.

During classification, a qualified prospect sits down with a classifier who weighs Navy needs for specific rates (the quotas) with the desires, test scores and academic credentials of the individual. For example, if the Navy has slots available for aviation electricians in June and the individual wants to be an aviation electrician the classifier will, generally, fill the slot with the individual. However, even if there is an opening for an electrician and it is the individual's first choice, there may be an urgent need for another rate (e.g., nuclear power technicians). If the individual is also qualified for this billet, the classifier may try to sell it to the prospect. If the prospect does not seem interested, the classifier can offer incentive packages. The two prime incentive plans are The Navy College Fund and The Enlisted Bonus Program.

The Navy College Fund (NCF) provides \$30,000.00 to \$40,000.00 for college to qualified individuals who successfully complete training in the specified field. For example, in the Nuclear Field, the Navy will pay \$40,000.00 and for Aviation Electronics, \$30,000.00. The Enlisted Bonus Program (EB) provides cash ranging from \$1000.00 (Aviation Electricians) to \$12,000.00 (Nuclear Field) for those who complete training. (CNRC, Code 20, 1997, BUPERS MSG DTG 091131Z Dec 1997) A prospective recruit may choose one, but not both, of these plans.

Classifiers do whatever they can to funnel individuals to the proper pipelines but will not do so at the expense of losing the recruit. If a prospect is qualified then he or she may be enlisted with no job assignment. In this case, classification is delayed and the

enlistment still occurs. Once the classification phase is complete the individual is enlisted in the Naval Reserve until he or she ships to bootcamp. The enlistment often occurs immediately following classification, which is usually the same day as the physical.

The final category of personnel to be discussed is those who qualify for waivers. In each of the previous phases, interviews and physicals may have found some trait or historical fact that makes the recruit generally unacceptable. In these cases, the recruiter may apply for a waiver of standards for the individual. CNRC evaluates these waivers on a case-by-case basis and may deem the candidate qualified. Waivers for prior drug use, physical impairments and prior legal problems are common.

## C. ENLISTMENT

### 1. Delayed Entry Program

After enlistment, recruits take one of two paths. If scheduled to begin bootcamp within 30 days, they are categorized as direct shippers and simply wait to be shipped to bootcamp. If they are not scheduled for bootcamp within 30 days, they enter the Delayed Entry Program or DEP. Individuals in the DEP attend monthly meetings and are tracked by their recruiter or a recruiting representative. While in DEP, they are expected to exercise and prepare for bootcamp but are not formally required to do anything. DEP is the first place in which qualified individuals attrite. Generally, the individual simply fails to report to bootcamp or quits, but a variety of other reasons have been identified. The categories in Figure 4 represent aggregates of the actual DEP attrite codes furnished by CNRC. The data was a set of 21332 DEP attrites (out of 112275 contracts) who dropped

out between July 1995 and October 1997. "Admin" attrites reflect individuals who left DEP due to administrative errors such as a change in his or her bootcamp shipping date or reclassification due to the needs of the Navy. The "Drugs/Alcohol" attrites represent individuals who failed urinalysis or had alcohol addiction problems. The "Medical" attrites represent those who had unwaiverable medical problems such as Crone's disease.

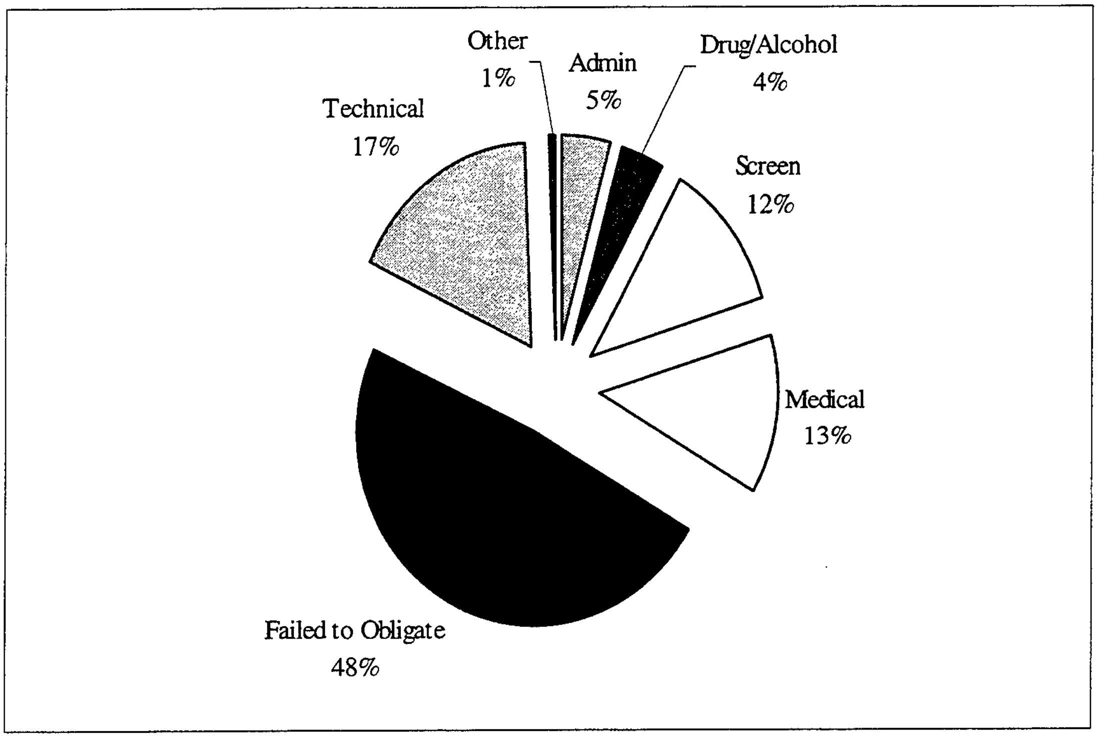


**Figure 4. DEP Attrition Breakdown**

The "Failed to Obligate" attrites simply quit. The "Screen" attrites represent individuals who had unacceptable and unwaiverable behavior in their past which was not discovered until DEP service began; quite often legal trouble falls into this category. Finally, the "Technical" category represents those individuals who became ineligible during DEP; pregnancy and death are included in this category. A complete breakdown of the aggregate categories and their associated attrition reasons can be found in

Appendix A. On average, 19% of the individuals who enter the DEP never entered bootcamp.

### 2. Indoctrination Training

Those individuals who do not attrite from the DEP ship to bootcamp. Bootcamp is conducted at the Recruit Training Center Great Lakes, Illinois (RTC). Indoctrination begins with a thorough medical screening, which includes urinalysis. While in bootcamp, recruits are volunteers and may quit at any time.

Indoctrination training is scheduled for eight weeks and ends by attrition or graduation for each individual. Upon graduation, the new recruit may either proceed to skills training (referred to as A-School) or directly to the fleet (if no skills training is required). If the individual attrites, he or she is sent home.

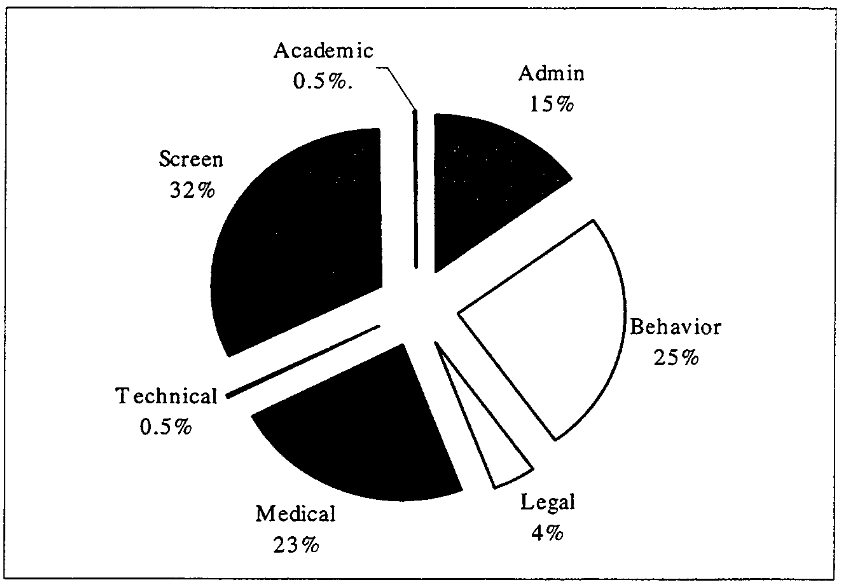


**Figure 5. RTC Attrition Breakdown**

Reasons for bootcamp attrition are as varied as those for DEP attrition and are summarized in Figure 5. The categories in Figure 5 represent aggregates of the RTC attrite codes used by the staff in Great Lakes. The "Academic" category represents

academic failure in the course work during the training program (including language deficiencies). The "Behavior" category describes actions by an individual during training which are not consistent with military service (e.g., sleepwalking, suicidal behavior, bedwetting). "Screen" encompasses the prior problems that were not evident during the recruiting process (e.g., failing the indoctrination urinalysis). The "Admin", "Technical" and "Medical" categories are similar to those described in the DEP attrition description. A complete breakdown of the aggregate categories and their associated attrition reasons can be found in Appendix B. On average, 13% of the individuals who entered bootcamp failed to graduate.

## D. COST ESTIMATION

### 1. Recruiting Costs

Estimating the cost expended on each recruit can be broken down into two distinct parts. The first estimate covers the recruiting process while the second process estimates the costs associated with shipping and bootcamp. CNRC derives the first estimate with the Planned Resource Optimization Model (PRO model) developed by Schmitz and Reinert (1995); Figure 6 summarizes the model.

The PRO model is designed to "estimate the costs of recruiting different types of individuals under different market conditions"(Schmitz and Bohn, 1996). Additionally, it provides CNRC with an optimal resource allocation schedule and a "recruits per recruiter" goal schedule. Using this model with input parameters from February 1998 (unemployment rate, current number of recruiters etc.), sensitivity analysis for various hypothetical attrition rates was performed. The results are summarized in Table 1.

The cells in Table 1 represent, in thousands of dollars, the cost to recruit an individual of a given cell type under varying hypothetical attrition rates ranging from 19% to 0%. For example, when attrition decreases from 17% to 15%, the cost to recruit

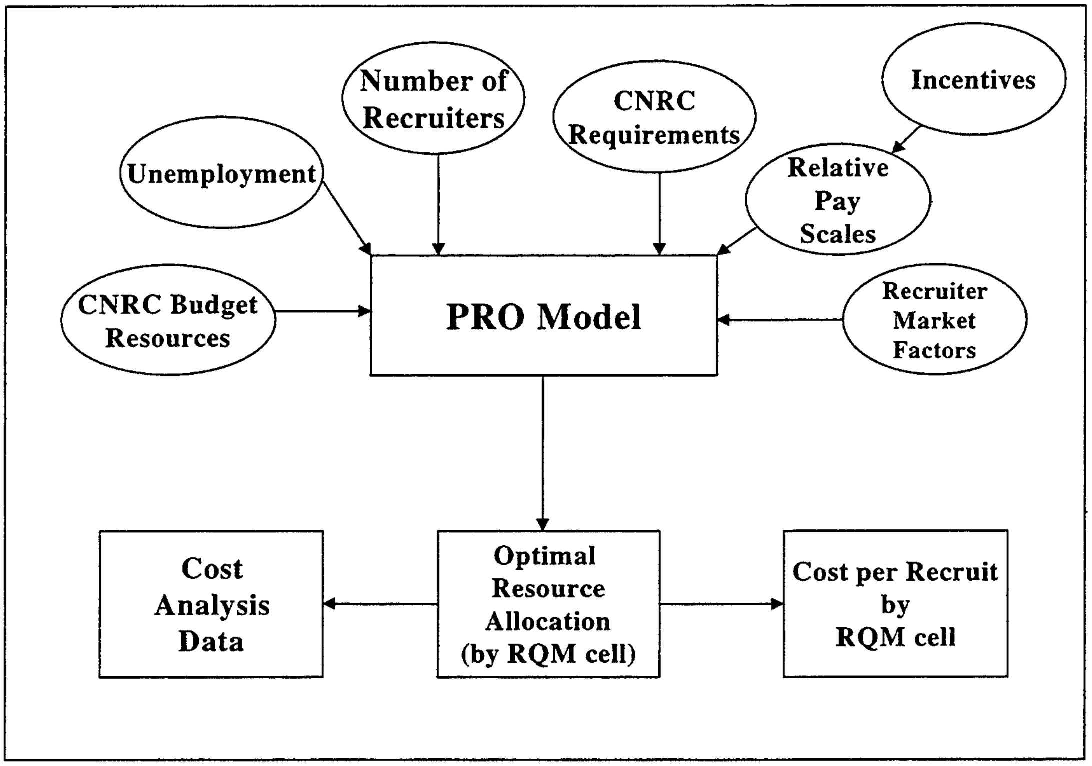


**Figure 6. Planned Resource Optimization Model**

A-cell individuals drops from $6900.00 to $6700.00. With the current state of unemployment (20 year low), it makes sense that it is more expensive to recruit talented A-cell individuals than B-cells, as the former can more easily find employment in the civilian sector. The second highest recruiting cost is C-cell individuals; this is attributed to their higher than average attrition rate, which drives their relative costs up in the PRO model.

**Table 1. Cost Per Recruit/DEP Attrition Percentage**

| Cell | Hypothetical Attrition Rates | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $ x 1000 | 19% | 17% | 15% | 13% | 11% | 9% | 7% | 0% |
| A-Cell | 7.1 | 6.9 | 6.7 | 6.6 | 6.5 | 6.3 | 6.2 | 5.9 |
| B-Cell | 5.7 | 5.7 | 5.4 | 5.3 | 5.3 | 5.1 | 5 | 4.7 |
| C-Cell | 6.7 | 6.6 | 6.4 | 6.3 | 6.1 | 6 | 5.8 | 5.6 |

From a cost standpoint, the $200.00 savings (per A-cell recruit) realized when the attrition rate is reduced from 17% to 15% results in a potential cost savings of $7,000,000.00 per year ($200.00 * 35,000 A-Cells =$7,000,000.00). The cost incurred during the recruiting process must also include DEP management costs. CNRC estimates that current management practices, which involve monthly contact and special events, result in a $50.00 per month expenditure per recruit (Schmitz and Bohn, 1996).

### 2. Bootcamp Costs

Jacklich (1998) recently estimated the costs associated with sending an individual to bootcamp. Individuals who fail the initial drug screening spend an average of nine days at Great Lakes. The nine day average cost (food, lodging, clothes, etc.), when combined with the cost of the plane ticket to RTC and the bus ticket home, results in an expenditure of $1200.00 per attrite. Depending upon the geographical origin of the new recruit, this amount can be as low as $900.00 and as high as $1500.00 (Jacklich, 1998). Analysis of RTC attrition data indicates that the average amount of time all attrites (including drug attrites) spend in RTC is 12 days but Jacklich's cost estimate is a useful lower bound.

Using Jacklich's estimate, sensitivity analysis with respect to varying attrition rates was performed. A 1.0% decrease in the RTC attrition rate increases the average

number of RTC completions by 583 recruits per year. Weighting this estimate by RQM cell type and multiplying by the relative costs yields the results summarized in Table 2.

**Table 2. RTC Savings**

| Parameter/Cell | A-Cell | B-Cell | C-Cell | Total |
|---|---|---|---|---|
| Number Recruits | 350 | 29 | 204 | 583 (Average) |
| Cost Multiplier** | $5,900.00 | $4,700.00 | $5,600.00 | N/A |
| Savings (in Millions) | $2.06 | $0.14 | $1.14 | $3.34 |

**Based upon Zero DEP attrition (hypothetical lowest cost)

## E. PREVIOUS RESEARCH

The high cost of recruiting an individual and sending him or her to bootcamp illustrates the need for minimizing unplanned losses. The issue is not new; it has been the focus of numerous studies. This section summarizes some of the prior research.

In 1995, Martin published a dissertation analyzing Army Attrition. He modeled first term attrition using contingency tables and logistic multiple regression models. Once the models were built they were tested with a range of "goodness of fit" diagnostics. Prior to modeling, Martin partitioned his data into two sets, one to build the model and one with which to test it. This process was designed to avoid over-fitting. His results broke individuals into two groups: high-risk and low-risk. Included in the high-risk category were overweight males, males with a history of problems with civil authorities, enlistees who signed up to "change their life," and high school drop outs (non-grads). Included in the low-risk category were minorities, females over 21 years of age, male college graduates, individuals with an AFQT over 65, and individuals who indicated they were interested in advanced education (Martin 1995).

Another study was a thesis by Murray (1985) which studied DEP attrition for the Navy. Murray employed several logistic regression models in an effort to predict DEP

attrition. She found that non-grads, individuals with high AFQT scores (above 65), individuals with long DEP stays (over 7 months), and individuals over 21 had a higher propensity to attrite (Murray 1985). Her findings contrast with those of Martin.

Matos conducted a third study in 1994. Matos analyzed the Navy's Delayed Entry Program and the effects of an individual's time spent in DEP. He employed a log-linear regression model, contingency tables and conditional probability theory to describe DEP attrition. His finding concluded that an increase in DEP length resulted in an increase in DEP attrition but decreases in fleet first-term attrition. He also found that non-grads had a higher propensity to attrite than high school graduates. (Matos 1994).

Bohn and Schmitz published an analysis in 1996 for CNRC. This study used OLS regression and logistic regression models to analyze DEP and RTC attrition factors. Bohn and Schmitz subdivided the recruit pool into two categories, work force and high school seniors. They assert that an individual who is recruited directly from the work force is different than an individual who signs up as a senior in high school. In the DEP analysis, they found that AFQT was inversely related to attrition, that seniors with dependents are more likely to attrite than those without, that Hispanics are more likely to attrite than non-Hispanics, that age is directly correlated with attrition, and that long DEP time leads to higher attrition among women. In the RTC analysis, Bohn and Schmitz found that non-grads have a higher likelihood to attrite, that AFQT scores are inversely related with attrition and that older individuals have a higher propensity to attrite. Bohn and Schmitz also formulated an optimization model for DEP duration, which minimizes DEP attrition and RTC attrition.

Finally, there is an analysis published by Quester, MacIlvaine and Barfield in 1997 for The Center for Naval Analysis (CNA). Their study used OLS regression and descriptive statistics to analyze RTC attrition. This analysis incorporated data from a new survey (known as the SHIP survey) which is being administered to all accessions at bootcamp. This survey added possible predictors such as smoking and exercise to the data set and subsequent analysis. The study reported that non-smokers, A-cell candidates, Asians, recruits with no enlistment waivers and recruits who accessed through the DEP (rather than shipping directly to RTC) were less likely to attrite than others. (Quester et. al 1997)

The review of previous studies shows some common threads in analysis and results. Most previous research has employed logistic regression and most previous research found that A-cell candidates, candidates with some DEP exposure, and minorities were less likely to attrite. The availability of the new SHIP data enabled Quester et al. to explore many other potential predictors with interesting results. The SHIP data (updated through Dec 1997) were available for this study.

## F. RESEARCH GOALS/ HYPOTHESES

Starting with previous research and using both CNRC personnel data and CNA SHIP data, this paper will try to further explain DEP and RTC attrition. The analysis will employ logistic (logit) regression techniques for comparison but will focus on classification tree methodology as a means to explain the attrition data. Specific hypotheses are that:

- Individuals who smoke and do not exercise have a higher propensity to attrite from RTC;
- A-Cell individuals have a lower propensity to attrite from both DEP and RTC;
- Individuals who sign up for the Navy College Fund or EB program have a lower propensity to attrite from DEP and RTC.

Given the above hypotheses, this analysis also has the following research goals:

- To identify, *post hoc*, other significant predictive factors (not found in previous research);
- To compare and contrast the logit regression and the classification tree methodologies for this type of data set;
- To address the policy implications of the resulting predictive model.

# II. METHODS

## A. DATA

Data for this analysis were provided by CNRC, Code 20, who merged data from CNA, RTC and CNRC databases. The data consisted of all individuals who were scheduled to report to RTC between October 1995 and December 1997. There were a total of 130,486 records in the data set, which was sorted by individual Social Security Number.

The data was imported to Microsoft Access for initial analysis and validation. Access was first used to search for null fields and bad data. Approximately 12,000 of the records had more than one null field; several of these had more than 4 null fields. To avoid potential problems with the analysis, the records with more than one null field were removed from the data set. There was concern that, in doing so, the data set would be compromised, so before assuming the null records were random occurrences, each column of the null set was plotted to check for uniformity and conformity with the remaining data set. For example, the number of null fields was plotted for each NRD to ensure that no single NRD or Area was consistently failing to input the data. Further, binomial probability hypothesis tests were used to compare categorical variables. This analysis identified several columns (variables) which were not complete; the data was not collected for DEP attrites. As a result, many of the variables available for RTC analysis were not available for DEP analysis. The variables available for DEP analysis are marked in Table 3 with a "*".

**Table 3: Data Descriptions**

| Variable | Description |
|---|---|
| SSN | Individual's Social Security Number |
| AGE* | Individual's Age in years (at time of Enlistment) |
| MALE* | Binary (0,1), 1 if Male |
| FEMALE* | Binary (0,1), 1 if Female |
| WHITE* | Binary (0,1),1 if White |
| BLACK* | Binary (0,1), 1 if Black |
| HISPANIC* | Binary (0,1), 1 if Hispanic |
| ASIAN* | Binary (0,1), 1 if Asian or Pacific Islander |
| NRD* | 3 Digit Code Representing Recruiting District of the Individual |
| SENIOR* | Binary (0,1), 1 if Individual was a Senior in High School upon Enlistment |
| PROGRAM-1* | Initial Rating- Assigned (String) |
| PROGRAM-2* | Final Rating- Assigned (String) |
| BONUS* | Binary (0,1), 1 if Individual Signed up for EB |
| NAVY COLLEGE FUND* | Binary (0,1), 1 if Individual Signed up for Navy College Fund |
| NON-GRAD* | Binary (0,1), 1 if Individual did not Graduate From High School |
| HIGH SCHOOL GRAD* | Binary (0,1), 1 if Individual Graduated from High School (NO-GED) |
| GED* | Binary (0,1), 1 if Individual Graduated with GED |
| DEP ATTRITION CODE | 3 Letter Code Assigned by CNRC to Categorize a DEP Attrite |
| RTC ATTRITION CODE | 3 Digit Code Representing RTC Attrition Category |
| DEP SCHEDULE* | Number of Days Individual was Scheduled for DEP |
| DEP DAYS | Number of Days Actually Spent in DEP |
| DEPENDENTS* | Number of Dependents |
| SHIPPING MONTH | Month Individual Shipped to RTC |
| ATTRITION MONTH | Month Individual Attrited from either DEP or RTC |
| CRIME WAIVER | Binary (0,1), 1 if a Waiver was Granted for Previous Criminal Behavior |
| DRUG WAIVER | Binary (0,1), 1 if a Waiver was Granted for Previous Drug Use |
| MEDICAL WAIVER | Binary (0,1), 1 if a Waiver was Granted for a Medical Condition |
| OTHER WAIVER | Binary (0,1), 1 if a Waiver was Granted for Any Other Reason |
| SMOKE | Binary (0,1),1 if Individual Indicated on SHIP Survey : Smoker |
| CHEW | Binary (0,1), 1 if Individual Indicated on Ship Survey: Used Smokeless Tobacco |
| RUNJOG | Binary (0,1), 1 if Individual Indicated on Ship Survey: Ran orJogged at least 3 Times a Week |
| DEP ATTRITE** | Binary (0,1), 1 if Individual Attrited from DEP |
| RTC ATTRITE** | Binary (0,1), 1 if Individual Attrited from RTC |
| JOBCHANGE | Binary (0,1), 1 if PROGRAM1=PROGRAM2 |

*Indicates variable was available for DEP analysis

**Indicates dependent variable

The search results and analyses of the variables indicated there was no reason to believe the null field occurrences were not random events (with respect to their

variables). Consequently, the individual's records (rows) with more than one null field were removed from the data set.

Once the data were removed, the remaining data were randomized. A column of random numbers, uniformly distributed between 0 and 1, was added to the set and Access sorted the data into two parts (random < .5 and random > .5). These partitions of the data set produced two sets containing approximately 60,000 records each. One set was used to build the models, the other was saved to test their predictive power. Once partitioned, the model building data were further subdivided to exclude DEP attrites from RTC analysis. The final data consisted of two partitioned data sets for DEP and RTC attrition analysis.

## B. LOGISTIC (LOGIT) REGRESSION

Previous research indicated that logistic, or logit, regression is a widely used technique for attrition analysis. As with other regression techniques, logit regression models a dependent variable by a linear combination of many independent variables. In attrition analysis, the dependent variable is categorical (i.e., whether or not a recruit attrites) and researchers are interested in the probability a person with a given set of characteristics will attrite. Since the outcome is a probability and bounded by zero and one, OLS regression is not suitable. Logit regression, however, will result in "predictive values which correspond to the probability of a positive (attrition) outcome" (Martin, 1995). The logistic model is defined by

$$\Pr [Y_i = 1 | X_i] = 1 / (1 + \exp [-(X_i^T \beta)])$$

where $Y_i$ is the dependent variable for recruit "i", DEP or RTC attrition, and $X_i$ represents the vector of independent variables (characteristics) of recruit "i" (male, GED, etc.). $\beta$ represents the vector of unknown regression coefficients for the model.

Using S-Plus (Mathsoft Inc., 1995), DEP and RTC data were modeled using logit regression. The first step was to build a model using all of the potential predictors (Table 3) and some possible interactions. The interactions are listed in Table 4.

**Table 4: Interactions**

| |
|---|
| BONUS & GED |
| NCF & GED |
| SMOKE & RUNJOG |
| CHEW & RUNJOG |
| CRIME WAIVERS & DRUG WAIVERS |

The interaction "BONUS & GED" was incorporated to look at possible motivation levels among GED entrants. "NCF & GED" was also incorporated to look at educational motivation among GED entrants. "SMOKE & RUNJOG" and "CHEW & RUNJOG" will examine whether the effect of tobacco use is different for runners than for non-runners. The waiver interaction is included to see if these two factors interact.

With all main effects and these interactions included, the full model was estimated and then the least significant variables were deleted (one at a time). The absolute t-values of the coefficients were computed and the coefficient corresponding to the smallest of these was deleted if its t-ratio was insignificant with $\alpha = .05$. The model was rebuilt and the process repeated until all coefficients had t-values which were significant with $\alpha = .05$. The goal was to build a statistically sound model with the fewest predictive variables.

Step-wise variable removal can produce questionable t-values in the resulting model so critical levels were adjusted using the Bonferroni inequality method. Ten variables were removed in the RTC model (among them were interactions) and four variables were removed in the DEP model. $\alpha$ was adjusted to 0.05/10 = .005 for the RTC model and to 0.05/4 = 0.0125 for the DEP model.

## C. CLASSIFICATION TREES

An alternative to logit regression is to use classification trees to describe the structure of the data. (Brieman et. al, 1984) Classification trees are similar to regression in that they model a dependent variable by the values of many independent variables. A classification tree is one where the dependent variable is categorical. Trees for continuous responses are referred to as regression trees. Fitting a tree model is a recursive procedure resulting in terminal nodes or "leaves" containing groups of cases with similar values in their independent variables and differences in the dependent variables, which reflect response probabilities.

The process begins with a parent node. This node has a "purity measure" with respect to the dependent variable. This purity measure is defined by S-Plus as deviance. The deviance formula follows:

$$\text{Deviance}_i = -2 * \sum_k (n_{ik} * \log (p_{ik}))$$

where "i" labels the node, "k" labels the classes in the node (here these are "attrite" or "no attrite"), "$n_{ik}$" represents the number of cases with class "k" in node "i" and "$p_{ik}$" is the multinomial probability associated with node "i" and class "k". The total deviance of the final tree is the sum of the leaf deviances. For each node, S-Plus looks at every

variable and every possible binary split within that variable and chooses the variable and split that brings about the maximum reduction in deviance at each stage, splitting the node into two children nodes. Each pair of child nodes has a combined deviance that is no larger than that of their parent. (Venables & Ripley, 1994)

In the attrition analysis, the initial parent node, or root, will contain all of the records in the data set. In the case of binary or categorical independent variables the splits are pre-determined by the variable (e.g., male or female in the binary case, WHITE or BLACK and ASIAN in the categorical case). In the case of continuous independent variables, the possible splits depend upon the data representation. For example, the age variable is tracked with a precision in tenths of a year; S-Plus will look at each possible split between tenths (e.g., if the data is 22.6 and 22.7 years, S-Plus will analyze the split between values, i.e., above 22.65 and below 22.65). When the program has found the best split (biggest reduction in deviance) for each variable, it will choose the best split across all variables. The procedure is repeated for each child. Figure 7 depicts a hypothetical example.

The tree algorithm often results in over-fitting the data, especially with large data sets. To compensate for this, S-Plus provides methods to reduce the size of the tree to an optimal predictive size. Cross-validation identifies the optimal-size tree and pruning enables the analyst to choose a tree size by selecting the number of terminal leaves.

Cross-validation repeatedly grows and prunes trees. The data is randomly split into ten sets or partitions. A sequence of trees (sizes 2,3,4…etc.) are grown with all but one of the data partitions; the remaining partition is used to test the predictive powers of the trees; the deviance of each tree is computed for the partition left out. The quality of

the tree is then evaluated for a range of possible sizes. This process is repeated for the other partitions and the minimum deviance, of the ten partitions, for each size tree can be

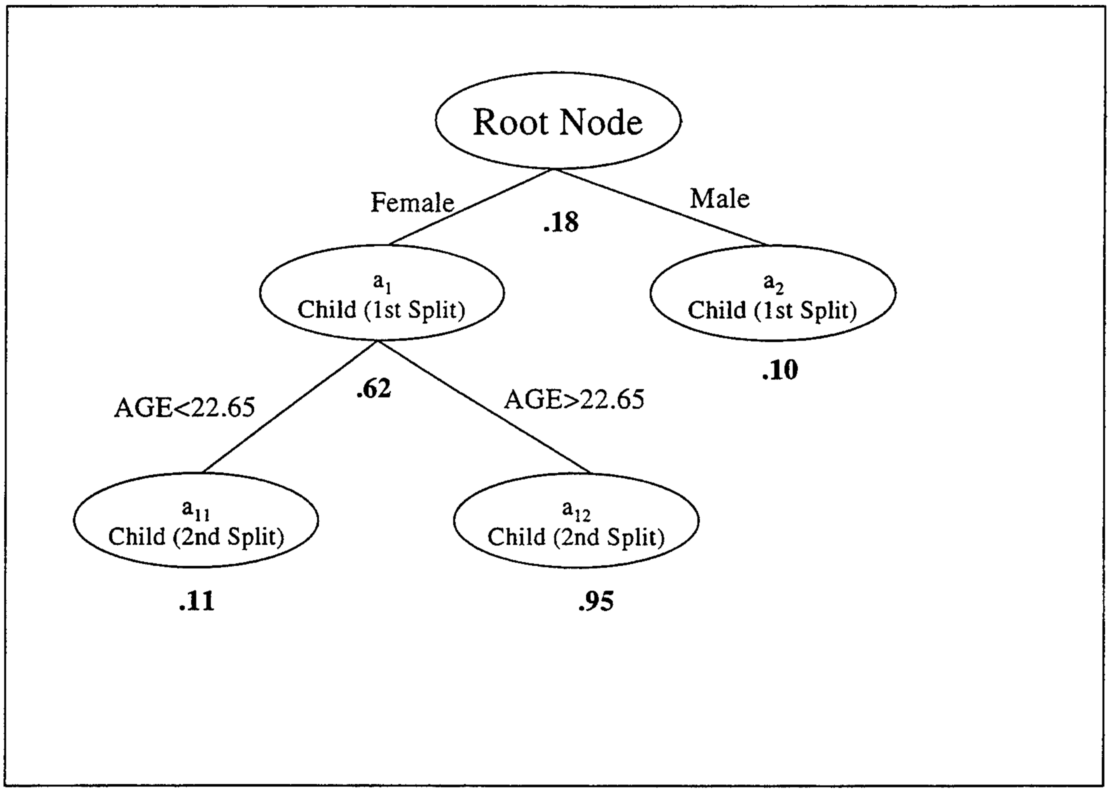


**Figure 7. Hypothetical Tree Example**

compared to the model size. The optimal size tree is determined by plotting model size versus minimum deviance and finding the minimum of these deviances.

Interpreting the results of the tree is accomplished by reading the probabilities in the terminal leaves. For example, Figure 7 would indicate that 11% of the women under 22.65 years of age would attrite (hypothetically). Ease of interpretation is a key benefit of tree-based models. Using the tree functions within S-Plus, a classification tree model was developed for the two attrition cases (DEP and RTC). Cross-validation was used to determine the optimal size and the trees were pruned accordingly.

## D. PREDICTION

Both the logit model and the classification tree result in probability estimates for attrition. To test the predictive power of the models, the data sets were randomly partitioned; half of the data was not used in the model development. For testing, these remaining partitions were run through the models and if the probability estimate was above a pre-determined decision threshold, the individual was scored as an attrite. For example, if the fitted value from the logistic regression model was .7 and the threshold was .69, the individual with a .7 probability of attrite was predicted to attrite. For the tree models fitted probabilities were obtained by using the "predict()" function built into S-Plus. The "predict()" function uses the model to derive the probability of both positive (attrition) or negative (non-attrition) responses for a given data set. In the attrition analysis, each record (row) was fitted with a predicted probability and this probability was compared to derived threshold. A record of predicted attrites was kept and compared to the actual data records.

Correct predictions fell in to two categories: attrites and non-attrites. Correct attrite predictions were those where the model first calculated a fitted value (probability); if the value was above the optimal threshold and the individual actually attrited, it was counted as a correct attrite prediction. Correct non-attrite predictions were those where the fitted value was below the threshold and the individual did not attrite. The sum of these two types of predictions was recorded. The final result was a number of correct predictions for each the model.

The decision threshold was developed using the fitted values from each model and the actual attrition values from the data used to build the models. A simple

optimization program was constructed using JAVA 1.1.4. The program read in the actual and fitted values for each record in the data set and walked through a preset number (200) of possible probability thresholds for the attrite decision. Several step sizes were tried and it was determined that a step size of 0.005 provided sufficient accuracy. A count of correct predictions was made for each threshold and the probability associated with the maximum number of correct predictions was identified for each model. The code for the program is listed in Appendix C. Figure 8 shows plots of threshold versus number of correct predictions for each model while Table 5 lists the optimal decision thresholds for each model.

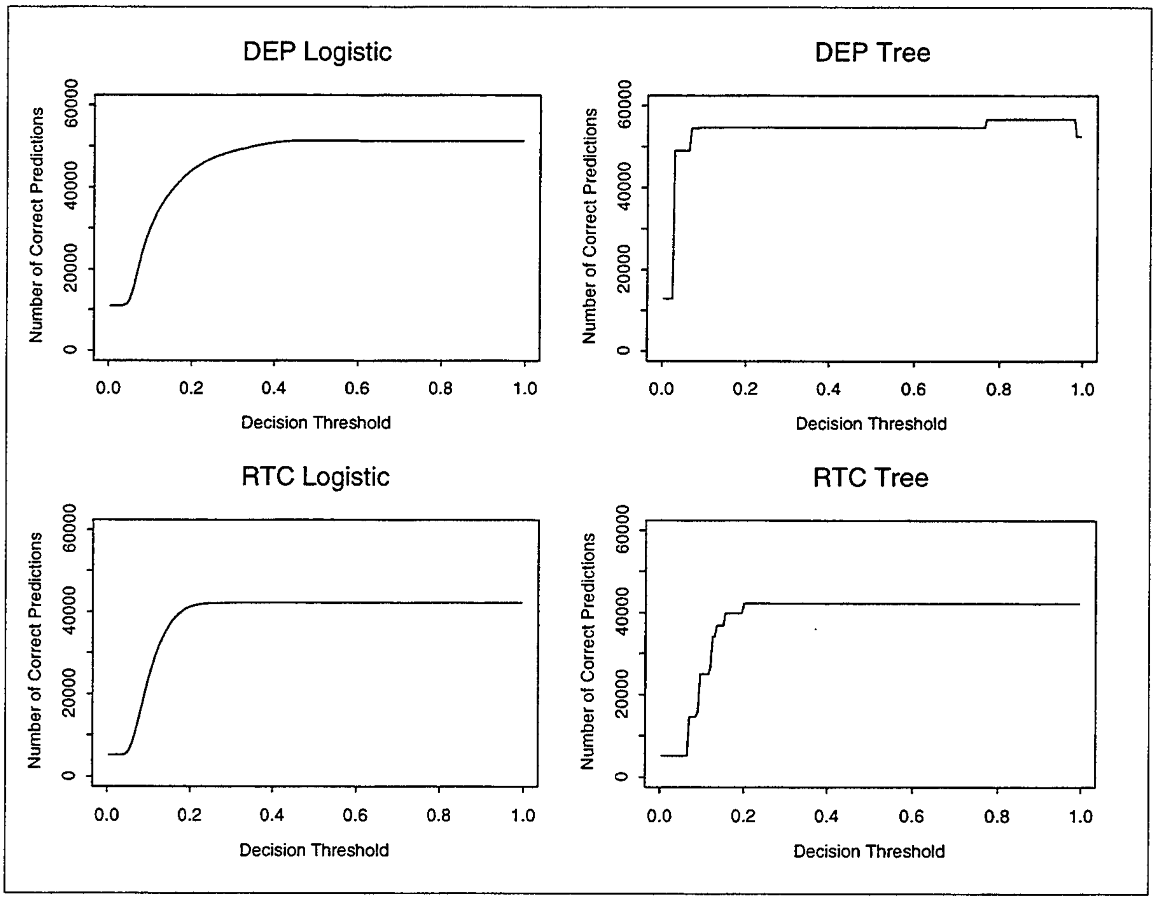


**Figure 8. Model Threshold Plots**

**Table 5. Optimal Thresholds**

| Model | Optimal Threshold |
|---|---|
| DEP Logistic | 0.54 |
| DEP Tree | 0.77 |
| RTC Logistic | 0.33 |
| RTC Tree | 0.2 |

# III. RESULTS

## A. DEP LOGISTIC MODEL

The logistic model for the Delayed Entry Program is summarized in Table 6. Model significance can be assessed by comparing the difference between the null deviance and residual deviance with a $\chi^2$ with eleven degrees of freedom (the number of parameters in the model). (Venables and Ripley, 1994) This approximation shows the model would be significant at very high confidence levels (58284.38 - 52397.63 = 5886.75; this is compared to a $\chi^2$ (11), which has an expected value of 11 and standard error of 3.31 ).

**Table 6. DEP Logistic Model Summary**

| Variable | Value | Std. Error | t value |
|---|---|---|---|
| (Intercept) | -3.391 | 0.112 | -30.409 |
| AFQT | -0.002 | 0.001 | -2.552 |
| AGE | 0.054 | 0.005 | 11.658 |
| MALE | -0.463 | 0.027 | -17.374 |
| WHITE | 0.176 | 0.029 | 6.114 |
| BLACK | 0.107 | 0.036 | 2.989 |
| SENIOR | -0.351 | 0.031 | -11.179 |
| BONUS | -0.259 | 0.042 | -6.126 |
| NCF | -0.197 | 0.034 | -5.825 |
| GED | 0.168 | 0.075 | 2.237 |
| SCHEDDEP | 0.008 | 0.00001 | 59.912 |
| JOBCHANGE | -0.287 | 0.033 | -8.787 |
| Null Deviance: 58284.38 on 62252 degrees of freedom<br>Residual Deviance: 52397.63 on 62241 degrees of freedom | | | |

The factors that significantly ($\alpha$ = .0125) increase the probability of attrition with an increase in their value are AGE, two races (WHITE and BLACK), Education Level (GED), and Time Scheduled for DEP (SCHEDDEP). The factors that significantly

decrease the probability of attrition with an increase in their value are AFQT score, sex (MALE), enlisting as a senior in high school (SENIOR), taking an enlistment bonus (BONUS ) and changes in future billet assignments (JOBCHANGE). All other variables listed in Table 3, and interactions from Table 4, were removed for insignificance.

## B. DEP TREE MODEL

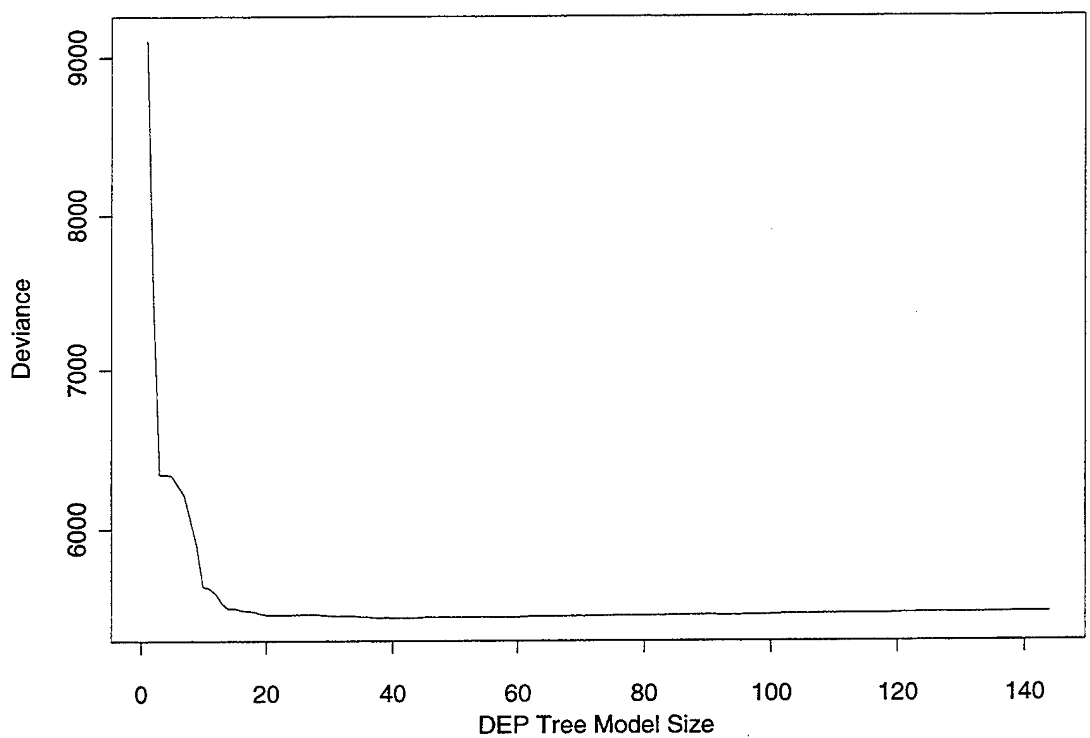


**Figure 9. DEP Tree Size vs Deviance**

Cross-validation identified an optimal tree with 52 terminal nodes. The large size of this tree made it very difficult to interpret as it had over ten levels of splits. As a result, the cross validation data (model size and deviance) was analyzed to see if there was an alternative size tree with similar deviance and predictive power. Figure 9 shows

the relationship between deviance and model size. The deviance (as depicted in Figure 9) is almost flat once model size grows above 20 terminal nodes. The actual difference in deviance between the tree with 20 nodes and the tree with 52 nodes is 5460 – 5440 = 20. The smaller tree, with 20 terminal nodes, was much easier to interpret (since it had only six levels). Further analysis indicated that the 20-node tree predicted with the same level of accuracy as the 52-node tree. The model selected and analyzed in this paper contained 20 terminal nodes (the code for all trees, including the 52-node tree, can be found in Appendix D).

The DEP tree model with 20 terminal nodes is depicted in Figure 10. The number inside each node is the attrition probability for all of the cases within the node. The root shows an attrition probability of 0.18. Rectangular nodes are terminal nodes and the number of cases in the node is listed beneath the node. The first split divides the cases into two sets (those with high school degrees and those without high school degrees). If the individual has no high school degree, is not a senior upon enlistment, and has an AFQT score below 49.5, he or she has a 0.76 probability of attrition (for scheduled DEP durations less than 121.5 days) or a 0.97 probability of attrition (for scheduled DEP durations above 121.5 days). If an individual has no high school degree, is a senior upon enlistment, but earns a GED or fails to graduate, he or she has an attrition probability of 0.98 . If an individual has a high school degree, is female, and is scheduled for DEP less than 75.5 days, she has a 0.06 attrition probability. Other specific cases can be evaluated by using Figure 10.

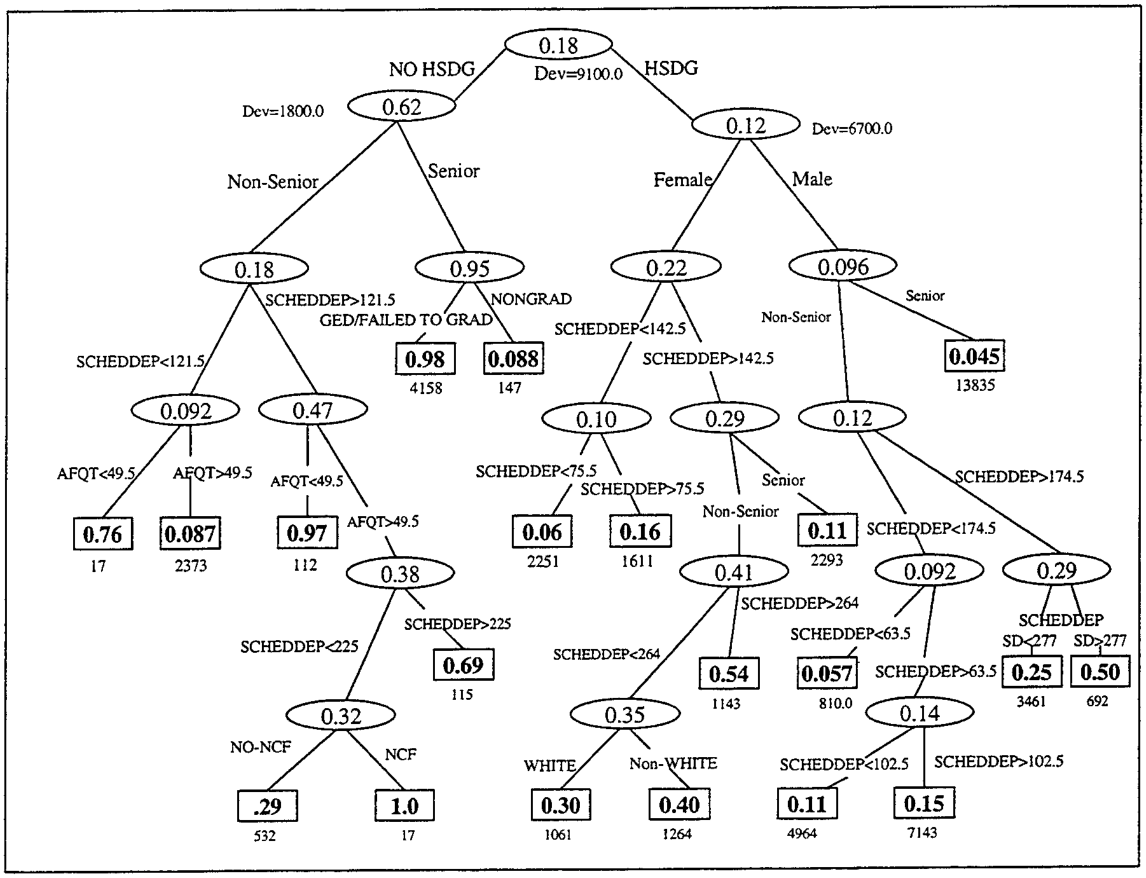


**Figure 10. DEP Tree Model**

## C. RTC LOGISTIC MODEL

The RTC logistic model is summarized in Table 7. The $\chi^2$ statistic (as discussed in section III.A) has 15 degrees of freedom and shows the model's significance (33135.88 - 32320.22 = 815.66; this is compared to a $\chi^2$ (15), which has an expected value of 15 and standard error of 3.87). The factors that increase the probability of attrition ($\alpha = .005$) with an increase in value are AGE, two races (WHITE and BLACK), two educational levels (NONGRAD and GED), time scheduled in DEP (SCHEDDEP), tobacco use (SMOKE only), waivers (CRIME and OTHER), changes in job, or

**Table 7. RTC Logistic Model Summary**

| Variable | Value | Std. Error | t value |
|---|---|---|---|
| (Intercept) | -2.151 | 0.131 | -16.391 |
| AFQT | -0.010 | 0.001 | -12.543 |
| AGE | 0.029 | 0.005 | 5.430 |
| WHITE | 0.211 | 0.040 | 5.313 |
| BLACK | 0.223 | 0.047 | 4.696 |
| NONGRAD | 0.316 | 0.093 | 3.417 |
| GED | 0.260 | 0.074 | 3.531 |
| SCHEDDEP | -0.001 | 0.0001 | -6.427 |
| CRIME | 0.183 | 0.048 | 3.821 |
| OTHER Waiver | 0.213 | 0.049 | 4.353 |
| SMOKE | 0.364 | 0.040 | 9.021 |
| CHEW | 0.166 | 0.081 | 2.061 |
| RUNJOG | -0.344 | 0.038 | -9.149 |
| JOBCHANGE | 0.116 | 0.043 | 2.718 |
| SMOKE:RUNJOG | 0.189 | 0.063 | 3.018 |
| CHEW:RUNJOG | -0.286 | 0.127 | -2.253 |

Null Deviance: 33135.88 on 47464 degrees of freedom
Residual Deviance: 32320.22 on 47449 degrees of freedom

program, classification (JOBCHANGE), and the interaction between SMOKE and RUNJOG. The factors which decrease the probability of attrition with an increase in their value are AFQT score, time scheduled to be in DEP (SCHEDDEP), RUNJOG, and the interaction between CHEW and RUNJOG. Other variables and interactions were removed from the model for insignificance.

## D. RTC TREE MODEL

The RTC tree model is depicted in Figure 11. Cross-validation identified an optimal tree with nine terminal nodes and this size was used for the ensuing analysis. The root node indicates an overall probability of attrition of 0.11. The first split divides the cases into two sets: smokers and non-smokers. Smokers with AFQT scores below

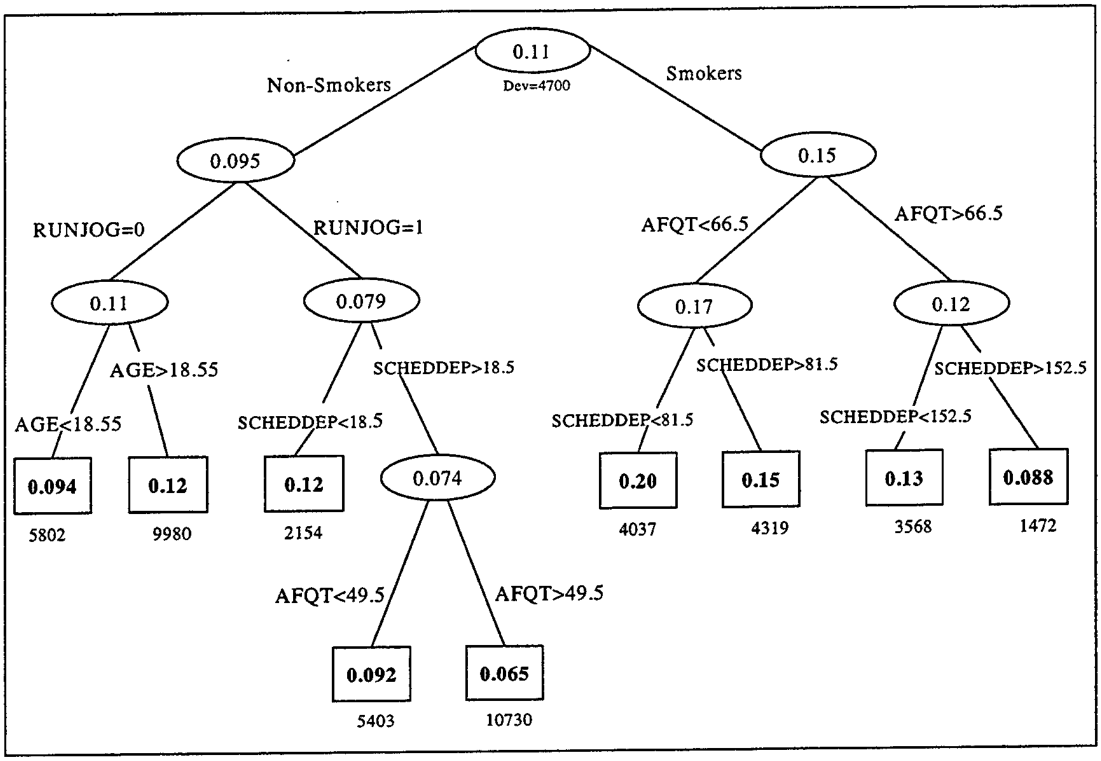


**Figure 11. RTC Tree Model**

66.5 who were scheduled for DEP for less than 81.5 days have the highest probability of attrition (0.20). Non-smokers who run or jog more than three times a week, with AFQT scores above 49.5 and a scheduled DEP duration greater than 18.5 days, have the lowest probability of attrition (0.065). Other specific cases can be evaluated using Figure 11.

## E. PREDICTION

The prediction methodology, discussed in section II.D, established the optimal decision thresholds depicted in Table 5. Using the data which was held out (i.e., not used to build the models), the thresholds depicted in Table 5, and the "predict ()" function within S-Plus, fitted values for each model were obtained. As discussed earlier, the fitted values (probabilities) were compared to the appropriate threshold and scored accordingly.

A summary of the prediction results is depicted in Table 8. The net gain column represents the difference between using the model and not using the model. "Not using the model" means that all of the predictions are "no attrite;" the values in that column reflects actual attrition. For example, the DEP logistic model predicts 49229 of the 61947 accessions correctly. If the model is not used, CNRC predicts every individual to complete (e.g., all 61947) and is correct for 49158 of the accessions; the remaining 12789 attrite. For the DEP logistic model, using the model resulted in 71 more correct predictions than not using the model. For the DEP tree, the model had 3954 more correct predictions. Neither RTC model had any impact on the predictive outcome (i.e., no additional correct predictions were made).

**Table 8. Model Prediction Summary**

| Model | Threshold | Number In Test Set | Actual Number of Attrites | Number of Correct Predictions | With Out Model | Net Gain With Model |
|---|---|---|---|---|---|---|
| **DEP Logistic** | 0.54 | 61947 | 12789 | 49229 | 49158 | 71 |
| **Dep Tree** | 0.77 | 61947 | 12789 | 53112 | 49158 | 3954 |
| **RTC Logistic** | 0.33 | 47465 | 5279 | 42186 | 42186 | 0 |
| **RTC Tree** | 0.2 | 47465 | 5279 | 42186 | 42186 | 0 |

Further analysis of the predictions indicated that the improvements realized by both of the DEP models reflected correct attrite predictions. While the 71 DEP logistic attrite predictions can not be attributed to a specific set of characteristics, the 3954 DEP tree attrite predictions fall exclusively in a single node of the tree. This node is defined by individuals with no high school degree, who are seniors upon enlistment, but fail to graduate (or they get a GED) from high school. This node has an attrition probability of 0.98. There were a total of 4158 cases in this node.

# IV. CONCLUSIONS

## A. ADDRESSING THE HYPOTHESES

It was hypothesized in section I.F. that individuals who do smoke and do not exercise have a higher probability to attrite from RTC. The RTC logistic model indicates that smoking increases the probability of attrition. The RTC logistic model also shows that exercise (represented by RUNJOG) decreases the probability of attrition (coefficient=-0.34). The RTC logistic model indicates that individuals who smoke and do not exercise have a higher propensity to attrite than those who exercise and do not smoke. The RTC tree model confirms the increased probability of attrition for smokers, but never splits on RUNJOG for smokers.

The second hypothesis asserted that A-cell individuals have a lower propensity to attrite (from both RTC and DEP) than B-cell, C-cell or D-cell individuals. The DEP logistic model confirms the inverse relationship between AFQT score and DEP attrition as well as the direct relationship between GED and DEP attrition, but it does not find NONGRAD to be significant. The results of the DEP model are inconclusive with respect to A-cells. The RTC logistic model clearly supports the hypothesis. The DEP tree model consistently shows that AFQT and DEP attrition are inversely related and also shows that not graduating from high school will result in a higher probability of attrition. The RTC tree model illustrates the inverse AFQT relationship but is inconclusive because it never splits on educational level.

The third hypothesis states that individuals who take an incentive package are less likely to attrite from DEP or RTC than those who do not take an incentive package. The

DEP logistic model supports the hypothesis for both the NCF and EB. The DEP tree model splits once on NCF (those with no high school degree, who are not seniors upon enlistment, who have a scheduled DEP duration between 121.5 and 225 days, and who have AFQT scores above 49.5) and this directly contradicts the hypothesis, but the relevant terminal node contains only 17 cases. The RTC logistic model does not find either variable significant and the RTC tree model never splits on incentive packages. With the exception of the DEP logistic model the results are inconclusive.

## B. ADDRESSING THE RESEARCH QUESTIONS

### 1. Other Predictive Factors

There were several predictive factors identified by the models not addressed in the previous section. First is the DEP tree node with the 0.98 DEP attrition probability and the exceptional predictive power (the node predicted 3954 attrites correctly). The node can be summarized as individuals who enlist as seniors in high school but fail to graduate (they may get a GED) from high school.

Research, into this node, indicates that these attrites actually fall into two categories: "Fail to Grad" and "Fail to Obligate" (see appendix A). "Fail to Grad" categorizes individuals who are disqualified by the Navy for failing to graduate from high school. As discussed in section I.C. the Navy has a cap on NON-GRADS of 5% which is more stringent than the congressional mandate of 10%. A recruiter will realize that an individual is not going to graduate from high school, in summer, which falls in the fourth quarter of the fiscal year. At that time, the quota for NON-GRADS will usually be full (or close to full) and these individuals will be lost.

The second group, and vast majority of these individuals (98%), are actually individuals who "Fail to Obligate" or quit from DEP. When individuals enter the DEP as high school seniors, they are given an educational code of "P" (probable graduate). When the individual drops out of the DEP before high school graduation (assuming he signs up as a senior in high school), the individual's education code is not updated and the tree classification does not reflect the actual educational level. There is no predictive power in this node.

The DEP logistic model and the DEP tree model both show that females have a higher propensity to attrite from DEP than males, but this is not the case in RTC. Variables such as SCHEDDEP and AFQT simply confirm previous research for both DEP and RTC attrition. Finally, all of the models indicate that AGE is directly related to attrition.

**2. Comparing Methodologies**

Comparing and contrasting the two methodologies is the second research question. Given the categorical nature of the data and binary response, both classification trees and logistic regression were well-suited for the problem. The models produced consistent probabilistic outcomes that center on the actual attrition rates. While the RTC models were similar, the DEP tree model did reveal structure within the data, which was not discernible from the logit model. While both provide insight into attrition, neither method was able to explain the phenomenon fully. The true structure of the phenomenon may not be discernible from the current data sets and it is recommended that CNRC and CNA continue to collect new data (similar to the SHIP survey) in hopes that models with better predictive power can be developed.

### 3. Policy Implications

Policy decisions related to this type of study often include adopting new screening procedures which exclude individuals with certain sets of characteristics. All of the models identified variables with statistical significance or structure and these both provide insight into attrition probabilities. Without exception, the models predicted poorly. Despite strong statistical significance, these results should not be used to predict an individual's outcome and possibly exclude him or her from service. These results do, however, improve our understanding of group behavior, which can be beneficial in aggregate forecasting, simulation and decision modeling.

# APPENDIX A. DEP ATTRITION CODES

This appendix contains a table with all of the actual DEP attrition codes and the corresponding sub-categories referred to in Figure 4.

| Attrition Code | DEP ATTRITION REASON | Figure 5 Category |
|---|---|---|
| RAA | Below 3.0 Grade Reading lvl | Academic |
| RAB | Reading Training failure | Academic |
| RAC | Test Failure (Art Graduate) | Academic |
| RAD | Test Failure (Reader) | Academic |
| RMA | Non-Adaptability | Personality |
| RMB | Lack of Motivation | Personality |
| RMC | Functionally Inadequate | Academic |
| RMD | Non-Swimmer | Medical |
| RPA | Situational Reactions | Personality |
| RPB | Personality Disorder | Personality |
| RUD | Non-Military | Personality |
| RCA | Drugs-Prior Service | Screen |
| RCB | Homosexual-Prior Service | Screen |
| RCC | Arrest Record-Prior Service | Screen |
| RCD | Previous Service | Screen |
| RDA | Orthopedics | Medical |
| RDB | Podiatry | Medical |
| RDC | General Surgery | Medical |
| RDD | Urology | Medical |
| RDE | Ophtalmology | Medical |
| RDF | Neurology | Medical |
| RDG | Dermatology | Medical |
| RDH | Internal Medecine | Medical |
| RDI | ENT | Medical |
| RDJ | Psychiatry | Medical |
| RDK | Other-Medical | Medical |
| RGA | Erroneous Enlistment | Screen |
| RGB | Minority | Technicality |
| RGC | Death | Technicality |
| RGD | Pregnancy | Technicality |

| Attrition Code | DEP ATTRITION REASON | Figure 5 Category |
|---|---|---|
| RGE | Enuresis | Technicality |
| RGF | Failed to Graduate | Technicality |
| ROA | Rescind Recommendation | Admin |
| ROB | Refuses to Obligate | Failed to Obligate |
| ROC | Member Reached E-4 | Screen |
| ROD | Earlier Class Date Not Available | Admin |
| ROE | Schedule Precludes Attendance | Admin |
| ROF | Desires TAD vice PCS Orders | Admin |
| ROG | Desires PCS vice TAD Orders | Admin |
| ROH | No Longer Desires to Convert | Admin |
| ROI | Change in Shipping date | Admin |
| ROJ | Change in Occ Spec/Rating Selected | Admin |
| ROK | Change in Program Selected | Admin |
| ROL | Change in Fleet Assignment | Admin |
| ROM | Change in Term of Enlistment | Admin |
| RON | Declined Enlistment | Failed to Obligate |
| RXA | Miscellaneous | Other |
| RCE | MEPS Drug Positive | Drug/Alcohol |
| RCF | MEPS Alcohol Positive | Drug/Alcohol |
| RCG | MEPS Drug and Alcohol Positive | Drug/Alcohol |

# APPENDIX B. RTC ATTRITION CODES

This appendix contains a table with all of the actual RTC attrition codes and the corresponding sub-categories referred to in Figure 5.

| Attrition Code | Academic/ Non-Academic | RTC ATTRITION REASON | *In Service (I) / *Prior to Svc (P) | Figure 6 Category |
|---|---|---|---|---|
| 84 | Academic | Lack of language proficiency | I | Academic |
| 135 | Non-Academic | Motivational drop on request | I | Admin |
| 149 | Non-Academic | Admin Hardship | I | Admin |
| 158 | Non-Academic | Pregnancy | I | Technicality |
| 167 | Non-Academic | Medical-Orthopedic | I | Medical |
| 168 | Non-Academic | Medical-Orthopedic | P | Medical |
| 169 | Non-Academic | Medical-Podiatry | I | Medical |
| 170 | Non-Academic | Medical-Podiatry | P | Medical |
| 172 | Non-Academic | General Surgery | P | Medical |
| 173 | Non-Academic | Medical-Urology | I | Medical |
| 174 | Non-Academic | Meducal-Urology | P | Medical |
| 175 | Non-Academic | Medical-Opthalmology | I | Medical |
| 176 | Non-Academic | Medical-Opthalmology | P | Medical |
| 177 | Non-Academic | Meducal-Neurology | I | Medical |
| 178 | Non-Academic | Meducal-Neurology | P | Medical |
| 180 | Non-Academic | Medical-Dermatology | P | Medical |
| 181 | Non-Academic | Medical-Internal | I | Medical |
| 182 | Non-Academic | Medical-Internal | P | Medical |
| 183 | Non-Academic | Medical-ENT | I | Medical |
| 184 | Non-Academic | Medical-ENT | P | Medical |
| 185 | Non-Academic | Medical-Gynecology | I | Medical |
| 186 | Non-Academic | Medical-Gynecology | P | Medical |
| 188 | Non-Academic | Psych-Suicidal Behavior | P | Behavior |
| 189 | Non-Academic | Psych-Suicidal Behavior | I | Behavior |
| 190 | Non-Academic | Psych-Excel Suicidal Behavior | I | Behavior |
| 191 | Non-Academic | Psych-Excel Suicidal Behavior | P | Behavior |
| 192 | Non-Academic | Psych-Personality Disorder | I | Behavior |
| 193 | Non-Academic | Psych-Enuresis | I | Behavior |
| 194 | Non-Academic | Psych-Sleepwalk | I | Behavior |

| Attrition Code | Academic/ Non-Academic | RTC ATTRITION REASON | *In Service (I) / *Prior to Svc (P) | Figure 6 Category |
|---|---|---|---|---|
| 195 | Non-Academic | Psych-Suicidal Situational Reaction | I | Behavior |
| 197 | Non-Academic | Medical-Not Aquatically Adaptable | I | Academic |
| 199 | Non-Academic | Legal-Civilian Conviction | I | Legal |
| 200 | Non-Academic | Legal-Deserter | I | Legal |
| 202 | Non-Academic | Legal-Breach of Contract | I | Legal |
| 203 | Non-Academic | Legal-Misconduct | I | Legal |
| 205 | Non-Academic | Legal-Homosexual | I | Legal |
| 206 | Non-Academic | Legal-Drugs | I | Legal |
| 207 | Non-Academic | Non-Training Related Death | I | Technicality |
| 208 | Non-Academic | Training Related Death | I | Technicality |
| 209 | Non-Academic | Suicide | I | Technicality |
| 212 | Non-Academic | PRT Failure | I | Admin |
| 215 | Non-Academic | Erroneous Enlistment | P | Admin |
| 216 | Non-Academic | Erroneous Enlistment-Best | P | Admin |
| 217 | Non-Academic | Erroneous Enlistment-Nav/Af | P | Admin |
| 217 | Non-Academic | Erroneous Enlistment-Motivation | I | Admin |
| 218 | Non-Academic | Under Age Enlistment | P | Screen |
| 220 | Non-Academic | Drug Screen-Non-CNBS | P | Screen |
| 221 | Non-Academic | Drug Screen-CNBS | P | Screen |
| 222 | Non-Academic | Drug Screen | P | Screen |
| 223 | Non-Academic | Homosexual | P | Screen |
| 224 | Non-Academic | Arrest Record | P | Screen |
| 226 | Non-Academic | Undisclosed Military Service | P | Screen |
| 311 | Non-Academic | Other | P | Screen |
| 320 | Non-Academic | Negative Military Attitude | I | Admin |
| 366 | Non-Academic | Medical-Other | I | Medical |
| 367 | Non-Academic | Medical-Other | P | Medical |
| 368 | Non-Academic | Not Adaptable to Military Life | I | Admin |
| 625 | Non-Academic | Drug Dependancy | I/P | Screen |

# APPENDIX C. JAVA SOURCE CODE

This appendix contains the JAVA 1.1.4 source code for the optimal threshold model. The model was run using step sizes of .0001, .0005, .001 and .005. There was no difference in threshold value and performance using the larger step size. The code follows, comments are preceded by "//"and are in bold print:

```
// java import classes for input and output methods

import java.util.*;
import java.io.*;

public class FindBest{

//instance variable and array declarations

//array to store predicted (fitted) values
   private double[] pred;

//array to store actual attrite (0/1)
   private double[] actual;

//array to store the number of correct predictions for each
//step
   private double[] corrPred;

//array to store the corresponding threshold for the number
of //correct predictions in corrpred
   private double[] probs;


   private int observations;

// counting variables
   private double difference;
   private double countActual;
   private double countCorrect;
   private double countWrong;
   private double bigCount;

//alias probability variable
   private double cutProb;
```

```
//input object variables to read  files

   private BufferedReader inStream;
   private String fileName;


//constructor
public FindBest(String file, int obs,double step){

// variable initialization

   fileName=file;
   bigCount=0;
   observations=obs;
   pred=new double[observations];
   actual=new double[observations];
   corrPred=new double[200];
   probs=new double[200];

//since fitted and actual values are bounded by zero and
//one the arrays are set to 2 to trigger errors
   for (int i=0;i<obs;i++){
     pred[i]=2;
     actual[i]=2;
   }
   countActual=0;
   countCorrect=0;
   countWrong=0;
   cutProb=0;
   difference=0;
   int j=0;
   inStream=null;

 // this reads the file and fills up the arrays
   try{
      inStream=new BufferedReader(new FileReader(file));
   }
   catch (IOException e){
     e.printStackTrace();

   }
   try{
     String line1=inStream.readLine();
      for(String line=inStream.readLine();line!=null;
```

```
          line=inStream.readLine()){
        bigCount++;
        StringTokenizer s=new StringTokenizer(line);
        while (s.hasMoreTokens()){
          pred[j]=Double.valueOf(s.nextToken()).doubleValue
          ();
           actual[j]=Double.valueOf(s.nextToken()).
           doubleValue();
        }
        j++;
      }
    }
    catch(NumberFormatException e){

    }
     catch(IOException e){
        e.printStackTrace();
     }
    }
```

```
// the following method steps through the arrays and does a
// correct prediction count for the given threshold, num is
// the index for the threshold, the threshold and count are
// then stored in a different array.

    private void doCount(double threshold,int num){
       double d2=0;
       double sum=0;
       double countRight=0;
       double countTot=0;
       double setProb=0;
       double countWrong=0;
       double countStay=0;
       double countStayWrong=0;
       double countStayTot=0;
       for(int i=0;i<observations-5;i++){
         if ((pred[i]>threshold) && (actual[i]==1)){
              countRight++;
         }
         if (actual[i]==1){
           countTot++;
         }
         if ((pred[i]>threshold) && (actual[i]==0)){
              countWrong++;
         }
         if((pred[i]<=threshold)&&(actual[i]==0)){
```

```
          countStay++;
        }
        if((pred[i]<=threshold)&&(actual[i]==1)){
          countStayWrong++;
        }
        if(actual[i]==0){
           countStayTot++;
         }
     }
      difference=countRight-countWrong;
      d2=countStay-countStayWrong;
      sum=countRight+countStay;
      corrPred[num]=sum;
      probs[num]=threshold;
      System.out.println("Sum/Prob/Tot=,"+sum+
     ","+threshold+","+countTot+","+countStayTot+","+fileNa
     me            );
   }
// the following method steps through the array of correct
// predictions and finds the threshold with the highest
// number of correct predictions and prints it out

   private void findOptimum(double s){
     double temp=0;
     int count=0;
     for(int i=0;i<(200-1);i++){
        if(corrPred[i]>temp){
            temp=corrPred[i];
          count=i;
        }
     }
     System.out.println(fileName+","+"Correct Predictions =
     "+temp+" , "+"  Threshold = "+probs[count]+","+
     bigCount);
   }

// the main method implements the program

   public static void main(String[]args){

//files to be read
     String file1="G:/depcut.txt";
     String file2="G:/dtreecut.txt";
     String file3="G:/rtccut.txt";
     String file4="G:/rtreecut.txt";
```

```
// the number of records in the appropriate data sets
      int ob1=61947;
      int ob2=61947;
      int ob3=47465;
      int ob4=47465;

// step size

      double step=.005;

// create an object for each model

      FindBest depLogCut=new FindBest(file1,ob1,step);
      FindBest depTreCut=new FindBest(file2,ob2,step);
      FindBest rtcLogCut=new FindBest(file3,ob3,step);
      FindBest rtcTreCut=new FindBest(file4,ob4,step);

// loop through the steps from zero to one for each model
      int j=0;
      for(double i=0;i<1;i+=step){

        depLogCut.doCount(i,j);
        j++;
      }
      j=0;
      for(double i=0;i<1;i+=step){
        depTreCut.doCount(i,j);
        j++;
      }
      j=0;
      for(double i=0;i<1;i+=step){
        rtcLogCut.doCount(i,j);
        j++;
      }
      j=0;
      for(double i=0;i<1;i+=step){
        rtcTreCut.doCount(i,j);
        j++;
      }
// find the best threshold for each model
      depLogCut.findOptimum(step);
      depTreCut.findOptimum(step);
      rtcLogCut.findOptimum(step);
      rtcTreCut.findOptimum(step);
   }
```

# APPENDIX D. TREE MODEL OUTPUT

This appendix contains the actual S-Plus version 3.3 tree outputs for the DEP trees with 20 and 52 terminal nodes and the RTC tree with nine terminal nodes. Each row contains the node split, the number of cases in the node, the deviance at the node, and the "probability of attrite" at that node. A "*" denotes a terminal node.

## 1. DEP TREE, 20 TERMINAL NODES

```
split, n, deviance, yval
     * denotes terminal node

 1) root 62253 9103.000 0.17790
   2) HSDG<0.5 7471 1752.000 0.62440
     4) SENIOR<0.5 3166  474.400 0.18350
       8) SCHEDDEP<121.5 2390  198.900 0.09163
        16) AFQT<49.5 17    3.059 0.76470 *
        17) AFQT>49.5 2373  188.100 0.08681 *
       9) SCHEDDEP>121.5 776  193.100 0.46650
        18) AFQT<49.5 112    1.964 0.98210 *
        19) AFQT>49.5 664  156.400 0.37950
          38) SCHEDDEP<225 549  118.500 0.31510
            76) NCF<0.5 532  110.300 0.29320 *
            77) NCF>0.5 17    0.000 1.00000 *
          39) SCHEDDEP>225 115   24.730 0.68700 *
     5) SENIOR>0.5 4305  209.700 0.94870
      10) NONGRAD<0.5 4158   85.180 0.97910 *
      11) NONGRAD>0.5 147   11.850 0.08844 *
   3) HSDG>0.5 54782 5658.000 0.11700
     6) MALE<0.5 9623 1634.000 0.21680
      12) SCHEDDEP<142.5 3862  354.600 0.10230
        24) SCHEDDEP<75.5 2251  127.800 0.06042 *
        25) SCHEDDEP>75.5 1611  217.400 0.16080 *
      13) SCHEDDEP>142.5 5761 1195.000 0.29350
        26) SENIOR<0.5 3468  841.600 0.41440
          52) SCHEDDEP<264.5 2325  531.700 0.35400
           104) WHITE<0.5 1061  223.900 0.30250 *
           105) WHITE>0.5 1264  302.600 0.39720 *
          53) SCHEDDEP>264.5 1143  284.200 0.53720 *
        27) SENIOR>0.5 2293  225.900 0.11080 *
     7) MALE>0.5 45159 3908.000 0.09568
      14) SENIOR<0.5 31324 3266.000 0.11820
        28) SCHEDDEP<174.5 27171 2267.000 0.09186
```

```
        56) SCHEDDEP<63.5 15064  811.800 0.05716 *
        57) SCHEDDEP>63.5 12107 1414.000 0.13500
         114) SCHEDDEP<102.5 4964  471.800 0.10640 *
         115) SCHEDDEP>102.5 7143  935.400 0.15500 *
      29) SCHEDDEP>174.5 4153  856.600 0.29090
        58) SCHEDDEP<277.5 3461  648.300 0.24960 *
        59) SCHEDDEP>277.5 692  173.000 0.49710 *
    15) SENIOR>0.5 13835  589.500 0.04460 *
```

## 2. DEP TREE, 52 TERMINAL NODES

```
1) root 62253 9103.000 0.17790
  2) HSDG<0.5 7471 1752.000 0.62440
    4) SENIOR<0.5 3166  474.400 0.18350
      8) SCHEDDEP<121.5 2390  198.900 0.09163
       16) AFQT<49.5 17    3.059 0.76470 *
       17) AFQT>49.5 2373  188.100 0.08681
         34) SCHEDDEP<52.5 1865  114.900 0.06595 *
         35) SCHEDDEP>52.5 508   69.440 0.16340 *
      9) SCHEDDEP>121.5 776  193.100 0.46650
       18) AFQT<49.5 112    1.964 0.98210 *
       19) AFQT>49.5 664  156.400 0.37950
         38) SCHEDDEP<225 549  118.500 0.31510
           76) NCF<0.5 532  110.300 0.29320
            152) GED<0.5 238   55.720 0.37390 *
            153) GED>0.5 294   51.730 0.22790 *
           77) NCF>0.5 17    0.000 1.00000 *
         39) SCHEDDEP>225 115   24.730 0.68700 *
    5) SENIOR>0.5 4305  209.700 0.94870
     10) NONGRAD<0.5 4158   85.180 0.97910 *
     11) NONGRAD>0.5 147   11.850 0.08844 *
  3) HSDG>0.5 54782 5658.000 0.11700
    6) MALE<0.5 9623 1634.000 0.21680
     12) SCHEDDEP<142.5 3862  354.600 0.10230
       24) SCHEDDEP<75.5 2251  127.800 0.06042 *
       25) SCHEDDEP>75.5 1611  217.400 0.16080
         50) SENIOR<0.5 1414  206.400 0.17750
          100) JOBCHANGE<0.5 1127  180.100 0.19960
            200) WHITE<0.5 538   73.610 0.16360 *
            201) WHITE>0.5 589  105.100 0.23260
              402) SCHEDDEP<101.5 206   27.710 0.16020 *
              403) SCHEDDEP>101.5 383   75.760 0.27150 *
          101) JOBCHANGE>0.5 287   23.640 0.09059 *
         51) SENIOR>0.5 197    7.675 0.04061 *
     13) SCHEDDEP>142.5 5761 1195.000 0.29350
```

```
      26) SENIOR<0.5 3468  841.600 0.41440
        52) SCHEDDEP<264.5 2325  531.700 0.35400
         104) WHITE<0.5 1061  223.900 0.30250
           208) JOBCHANGE<0.5 910  199.400 0.32420
             416) SCHEDDEP<212 548  110.700 0.28100 *
             417) SCHEDDEP>212 362   86.080 0.38950 *
           209) JOBCHANGE>0.5 151   21.520 0.17220 *
         105) WHITE>0.5 1264  302.600 0.39720
           210) JOBCHANGE<0.5 1066  258.600 0.41370
             420) SCHEDDEP<148.5 57   11.050 0.26320 *
             421) SCHEDDEP>148.5 1009  246.100 0.42220
               842) AGE<18.85 290   66.700 0.35860 *
               843) AGE>18.85 719  177.800 0.44780
                1686) AFQT<77.5 520  129.600 0.47310
              3372) SCHEDDEP<251.5 476  118.900 0.48950
           6744) SCHEDDEP<223.5 350   86.670 0.45140 *
            6745) SCHEDDEP>223.5 126   30.360 0.59520 *
             3373) SCHEDDEP>251.5 44    9.159 0.29550 *
                1687) AFQT>77.5 199   46.970 0.38190 *
           211) JOBCHANGE>0.5 198   42.210 0.30810 *
        53) SCHEDDEP>264.5 1143  284.200 0.53720
         106) SCHEDDEP<318.5 615  153.600 0.48460 *
         107) SCHEDDEP>318.5 528  126.900 0.59850 *
      27) SENIOR>0.5 2293  225.900 0.11080 *
   7) MALE>0.5 45159 3908.000 0.09568
    14) SENIOR<0.5 31324 3266.000 0.11820
      28) SCHEDDEP<174.5 27171 2267.000 0.09186
        56) SCHEDDEP<63.5 15064  811.800 0.05716
         112) SCHEDDEP<6.5 1454   32.250 0.02270 *
         113) SCHEDDEP>6.5 13610  777.600 0.06084
           226) AGE<23.15 11287  596.600 0.05599 *
           227) AGE>23.15 2323  179.500 0.08437 *
        57) SCHEDDEP>63.5 12107 1414.000 0.13500
         114) SCHEDDEP<102.5 4964  471.800 0.10640 *
         115) SCHEDDEP>102.5 7143  935.400 0.15500
           230) AGE<18.35 904   79.430 0.09735 *
           231) AGE>18.35 6239  852.600 0.16330
             462) AGE<25.25 5771  767.200 0.15790
               924) HISP<0.5 4902  678.200 0.16590
                1848) AFQT<35.5 512   49.870 0.10940 *
                1849) AFQT>35.5 4390  626.500 0.17240 *
               925) HISP>0.5 869   86.950 0.11280 *
             463) AGE>25.25 468   83.080 0.23080 *
      29) SCHEDDEP>174.5 4153  856.600 0.29090
        58) SCHEDDEP<277.5 3461  648.300 0.24960
         116) BLACK<0.5 2977  533.800 0.23410
```

```
          232) AGE<22.65 2436  412.400 0.21590
            464) SCHEDDEP<224.5 1756  272.300 0.19190
              928) WHITE<0.5 609   75.990 0.14610 *
              929) WHITE>0.5 1147  194.400 0.21620 *
            465) SCHEDDEP>224.5 680  136.500 0.27790
              930) AFQT<85.5 578  123.600 0.30970 *
              931) AFQT>85.5 102    9.020 0.09804 *
          233) AGE>22.65 541  117.000 0.31610
            466) SCHEDDEP<264.5 520  110.000 0.30380 *
            467) SCHEDDEP>264.5 21    4.952 0.61900 *
        117) BLACK>0.5 484  109.400 0.34500
          234) AGE<18.15 36    3.556 0.11110 *
          235) AGE>18.15 448  103.700 0.36380 *
       59) SCHEDDEP>277.5 692  173.000 0.49710
        118) SCHEDDEP<312.5 303   73.770 0.41910 *
        119) SCHEDDEP>312.5 389   95.950 0.55780
          238) AGE<23.8 343   85.490 0.52770 *
          239) AGE>23.8 46    7.826 0.78260 *
   15) SENIOR>0.5 13835  589.500 0.04460
     30) SCHEDDEP<362.5 13673  562.700 0.04300
       60) SCHEDDEP<169.5 3499   79.120 0.02315 *
       61) SCHEDDEP>169.5 10174  481.700 0.04983 *
     31) SCHEDDEP>362.5 162   23.810 0.17900 *
```

### 3. RTC TREE, 9 TERMINAL NODES

```
1) root 47465 4692.0 0.11120
   2) SMOKE<0.5 34069 2918.0 0.09460
     4) RUNJOG<0.5 15782 1571.0 0.11220
       8) AGE<18.55 5802  495.4 0.09428 *
       9) AGE>18.55 9980 1073.0 0.12250 *
     5) RUNJOG>0.5 18287 1338.0 0.07946
      10) SCHEDDEP<18.5 2154  226.3 0.11930 *
      11) SCHEDDEP>18.5 16133 1107.0 0.07413
        22) AFQT<49.5 5403  450.5 0.09180 *
        23) AFQT>49.5 10730  654.3 0.06524 *
   3) SMOKE>0.5 13396 1740.0 0.15350
     6) AFQT<66.5 8356 1198.0 0.17350
      12) SCHEDDEP<81.5 4037  635.4 0.19570 *
      13) SCHEDDEP>81.5 4319  559.1 0.15280 *
     7) AFQT>66.5 5040  533.1 0.12020
      14) SCHEDDEP<152.5 3568  413.2 0.13370 *
      15) SCHEDDEP>152.5 1472  117.7 0.08764 *
```

# LIST OF REFERENCES

Bohn, Don and Schmitz Edward, Reducing Attrition, Commander Naval Recruiting Command, Arlington Va.,1996.

Breiman, Leo; Friedman, Jerome H.; Olshen, Richard A.; and Stone, Charles J., Classification and Regression Trees, Belmont, California: Wadsworth, Inc., 1984.

CHIEF OF NAVAL PERSONNEL Letter 113 Ser 222c/7057736 Dated 15 Jul 97. Subject: Quality Standards.

CHIEF OF NAVAL PERSONNEL Message Dated 091131Z Dec 97. Subject: Navy College Fund/ Enlisted Bonus Program.

Clark, Linda A. and Pregibon, Daryl, "Tree-Based Models", Statistical Models in S, edited by John M. Chambers and Trevor J. Hastie, Pacific Grove, Ca.: Wadsworth and Brooks, Inc., 1992.

CNRC, Code-20, Commander Naval Recruiting Command Brief, Commander Naval Recruiting Command, Arlington Va., 1997.

Jacklich John J., Minimizing Drug Related Attrition Costs For Incoming Naval Recruits, Naval Postgraduate School, Monterey Ca., 1998.

Martin, Thomas, Who Stays,Who Leaves, An Analysis of First Term Army Attrition, Rand Graduate School, Santa Monica Ca.:Rand, 1995.

Matos, RafaelE., U.S. Navy's Delayed Entry Program: Effects of it's duration on Long-Term Attrition, Naval Postgraduate School, Monterey Ca., 1994.

Murray, Margaret, Navy Delayed Entry Program Attrition Analysis, Naval Postgraduate School, Monterey Ca., 1985.

Quester, Aline O.; MacIlvaine, Martha E.; and Barfield, Lisa, Bootcamp Attrition and Measures of Accession Quality, Center For Naval Analysis, Alexandria, Va., 1997.

Rovenstine, Michael J., Classification Analysis of Vibration Data from SH-60B Helicopter Transmission Test Facility, Naval Postgraduate School, Monterey Ca., 1997.

Schmitz, Edward and Reinert Carl, Planned Resource Optimization Model, Commander Naval Recruiting Command, Arlington Va. 1995.

Venables, W.N. and Ripley, B.D., Modern Applied Statistics with S-Plus, New York: Springer-Verlag, Inc., 1994.

# INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center ........................................ 2
   8725 John J. Kingman Road, Ste 0994
   Ft Belvoir, VA 22060-6218

2. Dudley Knox Libraray ........................................ 2
   Naval Postgraduate School
   411 Dyer Rd.
   Monterey, CA 93943-5101

3. Commander, Navy Recruiting Command ........................................ 1
   11 Balston Plaza
   Arlington, VA 22304-5000

4. Chief of Naval Education and Training ........................................ 1
   CNET 00
   250 Dallas St.
   Pensacola, FL 32508-5220

5. Professor Samuel E. Buttrey ........................................ 1
   Code OR/Sb
   Naval Postgraduate School
   Monterey, CA 93943-5002

6. Professor Harold J. Larson. ........................................ 1
   Code OR/La
   Naval Postgraduate School
   Monterey, CA 93943-5002

7. Lieutenant Bryant W. Knox. ........................................ 1
   100 Crestline Dr.
   Sneads Ferry, NC 28036