

AD-A114 769

GENERAL ELECTRIC CO WASHINGTON DC TEMPO CENTER FOR A--ETC F/G 15/5
AID LOSS PROBABILITY SYSTEM FOR ENLISTED (AID-E). VOLUME 1. EXE--ETC(U)
JAN 74 K H KIM, K R POWELL DAHC15-72-C-0345

UNCLASSIFIED

ODCSPER-73-2-A

NL

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

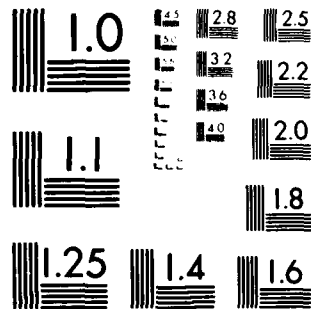
[]

[]

[]



END
DATE
FILMED
6 82
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

ODCSPER REPORT NUMBER 73-2-A

AD A114769

**AID LOSS PROBABILITY SYSTEM
FOR
ENLISTED (AID-E)
VOLUME 1
EXECUTIVE SUMMARY**

by
**Kwan H. Kim
Ken R. Powell**

Prepared for
**Department of the Army
Office of the Deputy Chief of Staff for Personnel**
by
**General Electric Company - TEMPO
777 14th Street, N.W.
Washington, D.C. 20005**
under
Contract Number DAHC 15 C 0345

January 15, 1974

**DTIC
SELECTED
MAY 20 1974
H**

**The findings in this report are not to be construed
as an official Department of the Army position.**

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED.

EXEMPTION STATEMENT A

**Approved for public release
Distribution Unlimited**

82 05 20 121

DTIC FILE COPY



DEPARTMENT OF THE ARMY
OFFICE OF THE DEPUTY CHIEF OF STAFF FOR PERSONNEL
WASHINGTON, D.C. 20310

DAPE-PBF

28 AUG 1975

SUBJECT: Army Manpower Prediction System (AMPS) Study

TO: SEE DISTRIBUTION

1. Attached are executive summaries of four manpower computer models developed by the Army during a three year study effort under contract to General Electric Co (TEMPO). The study was initiated in June 1972 with the objective of improving the Army's capability, given alternative situations and policies, to predict the active duty strength in terms of officers and enlisted personnel at specified dates in the future. The study was completed on 1 July 1975.

2. Overview of the models and proponent offices at Department of the Army for each are as follows:

a. Automatic Interaction Detector - Enlisted (AID-E) a loss rate generator that employs individual attributes (e.g., time left in service, educational level, mental group, etc.) as a basis to predict loss behavior. The AID technique, developed by the University of Michigan, is a powerful statistical routine which appears to have wide applicability in evaluating the causes for many types of human behavior. Proponent office at Department of the Army is the Manpower Programs Division, Directorate of Plans, Programs and Budget, Deputy Chief of Staff for Personnel (DCSPER) -- Mr. Petruzzi, OX 5-3857.

3 vols.

b. Automatic Interaction Detector - Officer (AID-O) a loss rate generator for officers comparable to AID-E. Proponent office is Officer Division, Directorate of Military Personnel Management, DCSPER -- MAJ Fishback, OX 7-6873.

2 vols.

c. Central Integrating Model - Enlisted (CIM-E) a force projection model capable of simulating the normal personnel life cycle (procurement, reenlistment, promotion, separation) of the enlisted population under varying sets of policies and situations. Force can be disaggregated by



DAPE-PBP

SUBJECT: Army Manpower Prediction System (AMPS) Study

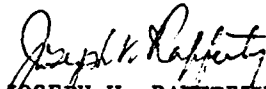
career management field, sex, race, educational level and mental group. Proponent office is Enlisted Division, Directorate of Military Personnel Management, DCSPER -- CPT Rose, OX 5-9419. 38015

d. Central Integrating Model - Officer (CIM-O) an officer force projection model similar to CIM-E except that CIM-O employs a modeling technique which permits many more types of disaggregations than the enlisted model. Proponent office is Officer Division, Directorate of Military Personnel Management, DCSPER -- MAJ Fishback, OX 7-6873. 38015

3. Further information on the models can be gained by contacting the proponents listed above.

FOR THE DEPUTY CHIEF OF STAFF FOR PERSONNEL:

4 Incl
as


JOSEPH V. RAFFERTY
Lt Colonel, GS
Contracting Officer's
Technical Representative

DISTRIBUTION:
OASA(M&RA)
DACS-DPM
Army Library (ASDIRS) 
DAPE-MP
DAPE-HR
DAPE-PBR
DAPC-PSS
National War College
Army War College
Industrial College of the Armed Forces
Armed Forces Staff College
Command and General Staff College
USMA

ODCSPER REPORT NUMBER 73-2-A

AID LOSS PROBABILITY SYSTEM

FOR ENLISTED (AID-E)

VOLUME I

EXECUTIVE SUMMARY

by

Kwan H. Kim
Ken R. Powell

Prepared for
Department of the Army
Office of the Deputy Chief of Staff
for Personnel

by

General Electric Company-TEMPO
777 Fourteenth Street, NW
Washington, DC 20005

under

Contract Number DAHC 15 72 C 0345

The findings in this report are not
to be construed as an official
Department of the Army position.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

January 15, 1974



FOREWORD

GE-TEMPO, after discussions with the Secretary of the Army and the Deputy Chief of Staff for Personnel, initiated a two-year project known as the Army Manpower Prediction System (AMPS) on June 30, 1972. Its purpose was to assist the Army in making more useful and accurate projections of manpower strength and composition. After six months of study, TEMPO submitted to the Army a Phase I report outlining areas in which the Army can significantly improve its manpower projections and recommending approaches to making such improvements.

The Army adopted a majority of these recommendations, and asked TEMPO to develop the following models in Phase II: the AID Loss Probability System for Enlisted (AID-E), the AID Loss Probability System for Officers (AID-O), the Central Integrating Model for Enlisted (CIM-E), and the Central Integrating Model for Officers (CIM-O).

The work on the AID-E was carried out between March and November of 1973.

This report in three volumes (1) describes TEMPO's analytical approach to development of the AID-E, (2) describes the AID-E itself, and (3) outlines procedures for operating the system.

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A	



ABSTRACT

The Department of the Army has many manpower projection models, each of which serves different purposes for different managers. Since the objectives of these managers are interrelated, it would be useful to the Army if the loss rates used in these models were not only reliable but also consistent, so that the projections of one model could easily be compared with the projections of others. In response to this need, GE-TEMPO has devised a system capable of generating several sets of loss rates from one basic set of loss rates. The generated rates can be used in both presently existing and future generations of manpower projection models with consistency and reliability. This system is called the AID Loss Probability System for Enlisted. The core of the system is a statistical technique which groups the actual enlisted population into several subpopulations according to personal characteristics most predictive of losses of individuals in each subpopulation. A loss probability is developed for each individual who belongs to a specific subpopulation and any desired loss format (e.g., grade by years of service) can be generated by summing over the appropriate individual characteristics. Since each individual's personal characteristics and his loss probability have a one-to-one correspondence, no matter how individuals are grouped for the purpose of generating loss rates (e.g., by grade, MOS, or any other relevant characteristic) each individual's characteristics and his loss probability are preserved. The loss probability of the group, however defined, is always the sum of the loss probabilities of actual individuals. It is apparent that this consistency renders flexibility for generating a variety of loss formats for all conceivable models. Furthermore, any historical probabilities can be easily adjusted to reflect current or proposed changes in manpower policy. This report describes the AID Loss Probability System, shows sample predictions developed from historical data, compares these predictions to actual loss experience, describes how the results of the system can be applied in other models, and explains how to use the system. The results of this study show that the system is an excellent method for predicting loss rates: its margin of error for each enlistment category is less than 1.5% in FY73 and less than 3.5% in FY70. Finally, the system is so versatile that it can be used for purposes other than predicting losses. For example, it can generate various reports such as distribution of age-group, marital status, MOS-imbalances, and other information contained in the EMF and Gain/Loss Transaction File. Also the system can tabulate and analyze information gathered by DOD/Army surveys.

VOLUME I

Executive Summary

TABLE OF CONTENTS

	Page
FOREWORD	i
ABSTRACT	ii
TABLE OF CONTENTS	iii
1.0 INTRODUCTION.....	1
1.1 Purpose of the AID-E.....	1
1.2 Scope of the AID-E.....	1
1.3 Organization of the Report.....	2
1.4 Participation of the Army.....	2
1.5 TEMPO Staff and Consultants.....	3
1.5.1 Phase I.....	3
1.5.2 Phase II: the AID-E.....	3
1.6 AID-E as a Means to Integrate Projection Models.....	4
1.6.1 Through Consistency.....	4
1.6.2 Through Accuracy.....	4
1.7 Assumption.....	4
1.8 Capabilities and Limitations.....	5
1.9 Resources Required to Operate the AID-E.....	6
1.10 Problems Encountered.....	7
1.11 Future Areas of Research.....	7
2.0 BASIC CONCEPT OF THE AID-E.....	8
2.1 Statistical Aspect.....	8
2.2 Logical Aspect.....	8
2.3 Overview of the AID-E.....	9
2.4 Method for Generating Loss Rates.....	12
2.5 The Variables.....	13
2.6 Causes of Loss by Enlistment Category.....	13
2.7 Use of the Loss Rates.....	17

	Page
3.0 APPLICATION OF THE AID-E AND INTERPRETATION OF RESULTS.	20
3.1 Introduction.	20
3.2 Data Used for Sample Outputs.	20
3.3 Results of AID Analysis Using FY70 Data.	20
3.3.1 All Losses (Based on FY70 Data).	20
3.3.2 Proportion of Losses by Cause to All Losses (Based on FY70 Data).	31
3.4 Sample Results of AID Analysis Using FY73 Data.	31
3.5 Actual vs. Predicted Losses in FY73 Using FY70 Data . .	32
3.6 Usefulness of the AID-E.	32
4.0 RECOMMENDATIONS.	47
APPENDIX: Definitions of Variables Used in Figures 2-13.	49

VOLUME I

LIST OF TABLES

	Page
TABLE 1: LIST OF VARIABLES IN THE AID-E	14
TABLE 2: TRANSACTION AND ENLISTMENT CATEGORIES BY SPN	15
TABLE 3: ENLISTED TRANSACTION CATEGORIES	18
TABLE 4: COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS) ENLISTMENT CATEGORY FT-AUS (FY 70)	27
TABLE 5: COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS) ENLISTMENT CATEGORY FT-RA (FY70)	28
TABLE 6: COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS) ENLISTMENT CATEGORY: CAREERIST INELIGIBLE FOR RETIREMENT (FY70)	29
TABLE 7: COMPARISON OF ACTUAL AND PREDICTED LOSSES ENLISTMENT CATEGORY: CAREERISTS ELIGIBLE FOR RETIREMENT (FY70)	30
TABLE 8: COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS) ENLISTMENT CATEGORY: FT-AUS (FY73)	41
TABLE 9: COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS) ENLISTMENT CATEGORY: FT-RA (FY73)	42
TABLE 10: COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS) ENLISTMENT CATEGORY: CAREERISTS (FY73)	43
TABLE 11: COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS) USING FY70 AID TREES TO PREDICT FY73 LOSSES: ENLISTMENT CATEGORY: FT-AUS	44
TABLE 12: COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS) USING FY70 AID TREES TO PREDICT FY73 LOSSES ENLISTMENT CATEGORY FT-RA	45
TABLE 13: COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS) USING FY70 AID TREES TO PREDICT FY73 LOSSES ENLISTMENT CATEGORY: CAREERISTS INELIGIBLE FOR RETIREMENT	46

VOLUME I

LIST OF FIGURES

	Page
FIGURE 1: AN OVERVIEW OF AID-E	11
FIGURE 2: AID TREE FOR FT-AUS (FY70)	22
FIGURE 3: AID TREE FOR FT-RA (FY70)	23
FIGURE 4: AID TREE FOR CAREERISTS INELIGIBLE FOR RETIREMENT (FY70)	24
FIGURE 5: AID TREE FOR CAREERISTS ELIGIBLE FOR RETIREMENT (FY70)	25
FIGURE 6: AID TREE FOR FT-AUS (L3, 13/L) (FY70)	33
FIGURE 7: AID TREE FOR FT-RA (L9/L) (FY70)	34
FIGURE 8: AID TREE FOR CAREER (L5-8/L) (FY70)	35
FIGURE 9: AID TREE FOR CAREER-RETIREMENT (L2, 3/L) (FY70)	36
FIGURE 10: AID TREE FOR FT-AUS (FY73)	37
FIGURE 11: AID TREE FOR FT-RA (FY73)	38
FIGURE 12: AID TREE FOR CAREERISTS (FY73)	39
FIGURE 13: AID TREE FOR CAREERISTS (L2/L) (FY73)	40

Manual for Planner/User

	Page
FOREWORD	i
ABSTRACT	ii
TABLE OF CONTENTS	iii
1.0 INTRODUCTION.	1
1.1 Purpose of the AID-E	1
1.2 Scope of the AID-E	1
1.3 Organization of the Report.	2
1.4 Participation of the Army.	2
1.5 TEMPO Staff and Consultants.	3
1.5.1 Phase I.	3
1.5.2 Phase II: the AID-E	3
2.0 BRIEF REVIEW OF THE AID-E	4
2.1 System Overview.	4
2.2 Computer Resource Considerations.	8
2.3 Comments and Considerations for Execution.	10
2.3.1 Program Execution Sequence.	10
2.3.2 Time Estimates for System Run.	12
3.0 DESCRIPTION OF SYSTEM COMPONENTS.	13
3.1 EMF Edit Program.	13
3.1.1 Program Narrative.	13
3.1.2 Input and Output of EMF Edit Program.	14
3.1.2.1 INPUT: EXTRACT OF EMF.	15
3.1.2.2 INPUT: GRADE-TABLE.	18
3.1.2.3 INPUT: MOS-TABLE.	18
3.1.2.4 INPUT: EDUCATION-TABLE.	19
3.1.2.5 INPUT: ASGTC-TABLE.	19
3.1.2.6 INPUT: STATUS-TABLE.	20
3.1.2.7 INPUT: AEISC-TABLE.	20
3.1.2.8 INPUT: ACFST-TABLE.	20
3.1.2.9 INPUT: DPACD-TABLE.	21
3.1.2.10 INPUT: ADRES-TABLE.	21
3.1.2.11 INPUT: SAMPLE SELECTION CONTROL CARD.	22
3.1.2.12 OUTPUT: EDITED EMF.	22
3.1.2.13 OUTPUT: PRINTED RECORD AND ERROR COUNT.	27

	Page
3.1.3 EMF Editing Routine.	33
3.1.3.1 PROCESS OF SAMPLE SELECTION. . .	34
3.1.3.2 SOCIAL SECURITY NUMBER (SSN). . .	36
3.1.3.3 MILITARY OCCUPATIONAL SPECIALTY (MOS).	37
3.1.3.4 MOS NUMBER (MOS-N).	37
3.1.3.5 SEX.	38
3.1.3.6 AGE.	38
3.1.3.7 RACE.	39
3.1.3.8 MARITAL STATUS (MARITAL).	40
3.1.3.9 NUMBER OF DEPENDENTS (DEPENDENTS).	40
3.1.3.10 GRADE.	41
3.1.3.11 AFQT PERCENTILE SCORE (AFQT). . .	42
3.1.3.12 MENTAL CATEGORY (MENTAL-CAT).	43
3.1.3.13 EDUCATIONAL LEVEL (EDUCATION).	44
3.1.3.14 PROFICIENCY PAY ELIGIBILITY (PRO-PAY).	45
3.1.3.15 ELIGIBILITY FOR VRB (VRB).	45
3.1.3.16 MAJOR COMMAND (MAJCOM).	46
3.1.3.17 FUNCTIONAL STATUS (STATUS).	47
3.1.3.18 AVERAGE EFFICIENCY INDEX (AEI). . .	48
3.1.3.19 TIME IN GRADE (TIG).	48
3.1.3.20 TOTAL MILITARY SERVICE (TMS) AND TOTAL ACTIVE MILITARY SERVICE (TAS).	50
3.1.3.21 EXPIRATION OF TERM OF SERVICE (ETS).	51
3.1.3.22 TERM OF ENLISTMENT OF SERVICE (TERMS).	53
3.1.3.23 DATE OF ESTIMATED RETURN FROM OVERSEAS (DEROS).	54
3.1.3.24 CAREER MANAGEMENT FIELD (CMF)	59
3.1.3.25 ENLISTMENT CATEGORY (ECAT).	59
3.1.3.26 PHYSICAL PROFILE (PHYSICAL).	59
3.1.3.27 RESIDENCE AT ENTRY (RESIDENCE). . .	60
3.1.3.28 MOS MISMATCH (MOS/JOB).	63
3.1.3.29 PROMOTION MOS (PROMOTION).	63
3.1.3.30 INCIDENCE OF AWOL (AWOL).	65
3.1.3.31 ETS-DEROS.	65
3.1.3.32 CONUS/OVERSEAS.	65
3.1.3.33 ERROR CODES.	66

	Page
3.1.4 Instructions for Computer Run: EMF Edit Program.	67
3.1.4.1 GENERAL DESCRIPTION.	67
3.1.4.2 ORDER OF CARDS IN JOB DECK.	67
3.1.4.3 CARD PREPARATION (SPECIFIC FOR THIS PROGRAM).	68
3.2 Gain/Loss Edit Program.	73
3.2.1 Program Narrative.	73
3.2.2 Input and Output of Gain/Loss Edit Program.	74
3.2.2.1 INPUT: EXTRACT OF GAIN/LOSS TRANSACTION FILE.	74
3.2.2.2 INPUT: GAIN TRANSACTION TABLE.	76
3.2.2.3 INPUT: LOSS TRANSACTION TABLE.	76
3.2.2.4 INPUT: SPN TABLE.	76
3.2.2.5 OUTPUT: EDITED GAIN/LOSS TRANSACTION FILE.	77
3.2.2.6 OUTPUT: TRANSACTION COUNTS.	78
3.2.3 Description of Gain/Loss Edit Program.	79
3.2.3.1 PROCESSES OF SAMPLING AND EDITING.	79
3.2.3.2 DEFINITIONS OF DATA ELEMENTS IN EDITED GAIN/LOSS TRANSACTION FILE.	81
3.2.4 Instructions for Computer Run: Gain/Loss Edit Program.	83
3.2.4.1 GENERAL DESCRIPTION.	83
3.2.4.2 ORDER OF CARDS IN JOB DECK.	84
3.2.4.3 CARD PREPARATION (SPECIFIC FOR THIS PROGRAM).	85
3.3 The Merge Program.	90
3.3.1 Program Narrative.	90
3.3.2 Input and Output of Merge Program.	91
3.3.2.1 INPUT: EDITED EMF.	91
3.3.2.2 INPUT: EDITED GAIN/LOSS TRANSACTION FILE.	92
3.3.2.3 OUTPUT: ENLISTED HISTORICAL FILE.	92
3.3.2.4 OUTPUT: RECORD COUNTS.	92
3.3.3 Instructions for Computer Program: Merge Program.	94
3.3.3.1 GENERAL DESCRIPTION.	94
3.3.3.2 ORDER OF CARDS IN JOB DECK.	95
3.3.3.3 CARD PREPARATION (SPECIFIC FOR THIS PROGRAM).	96

	Page
3.4 AID Statistical Routine.	100
3.4.1 Program Narrative.	100
3.4.2 Input and Output of AID Statistical Routine.	103
3.4.2.1 INPUT: ENLISTED HISTORICAL FILE	103
3.4.2.2 INPUT: CONTROL CARDS.	103
3.4.3 Description of AID Output.	107
3.4.3.1 BASIC OUTPUT.	108
3.4.3.2 TRACE OUTPUT.	110
3.4.3.3 BSS TABLE SHOWING EXPLANATORY POWER OF EACH VARIABLE.	111
3.4.3.4 PREDICTOR SUMMARY TABLE.	112
3.4.3.5 FORMULAS FOR COMPUTATION OF RESIDUALS.	115
3.4.3.6 PROFILE OF CLASS MEANS AND SLOPES.	115
3.4.3.7 ADDITIONAL FORMULAS.	116
3.4.3.8 SUMMARY OF AID RUNS.	118
3.4.4 Instructions for Computer Run: AID STATISTICAL Routine	121
3.4.4.1 GENERAL DESCRIPTION	121
3.4.4.2 ORDER OF CARDS IN JOB DECK.	122
3.4.4.3 CARD PREPARATION (SPECIFIC FOR THIS PROGRAM).	123
3.5 Rate Generator Program.	134
3.5.1 Program Narrative.	134
3.5.2 Logic of Rate Generator Program.	134
3.5.3 Instruction for Computer Run: Rate Generator Program.	135
3.5.3.1 GENERAL DESCRIPTION.	135
3.5.3.2 ORDER OF CARDS IN JOB DECK.	136
3.5.3.3 CARD PREPARATION (SPECIFIC FOR THIS PROGRAM).	138
3.5.4 An Annotated Example of Recoding in Rate Generator Program.	147
3.5.5 Description of Output by Rate Generator Program.	147
4.0 APPLICATION OF THE AID-E AND INTERPRETATION OF RESULTS.	150
4.1 Introduction to the Chapter.	150
4.2 Data Used for Sample Outputs.	150
4.3 Discussion of Alternative Methods for Generating Loss Probabilities by Cause.	151

	Page
4.4 Note on Sample Size.	158
4.5 Results of AID Analysis Using FY70 Data.	162
4.5.1 All Losses (Based on FY70 Data).	162
4.5.2 Proportion of Losses by Cause to All Losses (Based on FY70 Data).	174
4.5.3 Errors in FY70 Data.	174
4.6 Sample Results of AID Analysis Using FY73 Data.	179
4.7 Actual Losses in FY73 vs. Prediction Made from FY70 Data.	179
4.8 Usefulness of the AID-E.	180
4.8.1 Use of Loss Rates in Projection Models.	180
4.8.2 Other Uses of the AID-E: MOS Mismatch as an Example.	180

APPENDIX A INPUT-OUTPUT OF EMF EDIT PROGRAM

A1: GRADE-TABLE.	A1
A2: MOS-TABLE.	A2
A3: EDUCATION-TABLE.	A13
A4: ASGTC-TABLE.	A14
A5: STATUS-TABLE.	A15
A6: AEISC-TABLE.	A16
A7: ACFST-TABLE.	A17
A8: DPACD-TABLE.	A18
A9: ADRES-TABLE.	A19
A11: EMF ANALYSIS-RECORD COUNTS, FIRST TERMER.	A20
A12: EMF ANALYSIS-VARIABLE ERROR COUNTS, FIRST TERMER.	A21
A13: EMF ANALYSIS-RECORD COUNTS, CAREERISTS.	A25
A14: EMF ANALYSIS-VARIABLE ERROR COUNTS, CAREERISTS.	A26
A15: S360 JOB CARD.	A30
A16: LOGIC CHART FOR THE EMF EDIT PROGRAM.	A34

APPENDIX B INPUT-OUTPUT IN GAIN/LOSS EDIT PROGRAM

B1: GAIN TRANSACTION TABLE.	B1
B2: LOSS TRANSACTION TABLE.	B2
B3: SPN TABLE.	B3
B4: TRANSACTION COUNTS.	B5

	Page
APPENDIX C OUTPUT OF THE MERGE PROGRAM	
C1: UNMATCHED GAIN/LOSS TRANSACTION, FIRST TERMER.	.C1
C2: UNMATCHED GAIN/LOSS TRANSACTION, CAREERIST.	C2
C3: DISTRIBUTION COUNT BY TRANSACTION, FIRST TERMER.	C3
C4: DISTRIBUTION COUNT BY TRANSACTION, CAREERIST.	C4
 APPENDIX D INPUT-OUTPUT OF AID ROUTINE AND RATE GENERATOR PROGRAM	
D1: REQUIREMENTS FOR DISC WORK SPACE.	D1
D2: RECODE LANGUAGE.	D2
D3: FILTER CARDS.	D9
D4: DISCUSSION OF AID TREE RECORD FORM.	D11
D5: RECODE LIMITATIONS, WARNINGS, AND ERROR MESSAGES.	D14
D6: RATE GENERATOR PROGRAM ERROR MESSAGES. . .	D15
D7: ANNOTATED SAMPLE AID RUN.	D19
D8: PROGRAM OPTIONS FOR AID COMPUTATIONS.	D54
D9: ANNOTATED SAMPLE PRINTOUT OF RATE GENERATOR PROGRAM.	D89
 GLOSSARY.	 E1
INDEX.	F1

VOLUME II
Manual for Planner/User

LIST OF TABLES

	<u>PAGE</u>
TABLE 1: COMPUTER RESOURCE AND LANGUAGE REQUIREMENTS	9
TABLE 2: FORMAT OF CONDENSED EXTRACT FROM EMTR	16-17
TABLE 3: FORMAT OF SAMPLE SELECTION CONTROL CARD	23-24
TABLE 4: FORMAT OF EDITED EMF	25-26
TABLE 5: EMF ANALYSIS RECORD COUNTS	28
TABLE 6: EMF ANALYSIS VARIABLE ERROR COUNTS	29
TABLE 7: ERROR CODE BY TYPE OF ERROR	30-32
TABLE 8: DOD AND REGIONAL CODES FOR STATES .	61-62
TABLE 9: FORMAT OF EXTRACT FROM GAIN/LOSS HISTORY FILE	75
TABLE 10: FORMAT OF EDITED GAIN/LOSS TRANSACTION FILE	77
TABLE 11: TRANSACTION COUNTS	78-79
TABLE 12: TRANSACTION AND ENLISTMENT CATEGORIES BY SPN	82-83
TABLE 13: FORMAT OF ENLISTED HISTORICAL FILE .	92-93
TABLE 14: RECORD COUNTS: UNMATCHED GAIN/LOSS TRANSACTIONS	94
TABLE 15: RECORD COUNTS: DISTRIBUTION COUNT BY TRANSACTION	94

	<u>PAGE</u>
TABLE 16: ANALYSIS OF COVARIANCE	114
TABLE 17: SUMMARY OF AID RUNS	120
TABLE 18: AID TREE RECODE FORM	148
TABLE 19: COMPARISON OF ALL LOSSES BY ENLISTMENT CATEGORY	152
TABLE 20: PROPORTION OF LOSS CAUSES (TRANSACTION CATEGORIES FIVE TO EIGHT COMBINED BY ENLISTMENT CATEGORY)	153
TABLE 21: PROPORTION OF LOSS CAUSES WITHIN FT-RA	154
TABLE 22: COMPARISON OF DIRECT AND INDIRECT METHODS IN GENERATING LOSS PROBABILITIES	157
TABLE 23: SIZE OF AID SAMPLE: EXAMPLE	161
TABLE 24: ENLISTED TRANSACTION CATEGORIES . .	163-164
TABLE 25: COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS), ENLISTMENT CATEGORY: FT-AUS (FY 1970)	170
TABLE 26: COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS), ENLISTMENT CATEGORY: FT-RA (FY 1970)	171
TABLE 27: COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS), ENLISTMENT CATEGORY: CAREERISTS INELIGIBLE FOR RETIREMENT (FY 1970)	172
TABLE 28: COMPARISON OF ACTUAL AND PREDICTED LOSSES, ENLISTMENT CATEGORY: CAREERISTS ELIGIBLE FOR RETIREMENT (FY 1970)	173

	<u>PAGE</u>
TABLE 29: COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS), ENLISTMENT CATEGORY: FT-AUD (FY 1973)	185
TABLE 30: COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS), ENLISTMENT CATEGORY: FT-RA (FY 1973)	186
TABLE 31: COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS), ENLISTMENT CATEGORY: CAREERISTS (FY 1973)	187
TABLE 32: COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS) USING FY70 AID TREES TO PREDICT FY73 LOSSES, ENLISTMENT CATEGORY: CAREERISTS INELIGIBLE FOR RETIREMENT.	188
TABLE 33: COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS) USING FY70 AID TREES TO PREDICT FY73 LOSSES, ENLISTMENT CATEGORY: FT-RA.	189
TABLE 34: COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS) USING FY70 AID TREES TO PREDICT FY73 LOSSES, ENLISTMENT CATEGORY: FT-AUS.	190

VOLUME II
Manual for Planner/User
LIST OF FIGURES

	Page
FIGURE 1: AN OVERVIEW OF THE AID-E.	5-6
FIGURE 2: RELATION BETWEEN SAMPLE SIZE AND ITS STANDARD ERROR, AND CPU TIME IN MINUTES. . .	160
FIGURE 3: AID TREE FOR FT-AUS (FY70).	165
FIGURE 4: AID TREE FOR FT-RA (FY70).	166
FIGURE 5: AID TREE FOR CAREERISTS INELIGIBLE FOR RETIREMENT (FY70).	167
FIGURE 6: AID TREE FOR CAREER-RETIREMENT ELIGIBLE (FY70).	168
FIGURE 7: AID TREE FOR FT-AUS (L3, 13L) (FY70).	175
FIGURE 8: AID TREE FOR FT-RA (L9, 1/L) (FY70).	176
FIGURE 9: AID TREE FOR CAREER-INELIGIBLE FOR RETIREMENT (L5-8/L) (FY70).	177
FIGURE 10: AID TREE FOR CAREER-RETIREMENT ELIGIBLE (L2, 3L) (FY70).	178
FIGURE 11: AID TREE FOR FT-AUS (FY73).	181
FIGURE 12: AID TREE FOR FT-RA (FY73).	182
FIGURE 13: AID TREE FOR CAREERISTS (FY73).	183
FIGURE 14: AID TREE FOR CAREERISTS (L2/L) (FY73).	184

VOLUME III

Manual for Programmers/Systems Analyst

TABLE OF CONTENTS

	Page
FOREWORD	i
ABSTRACT	iv
TABLE OF CONTENTS	v
1.0 BRIEF REVIEW OF THE AID-E.	1
1.1 System Overview.	1
1.2 Computer Resource Considerations.	5
1.3 Comments and Considerations for Execution.	7
1.3.1 Program Execution Sequence.	7
1.3.2 Time Estimates for System Run.	9
2.0 INFORMATION AND CONSIDERATIONS NEEDED FOR RUNNING THE AID-E.	10
2.1 Libraries for Source and Load Modules.	10
2.2 EMF and Gain/Loss Edit Programs.	10
2.3 Merge Program.	11
2.4 AID Statistical Routine.	11
2.5 Rate Generator Program.	11
3.0 LOGIC FLOW CHARTS AND PROGRAM NARRATIVES FOR AID-E.	12
3.1 EMF Edit Program.	12
3.1.1 Program Narrative.	13
3.1.2 Input and Output of EMF Edit Program.	13
3.1.2.1 Input: Extract of EMF.	14
3.1.2.2 Input: Grade-Table.	17
3.1.2.3 Input: MOS-Table.	17
3.1.2.4 Input: Education-Table.	18
3.1.2.5 Input: ASGTC-Table.	18
3.1.2.6 Input: Status-Table.	19
3.1.2.7 Input: AEISC-Table.	19
3.1.2.8 Input: ACFST-Table.	19
3.1.2.9 Input: DPACD-Table	20

	Page
3.1.2.10 Input: ADRES-Table.	20
3.1.2.11 Input: Sample Selection Control Card. . .	21
3.1.3 EMF Edit Program Flow Charts.	21
3.2 Gain/Loss Edit Program.	26
3.2.1 Program Narrative.	26
3.2.2 Input and Output of Gain/Loss Edit Program. . .	26
3.2.2.1 Input: Extract of Gain/Loss Transaction File.	27
3.2.2.2 Input: Gain Transaction Table.	29
3.2.2.3 Input: Loss Transaction Table.	29
3.2.2.4 Input: SPN Table.	29
3.2.2.5 Output: Edited Gain/Loss Transaction File.	30
3.2.3 Gain/Loss Edit Program Flow Chart.	30
3.3 Merge Program.	31
3.3.1 Program Narrative.	31
3.3.2 Input and Output of Merge Program.	32
3.3.2.1 Input: Edited EMF.	32
3.3.2.2 Input: Edited Gain/Loss Transaction File.	33
3.3.2.3 Output: Enlisted Historical File.	33
3.3.2.4 Output: Record Counts.	33
3.3.3 Merge Program Flow Chart.	35
3.4 AID Statistical Routine.	35
3.4.1 Program Narrative.	35
3.4.2 Input and Output of AID Statical Routine.	36
3.5 Rate Generator Program.	37
3.5.1 Input and Output of Rate Generator Program. . .	37
APPENDIX A: PROGRAM LISTING OF EMF EDIT PROGRAM. . . .	A1
APPENDIX B: PROGRAM LISTING OF GAIN/LOSS EDIT PROGRAM.	B1
APPENDIX C: PROGRAM LISTING OF MERGE PROGRAM.	.C1
APPENDIX D: SOURCE PROGRAM LISTING OF THE SUBROUTINE TABLES OF THE RATE GENERATOR.	.D1
APPENDIX E: JCL PROCESURES FOR AID III AND RATE GENERATOR PROGRAM.	.E1
GLOSSARY.	.F1

VOLUME III
LIST OF TABLES

	Page
TABLE 1: COMPUTER RESOURCE AND LANGUAGE REQUIREMENTS.	5
TABLE 2: FORMAT OF CONDENSED EXTRACT FROM EMTR. . .	15-16
TABLE 3: FORMAT OF SAMPLE SELECTION CONTROL CARD. .	22-23
TABLE 4: FORMAT OF EDITED EMF.	24-25
TABLE 5: FORMAT OF EXTRACT FROM GAIN/LOSS HISTORY FILE.	.28
TABLE 6: FORMAT OF EDITED GAIN/LOSS TRANSACTION FILE.	30
TABLE 7: FORMAT OF ENLISTED HISTORICAL FILE.	.33

VOLUME III
LIST OF FIGURES

	Page
FIGURE 1: ALPS-E FLOW CHART.	2
FIGURE 2: LOGIC CHART FOR THE EMF EDIT PROGRAM. . . .	39
FIGURE 3: GENERAL LOGIC CHART FOR GAIN/LOSS EDIT PROGRAM.	62
FIGURE 4: FLOW CHART FOR MERGE PROGRAM.	65
FIGURE 5: RATE GENERATOR LOGIC FOR SUBROUTINE TABLES.	66

Chapter 1

INTRODUCTION

GE-TEMPO, after discussions with the Secretary of the Army and the Deputy Chief of Staff for Personnel, initiated a two-year project known as the Army Manpower Prediction System (AMPS) on June 30, 1972. Its purpose was to assist the Army in making more useful and accurate projections of manpower strength and composition. After six months of study, TEMPO submitted to the Army a Phase I report outlining areas in which the Army can significantly improve its manpower projections and recommending approaches to making such improvements.

The Army adopted a majority of these recommendations, and asked TEMPO to develop the following models in Phase II: the AID Loss Probability System for Enlisted (AID-E), the AID Loss Probability System for Officers (AID-O), the Central Integrating Model for Enlisted (CIM-E), and the Central Integrating Model for Officers (CIM-O).

The work on the AID-E was carried out between March and November of 1973.

This report (1) describes TEMPO's analytical approach to development of the AID-E, (2) describes the AID-E itself, and (3) outlines procedures for operating it.

1.1 Purpose of the AID-E

The purpose of the AID-E is to provide reliable and consistent loss rates (including reenlistment rates) for existing and future generations of manpower projection models. Use of these rates would integrate the Army's presently decentralized manpower projections.

1.2 Scope of the AID-E

The AID-E can generate loss rates (including reenlistment rates) of active enlisted personnel, both men and women, by enlistment category, by cause of loss, and by other personal characteristics. To illustrate the flexibility of the AID-E, the rates generated for inclusion in this report are in formats required by the CIM-E, PIA, and MAIL models. It should be kept in mind, however, that AID-E is capable of generating loss rates in virtually any format which may be needed by Army manpower planners.

The system does not include warrant officers, officers, reservists, or civilians serving the Army. Loss rates for officers will be generated by the AID-O model which will be completed by August 1974.

1.3 Organization of the Report

This final report on the AID Loss Probability System for Enlisted (AID-E) consists of three volumes with the following titles:

Volume I	Executive Summary
Volume II	Manual for Planner/User
Volume III	Manual for Programmer/Systems Analyst

Volume I, the Executive Summary, contains an overview and brief description of what the model is and what it does. The overview provides an understanding of the capabilities, limitations and assumptions of the AID-E; description of the system includes resource requirements and examples of the system's output.

Volume II, the Manual for Planner/User, is intended for action officers whose responsibility is to analyze manpower plans and programs and to make projections. It contains a detailed description of the system and the system logic. Detailed descriptions of input requirements are provided along with discussions of operating considerations and of proper interpretation of results. This volume provides the user with the information necessary to set-up input for computer runs and discusses various analytical approaches.

Volume III, the Manual for Programmer/System Analyst, contains the minute details of the AID-E system. The details of each computer program in the system are provided to facilitate modification of the system if, at some future date, modification becomes necessary.

A table of contents covering all three volumes has been included in each volume so that readers may locate sections most appropriate to their interests.

1.4 Participation of the Army

The Army assigned Col. Warren A. Spaulding, Lt. Col. Norman G. Blahuta, and Maj. John E. Pearson, Jr. to Phase I of the AMPS project on a full-time basis. Many other officers from the ODCSPER, notably Captain Roger W. Collins, gave very generously of their time.

During the development of the AID-E in Phase II, Lt. Col. Joseph V. Rafferty, Lt. Col. Robert Lewis, Maj. John E. Pearson, Capt. Roger W. Collins, and Lt. Richard Gudewicz were most helpful. Many officers and civilians in the ODCSPER, MILPERCEN, and USAMSSA also gave their assistance.

TEMPO is grateful for office space made available in the Pentagon to TEMPO staff throughout the project.

1.5 TEMPO Staff and Consultants

1.5.1 Phase I

The Phase I effort of the AMPS was assisted by Dr. Kwan H. Kim, G. E. Consultant. The TEMPO team, under the direction of Dr. Steve Enke and Mr. Terry L. Schilling, has included Dr. William E. McFarland, Jr., Mr. Richard A. Brown, Mr. Stephen Chabon, Mr. Bryan D. Hickman, Mr. James L. Johnston, and Miss Cynthia Bennett. TEMPO wishes to acknowledge assistance received from the Honorable Hadlai A. Hull, Assistant Secretary of the Army (Manpower and Reserve Affairs); Lieutenant General Walter T. Kerwin, Jr., former Deputy Chief of Staff for Personnel; General Paul D. Phillips (Ret.), Deputy Assistant Secretary of the Army (Manpower and Reserve Affairs); and many others.

1.5.2 Phase II: The AID-E

The AID-E was developed by Ken R. Powell and Terry L. Schilling of GE-TEMPO, with the advice and assistance of Dr. Kwan H. Kim, consultant. Nancy Karp of GE-TEMPO was helpful in programming the first version of the data edit program.

1. 6 AID-E as a Means to Integrate Projection Models

1. 6. 1 Through Consistency

The process of manpower planning and programming in the Army is organized into several directorates and divisions along many functional lines (usually interrelated in ultimate purpose). As a result, a high degree of decentralization has resulted and a variety of projection models have been developed. These models project enlisted strengths in a number of forms and the projected strengths usually differ from one another and lack compatibility. The disparities could be reduced if all models were to use the same consistent input parameters; a key parameter, of course, is loss rate. Since the AID-E is capable of generating loss rates in a form suitable for use in any model from a basic set of loss rates (i. e. "AID tree"), the use of the rates from AID-E would make the loss projections of all models consistent.

1. 6. 2 Through Accuracy

Since the need for new accessions is largely determined by losses incurred to existing strength, the use of accurate loss rates is fundamental to a credible manpower projection. Estimates of loss rates of Army personnel are developed in many forms from several sources; often they are expressed in very gross terms without regard to individual characteristics which are prone to both loss behavior and cause of loss. These rates frequently compete with one another in the policy-making process; it is impossible to validate which model can be trusted most. The preliminary results of this study show that the AID-E reliably produces loss rates of very great accuracy, varying from actual loss experience by only a small percentage. The Army's use of such accurate, validated rates would reduce the present state of confusion and integrate various projection models without loss of the advantages of the present decentralized manpower management.

1. 7 Assumption

The AID analysis is based upon a set of predictor variables representing personal and circumstantial characteristics contained in the EMF at a specific point in time. It does not consider any other variables. Thus states of war or peace, the unemployment rate in civilian industries, and enlisted men's attitudes toward the Army are not accounted for in the AID analysis. If the values of these and other potentially important variables have been stable over several years and are expected to be so in the immediate future, there is no need to weigh their impact on the AID analysis. However, if the values of such

omitted variables change frequently and become influencing factors on loss behavior in the future (or are expected to become so), predictions of losses based upon static EMF data alone may miss a shift in loss behavior over time.

One method of incorporating the variables representing war/peace, unemployment rate, etc., in the AID analysis is to add the values of these variables to the transaction data. More specifically, whenever a transaction occurs to a person at a certain date, the values of these variables as of that date can be added to his transaction record. These extra variables can be easily added to the transaction data during data editing and later used as predictor variables in the AID analysis.

In summary, the AID analysis is accurate in generating loss rates for use in predicting losses if (1) all changes in loss behavior are a result of changes in personal characteristics of the enlisted force as represented by the variables contained in the EMF and (2) there are no changes in potentially predictive variables not contained in the EMF.

1.8 Capabilities and Limitations

The AID-E is designed to handle a very large number of variables and of EMF records since the AID statistical routine is capable of analyzing simultaneously this large quantity of information in searching for the variables most predictive of loss behavior. The only limit on the number of variables which can be used is the content of the EMF itself.

The AID-E is designed to generate loss rates in any conceivable format with accuracy and consistency.

The AID analysis is best used to produce annual loss rates. Using the AID routine to produce loss rates covering a shorter period is not recommended for both practical and statistical considerations. This limitation is considerable in view of the fact that most projection models use monthly or quarterly loss rates. A great deal of care must be exercised in converting annual rates into monthly or quarterly rates. In order to preserve consistency in the use of AID loss rates

Estimation of the values of such variables is a major research task in itself.

in monthly and quarterly projection models, the Army cannot be too strongly urged to designate an official agency in charge of converting AID's annual loss rates into monthly and quarterly rates.

The AID-E loss rates do not agree with loss rates derived from the DCSPER-46 Report. However, the AID-E rates are correct for the CIM, PIA, MAIL and other models as long as these models add new accessions after losses are taken out from the beginning inventory. In fact, the loss rates developed by the AID-E take into account the 1st-month men, 2nd-men, 3rd-month men and so on. Therefore, the AID-E rates serve these models.

In order to make AID-E estimated losses agree with rates derived from the DCSPER-46 Report, the following modification on the AID-E must be added: New accessions during the fiscal year should be added to the EMF snapshot taken at the beginning of the fiscal year for the AID-E analysis. This augmented EMF inventory would then be analyzed by AID-E producing loss rates. However, these modified loss rates will not be correct for the CIM, PIA, and MAIL models that AID-E was originally designed.

1.9 Resources Required to Operate the AID-E

The AID-E requires extensive investment in (a) analyst's and system operator's time and (b) computer facilities and supplies. Below is a list of the estimated cost of various resources necessary to run the AID-E to predict loss rates and number of losses for each of the enlistment categories and loss causes listed in TABLE 2 for use in the CIM-E model:

Analyst	2 man-days
System Operator	5-10 man-days
CPU	375-575 minutes
Magnetic Tapes	7 reels
Other (Clerical and consulting, etc.)	5 man-days

Assuming this run is repeated four times a year and loss rates required by other models are also generated regularly, the system is obviously expensive. However, the budgetary savings which will result from (a) more accurate manpower projections which reduce costly errors in manpower management and (b) use of one accurate set of loss rates for all models will far outweigh these expenses.

1.10 Problems Encountered

During the development of AID-E, the normal technical problems were encountered such as computer limitations (e. g., core capacity, availability of CPU time, etc.), the selection of predictors and the "best" statistical approach, but these types of problems were solved. One problem the user of AID-E will encounter in supplying loss rates to other models is a clear concise definition of populations the loss rates are applied to. The logic of each manpower model is different and consequently requires different loss rates. For example, loss rates for a particular model may be applied to the enlisted population after reenforcements have been subtracted out of the population. In another model the same types of loss rates may be applied to the entire population. It is essential that the populations the rates will apply to are carefully defined.

1.11 Future Areas of Research

The AID-E system is not only a model for producing loss rates, but is also an excellent system to provide immediate information for various Army manpower studies. With the large number of data elements contained on magnetic tape, the system can be used to determine, for example, the influencing factors which cause MOS mismatch. Studies can be performed to determine the characteristics of individuals in each MOS and identify which characteristics are desirable to retain an individual in an particular MOS. Geographical origin, marital status, performance in previous MOS, and many other factors complicate manpower studies. The AID-E sorts these factors choosing the important characteristics for each type of study and uses these characteristics as the basis on which predictions can be made. Skillful use of AID-E can certainly benefit the Army in areas other than loss predictions.

**Copy available to DTIC does not
permit fully legible reproduction**

Chapter 2

BASIC CONCEPT OF THE AID-E

2.1 Statistical Aspect

Central to the AID is a statistical technique known as the Automatic Interaction Detector (AID)*, developed by the Survey Research Center of the Institute for Social Research at the University of Michigan. Given a list of explanatory (or predictor) variables, the AID ranks these variables, one at a time, in order of their importance in explaining loss behavior. This is accomplished in a series of steps using binary splits. At each step the population is divided into all conceivable pairs of two mutually exclusive subgroups for each predictor variable. After examining all such pairs for all variables, the pair which accounts for the greatest amount of variance from the average loss rate for that particular population group is chosen for the binary split. The AID then searches for the next most important split in the same manner as before. The AID stops splitting when there are no further subgroupings which can explain a statistically significant amount of variance within the limit set by the analyst. At each step, the AID provides the loss probability for each subgroup as well as the statistical significance of the split.

2.2 Logical Aspect

The AID is a logical way of analyzing losses for several reasons.

First, the AID groups a population into several subgroups according to homogeneity of personal and circumstantial characteristics which influence loss and reenlistment behavior.

Second, the AID is receptive to changes in Army strength and composition and characteristics of the force (e.g., educational level) over time. Such changes cause both growth and decay of certain population subgroups, affecting overall losses. The impact of these changes is automatically measured by the AID.

Third, the AID permits logical adjustment of rates where it is known that future experience will be different from past experience. For instance, if a three month early release program is planned for first term enlisted men in the coming year, it is possible to isolate the particular AID subgroups of men who are first-termers and whose

*All other modules (or subsystems) in the AID are designed either to prepare the data needed for the AID or to summarize (and generate) the AID results into suitable matrices for use in Army projection models.

ETS dates lie within the thirteenth to the fifteenth months, and to adjust their loss rates into line with similar groups whose ETS date is within 12 months. No changes have to be made to other groups.

Fourth, the AID-E provides a useful basis for understanding why losses are occurring and for developing remedial policies to influence particular types of losses. Knowledge of characteristics of each group is helpful in identifying specific target populations for special remedial policies. Special AID analysis can be made to study specific types of loss that are usually not separated for prediction purposes.

Fifth, the AID-F can handle large quantities of data, large numbers of variables, and variables which are expressed in non-quantifiable categories. These features permit the analyst to search for all available information with regard to loss and reenlistment behavior.

2.3 Overview of the AID-E

Shown in FIGURE 1 is a graphical representation of the AID-E. The system, verbally described, starts with two basic sources of data. The first is an EMF extract which contains approximately 40 elements of information for each enlisted man at a specific point in time. The second source of data is an extract of Gain/Loss Transaction Files. This extract contains approximately 10 elements of information on gain/loss experience of each enlisted man during a specified time period.* These two data extracts are edited in a form acceptable for AID statistical analysis by two computer programs--the EMF Edit Program and Gain/Loss Edit Program.

The two edited extracts are then merged by the Merge Program, producing a single data file referred to as the Enlisted Historical File. From this file containing records for each enlisted man, the analyst can choose an enlistment category.** Throughout the stages of data preparation, reports are produced with error counts (along with diagnosis of errors) which evaluate and summarize data quality.

*The time period chosen for the AID analysis is usually a year, but it can be varied. The important thing is that the beginning date of the Gain/Loss File must match the date of the EMF extract.

**The enlistment categories are First-term AUS, First-term RA, and Careerists. Alternatively, careerists can be divided into careerists eligible for retirement and careerists ineligible for retirement.

Next the Enlisted Historical File is analyzed by the AID statistical routine which generates loss rates by enlistment category and loss cause. The analyst can further recode the data at the time the AID statistical routine is executed and he can also exclude certain individual records and variables. The primary output of the AID routine is a collection of mutually exclusive statements of personal characteristics along with probabilities of loss for each statement. Each statement is a step-by-step description of a series of binary splits made upon successive predictor variables arranged in order of their importance in explaining losses. Since the statements are mutually exclusive, every individual in the population must belong to one (and only one) of these statements and its associated loss probability. The analyst can draw the collection of statements and associated probabilities as an upside-down tree, tracing each statement from the point at which the trunk first splits (the most important predictor variable for the collection) to the tip of the lowest branchlet (the last and least important predictor for a particular statement). After a critical review of the AID tree is made for logical consistency and accuracy of loss probabilities, the AID tree is transferred as input to the Rate Generator Program.

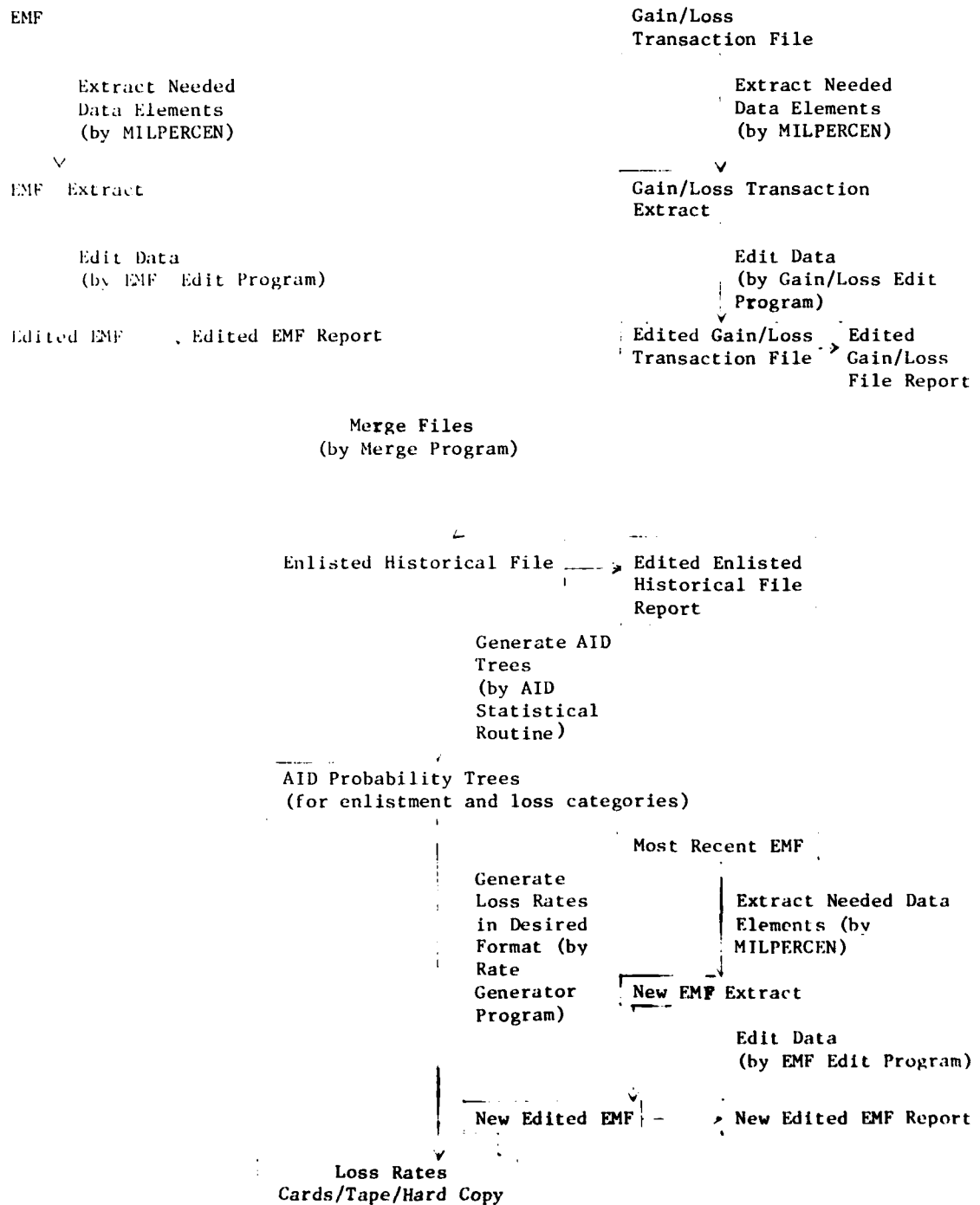
The Rate Generator Program then interrogates an extract* of the current EMF (edited by the previously mentioned EMF Edit Program), record by record. The purpose of the interrogation is (a) to compare the personal characteristics of each individual in that extract to the statements of characteristics which make up the AID tree and (b) to assign to each individual the loss probability associated with the statement that describes him.

In order to produce loss rates in a certain format, say by grade and years of service (YOS), the individual records in the current extract are then sorted by grade and YOS; the loss probabilities of all individuals in each grade-YOS cell are summed over; and the sum is divided by the number of the individuals in that cell to obtain a weighted average of loss probabilities. This weighted average is the loss probability for that cell. The Rate Generator Program can produce a matrix of loss probabilities in any format the analyst desires within the limit. **

*Again, this extract can be either the entire enlisted force or a random sample.

**First, the variables chosen for the matrix must be contained in the EMF extract. Second, the matrix accepts at maximum three variables at a time.

FIGURE 1 - AN OVERVIEW OF THE AID-E



2.4 Method for Generating Loss Rates

Two methods for generating loss rates may be used: direct and indirect.

The direct method is the method by which the AID routine generates the ratio of the number of losses from a specific cause to the number of persons in the population. The resulting AID tree is input to the Rate Generator Program. Loss rates in a desired format are produced by this program in the manner described at the end of the previous section.

The indirect method uses two AID trees. The first tree is produced by calculating the ratio of the number of all losses (i. e., irrespective of cause of loss) to the number of persons in the population. The second tree is produced by calculating the ratio of the number of losses from a specific cause to all losses. The two trees are transferred to the Rate Generator Program in which the following steps are taken to generate loss rates in a desired format. The characteristics of individuals in the current EMP are compared with the statements of personal characteristics which make up the two AID trees, and each individual is assigned to the loss rate associated with the two statements (one from each tree) which describe him. The two rates are then multiplied to produce a single rate for each individual. When this multiplication is done for all individuals, the loss rate for a cell specified by a loss format is generated by summing over the loss probabilities of all persons in that cell and the sum is divided by the number of persons in that cell.

The choice of one method over the other depends upon statistical considerations, the number of transaction categories to be analyzed, and the amount of computer time required. For example, the indirect method should be chosen when several loss causes are being analyzed and the number of losses from each cause is extremely small in comparison to the population being studied (since the indirect method produces a higher ratio of losses to population and therefore makes analysis more effective); when there is only one loss cause to be analyzed, the direct method is the only choice; and when there are many loss causes to be analyzed and the size of the population is very large, the indirect method saves computer run time.

In general, the reliability of the indirect method is greater (1) because two trees tend to produce more splits and (2) because the ratios are higher.

2.5 The Variables

The AID-E is capable of analyzing a very large number of variables at the same time. Most of them are used as predictor variables in AID analysis, but some are used for other purposes. For example, the Social Security number is used for sampling the population for the AID analysis; the error code variables, internally generated by the EMF Edit Program, are used in controlling inclusion or exclusion of certain variables and individual records in the AID analysis based upon the quality of data; and the MOS and CMF variables are used for sorting population in the Rate Generator Program.

Several variables resemble each other and seem redundant, but they serve useful purposes such as reconstructing missing data and cross-checking the validity of certain variables.

Chapter 3 contains sample results of AID analysis of actual Army data. In these results, some variables appear more often than others as loss predictors. The dominance of these variables should not establish the permanency of their importance. Certain other variables proved to be more important during a particular time period, becoming less important in other periods. It should also be noticed that some potentially useful variables were deliberately omitted from the analysis reported in Chapter 3 due to poor quality of the data. It is quite possible that, as the quality of data improves over time, some of the presently less predictive variables may become more useful. TABLE 1 lists the variables included in the AID-E.

2.6 Causes of Loss by Enlistment Category

Numerous computer runs of the AID were made using actual data to determine ways to improve reliability of the loss projection system, particularly by subgrouping losses by enlistment category and cause of loss.

After careful analysis of these runs, the Army and TEMPO agreed to analyze losses by the transaction categories (groups of loss causes) and enlistment categories listed in TABLE 2. The results of this analysis were used to generate a variety of loss rates in formats required by the CIM-E, PIA, and MAIL models. However, it should be borne in mind that the AID-E, as noted earlier in this report, is capable of generating loss rates in virtually any format needed by Army manpower planners.

TABLE 1

LIST OF VARIABLES IN THE AID-1

<u>Variable Number</u>	<u>Data Element</u>	<u>Abbreviation</u>	<u>Possible Value Range</u>
1	Social Security Number	SSN	0 - 9
2	Military Occupational Specialty	MOS	XXXXXX
3	3 Digit MOS	3MOS	0 - 999
4	Sex	SEX	0 - 9
5	Age	AGE	0 - 999
6	Race	RACE	0 - 9
7	Marital Status	MARITAL	0 - 9
8	Number of Dependents	DEPENDENTS	0 - 9
9	Grade	GRADE	0 - 9
10	AFQT Percentile Score	AFQT	0 - 99
11	Mental Category	MENTAL-CAT	0 - 9
12	Educational Level	EDUCATION	0 - 9
13	Proficiency Pay Eligibility	PRO-PAY	0 - 9
14	Eligibility for VRB	VRB	0 - 9
15	Major Command	MAJCOM	0 - 99
16	Functional Status	STATUS	0 - 9
17	Average Efficiency Index	AEI	0 - 9
18	Time in Grade	TIG	0 - 999
19	Total Military Service	TMS	0 - 999
20	Total Active Military Service	TAS	0 - 999
21	Expiration of Term of Service	ETS	0 - 99
22	Term of Enlistment	TERMS	0 - 99
23	Date of Estimated Return from Overseas	DEROS	0 - 99
24	Career Management Field	CMF	0 - 99
25	Enlistment Category	ECAT	0 - 9
26	Physical Profile	PHYSICAL	0 - 9
27	Residence at Entry	RESIDENCE	0 - 99
28	MOS Mismatch	MOS/JOB	0 - 9
29	Promotion MOS	PROMOTION	0 - 9
30	Incidence of AWOL	AWOL	0 - 9
31	TMS Error Code	E1	0 - 9
32	TAS Error Code	E2	0 - 9
33	Grade Error Code	E3	0 - 9
34	MOS Error Code	E4	0 - 9
35	TIG Error Code	E5	0 - 9
36	ETS Error Code	E6	0 - 9
37	Terms of Service Error Code	E7	0 - 9
38	DEROS Error Code	E8	0 - 9
39	Transaction Category	TRANSACTION	0 - 99
40	Process Date of Transaction	PROCESS	0 - 9999
41	Effective Date of Transaction	EFFECTIVE	0 - 9999
42	Term of Reenlistment or Extension	REUP/EXT TERM	0 - 999
43	Transaction Date minus EMF Date	TRANS-EMF	0 - 99

TABLE 2
TRANSACTION AND ENLISTMENT CATEGORIES BY SPN

<u>TRANSACTION CATEGORY & CODE</u>	<u>FT-AUS</u>	<u>FT-RA</u>	<u>CAREERIST</u>	<u>SPN</u>
O. Miscellaneous Transactions				
1. DFMC	X	X	X	SPN numbers are designated "FOR OFFICIAL USE ONLY" and hence are not listed in this volume. They may be found on pages 82-83 of Volume II Manual for Planner/User.
2. Retirement	-	-	X	
3. Physical Disqualification	X	X	-	
4. Hardship/Dependency	X	X	X	
5. Unfitness	X	X	X	
6. Unsuitability	X	X	X	
7. Misconduct	X	X	X	
8. ETS	X	X	X	
9. Reenlistment Losses	-	X	X	
10. Early Release	X	X	X	
11. Other Losses	X	X	X	

TABLE 2 (continued)

<u>TRANSACTION CATEGORY & CODE</u>	<u>FT-AUS</u>	<u>FT-RA</u>	<u>CAREERIST</u>	<u>SPN</u>
12. Return to Military Control	X	X	X	--
13. Extension	X	X	X	--
14. Reenlistment Gains	-	X	X	--
99. New Accessions	X	X	-	--

2.7 Use of the Loss Rates

In generating loss rates for use in different projection models, the user of AID-E must clearly understand the idiosyncrasies of each model with respect to (1) the time dimension of the loss rates it requires (e.g., annual rate, monthly rates, etc.), (2) its logic and simulation process, (3) its definitions of loss rates in the total population and in loss categories (i.e., SPN's), and (4) the dimensions of the loss format it requires. Failure to understand any of these idiosyncrasies will result in the generation of rates which are completely useless.

The CIM-E model, for example, is a monthly projection model which requires, by definition, monthly loss rates. CIM's logic for loss calculation is in the following sequence: To a monthly beginning inventory the attrition loss (Transaction Categories 1 through 7 and 10 to 11 in TABLE 2) are applied and the resulting attritions are taken out from the inventory, thus reducing the inventory. Now, to this remaining population, the ETS loss rates (Transaction Category 8 for the first-termers and careerists ineligible for retirement) are applied and the further reduction of the population caused by the ETS losses is made. If there are any further losses (e.g., special releases), they are also taken out after the ETS losses. All of the above processes in the CIM require defining and redefining of successive populations through use of SPN's for the purpose of generating appropriate loss rates. Another projection model may not necessitate such changes in the population base if the logic of that model demands a simultaneous take-out of all losses from the beginning inventory.

Finally, annual loss rates generated by the AID-E must be converted to monthly or quarterly rates with appropriate seasonality. This conversion should be made by experts who not only know past experience but also can accurately anticipate month-to-month future policies of the Army which may affect losses.

TABLE 3
ENLISTED TRANSACTION CATEGORIES

<u>AID Transaction Category</u>	<u>Transaction Category</u>	<u>SPN</u>
1. DFMC	DFMC	SPN numbers are designated "FOR OFFICIAL USE ONLY" and hence are not listed in this volume. They may be found on pages 163-164 of Volume II Manual for Enlisted/Inner
2. Retirement (Non-D)	Retired	
3. Disability Retirement	Disability Retirement	
4. Physical Disqualification	Physical Disqualification	
5. Hardship/Dependency	Hardship	
6. Unfitness	Unfitness	
7. Unsuitability	Unsuitability	
8. Misconduct	Misconduct	
	Punitive Discharge Good of Service	
9. School & Teaching	School & Teaching	
10. Early Release Res/NG	Early Release Res/NG	
11. Other Early Release	Early Release	
	Employment	
	OS Early Release	
12. ETS	ETS	
13. Other	Death	
	TBOWO	
	Failure of Promotion	
14. Reenlistment (immediate)	Both Losses and Gains	

TABLE 3 (Continued)

<u>AID Transaction Category</u>	<u>Transaction Category</u>	<u>SPN</u>
15. Research	Other Losses	All other & blank
16. Reenlistment (non- immediate)	Gains	-----
17. Returned to Milit- ary Control	Gains	-----

Chapter 3

APPLICATION OF THE AID-E INTERPRETATION OF RESULTS

3.1 Introduction

This chapter summarizes and interprets some of the results (including predicted vs. actual losses) produced by the AID-E. The reader should note that losses predicted by AID-E for FY 1973 vary from actual losses during this period by less than 1.5% and losses predicted for FY 1970 vary from actual losses by less than 3.5%. When the loss rates developed from FY 70 data were applied to FY 73 EMF, the margin of error was less than 2.4%.

Volume II of this report discusses the formulae used for the AID statistics, the meanings of these formulae, and the options available in running the AID routine. Chapter 4 of Volume II includes appropriate comments and suggestions on sample size, sampling errors, bias caused by deleting (or including) variables having errors, and methods of generating loss rates.

3.2 Data Used for Sample Outputs

For purposes of testing the reliability and validity of the AID-E both FYs 1970 and 1973 were used. For FY70 data, the date for the EMF Extract is as of July 1, 1969 and the time period covered for the extract of the Gain/Loss Transaction is from the beginning of July 1969 to the end of June 1970. For FY73 data, the date for the EMF is as of July 1, 1972, and the time period for the extract of the Gain/Loss File is from the beginning of July 1972 to the end of June 1973. In fact, GE-TEMPO used FY70 data primarily as a test and experimental data. In this chapter, the results of the AID runs using both FY70 and 73 data are presented.

3.3 Results of AID Analysis Using FY70 Data

TABLE 3 lists the enlisted transaction categories for which FY70 data was used by the AID.

3.3.1 All Losses (Based on FY70 Data)

FIGURES 2 through 5 are the AID trees drawn to reflect loss probabilities for four enlistment categories developed by the AID from a sample of the FY1970 enlisted population. The trees appear in the following order: FT-AUS, FT-RA, Careerists Ineligible for Retirement, and Careerists Eligible for Retirement.

In order to explore the degree to which losses predicted on the basis of these AID-developed probabilities approach actual loss experience during FY70, a second sample of the enlisted population in FY70 was drawn and the persons in this sample classified according to the sub-populations (clusters of personal characteristics) into which AID grouped the first sample. The loss probability for each sub-population generated by AID from the first sample was then applied to the number of persons in the second sample described by that sub-population to obtain "predicted losses" for each sub-population. The second sample was then reviewed again to determine the number of persons in that sample who were actually lost in the course of FY70.

Finally the AID trees developed from FY 70 data were used to predict losses in FY 73. The results are presented in Section 3.5. The results show that AID-E is very reliable in predicting losses.

FIGURE 2 - AID TREE FOR FT-AUS

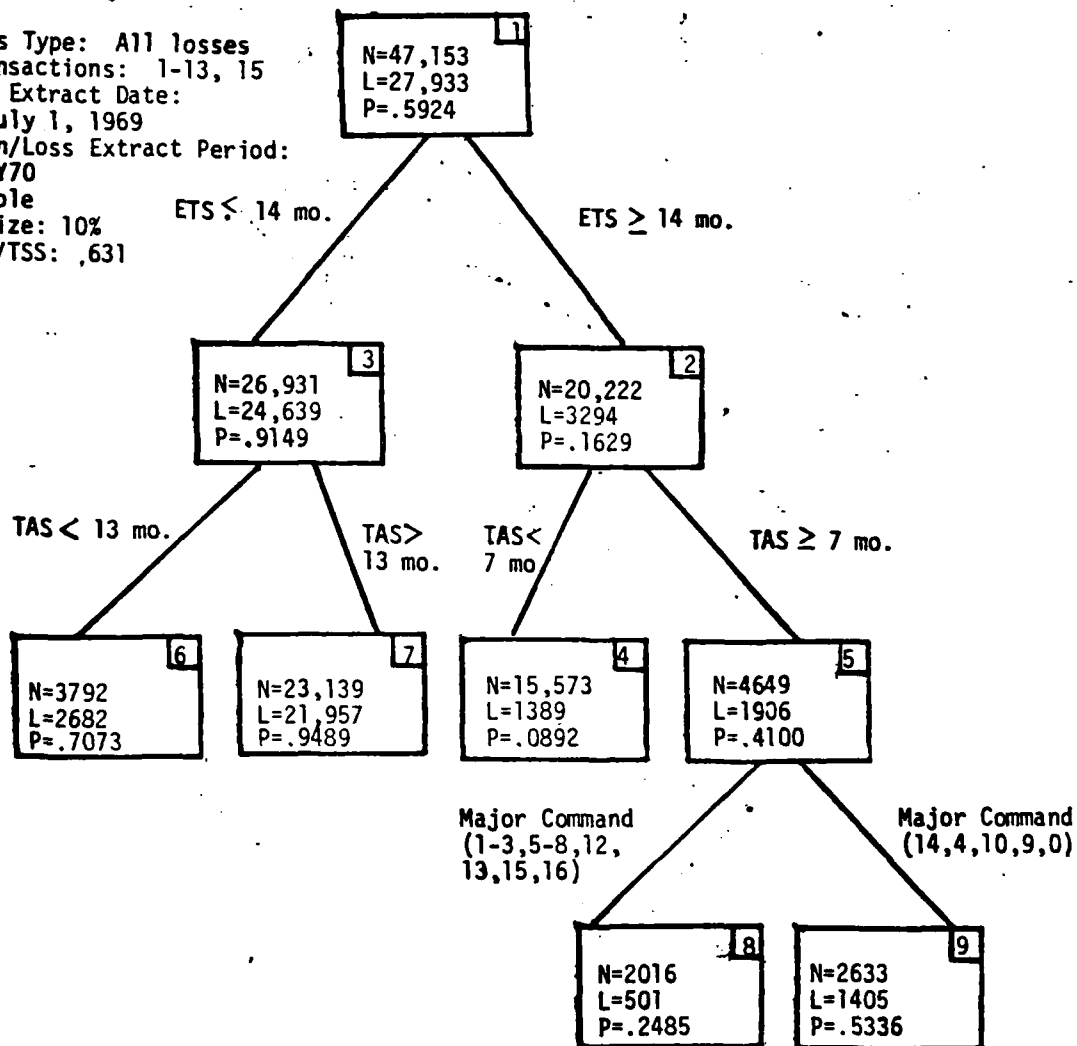
Loss Type: All losses
Transactions: 1-13, 15
EMF Extract Date:
July 1, 1969

Gain/Loss Extract Period:
FY70

Sample

Size: 10%

BSS/TSS: .631

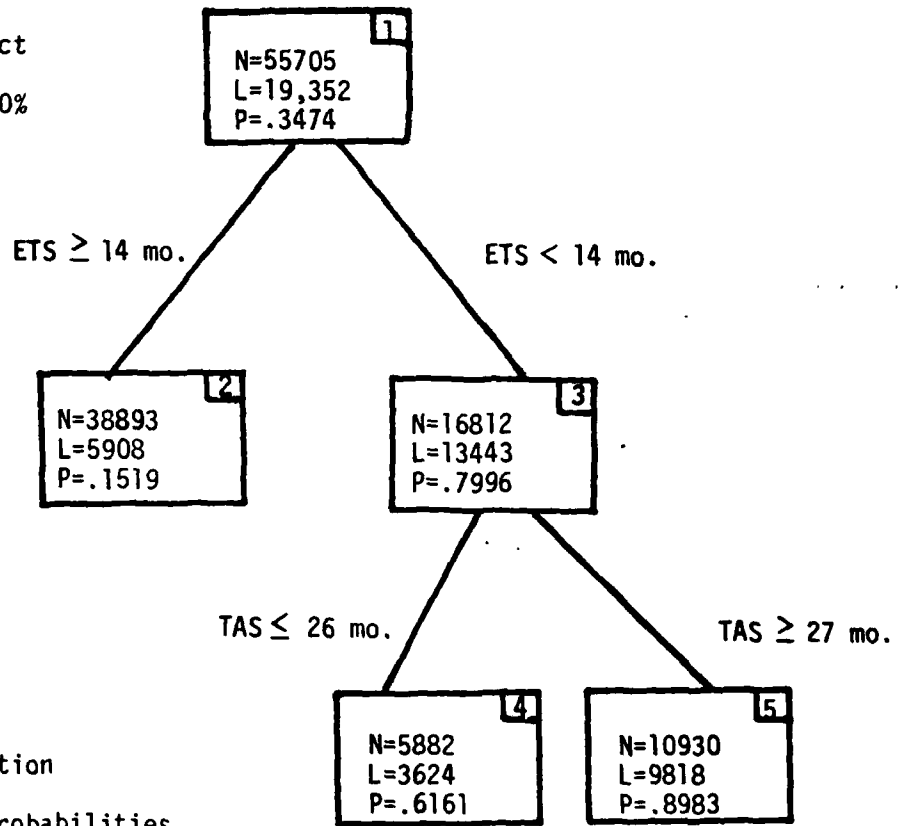


N = Population
L = Losses
P = Loss probabilities

See Appendix for definitions of variables.

FIGURE 3 - AID TREE FOR FT-RA

Loss Type: All losses
 Transactions: 1-13, 15
 EMF Extract Date:
 July 1, 1969
 Gain/Loss Extract
 Period: FY70
 Sample Size: 10%
 BSS/TSS: .416



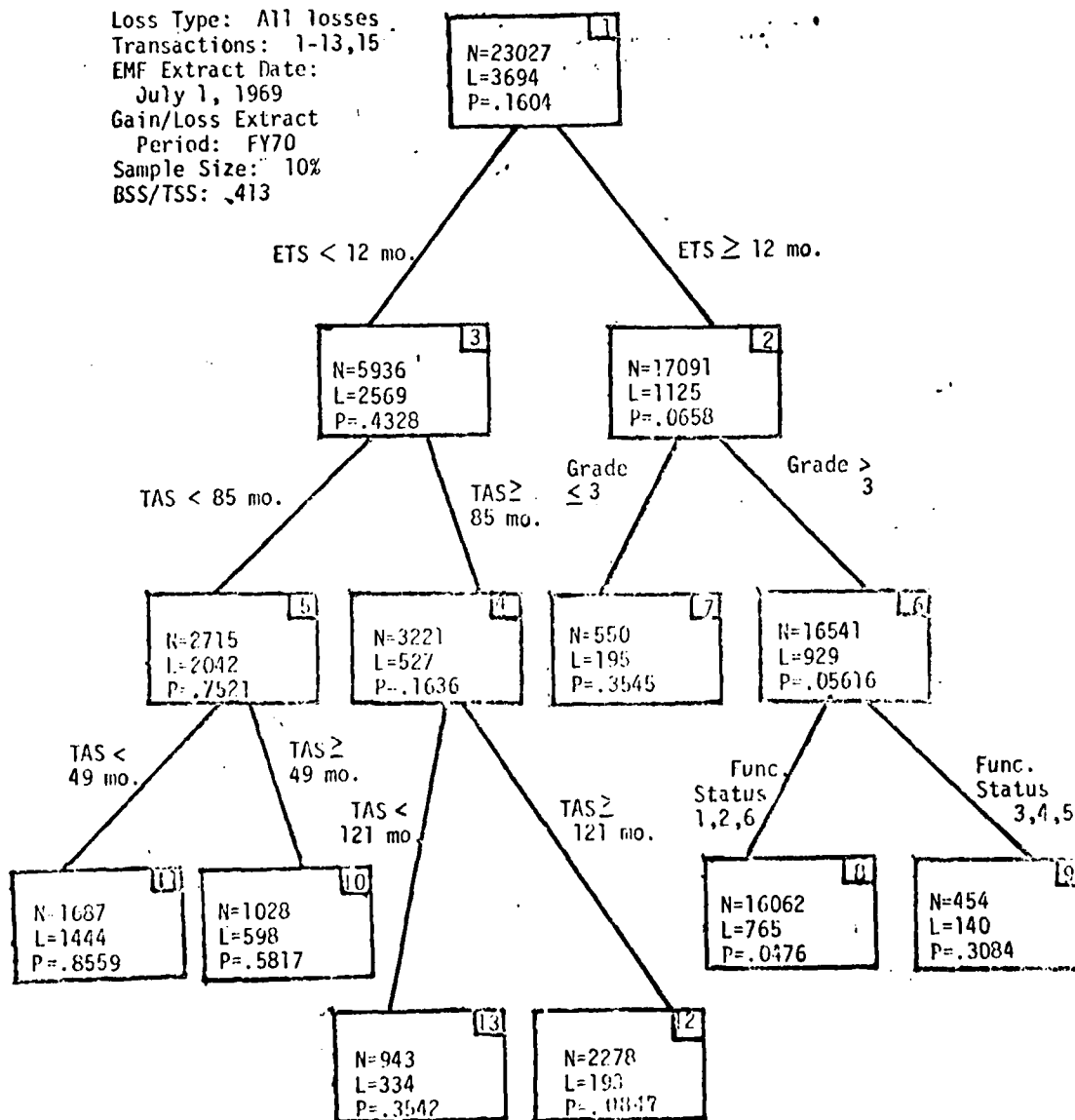
N = Population
 L = Losses
 P = Loss probabilities

See Appendix for definitions of variables

FIGURE 4- AID TREE FOR CAREERISTS
INELIGIBLE FOR RETIREMENT*

(FY 1970)

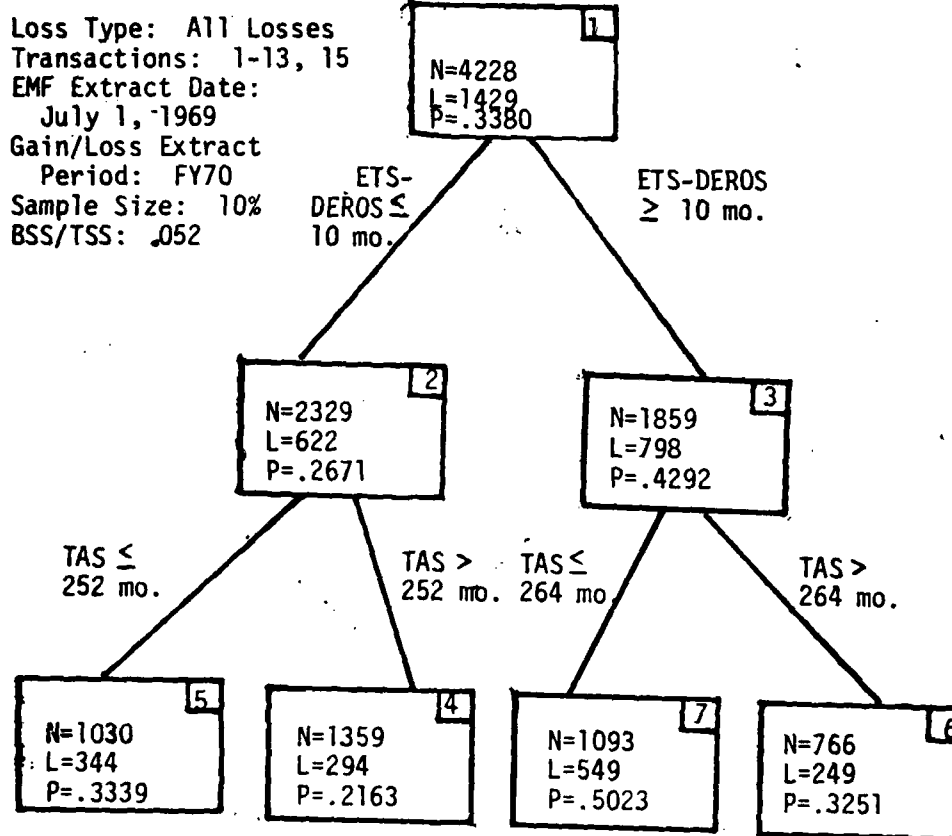
Loss Type: All losses
Transactions: 1-13,15
EMF Extract Date:
July 1, 1969
Gain/Loss Extract
Period: FY70
Sample Size: 10%
BSS/TSS: .413



N = Population
L = Losses
P = Loss probabilities

See Appendix for definitions of variables.

FIGURE 5 - AID TREE FOR CAREERISTS
ELIGIBLE FOR RETIREMENT*



N = Population
L = Losses
P = Loss probabilities

See Appendix for definitions of variables

TABLES 4 through 7 compare these actual and predicted losses in each sub-population defined by AID, as reflected in the trees shown in FIGURES 2 through 5. The transaction categories used in the AID trees and TABLES 4 through 7 are defined in TABLE 3. A few explanatory notes for these tables are made below:

- (a) The "BSS/TSS" is that proportion of the total sum of squares explained by a given number of AID splits (or sub-grouping)--a kind of correlation statistics.
- (b) The "MEAN" is the overall loss rate for the entire sample.
- (c) The "Cell Number" listed in each box of the second column of the tables corresponds to the AID tree cell numbers as shown in FIGURES 2 through 5.
- (d) The "Loss Probability" in the third column shows the probability attached to a terminal cell or to the characteristics belonging to those in the second column.
- (e) The "Number of Persons" in the fourth column indicates the number of persons in the second sample who belong to the characteristics listed in the corresponding box in the second column.
- (f) The "Predicted Losses" is the number of losses calculated by multiplying Column 3 by Column 4.
- (g) The "Actual Losses" is the actual number of losses incurred to the population in the second sample whose characteristics belong to those in Column 4.
- (h) The "Predicted Overall Loss Rate" at the bottom is the ratio of the total in Column 5 to that in Column 4.
- (i) The "Actual Overall Loss Rate" at bottom is the ratio of the total in Column 6 to that in Column 4.

TABLE 4
COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS)
ENLISTMENT CATEGORY: FT-AUS (FY70)

TRANSACTION CATEGORY: All losses (1-13, 15)

AID RUN DATE: 7/26/73

AID RUN NUMBER: E-7

SAMPLE SIZE FOR AID RUN:

Random or Not Random

Sample Number and Size: #1; 47,153 (10 percent)

SAMPLE SIZE FOR COMPUTING ACTUAL LOSSES (OR GAINS)

Comparison Run Date: 8/30/73

Random or Not Random

Sample Number and Size: #2; 47,220 (10 percent)

BSS/TSS: .631

MEAN: .5724

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Terminal Cell Number	Cell Number	Loss Probability	Number of Persons	Predicted Losses	Actual Losses	Percent Error
1	4	.0892	15,682	1396	1357	-3.1%
2	8	.2485	1,981	493	478	-2.9
3	9	.5336	2,672	1426	1403	-1.6
4	6	.7073	3,804	2689	2653	-1.4
5	7	.9489	23,081	21904	21872	-0.1
TOTAL			47,220	27,908	27,783	-0.3%

Predicted Overall Loss Rate = .5910

Actual Overall Loss Rate = .5884

*Actual - Predicted 100
Actual

**Major Command = 16, 5, 8, 6, 7, 12, 2, 1, 3, 13, 15

***Major Command = 14, 4, 10, 9, 0

TABLE 5
COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS)
ENLISTMENT CATEGORY: FT-RA (FY70)

TRANSACTION CATEGORY: All losses (1-13, 15)

AID RUN DATE: August 6, 1973

AID RUN NUMBER: E4

SAMPLE SIZE OFRAID RUN:

Random or Not Random

Sample Number and Size: Sample #1; 55,705 (10%)

SAMPLE SIZE FOR COMPUTING ACTUAL LOSSES (OR GAINS):

Comparison Run Date: August 29, 1973

Random or Not Random

Sample Number and Size: Sample #2; 55,427 (10%)

BSS/TSS: .416

MEAN: .3516

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Terminal Cell Number	Cell Number	Loss Probability	Number of Persons	Predicted Losses	Actual Losses	Percent Error
1	2	1520	38805	5898	5850	-0.8%
2	4	6161	5912	3642	3694	1.4
3	5	8984	10710	9622	9764	1.5
TOTAL			55,427	19,162	19,308	0.7%

Predicted Overall Loss Rate = .346

Actual Overall Loss Rate = .348

*Actual - Predicted 100
 Actual

TABLE 6
COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS)
ENLISTMENT CATEGORY:
CAREERIST INELIGIBLE FOR RETIREMENT (FY70)

TRANSACTION CATEGORY: All losses.

AID RUN DATE: August 6, 1973

AID RUN NUMBER: E2

SAMPLE SIZE FOR AID RUN:

Random or Not Random

Sample Number and Size: Sample #1; 23,027 (10%)

SAMPLE SIZE FOR COMPUTING ACTUAL LOSSES (OR GAINS):

Comparison Run Date: August 29, 1973

Random or Not Random

Sample Number and Size: Sample #3 - 20%

BSS/TSS: .413

MEAN: .1604

(1) Terminal Cell Number	(2) Cell Number	(3) Loss Probability	(4) Number of Persons	(5) Predicted Losses	(6) Actual Losses	(7) Percent Error
1	8	.0476	30805	1466	1569	6.6%
2	9	.3084	826	255	258	1.2
3	7	.3545	1101	390	376	-3.7
4	12	.0847	4550	385	405	4.9
5	7	.3542	1977	700	700	0.0
6	10	.5817	2041	1187	1141	-4.0
7	11	.8560	3306	2830	2782	-1.7
TOTAL			44,606	7213	7231	0.3%

Predicted Overall Loss Rate = .1617

Actual Overall Loss Rate = .1621

*Actual - Predicted
 Actual .1621

TABLE 7
COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS)
ENLISTMENT CATEGORY:
CAREERISTS ELIGIBLE FOR RETIREMENT(FY70)

TRANSACTION CATEGORY: All losses (1-13, 15)

AID RUN DATE: August 6, 1973

AID RUN NUMBER: E9

SAMPLE SIZE FOR AID RUN:

Random or Not Random

Sample Number and Size: Sample #1; 4248 (10%)

SAMPLE SIZE FOR COMPUTING ACTUAL LOSSES (OR GAINS):

Comparison Run Date: August 29, 1973

Random or Not Random

Sample Number and Size: Sample #2; 4355 (10%)

BSS/TSS: .052

MEAN: .3380

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Terminal Cell Number	Cell Number	Loss Probability	Number of Persons	Predicted Losses	Actual Losses	Percent Error
1	4	.2163	1433	310	316	1.9%
2	5	.3340	1059	354	381	7.1
3	6	.3251	781	254	262	3.1
4	7	.5023	1082	543	555	2.2
TOTAL			4355	1461	1514	3.5%

Predicted Overall Loss Rate = .3355

Actual Overall Loss Rate = .3476

*Actual - Predicted 100
Actual

3.3.2 Proportion of Losses by Cause to All Losses Using FY 70 Data

In Chapter 2, the two methods for generating loss rates were discussed. For some causes of loss, loss rates are calculated by the indirect method. Given the AID loss rates (L_i/P_i)* shown in FIGURES 2 through 5, the proportions of losses by cause (L_{ci}/L_i) will provide the loss rates for a specific cause (L_{ci}/P_i) when L_i/P_i is multiplied by L_{ci}/P_i .

FIGURES 6 through 9 shown four examples of AID trees made for L_{ci}/L_i for certain causes: an example is drawn from each of the four enlisted categories. The two sets of AID trees (one for L_i/P_i and another for L_{ci}/L_i) are multiplied in the Rate Generator Program to produce the loss rates by a specific cause (L_{ci}/P_i). An example of this multiplication is given in the ANNOTATED SAMPLE PRINTOUT OF RATE GENERATOR PROGRAM (Appendix D9) in Volume II. The predicted losses in fact have proven to be very close to actual losses using the indirect method.

3.4 Sample Results of AID Analysis Using FY73 Data

Figures 10 through 12 are the three AID trees drawn from the three AID runs made from FY73 data. The transaction categories used for these AID runs are listed in TABLE 3. No distinction is made between careerist-ineligible-for-retirement and careerist-eligible-for-retirement. The two are combined into a single group: careerist.

The first three AID trees (FIGURES 10 through 12) are loss rates for FT-AUS, FT-RA, and Careerist. The fourth tree (FIGURE 13) is the AID run made on the proportion of losses due to retirement of careerists. When, for example, the loss rates of careerist in FIGURE 12 are multiplied by the proportion of retirement losses in FIGURE 13, the retirement loss rates are generated by the indirect method for various subpopulations in the careerist population.

TABLES 8 through 10 show the actual and predicted losses.

* Notations: L=Number of Losses; P=Population; i=Enlistment Category; c=Cause of Loss

3.5 Actual vs. Predicted Losses in FY '73 Using FY 70 Data

Using the AID trees in FIGURES 2-5 (based on FY 70 data), predicted losses are made and compared with actual losses in FY 73. The results are presented in TABLES 11-13. The prediction errors are : 1.1% for AUS; 2.4% for first-term RA; and -1.5% for Careerists. Some of the errors are attributed to the quality of FY 70 data and to changes in definitions of several variables.

3.6 Usefulness of the AID-E

The Army and the Department of Defense have a long list of manpower projection models. All of the models need a consistent set of loss rates. The loss rates generated by the AID-E, for example, can be used for the CIM, PIA, and MAIL models. Also the AID-E can be used for generating various reports and analyses for Army's special requirements.

GE has already used the system to make a brief (two-day) study of the problem of the Army's MOS imbalances. MOS-mismatch is a considerable cause of the Army's inefficiency in using its most valued resource: manpower. The objective of GE's brief study was to identify the characteristics of an enlisted individual which "best" explain why he is an MOS-mismatch or loss to the Army. The problem deserves additional, careful analysis. In fact, the AID-E is a "generalized information generating system".

FIGURE 6 - AID TREE FOR FT-AUS (L3, 13/L) *

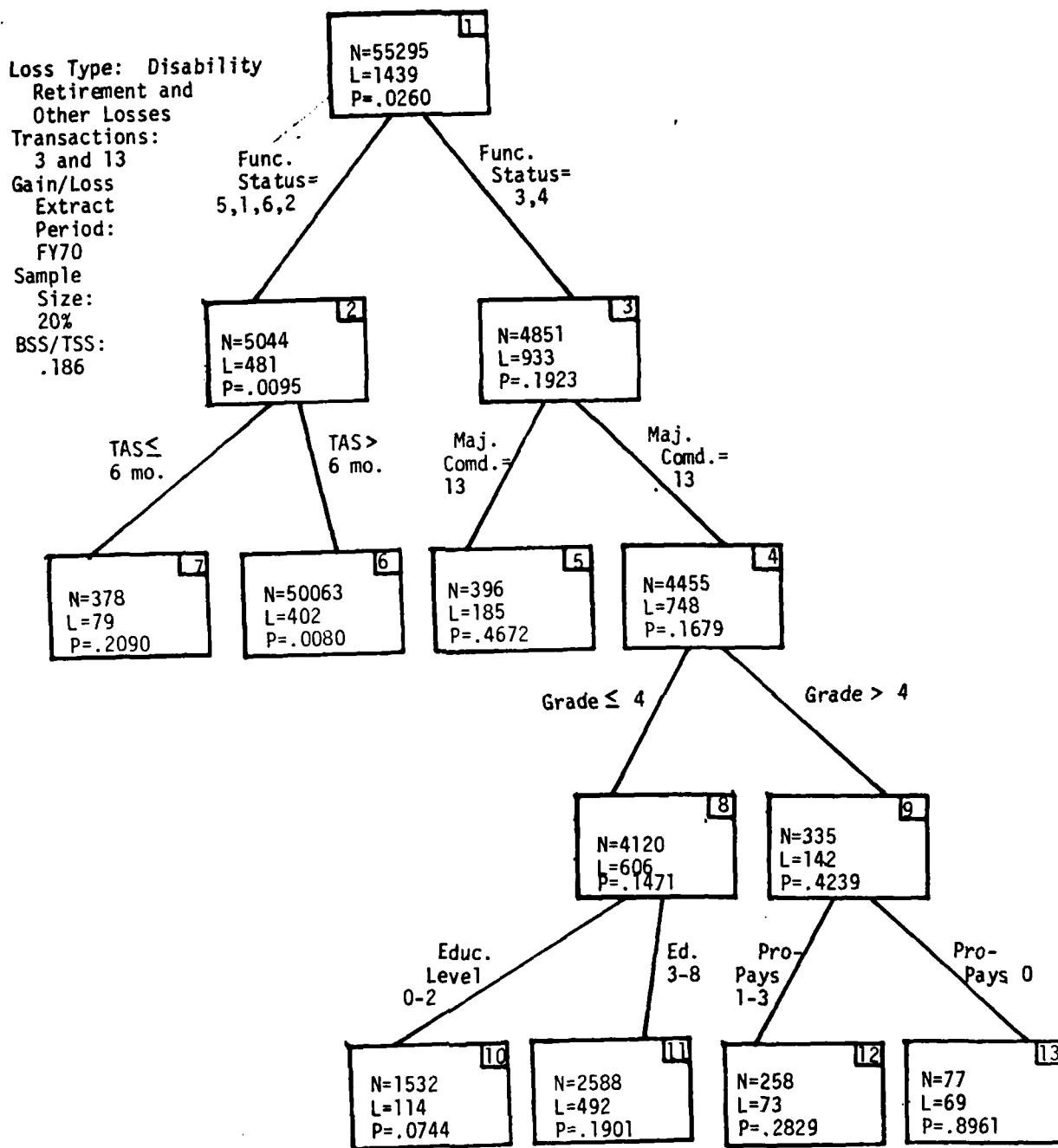
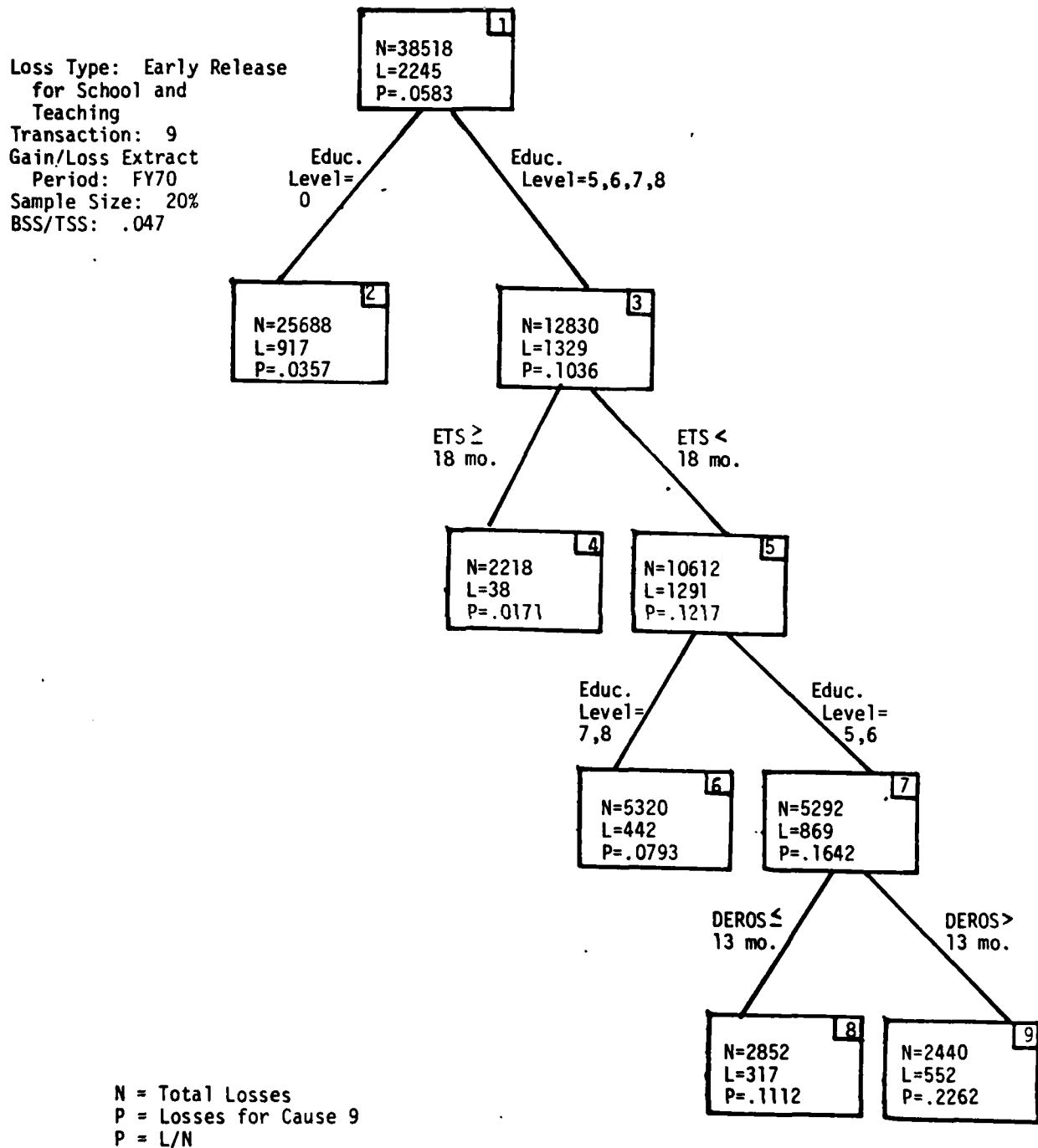
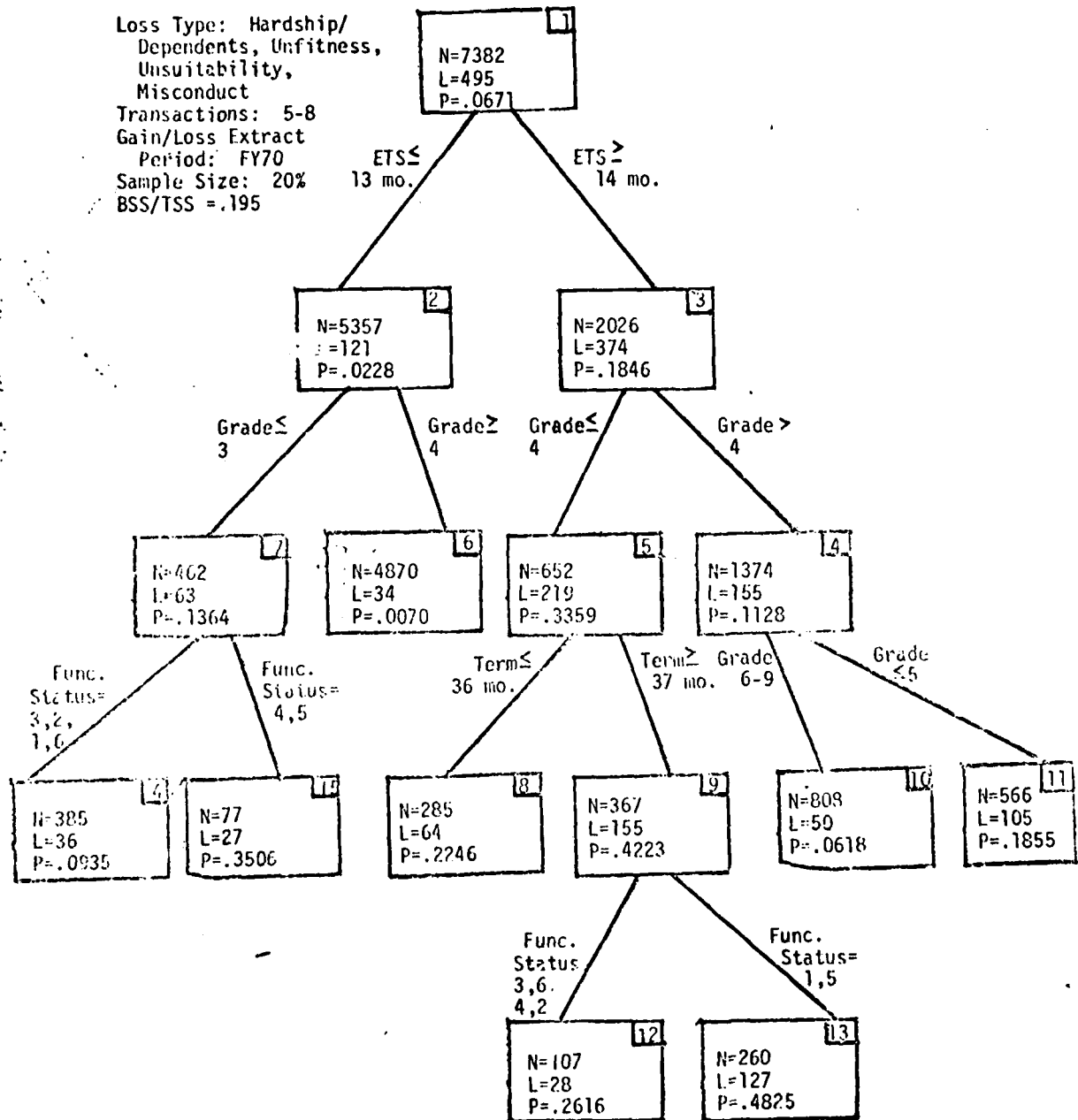


FIGURE 7 - AID TREE FOR FT-RA (L9/L)



See Appendix for definitions of variables.

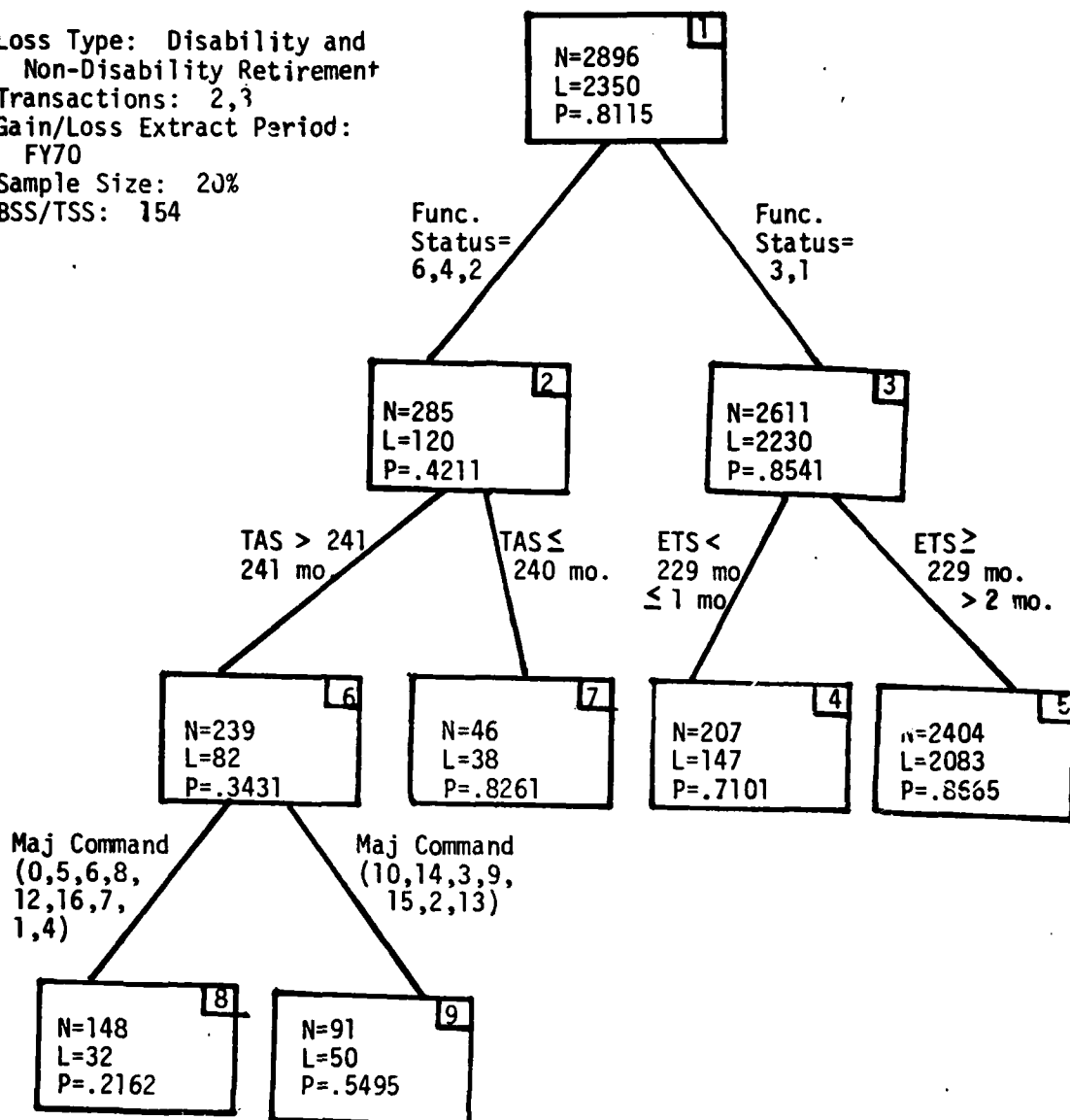
FIGURE 8— AID TREE FOR CAREER (L5-8/L) (FY 1970)*



*See Appendix for definitions of variables.

FIGURE 9 - AID TREE FOR CAREER-RETIREMENT (L2,3/L)

Loss Type: Disability and
Non-Disability Retirement
Transactions: 2,3
Gain/Loss Extract Period:
FY70
Sample Size: 20%
BSS/TSS: 154

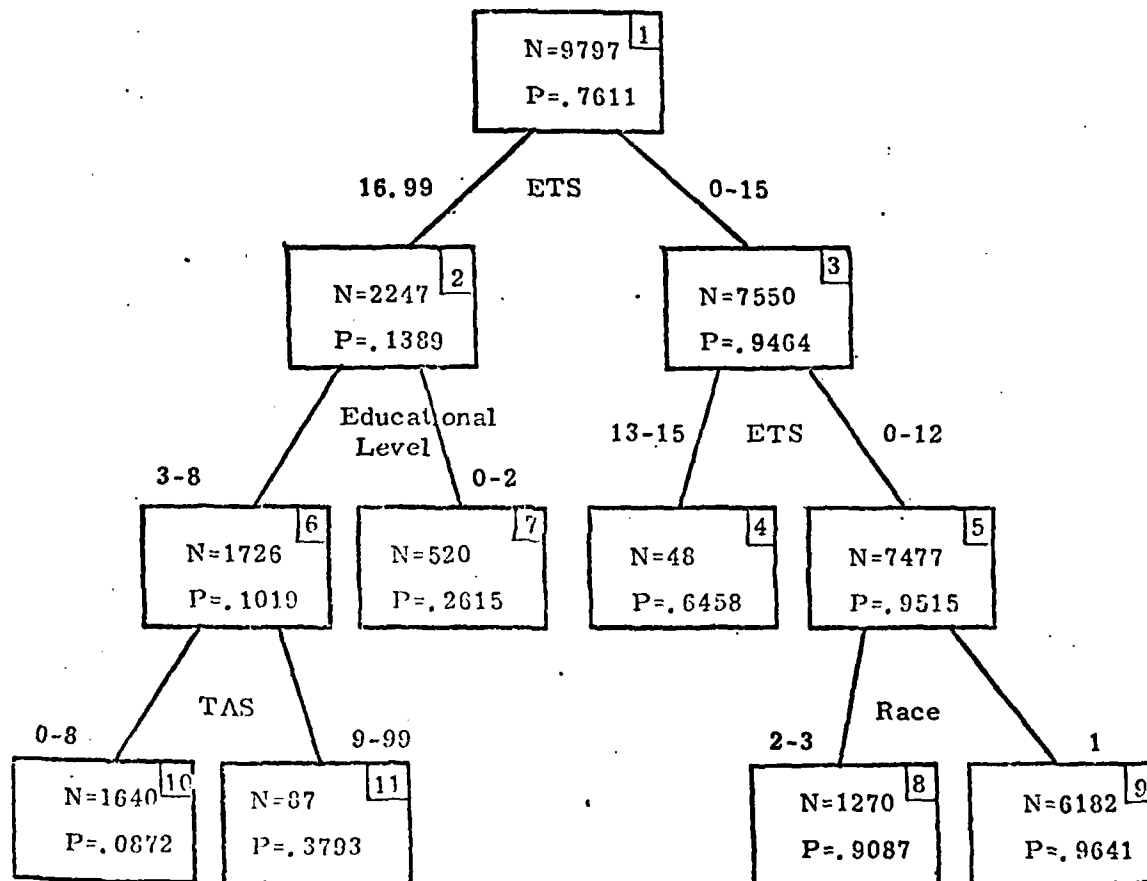


N = Total Losses
L = Losses for Cause 2.3
P = L/N

See Appendix for definitions of variables.

FIGURE 10: AID TREE FOR FT-AUS

Loss Type: All losses
 Transactions: 1-8, 10, 11
 EMF Extract Date: Start FY73
 Gain/Loss Extract Period: FY73
 Sample Size: 10%
 BSS/TSS: .649

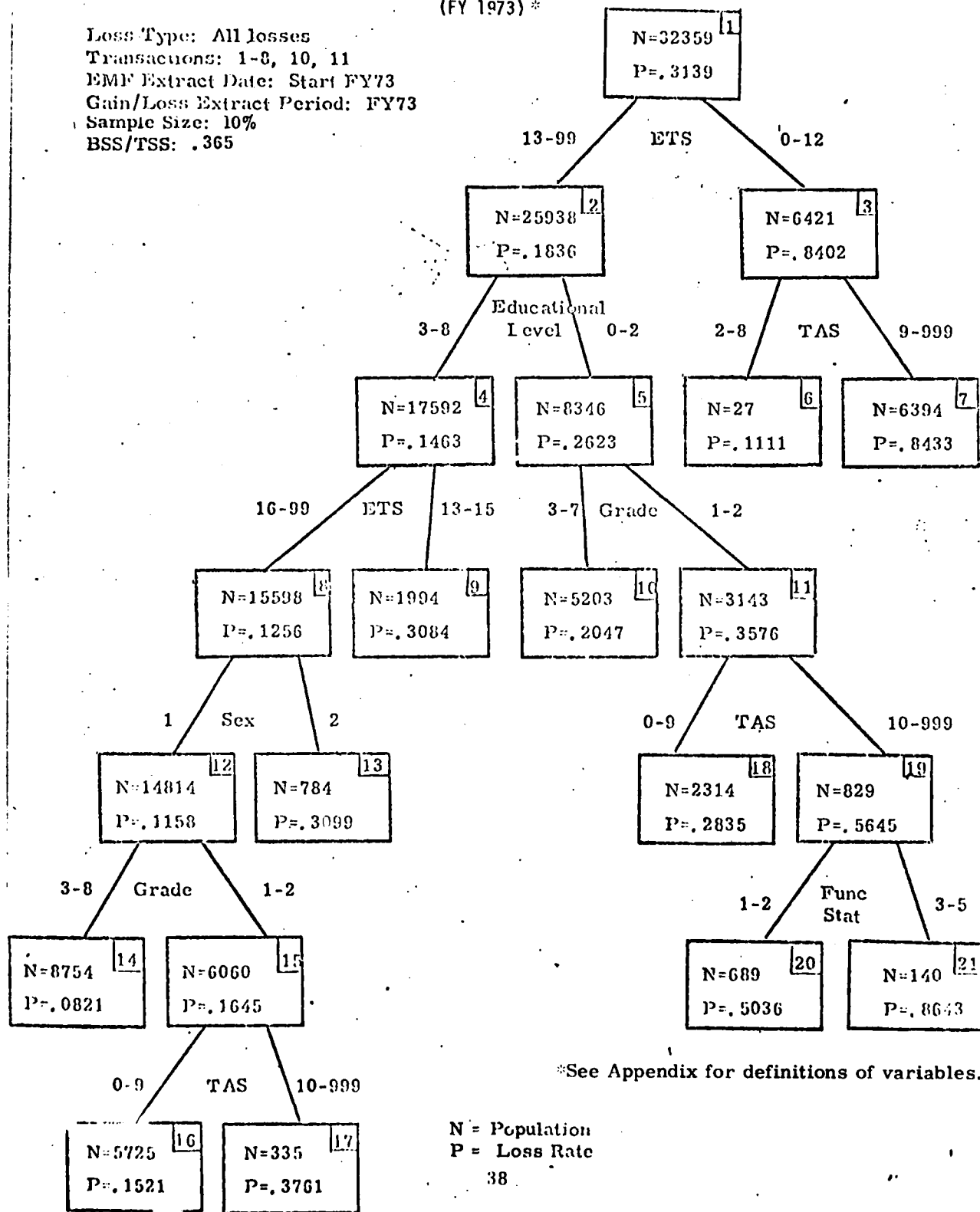


*See Appendix for definitions of variables

N = Population
 P = Loss Rate

FIGURE 11 - AID TREE FOR PT-RA
(FY 1973) *

Loss Type: All losses
Transactions: 1-8, 10, 11
EMF Extract Date: Start FY73
Gain/Loss Extract Period: FY73
Sample Size: 10%
BSS/TSS: .365



*See Appendix for definitions of variables.

FIGURE 12- AID TREE FOR CAREERISTS (FY 1973)

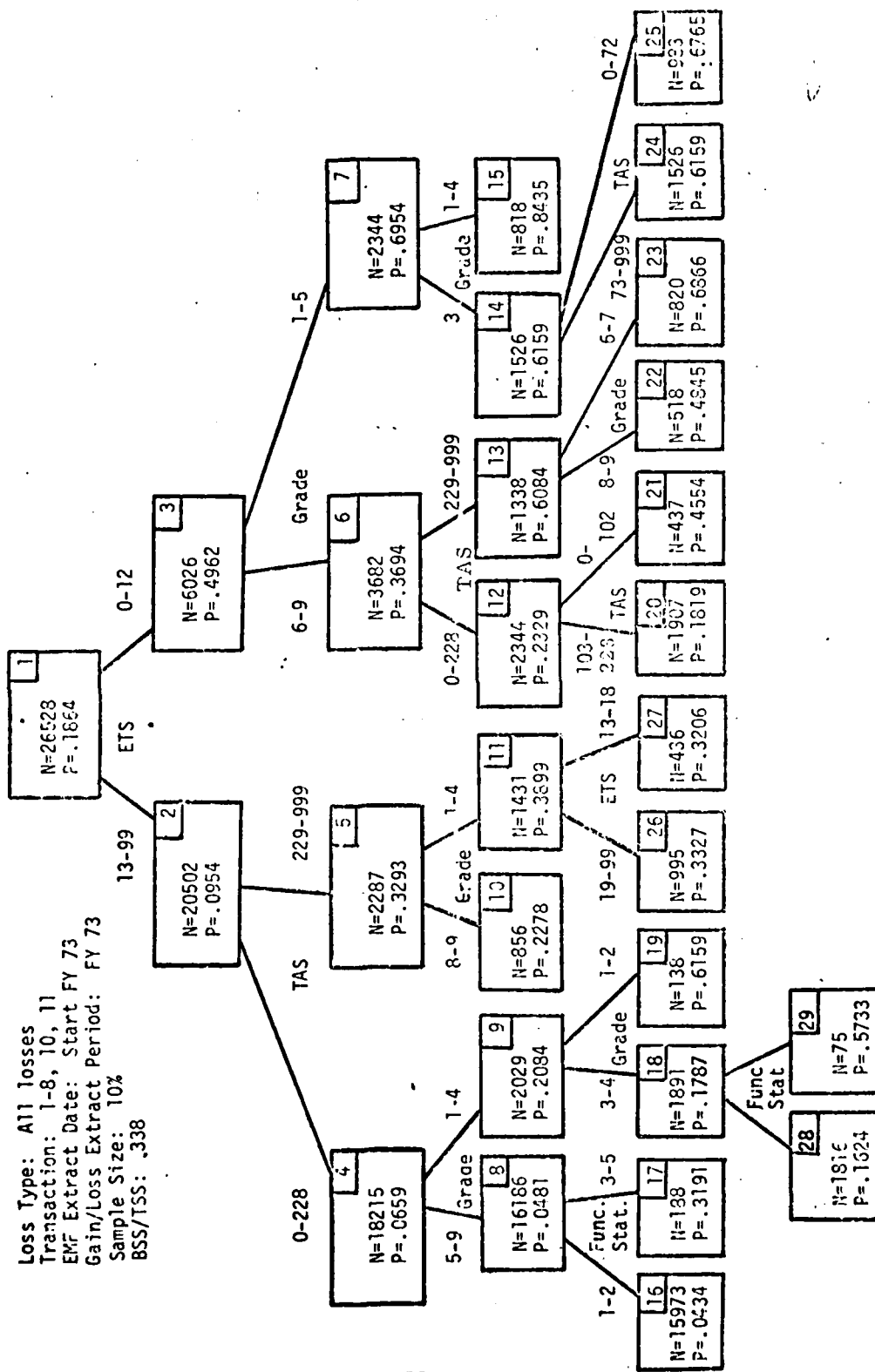
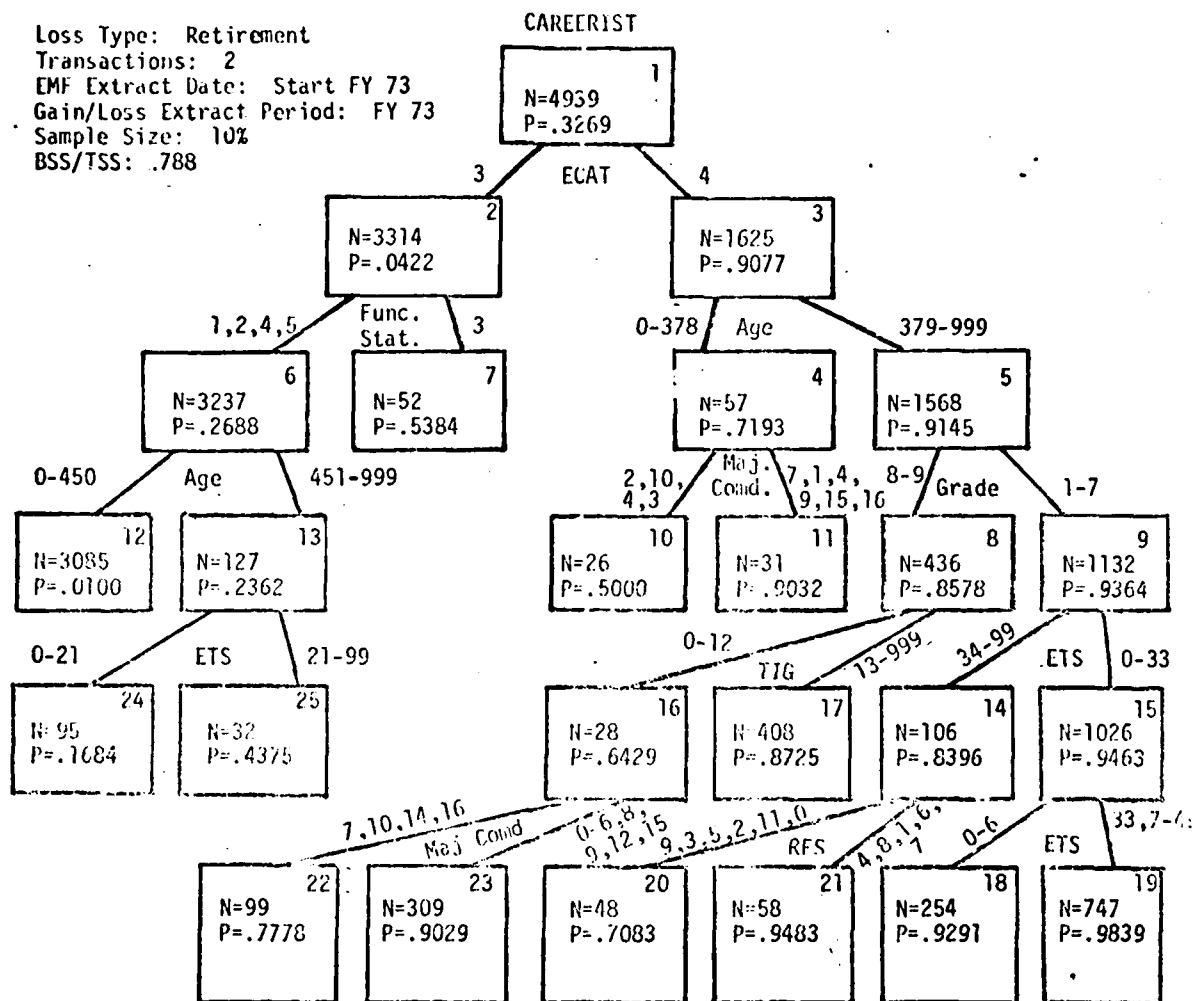


Table 18 contains the AID Tree Recode Form for this tree.
 *See Appendix for definitions of variables.

FIGURE 13- AID TREE FOR CAREERIST (L2/L)
(FY 1973) *

Loss Type: Retirement
Transactions: 2
EMF Extract Date: Start FY 73
Gain/Loss Extract Period: FY 73
Sample Size: 10%
BSS/TSS: .788



N = All losses
P = $\frac{\text{Losses by Retirement}}{\text{All Losses}}$

*See Appendix for definitions of variables.

TABLE 8
COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS)
ENLISTMENT CATEGORY: FT-AUS (FY73)

TRANSACTION CATEGORY: (L/P) All losses

AID RUN DATE: 15 November 1973

AID RUN NUMBER: L/P

SAMPLE SIZE FOR AID RUN:

Random or Not Random

Sample Number; Size; No. of Obs. Read in; Sample #5, 10% (9797)

SAMPLE SIZE FOR COMPUTING ACTUAL LOSSES (OR GAINS)

Comparison Run Date: 30 November 1973

Random or Not Random

Sample Number & Size: Sample #7; 10%

BSS/TSS: .649

MEAN: .7611

Terminal Cell Number	Cell Number	Loss Probability	Number of Persons	Predicted Losses	Actual Losses	Percent* Error
1	10	.0872	1588	138	143	+3.5%
2	11	.3793	72	27	18	-50.0%
3	7	.2615	490	130	126	-3.2%
4	4	.6458	55	36	30	-16.7%
5	8	.9087	1296	1178	1145	-2.9%
6	9	.9641	6111	5892	5852	-.7%
Total			9620	7401	7314	-1.2%

Predicted Overall Loss Rate = .7693

Actual Overall Loss Rate = .7603

$\frac{*Actual - Predicted}{Actual} \times 100$

TABLE 9
COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS)
ENLISTMENT CATEGORY: FT-RA (FY73)

TRANSACTION CATEGORY: All losses

AID RUN DATE: 15 November 1973

AID RUN NUMBER: L/P

SAMPLE SIZE FOR AID RUN:

Random or Not Random

Sample Number; Size; No. of Obs. Read in: Sample #5, 10% (32500)

SAMPLE SIZE FOR COMPUTING ACTUAL LOSSES (OR GAINS):

Comparison Run Date: 30 November 1973

Random or Not Random

Sample Number & Size: Sample #5; 10%

BSS/TSS: .365

MEAN: .3139

Terminal Cell Number	Cell Number	Loss Probability	Number of Persons	Predicted Losses	Actual Losses	Percent* Error
1	14	.0821	8529	700	691	- 1.3
2	6	.1111	21	2	5	+60.0
3	16	.1521	5886	895	931	+ 2.0
4	10	.2047	5346	1094	1005	- 8.9
5	18	.2835	2484	704	758	+ 7.1
6	9	.3084	1870	577	629	+ 8.3
7	13	.3099	783	243	229	- 6.1
8	17	.3761	288	108	86	-25.6
9	20	.5036	671	338	323	- 4.6
10	7	.8433	6474	5459	5472	+ .2
11	21	.8643	148	128	119	- 7.6
Total			32500	10248	10230	.2

Predicted Overall Loss Rate = .3153

Actual Overall Loss Rate = .3148

***Actual - Predicted**
Actual 100

TABLE 10
COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS)
ENLISTMENT CATEGORY: CAREERISTS (FY73)

TRANSACTION CATEGORY: All losses

AID RUN DATE: 25 November 1973

AID RUN NUMBER: L/P

SAMPLE SIZE FOR AID RUN:

Random or Not Random

Sample Number; Size; No. of Obs. Read in: Sample #5; 10% (26, 528)

SAMPLE SIZE FOR COMPUTING ACTUAL LOSSES (OR GAINS):

Comparison Run Date: 29 November 1973

Random or Not Random

Sample Number & Size: Sample #7; 10% (26, 528)

BSS/TSS: .338

MEAN: .1864

Terminal Cell Number	Cell Number	Loss Probability	Number of Persons	Predicted Losses	Actual Losses	Percent* Error
1	16	.0434	16002	692	694	+ .3
2	28	.1624	1876	305	358	+14.8
3	20	.1820	1885	343	321	- 6.8
4	10	.2278	850	194	171	-13.5
5	17	.3191	176	56	54	- 1.8
6	26	.3327	922	307	333	+ 7.8
7	21	.4554	433	197	184	- 7.1
8	22	.4846	486	236	233	- 1.3
9	24	.5064	503	255	243	- 4.9
10	27	.5206	424	220	179	-22.9
11	29	.5733	74	42	36	-16.6
12	19	.6159	137	84	76	-10.5
13	25	.6765	934	632	624	- 1.3
14	23	.6865	827	568	571	+ .5
15	15	.8435	754	636	658	+ 4.9
Total			26283	4767	4730	- .8

Predicted Overall Loss Rate = .1814

Actual Overall Loss Rate = .1800

* $\frac{\text{Actual} - \text{Predicted}}{\text{Actual}}$ 100 43

TABLE II

COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS)
 USING FY70 AID TREES TO PREDICT FY73 LOSSES
 ENLISTMENT CATEGORY: FT-AUS

TRANSACTION CATEGORY: All Losses

AID RUN DATE: July 26, 1973

AID RUN NUMBER: E-7 (L/P)

SAMPLE SIZE FOR AID RUN:

Random or Not Random

Sample Number & Size: No. of Obs. Read in : Sample #1,
 10% (47,153)

SAMPLE SIZE FOR COMPUTING ACTUAL LOSSES (OR GAINS):

Comparison Run Date: December 5, 1973

Random or Not Random

Sample Number & Size #7; (9,800)

BSS/TSS: .631

MEANS: .5724

Terminal Cell Number	Cell Number	Loss Probability	Number of Persons	Predicted Losses	Actual Losses	Percent Error
1	4	.0892	1590	142	160	111.2
2	8	.2485	524	130	75	73.3
3	9	.5336	159	88	43	-104.6
4	10	.7073	614	436	524	16.8
5	7	.9489	6913	6560	6472	1.3
TOTAL			9800	7356	7274	1.1

Predicted Overall Loss Rate = .7506

Actual Overall Loss Rate = .7422

$\frac{\text{Actual} - \text{Predicted}}{\text{Actual}} \times 100$

* Note: See Figure 2 and Table 4 for AID tree.

TABLE 12

COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS)
 USING FY70 AID TRESS TO PREDICT FY73 LOSSES
 ENLISTMENT CATEGORY: FT-RA

TRANSACTION CATEGORY: All Losses

AID RUN DATE: August 6, 1973

AID RUN NUMBER: E-4 (L/P)

SAMPLE SIZE FOR AID RUN:

Random or Not Random

Sample Number & Size: No. of Obs. Read in: Sample #1:
 55,705 (10%)

SAMPLE SIZE FOR COMPUTING ACTUAL LOSSES (OR GAINS):

Comparison Run Date: December 5, 1973

Random or Not Random

Sample Number & Size: Sample #7; 32,362 (10%)

BSS/TSS: .416

MEANS: .3516

Terminal Cell Number	Cell Number	Loss Probability	Number of Persons	Predicted Losses	Actual Losses	Percent* Error
1	2	.1520	25020	3803	3518	-8.1
2	4	.6161	4246	2616	3024	13.5
3	5	.8984	3096	2781	2442	13.9
TOTAL			32,362	9200	8984	2.4

Predicted Overall Loss Rate = .2843

Actual Overall Loss Rate = .2776

$\frac{\text{Actual} - \text{Predicted}}{\text{Actual}} \times 100$

Note: See Figures 4 & 5 and Tables 6 & 7 for Aid Tree.

TABLE 13

COMPARISON OF ACTUAL AND PREDICTED LOSSES (OR GAINS)
 USING FY70 AID TREES TO PREDICT FY 73 LOSSES
 ENLISTMENT CATEGORY: CAREERISTS INELIGIBLE FOR RETIREMENT

TRANSACTION CATEGORY: All Losses

AID RUN DATE: August 6, 1973

AID RUN NUMBER: E-2 (L/P)

Random or Not Random

Sample Number & Size; No. of Obs. Read in Sample #1; 23,027 (10%)

SAMPLE SIZE FOR COMPUTING ACTUAL LOSSES (OR GAINS):

Comparison Run Date: December 5, 1973

Random or Not Random

Sample Number & Size: Sample #7; 22,757 (10%)

BSS/TSS: .413

MEANS: .1604

Terminal Cell Number	Cell Number	Loss Probability	Number of Persons	Predicted Losses	Actual Losses	Percent Error
1	8	.0476	17908	852	806	-5.7
2	9	.3084	231	71	80	11.2
3	7	.3545	628	223	251	11.2
4	12	.0847	1651	140	161	13.0
5	7	.3542	387	137	137	0.0
6	10	.5817	1155	674	689	2.2
7	11	.8560	797	682	614	-11.1
TOTAL			22,757	2,779	2,738	-1.5

Predicted Overall Loss Rate = .1221

Actual Overall Loss Rate = .1203

$\frac{\text{Actual} - \text{Predicted}}{\text{Actual}} \times 100$

Note: See Figure 3 and Table 5 for AID Tree.

Chapter 4

RECOMMENDATIONS

- (a) The results of this study show that the AID-E is the best available method for predicting loss rates: its margin or error for each enlistment category is less than 1.5% in FY73 and less than 3.5% in FY70. For year-to-year predictions, the AID trees developed from FY70 data were used to predict losses in FY73. The results show that the prediction errors are less than 2.4%. For this and other reasons (among them, consistency and flexibility of the system), GE-TEMPO recommends that the AID-E be designated the official method for supplying loss rates to all the Army's manpower prediction models.
- (b) The AID-E system is best used to produce annual loss rates. Great care must be exercised in converting its annual rates into monthly or quarterly rates. In order to preserve consistency in the use of AID loss rates in all manpower models, the Army cannot be too strongly urged to designate an official agency in charge of providing AID loss rates for monthly, quarterly, and annual manpower prediction systems.
- (c) The AID-E is not only valuable in producing loss rates, but is also an excellent system to provide immediate information for various Army manpower studies. Any data elements contained in both the EMF and the Gain/Loss Transaction Files can be readily summarized, analyzed, and tabulated: for example, (1) the problem of AWOL and MOS imbalance/mismatch; (2) distribution of the enlisted population by age, sex, geographic origin, marital status, performance index, and others. GE-TEMPO recommends that the AID-E be designated as a key system to be used in manpower policy making studies. The system is capable of processing large quantities of data to produce accurate information in a timely manner. It should be noted that GE-TEMPO has already installed the AID-E in an Army facility.
- (d) The DCSPER-46 Report (the official report of the Army's manpower strength status) can be improved in reporting losses in a significant respect. It reports only the number of incidents of loss from various causes. If a person takes five DFMC's (drop from military control) and subsequently returns, five losses are reported. This practice is acceptable if the Army is interested simply in the number of incidents, but it obscures the true number of persons lost. GE-TEMPO recommends that, if the Army continues the present practice of counting incidents, the DCSPER-46 Report be modified to: (1) clearly distinguish permanent losses from incidents (temporary) of loss and (2) include the number of persons versus the number of incidents for loss categories such as DFMC where more than one loss for each individual can occur. The information for both number of incidents of loss and number of persons lost is of value depending on the purpose.

- (e) Testing of AID-E rates by GE-TEMPO has been most promising, but the evaluation of accuracy and validity of the AID-E loss rates should be continued by the Army. Such extensive testing and operational use of the AID-E system will highlight the strengths and shortcomings of rates generated. An office should be identified to document these strengths and shortcomings for user guidance.

APPENDIX

Definitions of Variables Used in Figures 2-13

ETS - Number of months remaining prior to expiration of term of Service

TAS - Number of months in total active military service

Major Command - The command or agency to which an individual is assigned.

<u>Code</u>	<u>Meaning</u>
01	First US Army
02	Third US Army
03	Fifth US Army
04	Sixth US Army
16	US Air Defense Command
14	The Adjutant General
05	US Army Alaska
15	Army Security Agency
14	Army Audit Agency
06	US Army, Southern Command (other than Antilles Cmd.)
06	US Army, Antilles Command
01	HQ's US Army Continental Army Command
13	US Army Strategic Communications Command
13	US Army Combat Development Command
14	Chief, Engineers
14	Office, Chief of Staff, Army
07	HQ's US Army, Europe
07	US Army Theater Army Support Command, Europe
07	SETAF
07	Berlin Command
07	Seventh US Army
13	US Army Intelligence Command
12	US Army Material Command
12	US Army Electronics Command
12	US Army Missile Command
12	US Army Tank-Automotive Command
12	US Army Munitions Command
12	US Army Aviations Systems Command
12	US Army Test & Evaluations Command
12	US Army Weapons Command
12	US Army Mobility Equipment Command
13	The Surgeon General
13	Military Traffic Management Terminal Service

<u>Code</u>	<u>Meaning</u>
14	Military District of Washington
08	HQ's US Army, Pacific
08	US Army, Hawaii
09	US Army, Japan
09	US Army, Ryukyu Islands
10	HQ's US Army, Vietnam
10	HQ's US Army Control, Thailand
11	US Army Forces, Taiwan
09	Eight US Army
14	Office of the Secretary of the Army
14	Dept. of Defense & US Army Elements & Joint Activities
14	Other Field Activities of the Army Staff
14	Chief, Support Services
00	Blank or Invalid

Functional Status - Status of active personnel by state of readiness for use in specific roles, missions, and activities.

<u>Code</u>	<u>Meaning</u>
1	US Army Air Defense Command (Admin. elements)
1	US Army Air Defense Command (Oper. elements)
1	STRAF-readiness capability (REDCAPE)
1	" " " "
1	" " " "
1	" " " "
1	Continental United States (operations)
3	Patients
2	En Route
4	Personnel entering military service
1	Special foreign activities
3	Foreign Patients
1	Joint Task Force
5	US Army Correctional Holding Detachment
2	Personnel in process of separation
2	Replacements
1	Reimbursable activities
2	Returnees for reassignment
4	Skills Development Base student
1	Overseas Troops
2	Transient
4	Students
9	Invalid or Blank

Grade - Pay Grade

<u>Code</u>	<u>Meaning</u>
1	E-1
2	E-2
3	E-3
4	E-4
5	E-5
6	E-6
7	E-7
8	E-8
9	E-9

ETS-DEROS - Number of months left in an individual's ETS after returning from overseas ETS.

Educational Level - The highest level of civilian education attained by an enlisted man.

<u>Code</u>	<u>Meaning</u>
0	No formal education
0	One year elementary school completed
0	Two years elementary school completed
0	Three years elementary school completed
0	Four years elementary school completed
0	Five years elementary school completed
0	Six years elementary school completed
0	Seven years elementary school completed
0	Eight years elementary school completed
1	One year High School completed (9th)
1	Two Years High School completed (10th)
2	Three years High School completed (11th)
2	Four years High School completed (12th) but did not graduate.
3	High School graduate (irrelevant to number of years completed)
4	General Education Development Level (High School)
5	General Education Development Level (College)
5	One year college completed
5	Two Years college completed (also graduate of 2 yr. college)

<u>Code</u>	<u>Meaning</u>
5	Three years college completed
5	Four years college completed but did not graduate
6	College graduate (irrelevant to years comp.)
7	Bachelor of Laws (LLB)
7	Doctor of Laws (LLD)
7	Juris Doctor (J.D.)
7	Doctor of Juridical Science (J.J.S)
7	Graduate work of one year or more completed, no degree
7	Masters Degree
7	Doctorate Degree
7	Other Professional degree
8	Blank or Invalid

Prop-Pays - Proficiency Pays which are the ratings given to enlisted personnel eligible to receive pays related to his skill or performance.

<u>Code</u>	<u>Meaning</u>
0	Withdrawal of Propay
1	P-1, the lowest level of skill-related propay
2	P-2, the second lowest level of skill-related propay
3	P-3, the highest level of skill-related propay
1	P-4, Superior Performance Pay
0	Unknown Propay

DEROS - Number of months remaining from the EMF date to the date on which an individual is eligible to return to CONUS or area of residence from overseas. A person not in overseas has zero month in DEROS.

Term - Term of service in terms of months.

Race - Race of an enlisted personnel.

<u>Code</u>	<u>Meaning</u>
1	Caucasian
2	Negro
3	Other & Unknown

Sex - Sex of an enlisted personnel.

<u>Code</u>	<u>Meaning</u>
1	Male
2	Female
3	Unknown
4	Error

ECAT - Enlistment category which assigns an enlisted man into one of the following codes:

<u>Code</u>	<u>Meaning</u>
1	First-term Army of the United States (draftees)
2	First-term Regular Army
3	Careerists ineligible for retirement
4	Careerists eligible for retirement

Age - Age of an enlisted man in terms of months.

Residence (RES) - State or geographical area of residence at the time of entry into the Army.

FIRST U.S. ARMY

<u>STATE</u>	<u>CODE</u>
Connecticut	7
Delaware	8
Kentucky	1
Maine	7
Maryland	8
Massachusetts	7
New Hampshire	7
New Jersey	2
New York	2
Ohio	0
Pennsylvania	2
Rhode Island	2

<u>STATE</u>	<u>CODE</u>
Vermont	7
Virginia	8
West Virginia	8

THIRD U.S. ARMY

<u>STATE</u>	<u>CODE</u>
Alabama	1
Florida	8
Georgia	8
Kentucky	1
Mississippi	1
North Carolina	8
South Carolina	8
Tennessee	1

FIFTH U.S. ARMY

<u>STATE</u>	<u>CODE</u>
Arkansas	4
Illinois	0
Indiana	0
Iowa	0
Kansas	5
Louisiana	4
Michigan	0
Minnesota	9
Missouri	5
Nebraska	5
New Mexico	3
Oklahoma	4
Texas	4
Wisconsin	5

SIXTH U.S. ARMY

<u>STATE</u>	<u>CODE</u>
Arizona	3
California	6
Colorado	3
Idaho	3
Montana	3
Nevada	6
North Dakota	9
Oregon	3
South Dakota	5
Utah	3
Washington	3
Wyoming	3

U.S. ARMY MILITARY DISTRICT OF WASHINGTON

	<u>CODE</u>
District of Columbia	2
Maryland	8
Virginia	8
Alaska	9
Hawaii	9

DATE
ILME
—8