NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

A FRAMEWORK FOR ARMY RESERVE RECRUITING ANALYSIS: ENLISTMENT TO INITIAL TRAINING

by

Harvey C. Denison

June, 2003

Thesis Advisor:	Susan M. Sanchez
Second Reader:	David H. Olwell

Approved for Public Release; Distribution is Unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** June 2003	**3. REPORT TYPE AND DATES COVERED** Master's Thesis

4. TITLE AND SUBTITLE: **A Framework for Army Reserve Recruiting Analysis: Enlistment to Initial Training**	**5. FUNDING NUMBERS**
6. AUTHOR(S) MAJ Harvey C. Denison	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of the Commander, U.S. Army Reserve 1421 Jefferson Davis Highway, Arlington, VA 22202	**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)*** Analysis of U.S. Army Reserve recruiting is conducted across the U.S. Army with data from the Recruit Quota System (REQUEST). A combination of partial manual data entry and a decided lack of tools for large scale data extraction make REQUEST difficult to use for analysis without an extensive knowledge of the system. In this thesis, I develop a process for screening, preparing, and evaluating REQUEST data for subsequent analysis. This process uses data mining software to progressively work through a series of rules that outline data inconsistencies, mark these records for exclusion and later investigation, and generate a "clean" dataset for analysis. I examine enlistments over a four year period with respect to Military Occupational Specialty and training program structure. Data from the Army Training Requirements and Resource System (ATRRS) are used to provide an overview of Initial Entry Training seat quotas and usage, and to confirm and/or update training dates in the REQUEST dataset. The joint examination of enlistments and training seats provides new insights into enlistment patterns.

Additional analysis is possible using demographic data provided by the U.S. Army Recruiting Command. I provide summaries of a few key demographic variables for various subsets of the enlistees, and discuss how similar analyses might prove useful for targeting recruiting efforts and incentives more effectively.

Good decisions require good data. This thesis is a start in providing a framework for generating quality USAR accession data for analysis.

14. SUBJECT TERMS USAR, recruiting, demographics, data mining, data preparation	**15. NUMBER OF PAGES** 162
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION 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

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for Public Release; Distribution is Unlimited

A FRAMEWORK FOR ARMY RESERVE RECRUITING ANALYSIS: ENLISTMENT TO INITIAL TRAINING

Harvey C. Denison
Major, United States Army
B.S., The Citadel, 1986

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 2003

Author: Harvey C. Denison

Approved by: Susan M. Sanchez
Thesis Advisor

David H. Olwell
Second Reader/Co-Advisor

James N. Eagle,
Chairman, Department of Operations
Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Analysis of U.S. Army Reserve recruiting is conducted across the U.S. Army with data from the Recruit Quota System (REQUEST). A combination of partial manual data entry and a decided lack of tools for large scale data extraction make REQUEST difficult to use for analysis without an extensive knowledge of the system. In this thesis, I develop a process for screening, preparing, and evaluating REQUEST data for subsequent analysis. This process uses data mining software to progressively work through a series of rules that outline data inconsistencies, mark these records for exclusion and later investigation, and generate a “clean” dataset for analysis.

I examine enlistments over a four year period with respect to Military Occupational Specialty and training program structure. Data from the Army Training Requirements and Resource System (ATRRS) are used to provide an overview of Initial Entry Training seat quotas and usage, and to confirm and/or update training dates in the REQUEST dataset. The joint examination of enlistments and training seats provides new insights into enlistment patterns.

Additional analysis is possible using demographic data provided by the U.S. Army Recruiting Command. I provide summaries of a few key demographic variables for various subsets of the enlistees, and discuss how similar analyses might prove useful for targeting recruiting efforts and incentives more effectively.

Good decisions require good data. This thesis is a start in providing a framework for generating quality USAR accession data for analysis.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ABBREVIATIONS, ACRONYMS, SYMBOLS

54B – Chemical Operations Specialist.

91S – Preventive Medicine Specialist.

95B – Military Policeman.

Accession – New recruit enlisted into the U.S. Army Reserve.

AIT – Advanced Individual Training: Second phase of training a new soldier receives to learn a specific specialty.

Applicant – Potential enlistee into the U.S. Army Reserve.

ARPERSCOM – U.S. Army Reserve Personnel Command: Headquarters responsible for personnel records, systems and management in the U.S. Army Reserve.

ATRRS – Army Training Requirements and Resource System: System that manages the listing of all individual training school classes and enrollments for the U.S. Army.

BCT – Basic Combat Training: First phase of training a new recruit must complete before learning a specialty and becoming a qualified member of the Army.

DA – Department of the Army.

DAPE-MPT – Department of the Army Personnel for Manpower, Personnel, and Training.

DCSPER – Deputy Chief of Staff for Personnel.

EB – Enlistment Bonus: Sum of money ($3,000/5,000/ 8,000) paid to some recruits over the course of an enlistment.

IET – Initial Entry Training: Training a new soldier must complete to be a soldier in the U.S. Army qualified in his or her assigned specialty.

IRR – Individual Ready Reserve: Encompasses several categories of Army Reservists who are not currently assigned to a deployable unit and not attending monthly drills.

Market Segments – Data from a commercial source that clusters markets by nine-digit zip code into 50 groups with similar demographics and buying patterns.

MGIB – Montgomery GI Bill: Educational fund available to any member enlisting in the U.S. Armed Forces that accrues for each month in service.

MGIB Kicker – Enlistment incentive used by the U.S. Army Reserve that adds an additional amount ($100/200/300) to the monthly amount set aside by the MGIB.

MEPS – Military Entry Processing Station: Place where applicants are tested, qualified and enlisted into the military.

MOS – Military Occupational Specialty: Specific specialty that a soldier is qualified in or will be qualified in once IET is complete.

NPS – Non Prior Service: Enlistee with no prior military service.

OCAR – Office of the Commander of the Army Reserve: Headquarters for the U.S. Army Reserve that makes policy and provides oversight for all U.S. Army Reserve forces.

OSUT – One Station Unit Training: IET for selected specialties that combine BCT and AIT into a single course at a single location.

PS – Prior Service: Enlistment where applicant had qualifying prior service in either the active or reserve components of the U.S. Army.

REAF – Reserve Enhanced Applicant File: Data source created and managed by the U.S. Army Recruiting Command combining various sources of accession data to include data from the Military Entry Processing Command, the Recruit Quota System, and commercial market segmentation data.

REQUEST – Recruit Quota System: System the U.S. Army Reserve uses to match vacant positions, available training seats, and applicants to enlist personnel into the U.S. Army Reserve.

SAMAS – Standard Army Manpower and Authorization System.

SLRP – Student Loan Repayment Program: Enlistment incentive in which a sum of money ($10,000/20,000) is used to repay the recruit's eligible student loans.

Split-Option Program – Special reserve component enlistment program where the individual attends BCT or phase 1 OSUT during one year, and AIT or phase 2 OSUT during the following year.

Straight-Through Program – Enlistment option where enlistee attends BCT and AIT in succession, or completes OSUT.

TAADS-R – Total Army Authorization Document System – Reserve: System that generates authorizations for each unit in the USAR.

Tier – Priority indicator for U.S. Army Reserve Units for resource allocation and readiness standards.

Training Seat – A slot to attend a training school for a given class starting for a given start date.

UIC – Unit Identification Code: Six digit code that uniquely identifies a unit, with Army units beginning with a W* and major or "parent" units ending with *AA.

USAR – United States Army Reserve.

USAREC – U.S. Army Recruiting Command: Organization that recruits all NPS accessions for the USAR.

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I would first like to thank Colonel Dennis Scott, Colonel Terry Lerch, Joyce Dyer, and Major Ward Litzenberg for having the Office of the Commander of the Army Reserve sponsor my research.

Secondly, I would like to thank Master Sergeant Patrick Sarley, the REQUEST NCO at the Army Reserve Personnel Command. Without his dedicated support, knowledge, and expertise I would not have had any data to support my research nor his vast experience with REQUEST to help me analyze the possible problems with the data. I owe him hearty thanks and a debt of gratitude.

I would also like to thank my advisor, Professor Susan Sanchez. Her expertise, commitment, and patience are in large part responsible for my research resulting in an intelligible finished product. I can't thank her enough.

Professor David Olwell, my second reader, deserves a great deal of credit for bringing me out of the "weeds" and re-focusing me on relating my thesis back to helping the Army Reserve. His keen eye and sharp mind were of immeasurable value to making this thesis relevant.

I could not have been anywhere near the mark on any of my analysis without the sage advice of Mr. Dennis Sullivan. He functioned as a tremendous source for knowledge of the U.S. Army Reserve Personnel systems, past and present. He also was my sounding board for ideas and a safety net to keep me "on the straight and narrow". He is a trusted counsel and a good friend.

My acknowledgements are not complete without thanks to my mentor on USAR accessions analysis, Mr. Charles Dalbec. His tremendous experience and knowledge of USAR personnel systems, recruiting, and REQUEST are a major reason I know anything about the USAR and recruiting. He also assisted in running down some data problems that were important in the overall scheme of my research. Thanks Chuck.

I would also like to thank Major Mike Kamei and my friend Rae Disney at U.S. Army Recruiting Command for the great assistance, tremendous cooperation, and good advice.

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

The U.S. Army Reserve fills a majority of its entry level positions in units across the Unites States through the efforts of the U.S. Army Recruiting Command and the Military Entry Processing Command. A Reserve enlistment is recruited to a specific position in a unit, he or she does not just join the Reserves. An applicant is only eligible to enlist in positions within a nearby unit for a Military Occupational Specialty (MOS) that has Initial Entry Training (IET) opportunities or "school seats" available. An applicant's choice is affected by the positions and MOSs available in local units, training seats available for the position specialty and starting date, enlistment incentives for different positions, the training program, and a range of others. The vacant positions by unit and by specialty, the availability of training, and the enlistment incentive are all aspects that are presented to the applicant by the guidance counselor at the Military Entry Processing Station (MEPS) from a system called the Recruit Quota System (REQUEST).

For an analyst, REQUEST is the source of choice to conduct analysis on new enlistments or accessions into the Army Reserve. But the REQUEST system is often populated by many duplicate records for a single accession, so generating a valid dataset for analysis is difficult. There are systems that "roll up" these data into a finite set such as the Reserve Component Manpower System (RCMS), but none offer insight into "how we got there." There is no understanding of the steps taken to produce this data,

what was lost and why, or what common problems were encountered. Given the complex nature of the REQUEST data, this thesis generates a reusable process to screen raw queries from the REQUEST data to generate a “clean dataset” with information about the preparation process, and uses the data to conduct a sample analysis relating REQUEST data to the IET data.

An important part of this process is the handling of the training program referred to as split-option training. Split-option training occurs when the two phases of IET are conducted separately, generally a year apart, as opposed to straight-through training in which both phases are conducted consecutively. The split-option training enlistments constitute a large portion of the duplicate and inconsistent records in REQUEST, and require more attention in the data preparation process.

The process dramatically reduces the number of duplicates and inconsistent records, and provides an overview of the number and types of problems screened out.

Additional data for IET training containing USAR quotas and inputs to training are included in the analysis to provide an overview of IET training by the different categories, and to corroborate the IET related data in REQUEST. The data are binned by month and examined with respect to the ratio of inputs to quotas (or quota usage) for various MOS by training program over time. The quota usage is used to identify those MOSs with consistently high quota usage, such as the Military Policeman (95B MOS), and some that have a consistently low usage, such as the Preventive Medicine Specialist (91S MOS). Seasonal

patterns were suggested with consistently low usage in February and consistently high usage in June and July. Split-option phase 1 training quota usage for Basic Combat Training (BCT) and phase 1 One Station Unit Training (OSUT) were found to be consistently high, yet the phase 2 quota usage rates much lower. Comparisons of phase 1 and phase 2 training inputs suggest an average completion rate for IET by split-option trainers to be low. The definite lack of scheduling of phase 1 split-option recruits for their phase 2 AIT or OSUT is a significant issue which is the primary cause for the low phase 2 split-option quota usage.

With a picture of IET training seat usage, the REQUEST data was analyzed to look at relationships between month of enlistment and month of the start of IET training. The average delay in days between enlistment and training start was added to the data fields for analysis. Once again, delays from the time of enlistment indicated a low density of enlistments for February, and a high density for the summer months.

Demographic data used by the U.S. Army Recruiting Command for marketing analysis, called the market segments, were added to the data available in REQUEST. These segments outline different commercial markets by various demographic characteristics, and are coded to an accession record depending on the expanded nine-digit zip code address of the applicant. These market segments, in conjunction with the training seat usage, delay from enlistment to training start, and quantitative variables such as age, AFQT score, and years of education can provide a picture of the accession population for a specialty.

Understanding the accession population demographics with respect to training seat usage can provide useful information with regards to the recruiting process, and provide insight into policy decisions such as enlistment incentives and training seat quota management.

Good data are necessary for good decisions. And as the data get aggregated, the aggregation process offers important information about the system. These insights can in turn be used for system improvements and to provide knowledge of the strengths and weaknesses of the data. The USAR needs to take advantage of the data mining capabilities outlined in this thesis to improve the data used to conduct analysis on accessions and training seat management in an integrated manner.

I. BACKGROUND

The United States Army Reserve (USAR) is a force provider, in that it is a source of units to meet missions assigned to the U.S. Army. These units are evaluated on, and must meet, certain readiness requirements in personnel, equipment, and training. In order to be ready to deploy, they must have trained personnel available. There are several ways units acquire the personnel they need, but the majority of personnel in the USAR are recruited into entry level positions by the U.S. Army Recruiting Command (USAREC). The topic of this thesis is to examine this process and understand the major influences that affect it.

The way the USAR operates with regard to manning is very different from the active component of the U.S. Army. The active component of the Army recruits the personnel they need, sends them to individual training, and then distributes them world-wide to the force as the Army needs. The USAR, on the other hand, recruits individuals into specific positions in specific units at specific locations.

The USAR recruits from two distinct populations, defined as Prior Service (PS) and Non Prior Service (NPS). The first population consists of individuals in the Individual Ready Reserve (IRR) who have already completed all initial training requirements to be a qualified soldier. These individuals have already served in either the active or reserve components of the U.S. Army. They are placed into a vacant position in a local unit and transferred from the IRR into the selected reserve. The second population has no prior Army experience or equivalent, and is recruited and inducted to the USAR with

appropriate initial training scheduled at time of the enlistment. The process for NPS Accessions is the focus of my analysis.

Here is how the NPS recruiting process works. A recruiter encourages a potential applicant to consider joining the USAR, and schedules the individual to visit the local Military Entry Processing Station (MEPS) to be evaluated physically and mentally for potential enlistment into the USAR. Once evaluated, the individual meets with a career guidance counselor, who assists the applicant in choosing a job position.

This process sounds relatively simple, but the portion where the applicant sits down with the guidance counselor to select a position is the key event of interest. The Guidance Counselor shows the positions available to the applicant using the Recruit Quota System (REQUEST). This system lists all positions in local reserve units, based on the current address zip code for the applicant, that are vacant and have an available Initial Entry Training (IET) school seat for the position's Military Occupational Specialty (MOS). The MOS is usually represented by a three-digit alphanumeric code (a list of U.S. Army MOS codes is attached in Appendix 1). The training school seat information is obtained through a link with the Army Training Requirements and Resource System (ATRRS). Also, some unit-MOS combinations will have an associated enlistment incentive associated with the position.

This presents several problems in recruiting new soldiers for the USAR. A potential enlistee to the USAR is limited in choice of MOS based on vacancies in units within 75 miles of their current address. This requirement

can be waived under certain conditions, but highlights the geographic problem associated with recruiting. The training availability can potentially limit the applicant's choices, and the incentive can also affect which position the applicant will choose.

The U.S. Army conducts IET at various locations across the United States. It is split into two portions: Basic Combat Training (BCT) and Advanced Individual Training (AIT). For some specialties, both portions are completed at the same location. This form of training is referred to as One Station Unit Training (OSUT). For classification purposes it is split up into two portions: phase 1 meeting the BCT Requirements, and phase 2 meeting the AIT requirements.

Additional complications are created by the split-option training program. Split-option trainees go to BCT (or phase 1 OSUT) in one summer, and their AIT (or phase 2 OSUT) the following summer. There are a number of issues associated with this program in terms of the scheduling of training and the entry of this information into REQUEST. These problems have caused difficulty in assembling the data necessary for the conduct of my analysis.

The three major elements listed above; i.e., unit location, training seat availability, and enlistment incentives, are the factors on the USAR side that affect the recruiting process. The other side of the recruiting piece relates to demographics and their effect on the enlistment choices.

The rest of this thesis is organized as follows. In Section II, I describe the methodology used to prepare for

and conduct the analysis. In Section III, I discuss the data sources and the data preparation process. Section IV provides an overview of the ATRRS IET training data and an analysis of the REQUEST based IET data as it relates to enlistments. In Section V, I look at some demographic data for the entire population, as well as for a few selected specialties. The last section contains recommendations and conclusions. There are four appendices which provide the descriptions for the USA Army MOSs (Appendix 1), the details for the REQUEST portion of the data preparation (Appendix 2),the data definitions for the accessions data (Appendix 3), and the market segment definitions (Appendix 4).

II. METHODOLOGY

To begin the analysis of the recruiting process, the first step 1s data collection and preparation. As the data sources are many and their quality is an issue, this is the major portion of my thesis work. During my thesis research, I visited the major organizations that have provided the data necessary. The data sources include training seat, recruiting, personnel, unit-specific data, and demographic data. I have chosen to work with recruiting data from fiscal year (FY) 1999 through the end of FY 2002. An additional year of data from FY 1998 was used to determine training seat availability for FY 1999 based on those who enlisted in FY 1998 but started training in FY 1999. The combination provides four years of accessions data and REQUEST based training data for analysis. The data preparation includes cleaning and validation of these data, as well as converting them into formats more amenable to analysis. A key product of my thesis is a process that can be implemented to assist in the preparation of data for future USAR recruiting accession data analysis, either by students, the Office of the Commander of the Army Reserve, or other organizations that conduct analysis on USAR recruiting.

The initial analysis of the ATRRS training seat data provides on overview of training seat quota availability and usage. The deeper analysis of training seat data uses REQUEST based training seat data to compare training seat usage over time relative to enlistment month. The time unit for the analysis is the month, so all data are binned

by month by FY for purposes of comparison and temporal analysis.

The initial demographic analysis of the NPS accessions for the USAR provides a summary of statistical information relevant to the recruits who have joined the USAR. The analysis then compares and contrasts some quantitative and qualitative demographic data for enlistees in three sample MOSs as well as the entire accession population. Additional possibilities for use of the demographic data are also discussed.

III. DATA PREPARATION

A. DATA SOURCES

To look at the recruiting process, I obtained data from a number of sources.

1. Headquarters, Department of the Army Personnel for Manpower, Personnel and Training (DAPE-MPT)

DAPE-MPT provided a quota and training input summary for each BCT, AIT, and OSUT class conducted for FY99-02. Mr. Alan Craig at the Department of the Army, Deputy Chief of Staff for Personnel, Manpower, Personnel and Training, provided the data.

2. U.S. Army Reserve Personnel Command (ARPERSCOM)

ARPERSCOM provided data that contained information on all NPS accessions from 1998 through 2002. The fields include the date of enlistment, the date(s) the recruit was scheduled for BCT and AIT, a field that identified whether or not this was split-option training, and a verified date that the applicant shipped to training. MSG Patrick Sarley at the Army Reserve Personnel Command, REQUEST Management Office, St. Louis, MO, queried the data out of the REQUEST system.

3. U.S. Army Recruiting Command (USAREC)

USAREC also provided data on USAR accessions. These data include each recruit’s date contracted to join the USAR, along with his/her MEPS testing data, demographic data, and the market segment. This market segment is obtained from a commercial source that has clustered every zip code+4 into one of 50 market segments that characterize demographics, purchasing habits, and so on. These data span all accessions from FY92 through end of FY02. Major Mike Kamei, with the Programs Analysis & Evaluation

directorate at Headquarters, USAREC at Fort Knox, KY, provided the data.

4. Office of the Commander of the Army Reserve (OCAR)

Major Ward Litzenberg in the Programs Analysis & Evaluation directorate at OCAR, Arlington, VA, provided additional data pertaining to USAR force structure, recruiting priorities, and USAR data.

B. DATA PREPARATION

Before conducting the analysis, I needed to integrate the data from the four sources listed above. My goal was to create a data preparation process that can be updated and reused as time progresses. For analysis purposes, I needed a table of unique SSNs for all accessions into the USAR from FY99 through FY02; another table with these same accessions binned by MOS, enlistment month and year, and BCT/phase 1 OSUT start month and year; and a third table of training seat quotas and inputs binned by month and FY (and by MOS for AIT and OSUT). Finally, using the USAR accessions data, I developed a matrix of training seat usage (FY99 through FY02) by delay in months between the enlistment date and the IET training start date.

I conducted the data preparation in four parts: the ATRRS data, the REQUEST Data, the integration of the REQUEST and Reserve Enhanced Applicant File (REAF) data into an accessions "master," and the aggregation of the accessions master into monthly bins for IET training start date and enlistment date comparisons.

I built the data preparation process using two software packages: Microsoft ACCESS™ and SPSS Clementine™ 7.1. Clementine™ 7.1, a data mining software application,

is the software I used to classify and integrate the data. Clementine is unique in that the operations performed on the data are represented as graphical objects on a computer screen “palette.” These operations are sequenced into data “streams,” where data flows from a source on the left, through connected operation “nodes,” and then to output nodes that are generally on the right of the palette. The operation nodes perform operations such as setting data field types (**Type**), sorting the records (**Sort**), filtering out selected fields (**Filter**), merging records on certain keys such as SSN (**Merge**), appending records together (**Append**), filling in records based on some criteria (**Filler**), and creating fields based on a criteria (**Derive**). Other operations include selecting records with distinct values to find or eliminate duplicate values on keys such as SSN (**Distinct**), and selecting records based on a criteria in one or more of the fields (**Select**). A collection of operations can be represented within a supernode. Input nodes are circles, output nodes are boxes, graphs are triangles, operations are hexes, and supernodes are stars.

Figure 1 is a sample data stream. During the discussion of data preparation of the REQUEST data and the integration of the USAREC and REQUEST data, I will present detailed diagrams for Clementine “streams” corresponding to different aspects of the data preparation process.

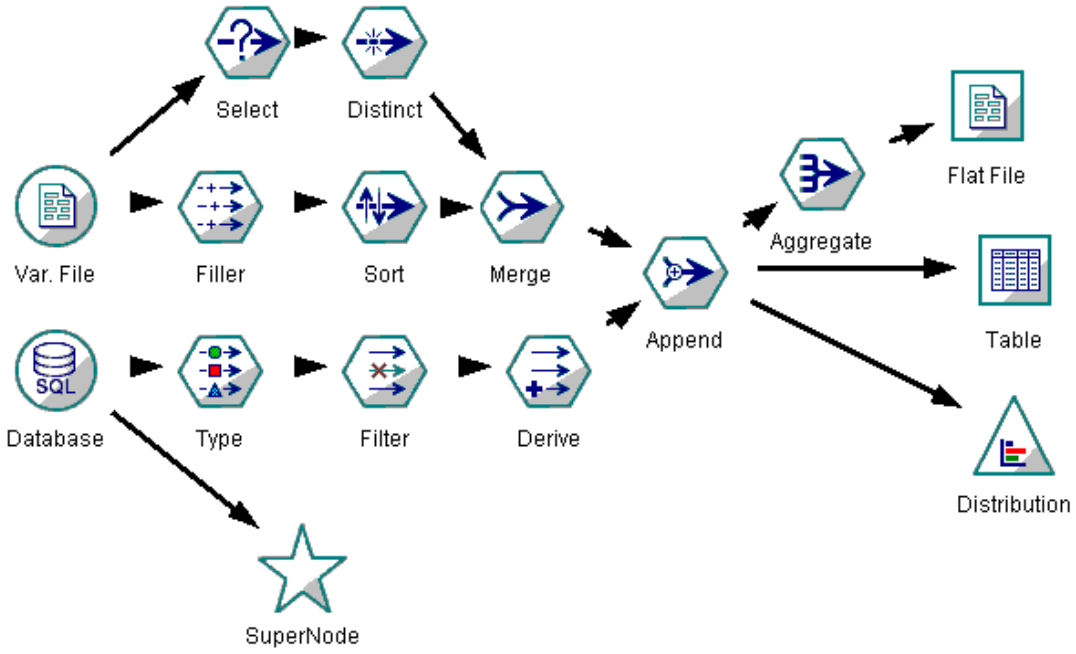


Figure 1. Sample Clementine Stream

Figure 1 is a sample stream that represents the data, operation, and output nodes connected with arrows. The data move through various operations until the output(s) are reached on the right side of the stream.

Collections of streams make up the processes, which are further collected into a project. This project organizes the streams that look at the data and perform the processing, as well as the output from the different streams. The project organization in Clementine is shown in Figure 2. The first part of the project contains streams and output used during the preliminary analysis under the folder labeled “data understanding.” The data preparation folder contains the streams that pertain to each of the parts of the process: REQUEST data preparation, REAF-REQUEST integration, and REQUEST Enlistment to Training Date Aggregation (not shown).

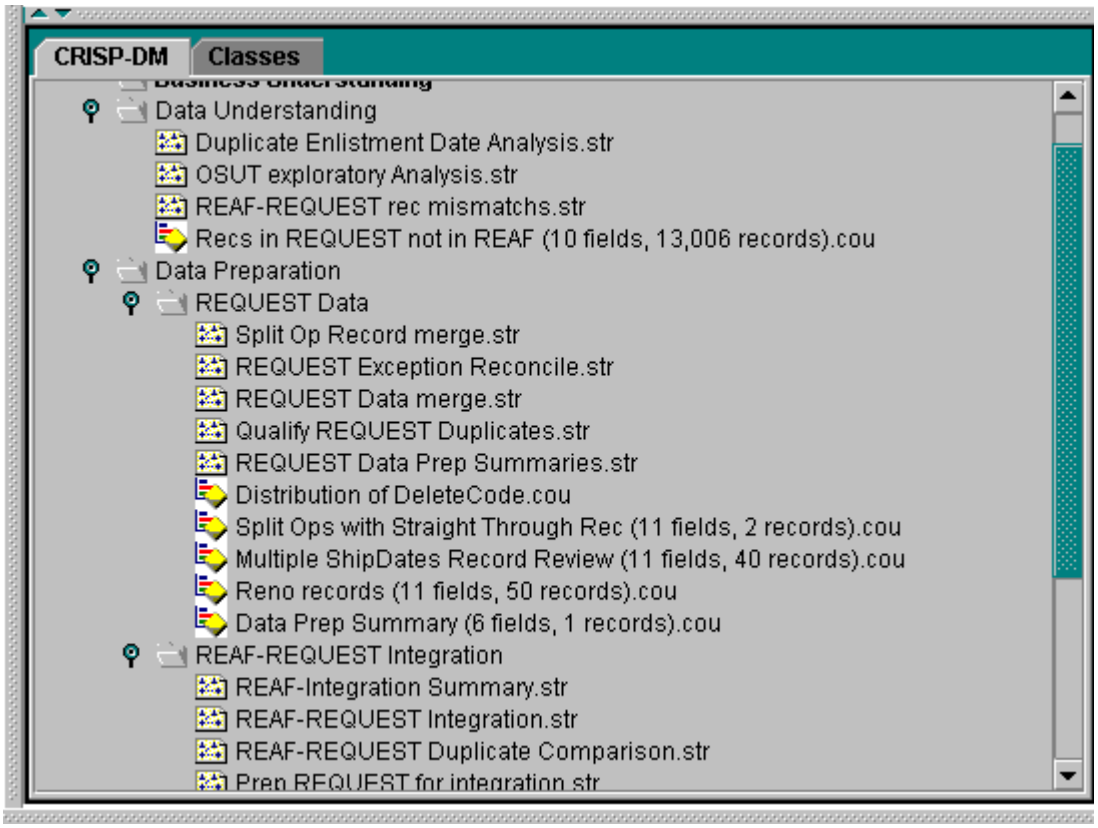


Figure 2. Clementine Project View

Figure 2 shows the project view in Clementine where each folder corresponds to a part of the data preparation process, and the items within each folder represent a stream or output from a stream.

1. ATRRS Data

Most soldiers go to AIT immediately after completing BCT. The AIT may be at a different location, or they may complete the entire training at one site (OSUT). In either case, this is called “straight-through ticket” training. There is an alternate program where the recruit completes BCT or phase 1 OSUT one year (typically in summer), and AIT or phase 2 OSUT the following year. This program is referred to as the “split-option” training program. The “split-option” program facilitates enlistment of

individuals who do not have the time to complete both phases of IET consecutively, such as high school juniors.

usarosut9903.xls

	A	B	C	D	E	F	G
1	CRS	Phase	REPORT DATE	START DATE	QS	QTA	NEW INPUTS
124	54B10-OSUT		001016	001020	MJ	59	64
125					MK	6	6
126			001113	001117	MJ	27	39
127					MK	14	13
128			010115	010119	MJ	29	40
129					MK	19	15
130			010226	010302	MJ	56	52
131					MK	41	24
132					MN	0	0
133			010326	010330	MJ	15	18
134					MK	16	5
135			010423	010427	MJ	12	3
136					MK	16	6
137			010723	010727	MJ	30	33
138					MK	16	16
139					MN	0	0
140			010723	010727	MP	0	1
141			010820	010824	MJ	33	41
142					MK	15	16
143					MP	1	0
144	54B10-OSUT (ST)	1	010611	010615	MJ	0	3
145					MK	0	5
146					MN	85	64
147					MP	40	27
148		2	010430	010504	MN	19	1
149					MP	10	0
150			010521	010525	MN	17	3
151					MP	5	3
152			010618	010622	MN	58	51
153					MP	11	11

usar00osut | usar01osut | usar02osut | usar03osut

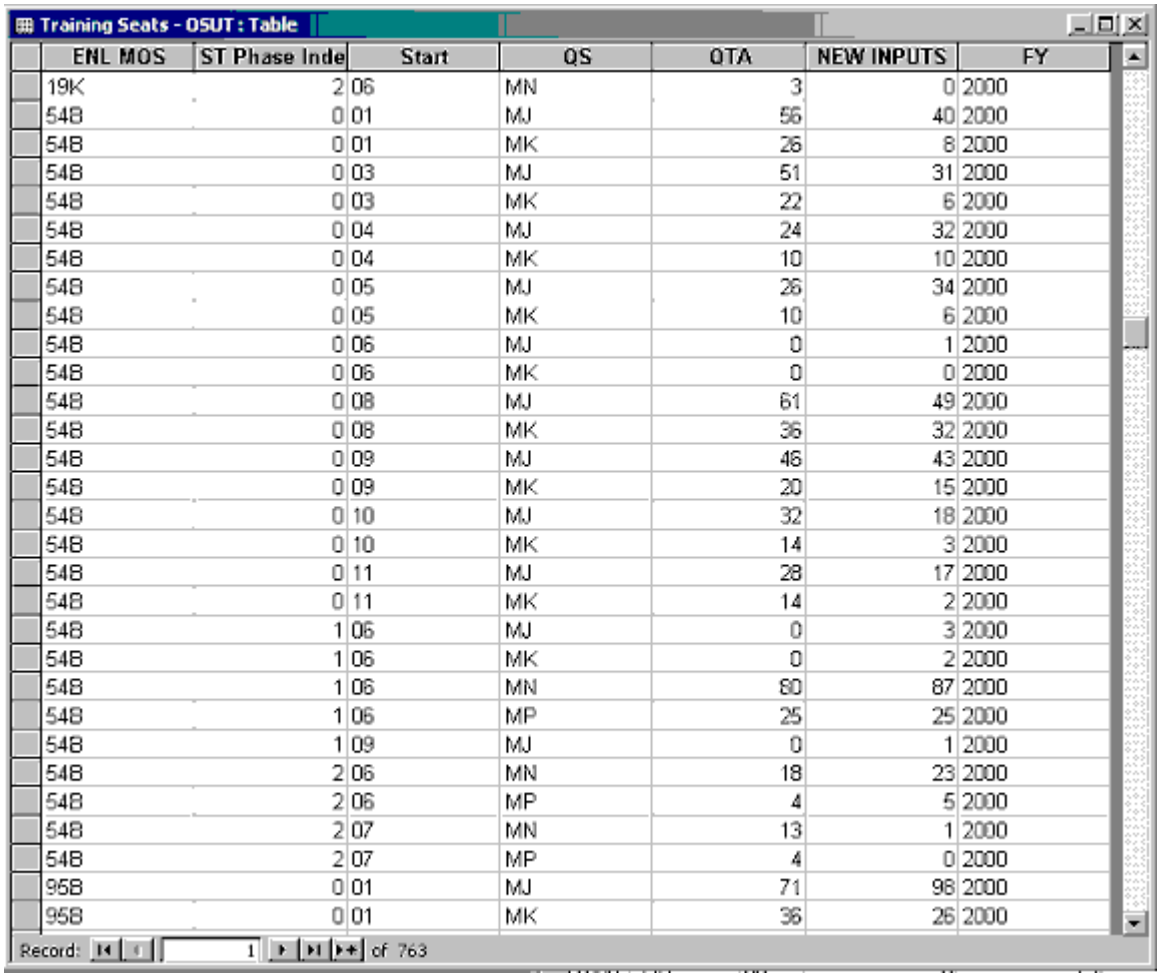
Figure 3. ATRRS Data Spreadsheet

Figure 3 shows the ATRRS data in the format received from the Department of the Army (DA), with each line representing a quota source with quotas and inputs for a particular class. CRS is the course name in ATRRS, the QS is the quota source (MJ is straight-through male, MK is straight-through female, MN is split-option male, and MP is split-option female), QTA is the quotas assigned, and NEW INPUTS is the number of individuals who actually started training.

The ATRRS data came in three Microsoft EXCEL™ spreadsheets derived from queries Mr. Craig at DA ran in

ATRRS. The data lists (by quota source) every BCT, OSUT and AIT training class in FY 1999-2003 with the number of quotas and training inputs for each class. The four quota sources refer to the training program (split-option or straight-through) and gender. Gender is a somewhat important quota management tool since some MOSs are male specific, and BCT classes are managed to a ratio per class of men and women. These quotas are assigned to the USAR by four quota sources: straight-through male (MJ), straight-through female (MK), split-option male (MN), and split-option female (MP). Grouping these quota sources by program equates split-option to a combination of MJ and MK, and equates straight-through to a combination of MN and MP.

In each training type's EXCEL spreadsheet, a fiscal year's data is represented by one worksheet, as shown for the OSUT classes in Figure 3. The three spreadsheets are linked into an ACCESS database, and each year's data are merged into a single table for each training type. For all the IET data, I changed each class report date to a month and fiscal year column. The result is three tables, each spanning FY 1999 through FY 2003 for their respective training type. These three tables are OSUT, BCT, and AIT, and contain both split-option and non-split-option training quotas and inputs. As an example, a portion of the OSUT training table is shown in Figure 4.



Training Seats - OSUT : Table

ENL MOS	ST Phase Inde	Start	QS	OTA	NEW INPUTS	FY
19K	2	06	MN	3	0	2000
54B	0	01	MJ	56	40	2000
54B	0	01	MK	26	8	2000
54B	0	03	MJ	51	31	2000
54B	0	03	MK	22	6	2000
54B	0	04	MJ	24	32	2000
54B	0	04	MK	10	10	2000
54B	0	05	MJ	26	34	2000
54B	0	05	MK	10	6	2000
54B	0	06	MJ	0	1	2000
54B	0	06	MK	0	0	2000
54B	0	08	MJ	61	49	2000
54B	0	08	MK	36	32	2000
54B	0	09	MJ	46	43	2000
54B	0	09	MK	20	15	2000
54B	0	10	MJ	32	18	2000
54B	0	10	MK	14	3	2000
54B	0	11	MJ	28	17	2000
54B	0	11	MK	14	2	2000
54B	1	06	MJ	0	3	2000
54B	1	06	MK	0	2	2000
54B	1	06	MN	80	87	2000
54B	1	06	MP	25	25	2000
54B	1	09	MJ	0	1	2000
54B	2	06	MN	18	23	2000
54B	2	06	MP	4	5	2000
54B	2	07	MN	13	1	2000
54B	2	07	MP	4	0	2000
95B	0	01	MJ	71	98	2000
95B	0	01	MK	36	26	2000

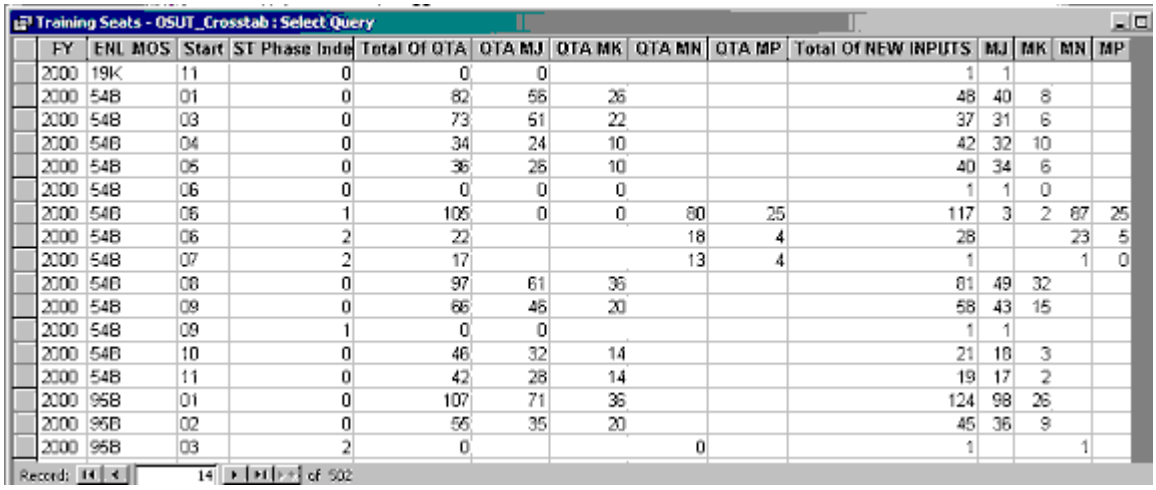
Record: 1 of 763

Figure 4. ACCESS OSUT Data Table

Figure 4 shows the data table created from the ATRRS input spreadsheets, combining all five fiscal years' data for all OSUT classes binned by month and year.

Additionally, for AIT and OSUT schools, the MOS of the training is substituted for the class name. The OSUT training table also has an additional field representing split training phase (since OSUT can be either phase 1 or phase 2). In the EXCEL spreadsheets, each line represents a single quota source for a particular class, which is how the queries in ATRRS output the data. To create a table where each record is one month of one FY with quota and inputs by source as entries for each record, I built a

cross tabulation query, shown in Figure 5. Records in this table are ready to use for the training analysis.



Training Seats - OSUT_Crosstab : Select Query

FY	ENL MOS	Start	ST Phase Inde	Total Of QTA	QTA MJ	QTA MK	QTA MN	QTA MP	Total Of NEW INPUTS	MJ	MK	MN	MP
2000	19K	11	0	0	0				1	1			
2000	54B	01	0	82	56	26			48	40	8		
2000	54B	03	0	73	51	22			37	31	6		
2000	54B	04	0	34	24	10			42	32	10		
2000	54B	05	0	36	26	10			40	34	6		
2000	54B	06	0	0	0	0			1	1	0		
2000	54B	06	1	105	0	0	80	25	117	3	2	87	25
2000	54B	06	2	22			18	4	28			23	5
2000	54B	07	2	17			13	4	1			1	0
2000	54B	08	0	97	61	36			81	49	32		
2000	54B	09	0	66	46	20			58	43	15		
2000	54B	09	1	0	0				1	1			
2000	54B	10	0	46	32	14			21	18	3		
2000	54B	11	0	42	28	14			19	17	2		
2000	95B	01	0	107	71	36			124	98	26		
2000	95B	02	0	55	35	20			45	36	9		
2000	95B	03	2	0			0		1			1	

Record: 14 of 502

Figure 5. ACCESS Crosstab Query for OSUT Data

Figure 5 shows the crosstab query results, combining quotas and inputs into a single record per month-year bin.

2. REQUEST Data

The REQUEST data came as a series of queries by FY. Each record contained the following information:

Social Security Number (SSN)

Military Occupational Specialty (MOS)

Split-option Training Phase

Enlistment Date

Basic Combat Training Start Date

Advanced Training Start Date

Ship Verification Date

During the exploratory analysis of the data, I uncovered some serious problems with the data. In particular, the REQUEST data contained multiple records for many SSNs. Some of these records are total duplicates, but most are partial duplicates with differing values in

various fields with conflicting information referencing a specific SSN. For example, there might be two records for the same SSN that differ only in the “BCT Start Date” field: one record has a date and the other is blank. The large number of partial duplicates greatly complicates determining the correct values for a specific SSN. I worked through several iterations of queries from ARPERSCOM with additional fields to distinguish the records from one another. It was time-consuming and difficult. The streams in Clementine (Figures 7-12) indicate how I added fields and iteratively “weeded out” duplicates. This process is discussed in more detail later in this section.

Consistency between fields and records is also a problem I confronted. None of the records for enlistees that attend OSUT have a BCT Start date, as the enlistees receive their advanced training in conjunction with BCT requirements. This problem compounds the split-option duplicate issue, as there are multiple values for the AIT start date for the same SSN, one for phase 1 and another for phase 2. The fact that some of the phase two records do not have an Alternate Phase Training field equal to 2 (denoting a phase 2 or AIT) compounds problems in differentiating the records and SSNs. There is also a problem with a large number of records missing training data (BCT and AIT start dates). Since any NPS recruit requires at a minimum a BCT or phase 1 OSUT date, identifying the initial date and the follow on dates is difficult for the split-option accessions. OSUT accessions in the straight-through program do not require a second date, but all other accessions do.

Additional problems discovered in these duplicate fields are records with missing information or illogical entries of data. Entries such as a ship date after the BCT start date, an enlistment date after the ship date, and so on, are some of the situations I encountered.

	IND_SSN	MOS	ALT_TNG_PH	EnlistmentDate	BCTDate	AITDate	ShipDate	OSUT	SO Flag
19611	[redacted]8458	88N	1	2003-03-27	2003-06-13			N	T
19612	[redacted]5199	51R	1	2001-02-24		2002-06-17	2001-05-31	N	T
19613	[redacted]5199	51R	1	2001-02-24	2001-06-08		2001-05-31	N	T
19614	[redacted]5199	51R	2	2001-02-24		2002-06-17	2002-06-05	N	T
19615	[redacted]5199	51R	2	2001-02-24	2001-06-08		2002-06-05	N	T
19616	[redacted]5093	54B	2	2001-02-16		2002-06-21	2002-06-11	O	T
19617	[redacted]5093	54B	2	2001-02-16		2001-06-15	2002-06-11	O	T
19618	[redacted]5093	54B	1	2001-02-16		2001-06-15	2001-06-05	O	T
19619	[redacted]5093	54B	1	2001-02-16		2002-06-21	2001-06-05	O	T
19620	[redacted]0173	62J	1	2003-04-18	2003-06-13			N	T
19621	[redacted]6575	57E	1	2001-01-10	2001-06-15		2001-06-05	N	T

Figure 6. Split-Option Duplicate Records Example

Figure 6 shows two split-option enlistees, one OSUT and one non-OSUT, each with four records. The third column indicates the training phase, and there should be exactly one record for each phase, not two as is highlighted.

The largest single source of partial duplicate records in the data was the split-option training program accessions. Anywhere from two to four records appeared for each split-option enlistee, sometimes as many as eight. The sample records in Figure 6 show two highlighted split-option accessions, each with four records matching their SSN: two phase 1 records and two phase 2 records. Each should have two records: one for their phase 1 school date during the year of enlistment, and another for their phase 2 school date during the following year.

By eliminating the duplicates with BCT or phase 1 OSUT listed for a phase two record, and the reverse, there should be only two records remaining. This is relatively easy for the non-OSUT enlistees, as the phase 1 records are

without a BCT date, so the phase 2 records without an AIT date could be deleted. This approach does not work with the OSUT enlistees, as both their phase 1 and phase 2 start dates are listed in the AIT start date field. The only way to tell is that the AIT start date for the phase 1 OSUT is usually one year prior to the phase 2 OSUT start date. By making a comparison with the OSUT records in days between the enlistment and AIT start dates of all records, the delay in days between enlistment and equivalent scheduled OSUT phase 1 start dates can be determined. Using duplicate OSUT records with both phase 1 and phase 2 scheduled, and a common non-null enlistment date, I derived a field that represented the number of days between the enlistment date and the AIT date. I then aggregated the records down to SSN with a minimum value and a maximum value in days. This minimum is the number of days from the enlistment to phase 1 start date, and the maximum the number of days from the enlistment date to the phase 2 start date. The largest minimum value was 280 days, and the smallest maximum value was 373 days.

By selecting all enlistment-to-AIT-start-date differences of greater than 335 to represent phase 2 and less than 335 to represent phase 1, the bogus OSUT split-option records can be identified and marked. I used 335 days as the cut off criteria because it works for the dataset used, and also represents the earliest a recruiter can prospect for most split-option enlistees. Potential applicants cannot be contacted by a recruiter until they begin their junior year of high school. Since 95% of all split options attend phase 1 OSUT or BCT in May, June, and July, and the earliest a recruiter can contract an

individual is in August, this means that the enlistment date to start date is something less than 11 months in the worst case. The data separated into two distinct groups since there were no start-date differences between 280 and 373 days.

I used the criteria specified above as the foundation for the rules to progressively screen out the duplicate records.

I assigned letter codes to each of the following reasons to assist in helping me determine why a record was marked for deletion. These codes are in order of evaluation. Once a record is marked, it is not evaluated further. A record marked for deletion will only have a single deletion code.

A: Duplicate record with blank or null BCT date and AIT date.
B: Straight-through accession with more than 1 duplicate record and BCT date before ship verification date.
C: Straight-through accession with more than 1 duplicate record and a BCT or AIT date prior to the enlistment date.
D: Spare.
E: Split-option duplicate record.
F: Spare.
G: Split-option OSUT MOS phase 1 record with an AIT date at least 335 days later than the enlistment date.
H: Split-option OSUT MOS phase 2 record with an AIT date at most 335 days later than the enlistment date.
I: Split-option non-OSUT MOS phase 1 record with blank or null BCT date.
J: Split-option non-OSUT MOS phase 2 record with non-blank or non-null BCT date, or blank or null AIT date.
K: Non-duplicated SSN with null or blank BCT and AIT dates.
L: Split-option phase 2 record merged with a matching phase 1 record.
M: Split-option phase 2 record merged with a corresponding phase 1 record without a matching

enlistment date (one of the records had an erroneous enlistment date).

N: Duplicate straight-through record with blank or null BCT date, or blank or null enlistment date.

O: Duplicate straight-through record with a ship date at least 5 weeks earlier than the BCT date.

Using these rules, I constructed a series of streams in Clementine to mark each record the first time it meets these criteria for deletion, merge split-option accessions into a single record, provide a record summary of deletions, and create a file with the undeleted records for integration with the USAREC data. This is critical, because every duplicate that is left in the REQUEST data may have a corresponding duplicate in the REAF data, and could possibly magnify the number of duplicates during the integration.

I prepared the REQUEST data in four steps: merging the separate FY queries into a single file; qualifying the duplicate records and marking easily identifiable “bogus” records for deletion; merging split-option records into a single record; and reconciling as many of the records with duplicate enlistment and ships dates as possible.

The merge stream shown in Figure 7 appends the records from the four queries together, converts the date string to dates, flags (with a binary key) the split-option records and the MOSs that are associated with OSUT training, and generates lists of duplicate SSNs, SSNs without a ship date, and SSNs with duplicate records with differing enlistment dates.

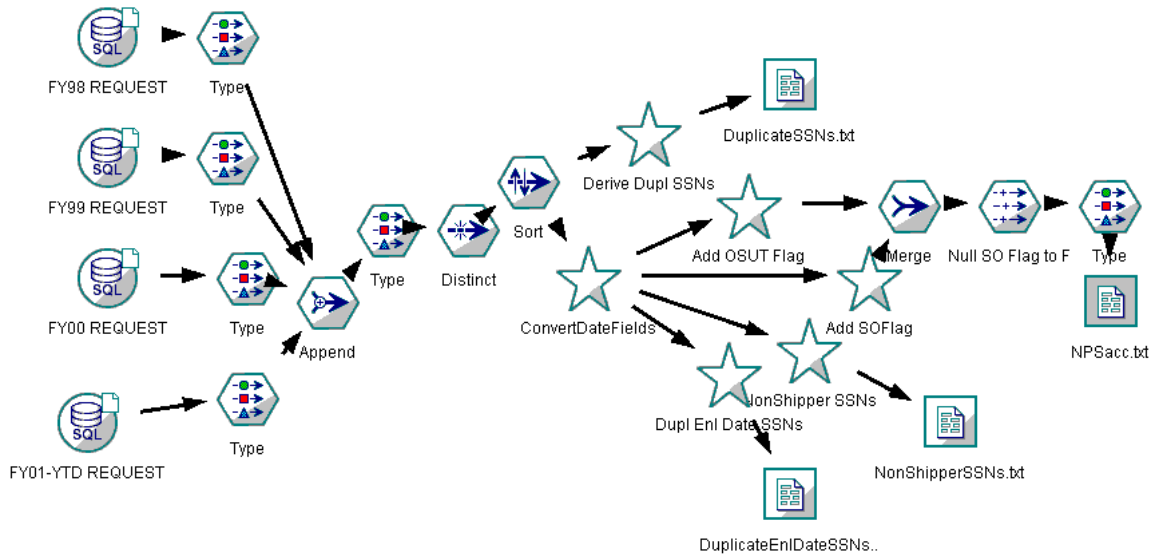


Figure 7. REQUEST Data Merge Stream

Figure 7 shows the stream that merges the four years of REQUEST data, converts the date fields, adds the flags for split-option and OSUT accessions, generates the duplicate tables, and creates the accessions table called NPSacc.txt on the right.

The “duplicate qualification stream” shown in Figure 8 starts with the merged accession file, NPSacc.txt, and the duplicate SSN output from the previous stream. This stream selects the records meeting the deletion criteria, codes each record, and then creates a file containing the records marked for deletion. This stream prepares the split-option records for merging by deleting the duplicates and leaving exactly two records for each: a phase 1 record and a phase 2 record. It also qualifies the unique records without BCT and AIT dates for deletion, and also qualifies duplicate straight-through records. The upper portion of the stream qualifies the duplicate straight-through or “non-split-option” records.

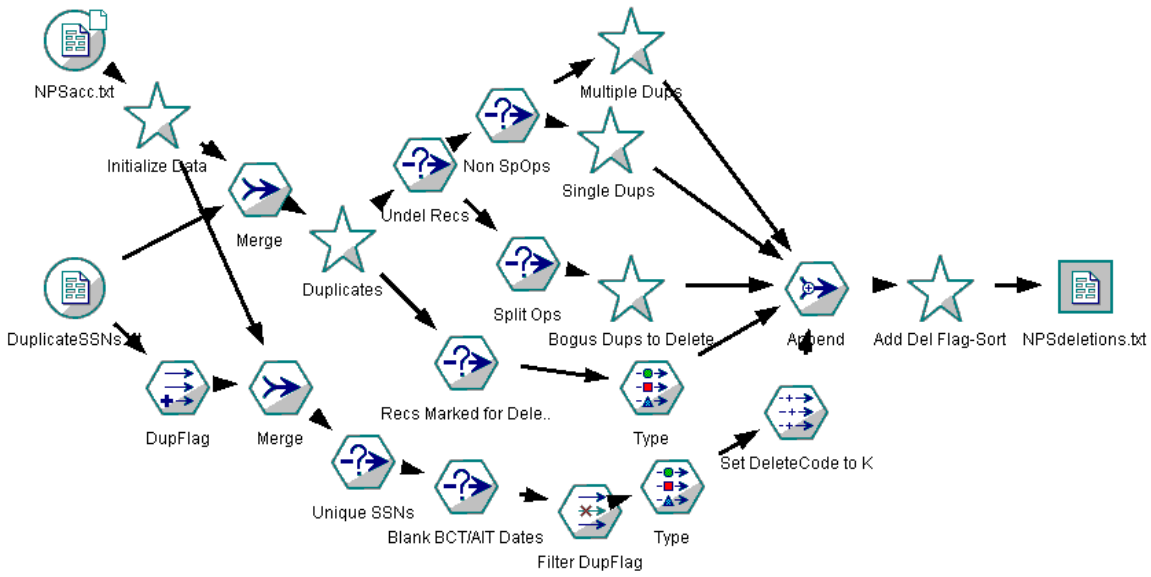


Figure 8. REQUEST Duplicate Qualification Stream

Figure 8 shows the duplicate qualification stream. This stream takes the merged REQUEST file and duplicates file, and qualifies the records based on the lettered criteria through a series of node operations. The records are flagged for deletion and output to a deletion file that catalogues all records marked for deletion.

In Figure 8, the supernode for the straight-through records with more than 1 duplicate is represented by a star node labeled **Multiple Dups**. Figure 9 illustrates the contents of that supernode or sub-stream.

Figure 8 also shows a supernode labeled **Bogus Dups to Delete**. The sub-stream for this supernode marks split-option records for deletion based on whether they are OSUT or non-OSUT, and is shown in Figure 10.

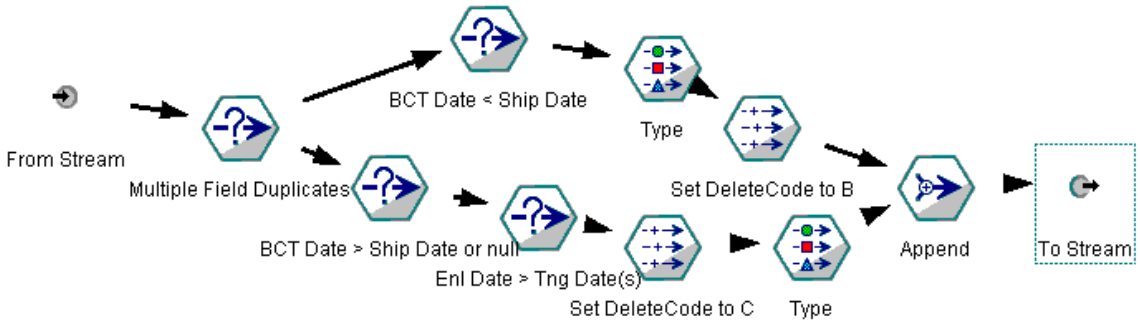


Figure 9. 'Multiple Dups' Supernode

Figure 9 shows the multiple duplicate qualification supernode. The data, which are straight-through duplicate records, enter from the stream on the left. Illogical records are selected, and then marked with a code. They are appended together and then passed back to the stream.

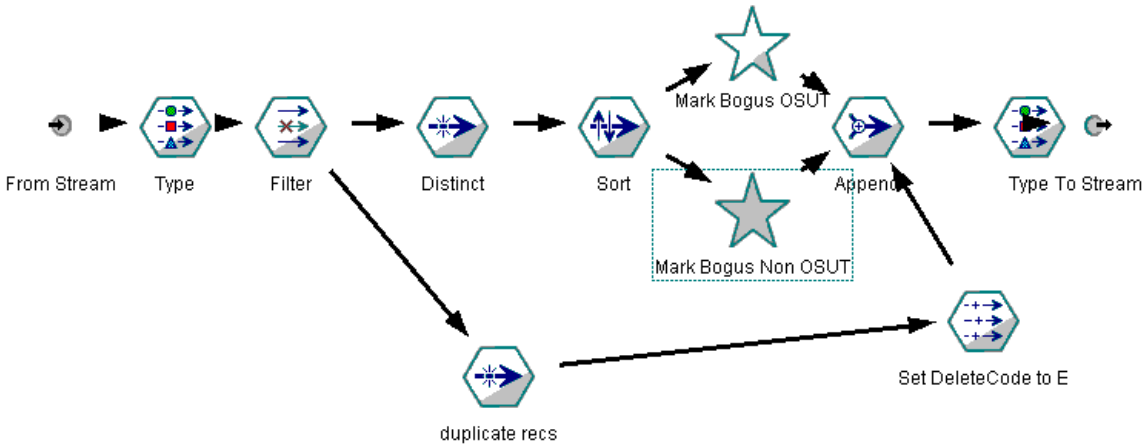


Figure 10. Split-Option Deletion Node.

Figure 10 shows the supernode that sorts the split-option records into OSUT and non-OSUT accessions, and then checks them for illogical entries. They are then marked, appended together and passed back to the stream.

Once the initial screening of duplicates is complete, the split-option records are merged into a single record. The split-option merge stream (shown in Figure 11) merges

the split-option records with exactly 1 record for each phase 1 and phase 2 with the same enlistment date.

The split-option records are then merged. First, two new fields, AITDate2 and ShipDate2, are appended to the phase 1 record. These fields are set equal to the values for the phase 2 record's AIT date and ship date, respectively. The phase 2 record is then marked for deletion. These marked records are added to the original list of records marked for deletion, and the merged split-option records are stored in a flat file for later integration into the file for analysis.

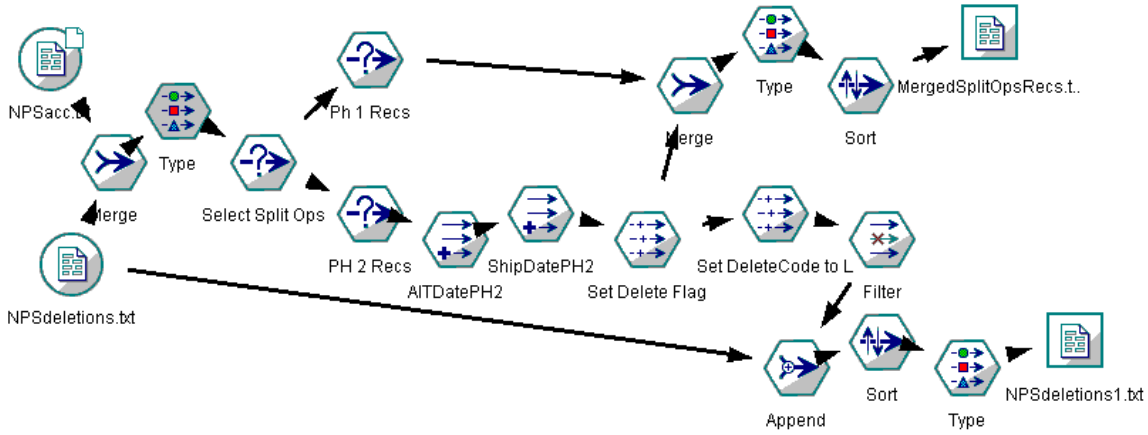


Figure 11. Split-Option Merge Stream

Figure 11 shows the split-option merge stream. This stream takes the multiple split-option records and creates a single record with two additional fields containing the phase 2 training start date and ship date. The data are merged into the phase 1 record, and the phase 2 record is then marked for deletion.

The last stream is used to qualify duplicate records addressing the records with the same SSNs and multiple values for the date fields. These represent the most

difficult records to differentiate correct from incorrect. Most are simply identified for later. The records identified for later include records with multiple ship dates and multiple enlistment dates. These records are in small enough groups to reconcile "by hand." For the records with duplicate enlistment dates that have identical BCT and AIT dates, I chose to merge using the first of the enlistment dates and to mark the additional records(s) for deletion. If they were split-option records, they were merged using the same process as outlined in the merge split-option stream in Figure 11.

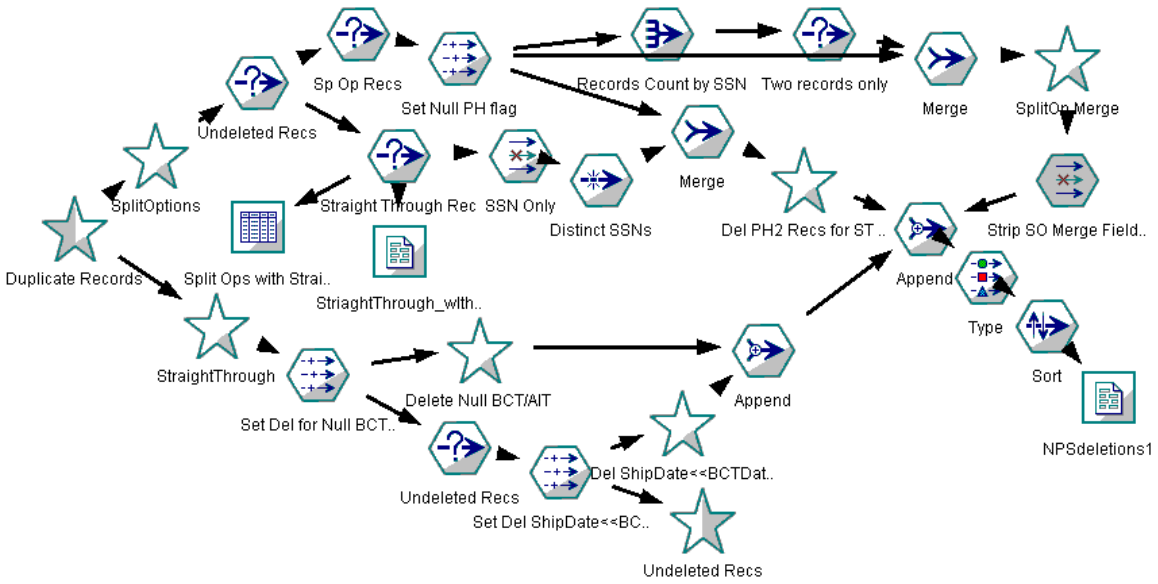


Figure 12. Duplicate Reconciliation Stream

Figure 12 shows the last duplicate screening stream. This stream tries to reconcile duplicate records with differing enlistment dates for the same SSN. It also marks for deletion any record that is left that is a non-OSUT straight-through without a BCT date or AIT date, and identifies SSNs that have records matching straight-through and split-option criteria. Any split-option records identified are merged using the same process in the merge split-option stream of Figure 11.

The three major products of these streams are the file with all the records (NPSacc.txt), a file containing all records marked for deletion with a deletion code (NPSdeletions1.txt), and a file with the merged split-option records (MergedSplitOpRecs.txt). There are several minor products that collect unqualified duplicate records for SSNs with duplicate ship dates, duplicate enlistment dates, and SSNs with both split-option and straight-through records.

The records are merged and the undeleted records with the merged split-option records are passed on to a new file in preparation for integration with the data from USAREC. That stream is shown in Figure 13.

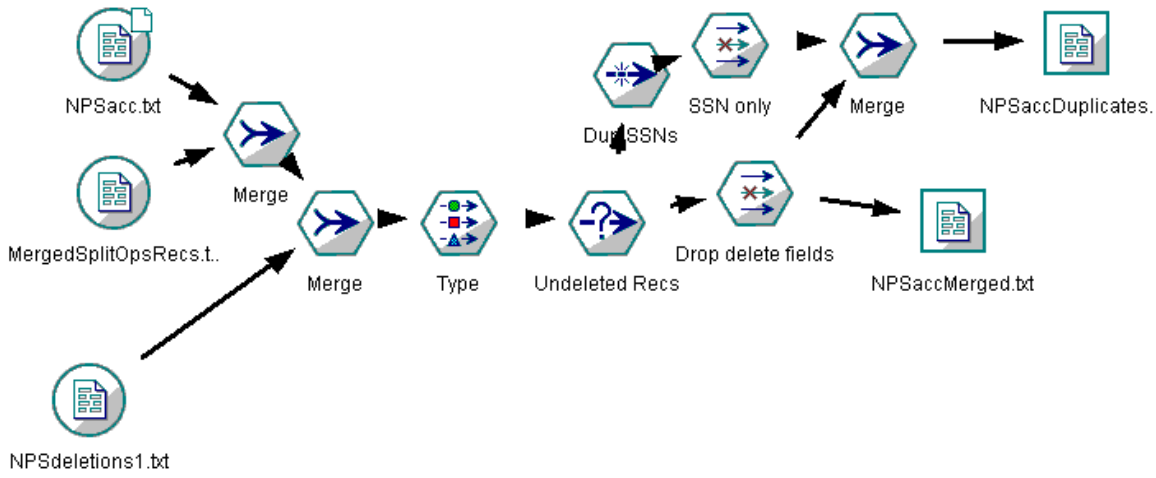


Figure 13. REQUEST-REAF Integration Preparation Stream

Figure 13 shows the last step in preparing the REQUEST data. This stream merges the merged split-option records with the accessions file and the deleted records file. The undeleted records are selected, the delete flags filtered, and the results stored in the NPSaccMerged.txt file that represents the undeleted screened files ready for analysis.

The screening effectiveness is measured by the number of records deleted, the reasons for deletion, and the duplicates remaining undeleted. The merged input from REQUEST totaled 87,598 records with 72,156 unique SSNs and 15,442 duplicate records. If these data were to be used without filtering the duplicates, or just as bad, arbitrarily deleting the duplicates, any analysis centered on the contents of the records would certainly be skewed. Since I am planning on using these data to conduct a temporal analysis with the training fields in REQUEST, fidelity of the data entries is as important as having the "right numbers." Accepting the amount of error represented by 15,442 duplicates would certainly cause my data to have an unacceptably high relative error when compared with the ATRRS data.

The last portion of the REQUEST data preparation is to evaluate how the process performed to reduce the duplicate entries, determine how many records were marked for deletion and for what reason, and how many SSNs were eliminated from the dataset to be used for analysis.

I used the stream shown in Figure 14 to aggregate the results through comparison with the deleted records, and generate a distribution graph of the delete codes as well as a small record summary, both shown in Figure 15.

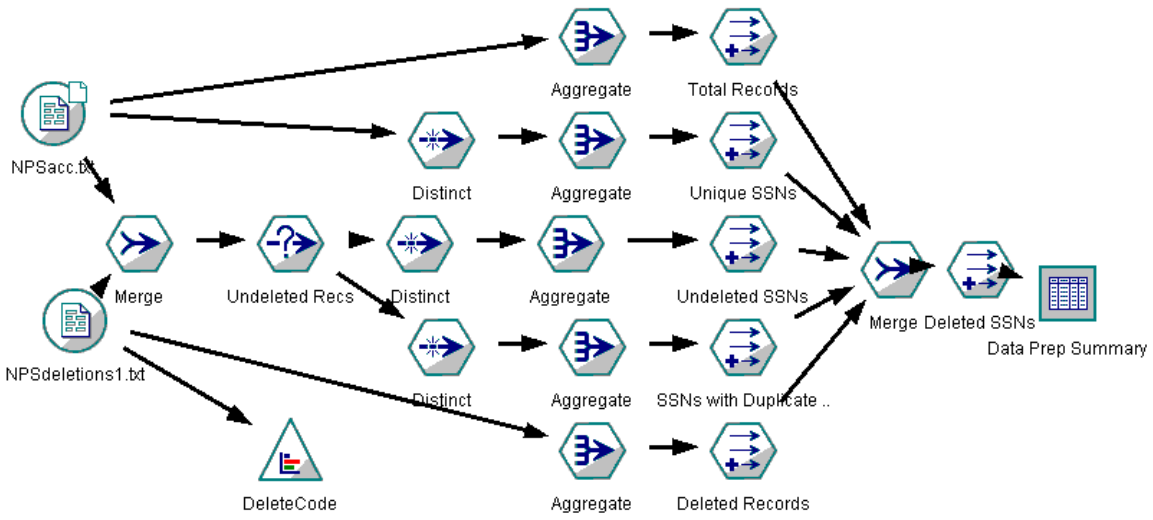


Figure 14. REQUEST Data Prep Summary Stream

Figure 14 shows the stream that generates a single record summary of the records, the deletions, and the remaining duplicates. It generates a proportion graph of the deletion codes as well.

One interesting item to note is that 2,546 records were deleted for having blank or null training data. These records represent unique SSNs. Compare this to the total unique SSNs deleted, as shown in the summary table in Figure 15. That means that the screening process deleted 2,615-2,546 or 69 unique SSNs. These 69 SSNs had multiple records, but either had key fields still blank or null in all the partial duplicate records or had illogical field values. For example, it might be that two records had the same enlistment date, yet only one had a BCT date that predated the enlistment date. The results from the preparation summary can be a starting point for analysis into the systematic errors and potentially lead to improvements in the data process.

	Total Records	Unique SSNs	Undeleted SSNs	SSNs with Duplicate Records	Deleted Records	Deleted SSNs
1	87598	72156	69541	19	18038	2615

Val...	Proportion	%	Count
A		2.15	387
B		0.06	11
C		0.02	4
E		0.04	7
G		4.21	760
H		4.55	821
I		17.21	3105
J		17.32	3124
K		14.11	2546
L		29.45	5313
M		0.01	2
N		10.79	1946
O		0.07	12

Figure 15. REQUEST Data Prep Summary

Figure 15 represents two outputs for the data prep summary stream. The single line output represents the number associated with the data input and output. The number of records, unique SSNs, unique SSNs remaining after preparation process, SSNs with duplicate records remaining after the preparation process, SSNs with duplicate records, records deleted by the preparation process, and unique SSNS deleted by the process. The distribution graph shows the associated deletion codes and how many records were marked with that particular code. Code K represents unique SSNs deleted due to null data fields.

3. USAREC Data and Integration with the REQUEST Data

The Reserve Enhanced Applicant File (REAF) provided by USAREC is the primary file for demographic data that contains the merged data from REQUEST, MEPS, and USAREC specific data (recruiting station, recruiter, market segment, etc). Although this is not the “official” record, it is derived from REQUEST, and I used it during the data cleaning process to correct known deficiencies in the REQUEST data.

The preparation of these data included generating an extract of the required information for FY 98 – FY 02.

This extract is a complete subset of the REAF for the listed years.

Like the REQUEST data that it uses as a source, the REAF data include a large number of duplicate or partial duplicate records. For the time period extracted, there were 9,774 duplicate records out of 106,600 total records.

Since the purpose of the REAF data is to provide demographic data and function as a source to fill in some of the blank and invalid entries, purging the duplicates was slightly less difficult. By examining the data I found that most duplicates were a function of differences in contract date, age differences, blank fields in one record with a non-blank in another record, differences in education level, and whether the individual was a high school graduate.

The important fields for merging the data, the SSN, MOS and vacancy control number (which corresponds to the matching REQUEST record) were consistent throughout the records. Merging the records from the REAF on these fields with the prepared REQUEST data output reduced the number of SSNs with a duplicate record from 4,848 to a single entry. This process is shown in Figure 16. Without understanding the exact process that USAREC used for the integration of their data sources to construct the REAF, it is difficult to assess the loss of accuracy in the REAF-REQUEST integration. The substitution of blank fields with populated fields, along with collapsing the data to a single record for each SSN, are improvements over the original REAF data with regard to integrating the data with the prepared REQUEST data output. For fields with multiple values in REAF data duplicate records, the latest of the

multiple records with either a more recent contract date or applicant age was the value used for that field in the merged record.

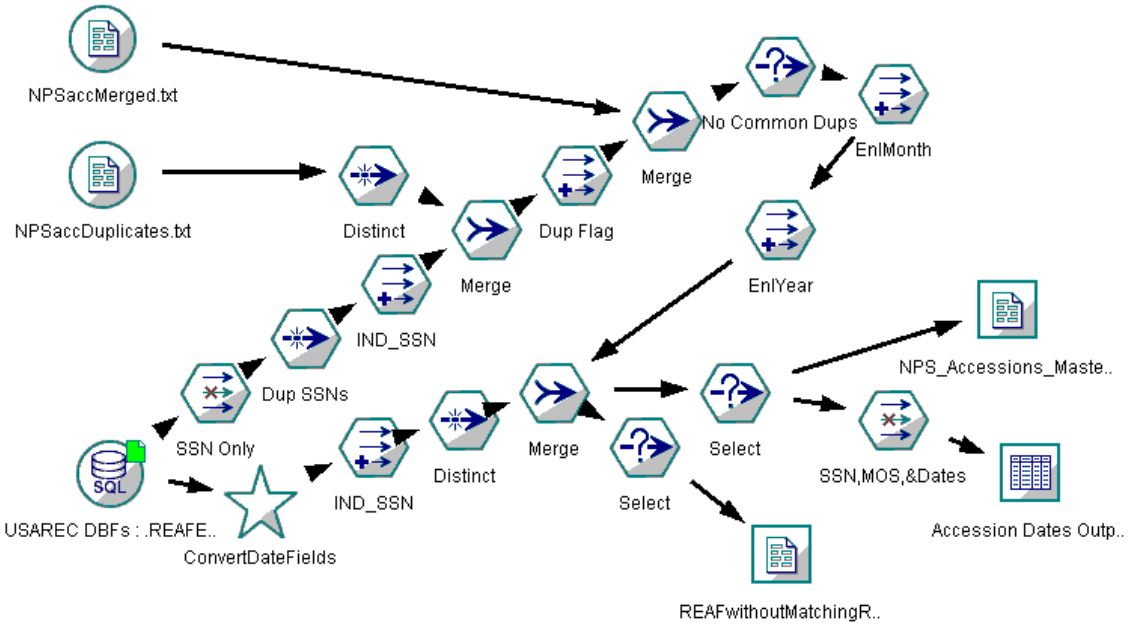


Figure 16. REQUEST REAF Integration Stream

Figure 16 shows the stream that merges the data from the REQUEST data preparation with the demographic data from the REAF file. It does not integrate records with an SSN that has duplicate records in both data sources, in order to prevent creation of additional duplicates.

4. Aggregation by Enlistment Date and Training Dates.

In the final phase of the data preparation, I take the merged records and build aggregated tables by enlistment month, enlistment FY, training start month, training start FY, and MOS.

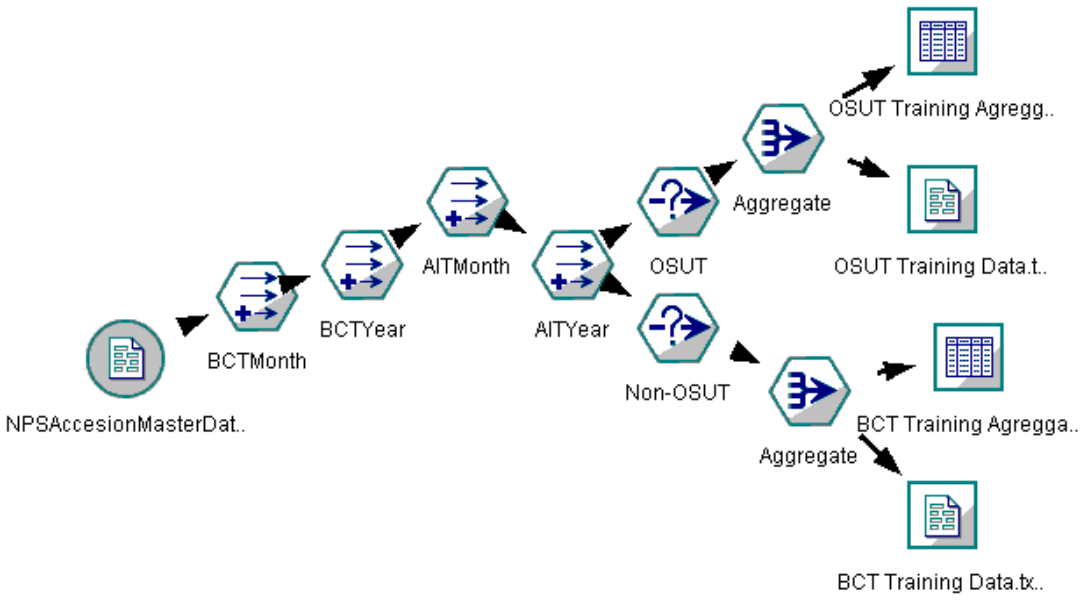


Figure 17. Aggregation by Enlistment and Training Dates

Figure 17 shows the stream that separates and aggregates the REQUEST-REAF integrated data by OSUT and non-OSUT, as the start date for training differs between these two training types.

Figure 17 shows the aggregation with month and FY added from appropriate training date fields.

To check the validity of the aggregated REQUEST source data by enlistment date and training start date, I compared the results with the binned training input data. In theory, the number of personnel listed as training inputs in ATRRS for a particular training date should correspond to the same number of USAR accessions listing that training date in the REQUEST-REAF data.

	MOS	SO Flag	EnlMonth	EnlYear	AITMonth	AITYear	Record_Count
1140	95B	F	1	2000	5	2000	10
1141	95B	F	1	2000	6	2000	1
1142	95B	F	1	2000	7	2000	1
1143	95B	F	1	2000	8	2000	2
1144	95B	F	1	2000	9	2000	18
1145	95B	F	1	2000	11	2000	1
1146	95B	F	10	2000	10	2000	2
1147	95B	F	10	2000	11	2000	17
1148	95B	F	10	2000	1	2001	6
1149	95B	F	10	2000	2	2001	10
1150	95B	F	10	2000	3	2001	1
1151	95B	F	10	2000	4	2001	2
1152	95B	F	10	2000	5	2001	1
1153	95B	F	10	2000	7	2001	28
1154	95B	F	10	2000	8	2001	9
1155	95B	F	10	2000	10	2001	3
1156	95B	F	11	2000	1	2001	1
1157	95B	F	11	2000	2	2001	8
1158	95B	F	11	2000	3	2001	3
1159	95B	F	11	2000	7	2001	5

Figure 18. Table of Aggregated OSUT REQUEST Data

Figure 18 shows the table generated by the aggregation by enlistment date and training start date stream. It contains the record count (number of SSNs) for each MOS by enlistment date and training start date.

I compared the results between binned months since this is how the data were aggregated. The overall numbers are comparable with a mean absolute relative error of 11.5%. The highest single absolute relative error between the REQUEST and ATRRS summary data by monthly bin was 56.0% in June 1999. The next largest variation was 26.6% in August 1999.

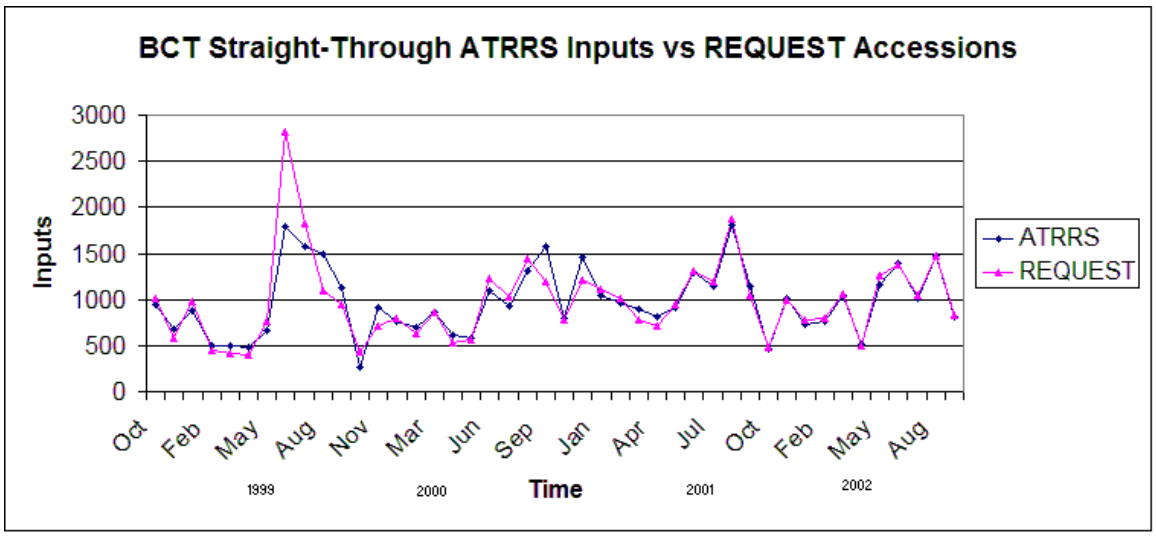


Figure 19. ATRRS Inputs vs. REQUEST Accessions

Figure 19 shows the graph of the summed accessions by training start date for BCT graphed against the sum of ATRRS inputs and quotas. The ATRRS and REQUEST data initially have distinct differences, which get progressively smaller as the time moves from 1999 to 2002.

Recall that by binning the data, there is a certain loss of resolution into the flow over time, so the number of inputs for a month is in part dependent on the number of BCT training class report dates that fall within the calendar month, and the number of quotas for each class. This problem may surface in the form of wild variation, particularly during the summer where the number of straight-through inputs per class ranges from 250 to 550.

Table 1 shows the number of classes per month. FY99 had only four classes in the July bin, where all the subsequent years had five. The reverse is true with regards to August. This accounts for, in part, for the large deviation of the FY 99 data in July and August, but it does not account for the sheer number of inputs in June not reflected in the REQUEST data.

Table 1. Number of Classes and Average Class Sizes

Table 1 shows the number of classes and average class size from ATRRS data. This represents the number of classes per bin.

	# Classes		# Classes		# Classes		# Classes	
Month	FY99	Avg	FY00	Avg	FY01	Avg	FY02	Avg
01	4	220	5	153	4	260	4	181
02	4	123	4	172	4	240	4	190
03	5	100	4	217	4	222	4	262
04	4	119	4	152	5	163	5	103
05	5	131	5	117	4	228	4	292
06	4	449	4	274	4	323	4	347
07	4	393	5	185	5	228	5	201
08	5	297	4	326	4	451	4	370
09	5	225	5	314	4	287	4	203
10	3	316	3	87	5	161	5	91
11	3	227	4	226	5	292	4	254

If we look only at the 2000-2002 data, the standard error drops to 7.8% with the single highest deviation being 23.7% in September of 2000. The mean relative error for each year gets progressively smaller, with the 1999, 2000, 2001 and 2002 mean relative errors being 22.8%, 13.5%, 7.3% and 3.1% respectively. With the better fit for the 2000 and later data, I will restrict the comparison of enlistment dates to training start dates to FY or calendar year 2000 and later.

C. DATA PREPARATION SUMMARY

The main purpose of the data preparation was to build a process for screening and integrating different data sources to provide information useful in examining the recruiting process and usage of IET training seats. Identification of records with data consistency issues,

whether between fields or records, and the ability to classify them for further analysis or exclusion from the data for analysis is the primary way to achieve this purpose. The collection of the records excluded can also provide a source of information about errors either with the data process or the data itself.

The fact that the REAF contained over 9,700 partial duplicate records for the four years I looked at is an indicator that there are few methods available for screening the erroneous duplicates records for SSNs out of REQUEST based data used to analyze the USAR recruiting process. By identifying the duplicate records, identifying possible errors, and marking known errors, the process outlined in this chapter provides a clean starting point for conducting analysis.

Without performing the preparation outlined above, then there is the potential to seriously degrade any USAR source recruiting analysis, particularly with regard to the split-option program. If I could not identify unique individuals with the correct information, then my analysis would be suspect.

The source of these errors is unknown in many cases. Some originate at the data entry point. Since some of the data in the REQUEST system is input at a terminal at the Military Entry Processing Station (MEPS), there is the possibility of human input error. The occurrences of multiple enlistment dates and ship dates are in part due to multiple visits to the MEPS. I checked several records with LTC (Retired) Charles Dalbec, Senior Personnel Analyst with Resource Consultants Inc. under contract to the U.S. Army Reserve Command G-1, and in each case the additional

ship date or enlistment date was due to an enlistee who "renegotiated" his contract. This "renegotiation" involved a change in training dates. In several cases, the day that they entered the MEPS to change the dates was entered as a ship date, although they did not "get on the bus" and go to IET. In other cases this date was entered in the verify enlistment date field.

In other cases, it may be that that the software used to conduct these queries from REQUEST, called FOCUS, may generate duplicate records for any SSN with multiple and/or conflicting values for a queried field. I cannot confirm this without testing the system, but it is a possibility.

In any case, the process identifies problem data records for further analysis as to the possible source of the error. This analysis could prove useful in efforts to engineer improvements to REQUEST.

The errors contained in the dataset created for this analysis can be further reduced with additional data sources. If further comparisons are made from the Total Army Personnel Database – Reserve (TAPDB-R), and ATRRS by individual SSNs, the null and inconsistent records could be identified and corrected. Mistyped SSNs could be checked against TAPDB-R, and training dates and school attendance could be confirmed using by SSN ATRRS data.

The process for merging this data is contained within the Clementine project. It can easily be modified to accommodate additional data sources and updated data for further use in preparation for future USAR accessions analysis.

In order to use the REQUEST preparation, there is a requirement to have the Clementine™ 7.1 software. Anyone using this process needs to have working knowledge of the REQUEST. Since the input is from FOCUS queries, anyone wishing to prepare REQUEST data for USAR recruiting analysis needs to have access to REQUEST, or to personnel who have access. In either case, knowledge on how to use FOCUS to query the data is required. With REQUEST access and availability of Clementine™ 7.1, the process can be constructed following the stream diagrams in this document and the node specifics listed in Appendix 3.

The integration with the REAF requires an additional data source from USAREC, the REAF. The REAF can be obtained through the HQ, USAREC Programs Analysis and Evaluation branch. A database software package such as Microsoft ACCESS™ or FOXPRO™ may be necessary to work with the REAF, as it is a very large file, and it is best to extract what data is needed prior to integration with the REQUEST data.

IV. TRAINING SEAT OVERVIEW

I analyzed the training seat data two ways: an exploratory overview of the training seat data provided by the Department of the Army using EXCEL, and a further analysis of the data with respect to the recruiting process by month of enlistment and start date for BCT or phase 1 OSUT.

A. ATRRS DATA OVERVIEW

The binned training seat data are organized into tables by month by FY comparing available quotas by type (merged by gender) and the associated training inputs.

1. BCT Data

The starting point for the training seat overview is BCT. BCT represents the point of entry into the system for new enlistees except for OSUT MOS, as it marks the official beginning of their IET training. The start of BCT marks the junction between recruiting and training.

a. Straight-Through Training

Straight-through training represents the standard training program for training new recruits, and is the major source of newly trained soldiers in the USAR.

Table 2. ATRRS BCT Quotas and Inputs

Table 2 lists aggregates by FY the BCT quotas and inputs for fiscal years 1999 through 2003.

FY	Total Of QTA	QTA ST	QTA SO	Total INPUTS	ST	SO
1999	17365	13662	3703	12954	10590	2364
2000	17904	14524	3380	12837	9575	3262
2001	17760	14956	2804	14751	12368	2383
2002	18308	15696	2612	12761	10368	2393
2003	17574	15096	2478	N/A	N/A	N/A

Looking at the distribution of these seats through the year in Figure 20, the high quota months for straight-through training are July, August and January.

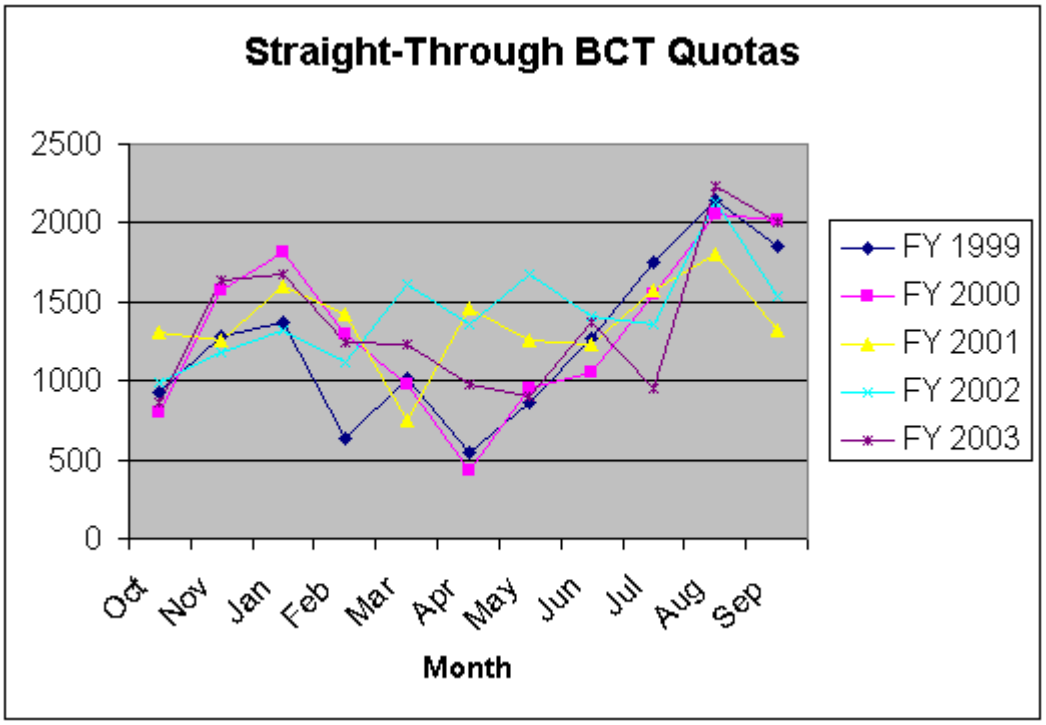


Figure 20. Straight Through BCT Quotas by Month

Figure 20 overlays each fiscal year's straight-through BCT quotas by month.

Comparing the available quotas to the training inputs is how the training seat usage, or percent of seats used, is derived.

Straight-through training inputs over the four years are fairly consistent with respect to time, although the magnitude varies between years. The inputs, shown in Figure 21, peak in the summer and are lowest in February through April. The largest variation in the inputs was in the summer of 1999, where there was a heavy variation in

the training inputs. The binning had one less class in July of 1999 and one more class in August of 1999 than the other three years (see Table 1).

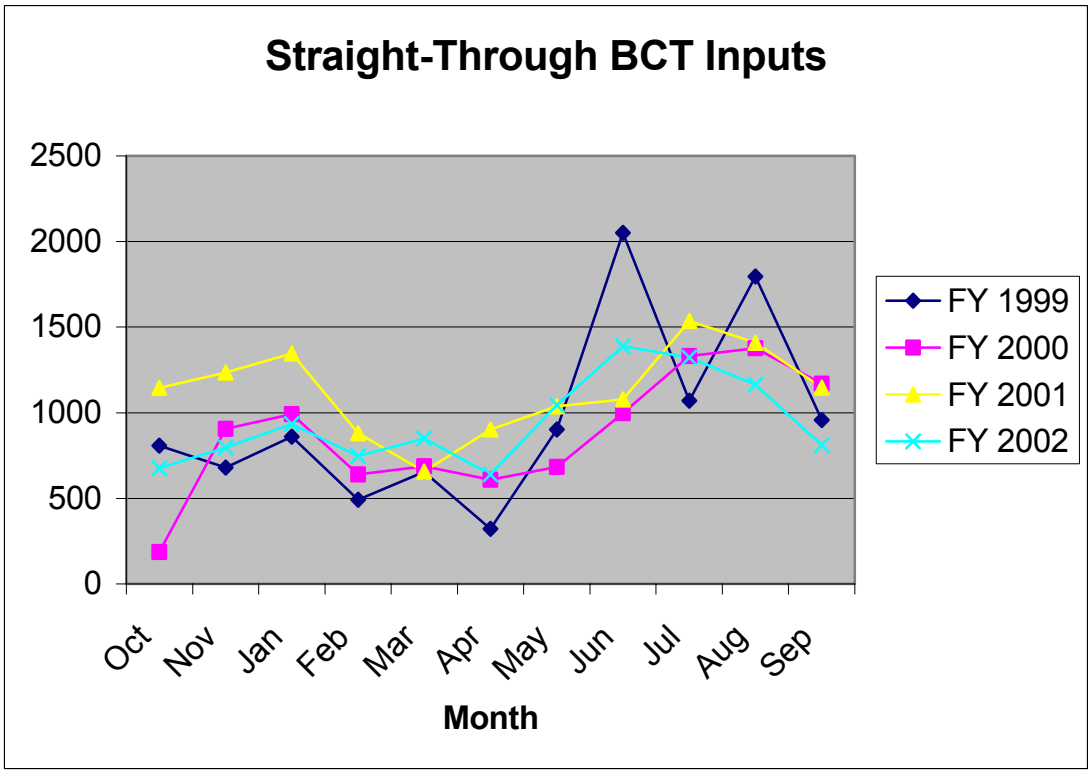


Figure 21. Straight-Through BCT Inputs

Figure 21 overlays the four years of binned ATRRS training inputs by month.

Comparing the quotas to the training input nets the training seat usage. Looking at the training seat usage over the last three years (in Figure 22), June and July were consistently the best in terms of usage and February the worst. During these low months over the past three years, the inputs varied between 600 and 800 inputs. Over that same time frame, the quotas have varied from 500 to 1500, resulting in the low usage for 2002, and large variations in 2000 and 2001. The 2003 quotas for this time

frame are between 1000 and 1300. Assuming 800 inputs for each month, the best that could be expected is an 80% usage rate.

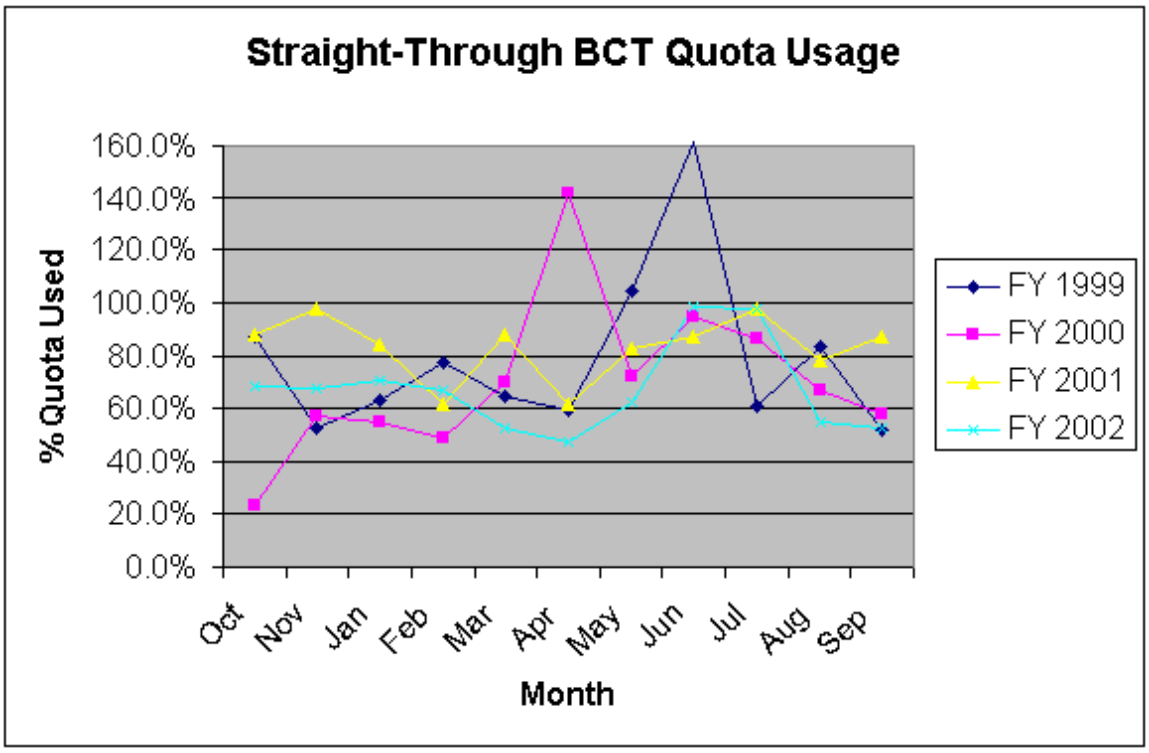


Figure 22. Straight-Through BCT Quota Usage

Figure 22 overlays the BCT % quota usage by month for all four years of ATRRS data.

Based on the provided training seat data, it appears that straight-through training seat usage is consistently better during June and July than during February and March.

b. Split-Option Training

The U.S. Army conducts split-option training primarily over the summer months, with the maximum number of USAR BCT quotas in June, as shown in Table 3 and Figure 23.

Table 3. Split-Option BCT Quotas

Table 3 compares the June and overall split-option BCT quotas.

FY	QTA SO	June	%June
1999	3703	2883	77.9%
2000	3380	2409	71.3%
2001	2804	1967	70.1%
2002	2612	2001	76.6%
2003	2478	1757	70.9%

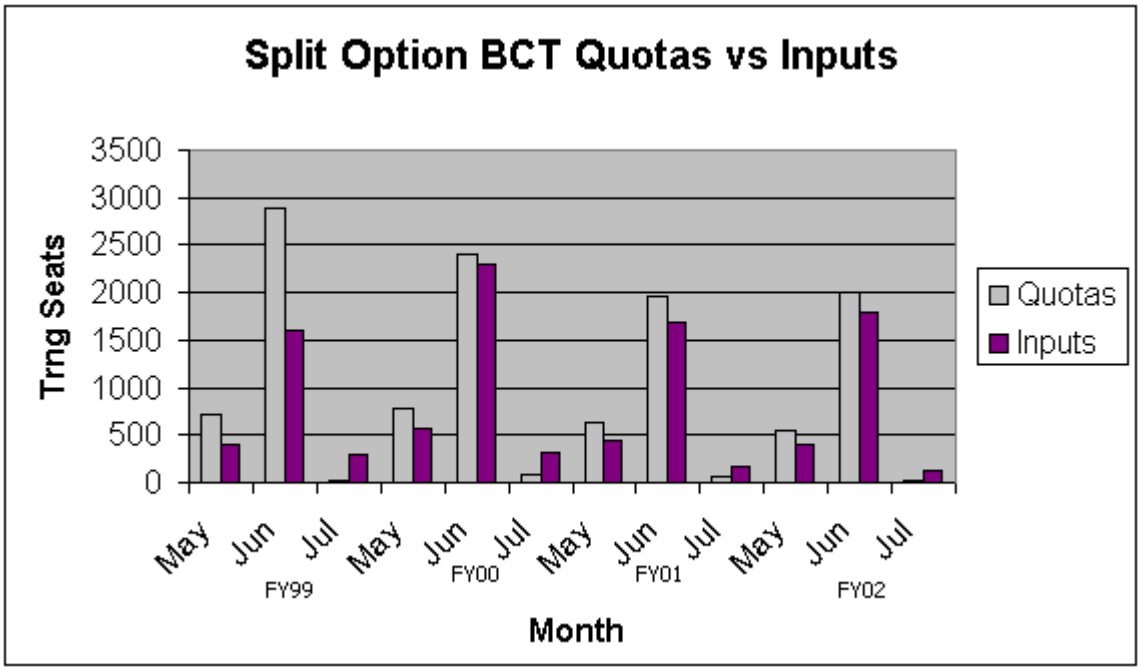


Figure 23. Split-Option BCT Quotas Versus Inputs

Figure 23 compares split-option quotas to inputs for May, June and July for the four years of ATRRS data.

With the exception of 1999, the split-option phase 1 BCT training seat usage has been at 85% to 97% usage.

The actual aggregate quota numbers for the four years are shown, by category, in Table 4. Note that with

the exception of 2000, the training inputs for split-option BCT average 2,380 plus or minus 18 inputs.

Table 4. Split-Option BCT Quotas and Inputs

Table 4 lists the quotas and respective inputs by quota source by fiscal year, with an overall percent quota usage.

FY	QTA MN	QTA MP	SO QTA	MN	MP	SO I	USAGE
FY1999	2596	1107	3703	1627	737	2364	63.8%
FY2000	2570	810	3380	2382	880	3262	96.5%
FY2001	2082	722	2804	1840	543	2383	85.0%
FY2002	2035	577	2612	1928	465	2393	91.6%

Split-option BCT quota usage for June matches that of the OSUT phase 1 in that it is a high usage month.

2. AIT Data

I looked at the AIT seat data from ATRRS with respect to the usage of quotas by source by year. Given that percent used is a limited usefulness in “low density” or MOSs with very few seats a year, I restricted evaluating those MOSs with more than 10 per year over the four years. I also looked at MOSs with at least 10 inputs in FY02, as some MOSs have been phased out or merged during the time frame of interest (1999-2002).

a. Straight-Through AIT Training

Straight-through AIT is conducted upon completion of BCT.

Looking at the MOS with usage rates of less than 65% over the four years 1999-2002, 14 specialties meet the stated criteria above. These specialties are shown in Table 5.

Table 5. Low Usage AIT MOS

Table 5 lists the low quota usage AIT MOS for the four year span from 1999 to 2002. Usage and average usage values are in percent.

ENL	1999			FY00			FY01			FY02			AVG
MOS	USE	QTA	INP	USE	QTA	INP	USE	QTA	INP	USE	QTA	INP	USE
98C	7.7	26	2	36.4	11	4	8.3	12	1	30.8	13	4	20.8
31P	29.6	27	8	20.8	24	5	45.0	20	9	28.6	14	4	31.0
97E	50.0	16	8	54.2	24	13	18.5	27	5	18.6	43	8	35.3
63H	18.2	33	6	73.7	19	14	41.2	17	7	15.4	13	2	37.1
35J	26.7	15	4	18.2	11	2	70.0	20	14	61.5	13	8	44.1
91S	26.7	45	12	51.1	90	46	47.8	69	33	54.2	72	39	44.9
63Y	34.6	26	9	23.5	17	4	55.6	9	5	78.6	14	11	48.1
31R	59.8	82	49	35.2	54	19	53.9	76	41	51.6	62	32	50.1
62H	52.8	36	19	49.0	51	25	48.5	68	33	66.7	30	20	54.2
92M	34.9	43	15	53.7	54	29	78.0	41	32	57.1	35	20	55.9
25R	72.7	11	8	76.2	21	16	37.0	27	10	50.0	12	6	59.0
96D	53.3	30	16	38.9	18	7	47.8	23	11	100.0	14	14	60.0
35E	26.3	38	10	65.1	43	28	87.1	70	61	63.8	47	30	60.6
88H	35.9	326	117	53.8	260	140	88.3	265	234	65.1	318	207	60.8

The common characteristic for the low usage MOSs is the low number of overall quotas. Only three of the low performing MOSs had more than 50 quotas in 2002. The low number of quotas is a reflection on the low overall density of the MOSs within the USAR, the limited potential number of locations, and the possible limited access to potential recruits. We will look at 91S (Preventive Medicine Specialist) in more detail later on.

The high performing MOSs, or those meeting the criteria and having an average quota usage rate in excess of 90%, are shown in Table 6.

Table 6. High Usage AIT MOSs

Table 6 lists the high usage MOSs for the four year span from 1999 to 2002. Usage and average usage values are in percent.

ENL	1999			FY00			FY01			FY02			AVG
MOS	USAGE	QTA	INP	USAGE	QTA	INP	USAGE	QTA	INP	USAGE	QTA	INP	USAGE
73D	73.1	26	19	71.0	31	22	950.0	2	19	84.4	32	27	294.6
75F	89.7	29	26	118.5	27	32	725.0	4	29	105.0	20	21	259.5
75H	155.5	182	283	88.6	246	218	304.1	122	371	115.8	221	256	166.0
75B	160.9	138	222	105.8	104	110	124.5	139	173	109.0	111	121	125.0
91E	112.7	150	169	114.8	128	147	158.1	43	68	110.5	76	84	124.0
73C	90.0	160	144	97.5	122	119	175.0	40	70	131.9	47	62	123.6
91D	105.6	144	152	99.5	210	209	135.1	77	104	101.0	98	99	110.3
92A	124.7	446	556	89.0	671	597	112.2	607	681	100.3	738	740	106.5
74B	118.9	53	63	79.8	84	67	152.9	34	52	74.2	62	46	106.4
92Y	114.4	263	301	96.8	411	398	107.0	473	506	104.3	234	244	105.6
38A	108.8	113	123	103.5	170	176	102.7	149	153	99.6	228	227	103.7
37F	121.0	105	127	105.7	87	92	106.0	83	88	78.8	259	204	102.9
77W	89.2	93	83	124.2	99	123	100.0	203	203	95.3	233	222	102.2
91K	111.1	36	40	59.1	93	55	116.4	55	64	122.0	50	61	102.2
51M	113.2	38	43	122.2	18	22	86.0	43	37	84.2	38	32	101.4
91X	87.8	41	36	89.5	76	68	101.0	103	104	124.3	37	46	100.6
91A	157.1	7	11	70.5	61	43	93.3	60	56	75.8	33	25	99.2
88N	73.0	141	103	101.2	169	171	121.4	187	227	99.6	271	270	98.8
25M	103.1	32	33	94.7	19	18	96.2	26	25	88.5	26	23	95.6
77F	78.9	331	261	97.0	536	520	111.3	577	642	94.9	846	803	95.5
91T	71.4	7	5	88.9	18	16	141.7	12	17	80.0	20	16	95.5
92G	89.7	348	312	90.1	433	390	99.8	515	514	98.0	356	349	94.4
96B	90.4	52	47	100.0	86	86	102.9	70	72	82.2	73	60	93.9
45B	100.0	21	21	81.6	38	31	100.0	18	18	91.7	12	11	93.3
31L	112.0	75	84	66.4	119	79	107.1	98	105	81.6	103	84	91.8
71L	101.4	587	595	67.2	696	468	87.4	824	720	108.3	780	845	91.1

Of the 26 higher-usage MOSs, only 9 had fewer than 9 quotas in 2002.

The usage rates highlight MOSs that would be interesting to look at in more detail from a demographic and recruiting perspective.

b. Split-Option AIT Training

The new split-option recruit attends AIT the year following his BCT. The new soldier must go to the MEPS and ship to AIT just as he or she did for their BCT training.

Until the soldiers complete their AIT, they are not deployable members of the USAR, and do not contribute to their assigned units' personnel readiness.

Before looking specifically at the split-option MOSs, I will compare overall phase 2 to phase 1 attendance.

Table 7. IET Completion Rate for Split-Options

Table 7 compares the phase 1 BCT inputs against the following year's phase 2 AIT inputs to estimate the IET completion rate for a fiscal year's split-option enlistments.

FY	BCT INP	FY	AIT INP	% COMPLETE IET
1999	2364	2000	1777	75.2%
2000	3262	2001	2059	63.1%
2001	2383	2002	1527	64.1%
2002	2393			

Table 7 shows that the estimated completion rate, based on comparing phase 1 inputs to the following year's phase 2 input, is less than 65% for each of the last two years.

Now looking at the split-option MOSs that had 20 or more quotas for 2002, only 5 MOSs had 80% or better average usage over the four year period. The overall average quota usage for phase 2 AIT is 65%, similar to the IET completion rate for the last two years. This indicates that the phase 2 quotas are similar in quantity to the phase 1 training inputs for the year prior, and only 65% of

the previous years training inputs return for phase 2 training.

One question that the training seat data cannot address is the reason for such a low usage rate of phase 2 AIT seats. But looking at the REQUEST data, I tracked the SSNs that did not ship. When looking at the number of phase 2 records with a phase 2 AIT date, the only records without ship dates were for the 2003 class dates. Summing the entire list of over 8,160 non-OSUT split-option records, only two records showed an AIT date without a ship date for an AIT starting in 2002 or earlier. The major problem appears to be lack of a scheduled date, as 3,559 records showed a null or blank for the phase 2 AIT start date. **It seems a significant proportion of split-option phase 1 trainees are not going to phase 2 training because they are not scheduled to go**.

Table 8. Split-Option AIT Quotas and Inputs FY01-02

Table 8 lists the split-option AIT MOS with 20 or more quotas for 2002, with quotas, inputs and usage by year. The last two columns provide the four-year overall annual usages and average quotas. The total row is the total for all MOSs.

ENL MOS	FY01 USAGE	QTA	INP	FY02 USAGE	QTA	INP	AVG USAGE	AVG QTA
88M1	79.6%	186	148	82.3%	186	153	77.6%	183.5
71L1	92.9%	126	117	105.1%	137	144	75.1%	121.5
77F1	78.8%	132	104	90.2%	122	110	80.2%	92.5
38A1	71.6%	74	53	52.3%	88	46	54.7%	77.0
63B1	43.5%	108	47	72.1%	86	62	49.7%	120.3
92A1	52.6%	190	100	108.8%	80	87	86.4%	127.0
75H1	84.3%	121	102	95.2%	63	60	73.0%	88.5
92G1	44.9%	127	57	96.6%	58	56	56.4%	80.3
88H1	64.5%	76	49	66.7%	54	36	55.1%	61.8
37F1	77.8%	45	35	72.2%	54	39	65.8%	44.0
88N1	78.5%	93	73	89.1%	46	41	75.9%	56.3
63S1	44.2%	77	34	104.8%	42	44	51.6%	72.3
52D1	59.0%	105	62	73.2%	41	30	57.1%	70.0
62E1	148.3%	29	43	76.3%	38	29	89.7%	39.8
62B1	47.6%	82	39	88.6%	35	31	68.4%	54.5
63W1	62.7%	67	42	100.0%	34	34	61.7%	46.3
92Y1	48.6%	142	69	145.5%	33	48	83.4%	108.5
51B1	24.2%	62	15	78.6%	28	22	64.5%	63.3
75B1	106.9%	29	31	125.0%	28	35	78.9%	42.8
31U1	62.0%	50	31	85.2%	27	23	62.5%	47.0
77W1	67.2%	61	41	134.6%	26	35	87.3%	39.0
62J1	35.4%	48	17	105.0%	20	21	67.5%	36.8
96B1	46.9%	32	15	75.0%	20	15	66.5%	23.5
Total	59.1%	3482	2059	92.0%	1659	1527	65.1%	2669.8

3. OSUT Data

The last training category is the OSUT enlistees. There are a small number of MOSs in the USAR that have their initial training conducted using OSUT. OSUT combines the aspects of starting the IET training path and receiving the advance training. We will examine the OSUT data like the BCT data, except that we break it out by MOS. There

are only a few OSUT MOS, and only five that involve more than 20 total quotas in one year. 11C (Indirect Fire Infantryman), 11H (Heavy Anti-Armor Weapon Infantryman), 13B (Cannon Crewmember), 19D (Cavalry Scout), and 19K (M1 Armor Crewman) are low density in terms of quotas and will not be examined. The OSUT programs for 71L (Administrative Specialist) and 63A (Abrams Tank Systems Maintainer) did not start until FY 2003 and will not be considered.

a. Straight-Through Training

The MOSs with 20 or more straight-through quotas form a small group, consisting of 11B (Combat Infantryman), 12B (Combat Engineer), 12C (Bridge Crewmember), 54B (Chemical Operations Specialist), and 95B (Military Police).

Of all the OSUT MOSs, 95B is the only one averaging over 90% usage, and 11B is the only one averaging less than 65%.

11B (Combat Infantryman) is interesting in that there is only one active infantry battalion in the USAR, which contains most, if not all, of the entry level positions. In the last two years, it totaled 136 inputs against 193 quotas (70.5% usage). It is the only OSUT MOS not averaging at least 80% quota usage.

95B (Military Policeman) is the core MOS for Military Police units which are positioned across the United States in many locations. 95B had 1,415 inputs against 1,513 quotas over the last two years, yielding a 93.3% usage rate.

54B (Chemical Operations Specialist) is the core MOS in chemical warfare units, as well as being present in

most other battalion level and larger units. Its overall usage is 86.1% over the last two years, with 762 inputs against 885 quotas.

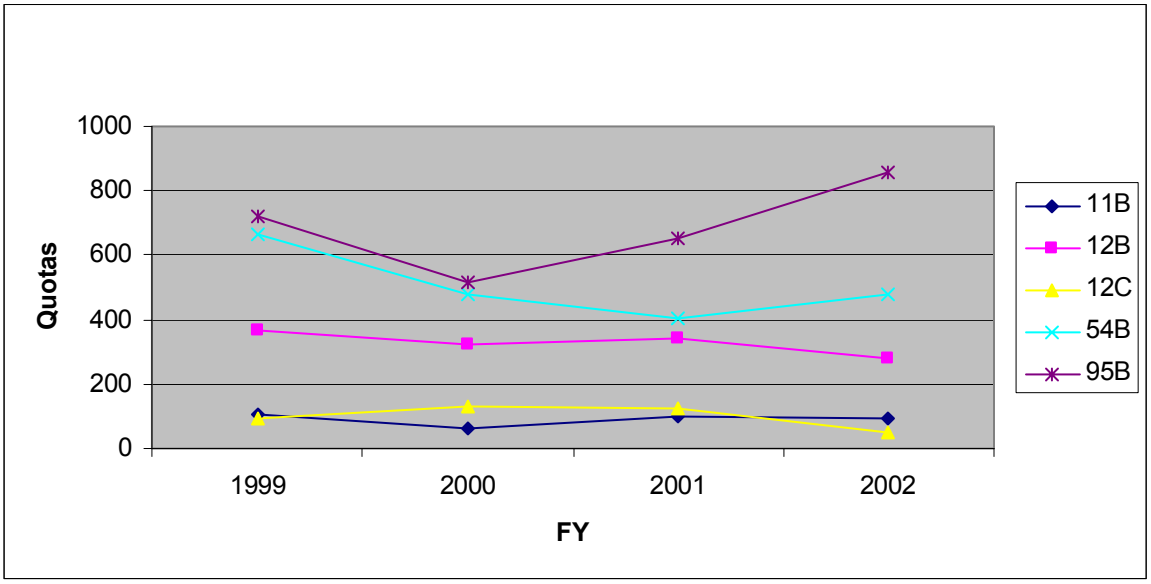


Figure 24. OSUT Straight-Through Quotas

Figure 24 shows the annual straight-through quotas for the major OSUT MOSs for fiscal years 1999 through 2002.

Table 9. OSUT Straight-Through Quotas and Inputs

Table 9 lists the five major OSUT MOS straight-through quotas and inputs for fiscal years 1999 through 2002.

ENL	1999			FY00			FY01			FY02		
MOS	USE	QTA	INP	USE	QTA	INP	USE	QTA	INP	USE	QTA	INP
11B	55.6%	108	60	89.2%	65	58	85.1%	101	86	55.4%	92	51
12B	53.8%	364	196	83.1%	320	266	86.3%	343	296	91.5%	282	258
12C	63.2%	95	60	67.4%	129	87	81.6%	125	102	106.4%	47	50
54B	53.2%	665	354	74.2%	476	353	98.8%	404	399	78.4%	481	377
95B	86.4%	723	625	99.8%	516	515	99.7%	653	651	90.9%	860	782

Of all the OSUT MOSs, and all others as well, no MOS has the same volume and usage as 95B. I will look at the 95B and 54B in detail later on.

b. Split-Option Phase 1 Training

Split-option phase 1 quotas have increased over the last four years, with 12B, 54B, and 95B having the largest density, as shown in Figure 25. The overall numbers are shown in Table 10.

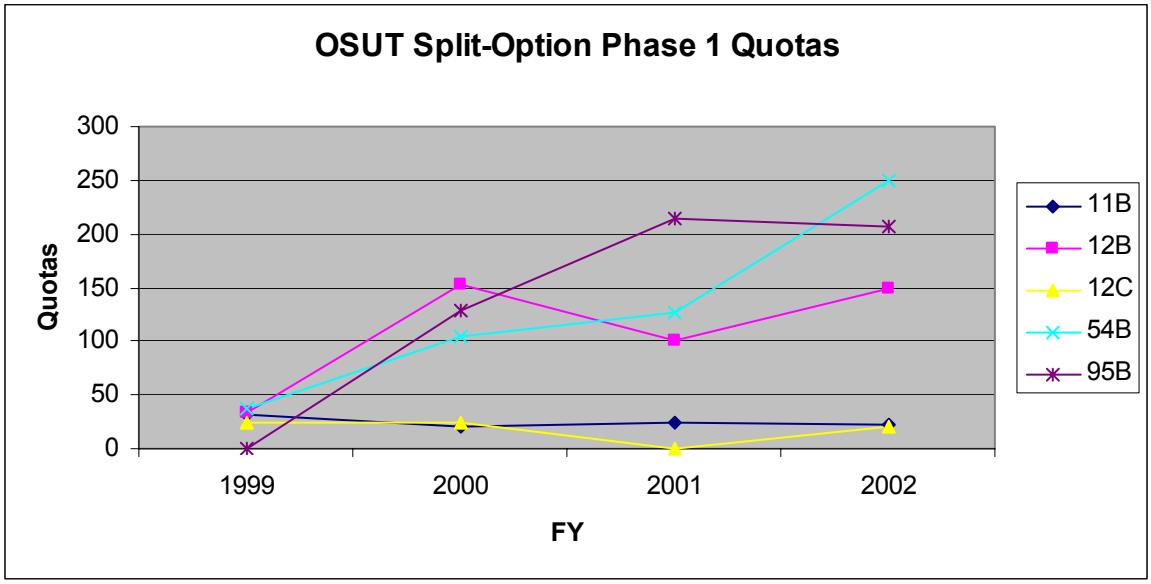


Figure 25. OSUT Split-Option Phase 1 Quotas

Figure 25 shows annual split-option phase 1 quotas for fiscal years 1999 through 2002.

Table 10. OSUT Split-Option Phase 1 Quotas and Inputs

Table 10 lists the annual quotas and inputs for phase 1 split-option OSUT for fiscal years 1999 through 2002.

ENL	1999			FY00			FY01			FY02			AVG
MOS	USE	QTA	INP	USE	QTA	INP	USE	QTA	INP	USE	QTA	INP	USE
12B	79.4%	34	27	63.2%	152	96	67.0%	100	67	87.3%	150	131	64.3%
12C	56.0%	25	14	92.0%	25	23	0.0%	0	0	66.7%	21	14	73.6%
54B	115.8%	38	44	106.7%	105	112	73.0%	126	92	92.4%	249	230	71.8%
95B	0.0%	0	6	117.8%	129	152	79.4%	214	170	99.5%	207	206	92.3%
Total	71.1%	128	91	91.2%	431	393	76.8%	465	357	93.4%	649	606	86.5%

Much like the split option BCT quotas, there is a relatively high average usage rate. There does not seem to be a problem getting enlistments using the split-option program, but since they are not a deployable asset to their unit until they complete phase 2, the phase 2 numbers tell us more about the effectiveness of the program.

c. Split-Option Phase 2 Training

Phase 2 split-option training quota usage shows a marked difference from the phase 1 training seat usage. The average usage is 25% less than the phase 1 average. The lower phase two usage is similar to the non-OSUT split-option figures for phase 1 and phase 2 usages. Of the 1,346 records in the REQUEST data for split-option OSUT trainees who had BCT date prior to 2003, 396 did not have a scheduled phase 2 AIT date. Similar to the non-OSUT phase 2 split-option usage, there appears to be a large population of phase 1 trainees not being scheduled for phase 2.

Table 11. OSUT Split-Option Phase 2 Quotas and Inputs

Table 11 lists the annual high-density OSUT split-option phase 2 inputs and quotas for fiscal years 1999 through 2002.

ENL	1999			FY00			FY01			FY02			AVG
MOS	USE	QTA	INP	USE	QTA	INP	USE	QTA	INP	USE	QTA	INP	USE
12B	73.5%	68	50	63.6%	44	28	69.6%	102	71	102.0%	51	52	75.8%
12C	50.0%	6	3	64.0%	25	16	53.6%	28	15	0.0%	0	0	57.6%
54B	24.3%	70	17	74.4%	39	29	57.5%	120	69	39.2%	186	73	45.3%
95B	40.4%	136	55	0.0%	0	2	82.6%	115	95	73.1%	145	106	65.2%
Total	44.6%	280	125	69.4%	108	75	68.5%	365	250	60.5%	382	231	60.0%

The next item to compare is the estimated IET completion rate. Calculated as a whole for the OSUT MOSs,

and separately by each MOS, the numbers in Table 12 are similar than the non-OSUT MOS IET completion, though slightly higher at 69% versus 65% for the non-OSUT.

Even the 95B MOS, which enjoys high usage rates for both straight-through and phase 1 split-option recruits, achieves only a 62% average phase 2 usage rate. There is a systemic problem for phase 2 split-options in that the apparent lack of scheduling is the major reason for low quota usage.

Table 12. OSUT Estimated IET Completion Rate

Table 12 lists the estimated IET completion rate for the split-option OSUT MOS enlistees who start phase 1 in fiscal years 1999 through 2001.

	1999 2000			2000 2001			2001 2002			AVG
MOS	PH1	PH2	% IET	PH1	PH2	% IET	PH1	PH2	% IET	% IET
12B	27	28	103.7%	96	71	74.0%	67	52	77.6%	79.5%
12C	14	16	114.3%	23	15	65.2%	0	0	0.0%	83.8%
54B	44	29	65.9%	112	69	61.6%	92	73	79.3%	69.0%
95B	6	2	33.3%	152	95	62.5%	170	106	62.4%	61.9%
Total	91	75	82.4%	383	250	65.3%	329	231	70.2%	69.2%

B. REQUEST-REAF INTEGRATED TRAINING DATA

The REQUEST-REAF data with the month and year coded for both the enlistment and training start date is used to establish if there is a relationship between enlistment date and training start date.

IET Straight Through REQUEST Aggr...

EnlYearMonth	IETYearMonth	SumOfRecord
1999-6	1999-11	28
1999-6	1999-6	36
1999-6	1999-7	270
1999-6	1999-8	216
1999-6	1999-9	161
1999-6	2000-1	42
1999-6	2000-2	7
1999-6	2000-3	14
1999-6	2000-5	4
1999-6	2000-6	126
1999-6	2000-7	1
1999-6	2000-8	2
1999-6	2000-9	2
1999-7	$null$-$null$	3
1999-7	1999-10	56
1999-7	1999-11	62
1999-7	1999-7	37
1999-7	1999-8	182

Record: 1 of 885

Figure 26. IET Straight-Through REQUEST Aggregate Table

Figure 26 is a portion of the Table generated by an ACCESS query to aggregate the IET from REQUEST data down to enlistment date-start date bins.

I aggregated the data by MOS, enlistment month, enlistment year, training start date, and training year again to eliminate the MOS, and have a resulting table with one entry per enlistment date-training start date combination, as shown in Figure 26. The results are then run through a second query to put the results in matrix form, as shown in Figure 27.

IET Straight Through REQUEST Aggregate_Crosstab : Crosstab Query

EnlYearMonth	1999-11	1999-1	1999-2	1999-3	1999-4	1999-7	1999-8
1997-10							
1997-11							
1997-12							
1998-1							
1998-10	2	2	12	8	2	193	37
1998-11				1	1	169	42
1998-12	8	213				137	82
1998-2							
1998-3							1
1998-4							
1998-5							
1998-6							
1998-7		4				113	
1998-8		5				132	6
1998-9		4		2		132	9
1999-1	4	844	161			176	156
1999-10	227						
1999-11	33						
1999-12							

Record: 1 of 68

Figure 27. IET Straight-Through Aggregate Crosstab

Figure 27 shows the table of results from the IET straight-through aggregate crosstabulation, which re-bins the date by enlistment date against IET start date.

Using ACCESS once again, I screened the data for null entries in either the IET training date or the enlistment date.

Each entry needs to have the training start date fields of month and year replaced by a value for months between enlistment and start date, starting with 0 for those who start during their month of enlistment. To accomplish this data transformation and also place the data into a matrix, I once again used an ACCESS crosstab query. The results for the straight-through and split-option recruits are shown in Figures 28 and 29, respectively.

IET Straight Through with months : Crosstab Query

EnlMonth	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	71	873	651	179	214	284	277	416	150	56	45		6	1			2
2	80	1337	380	302	288	303	529	175	91	54		15	3	1			6
3	232	1035	626	295	262	668	217	178	112		33	7	5	3			3
4	75	1000	282	246	560	251	195	145		95	18	9	3	1	2	3	
5	196	686	273	617	321	195	232		150	44	22	3	12	5	1		1
6	64	645	562	522	211	276		255	118	43	9	36	586	25	7	3	4
7	48	802	548	359	282		330	139	60	27	23	421	374	14	2	4	1
8	129	1000	589	461		425	141	91	36	102	295	375	113	6	2	4	
9	124	674	1071		559	229	139	79	115	336	329	224	24	3	5		2
10	57	682		300	210	99	66	104	158	237	176	30	13	7		3	2
11	140		394	351	146	56	80	172	192	166	35	17	16		1	1	
12		375	333	195	42	92	160	190	198	49	16	18			4		1

Record: 1 of 12

Figure 28. IET Straight-Through by Months Crosstab

Figure 28 shows the results of the IET straight-through by months crosstabulation, that further aggregates the data down to enlistment month by number of months out until starting IET, whether BCT or OSUT.

The matrix in Figure 28 reveals a null diagonal. This null diagonal represents December, as no IET training had a report date that qualified as December during the initial binning by IET training start date by month by year.

The relationship between the split-option enlistment month and the delay in months was unusual but not unexpected. The results in Figure 29 are organized the same as in Figure 28. The first thing that stands out is the null diagonal associated with December, just like that for the straight-through enlistments. The other is the diagonal with 60% of more of all the observations. That diagonal corresponded to June, which is the month in which 70% or more of all split-option enlistees begin training. The diagonals associated with May and June account for more than 90% of the observations.

IET Split Option Crosstab by month : Crosstab Query

EnlMonth	0	1	2	3	4	5	6	7	8	9	10	11	13
1		7	1	1	197	1329	5		1				
2		5	1	208	1454	12							
3			190	1239	17		5						
4	1	139	714	20		1							
5	55	521	25	3	5	4							1
6	110	21	10	4									
7	5	5	2		1		3			2	2	1	
8		8	1			2		2		40	8		
9	6	10	2		4	1		2	92	381	2	1	
10	2	1		2	4			48	303	14			
11			3	1	1		73	449	8				
12		3			1	109	567	3	2				

Record: 1 of 12

Figure 29. IET Split-Option by Month Crosstab

Figure 29 shows the results of the IET split-option by month crosstabulation, with enlistment month against delay in months until the start of phase 1 IET training.

The split-option results show that there is clearly a relationship between the enlistment month and the delay in months until training starts. In any column, between 92% (column 0) and 98% (column 4) of all the entries are in the two cells that correspond to May and June.

Unlike the split-option crosstabulation, the straight-through data shows no clear relationship other than the December null diagonal. To eliminate this null diagonal, I combined the December and November accessions into a single month. I imported the data from ACCESS into EXCEL, made the appropriate modifications to the matrix for combining November and December, and binned all the entries past 12 months into a combined column representing 12 or more months.

Once in EXCEL, I then build a table of proportions, with a second matrix representing the matrix of expected

values based on the assumption enlistment date and delay in months until the IET start date are independent, as shown in Figure 30.

Enl	**Original Values**													
Month	0	1	2	3	4	5	6	7	8	9	10	11	12 or more	
Jan	71	873	651	179	214	284	277	416	150	56	45	6	10	3232
Feb	80	1337	380	302	288	303	529	175	91	54	15	3	10	3567
Mar	232	1035	626	295	262	668	217	178	112	33	7	5	7	3677
Apr	75	1000	282	246	560	251	195	145	95	18	9	3	6	2885
May	196	686	273	617	321	195	232	150	44	22	3	12	8	2759
Jun	64	645	562	522	211	276	255	118	43	9	36	586	45	3372
Jul	48	802	548	359	282	330	139	60	27	23	421	374	26	3439
Aug	129	1000	589	461	425	141	91	36	102	295	375	113	13	3770
Sep	124	674	1071	559	229	139	79	115	336	329	224	24	14	3917
Oct	57	682	300	210	99	66	104	158	237	176	30	13	18	2150
Nov/Dec	112	1447	1110	881	493	606	394	178	70	32	457	960	71	6811
	1188	10181	6392	4631	3384	3259	2512	1729	1307	1047	1622	2099	228	39579
	Expected Values													
Jan	97	831	522	378	276	266	205	141	107	85	132	171	19	3232
Feb	107	918	576	417	305	294	226	156	118	94	146	189	21	3567
Mar	110	946	594	430	314	303	233	161	121	97	151	195	21	3677
Apr	87	742	466	338	247	238	183	126	95	76	118	153	17	2885
May	83	710	446	323	236	227	175	121	91	73	113	146	16	2759
Jun	101	867	545	395	288	278	214	147	111	89	138	179	19	3372
Jul	103	885	555	402	294	283	218	150	114	91	141	182	20	3439
Aug	113	970	609	441	322	310	239	165	124	100	154	200	22	3770
Sep	118	1008	633	458	335	323	249	171	129	104	161	208	23	3917
Oct	65	553	347	252	184	177	136	94	71	57	88	114	12	2150
Nov/Dec	204	1752	1100	797	582	561	432	298	225	180	279	361	39	6811
	1188	10181	6392	4631	3384	3259	2512	1729	1307	1047	1622	2099	228	

Figure 30. Tables of Proportion

Figure 30 shows the tables with the original and expected values, assuming independence of enlistment date and delay in months until starting IET training.

I then generated a matrix of the residuals or differences. Then I squared the differences and divided by the expected values in order to generate the values to test for independence. The resulting table is shown in Figure 31.

Squared differences/observed												
6.974	2.084	31.9	104.9	14.06	1.2	25.18	534.9	17.54	10.18	57.74	159.6	3.989
6.843	191.8	66.73	31.89	0.945	0.294	404.5	2.36	6.094	17.26	117.7	183.2	5.415
134	8.404	1.742	42.51	8.728	440.6	1.149	1.879	0.731	42.46	137	185.1	9.495
1.553	89.61	72.61	24.84	398	0.761	0.773	2.855	8E-04	44.56	100.9	147.1	6.786
154.7	0.792	66.84	268.1	30.7	4.558	18.48	7.208	24.36	35.62	107.1	123.3	3.92
13.68	57.02	0.557	41.17	20.73	0.01	7.849	5.83	41.96	72.11	75.57	927.1	33.67
29.54	7.717	0.099	4.678	0.493	7.744	28.79	54.19	65.98	50.79	556.5	201.3	1.934
2.217	0.943	0.647	0.896	32.7	92.47	91.88	100.6	4.065	382.3	314.7	37.8	3.499
0.351	110.4	303.8	22.12	33.49	104.4	115.7	18.4	330.1	490.2	25.1	162.5	3.251
0.88	30.07	6.423	6.867	39.14	69.64	7.72	43.72	388.1	249.5	38.32	89.5	2.545
41.8	53.1	0.091	8.868	13.71	3.638	3.39	48.02	106.7	121.9	113.4	992.6	25.72

Figure 31. Table of Squared Differences

Figure 31 shows the squared differences between the actual and expected squared, and divided by the expected.

Summing the differences and comparing to a χ^2 distribution with (11-1)*(12-1) degrees of freedom, the results were highly significant (p-value = 0.012). The probability of independence being small, I then compared the residuals to the expected values.

Using a proportion of 20% as the baseline to determine if there is an increased or decreased likeliness of an enlistment in a particular month to have a corresponding delay, I built a matrix of plusses and minuses. This is shown in Figure 32.

Month	0	1	2	3	4	5	6	7	8	9	10	11	12 or more
Jan	-		+	-	-		+	+	+	-	-	-	-
Feb	-	+	-	-			+		-	-	-	-	-
Mar	+			-		+				-	-	-	-
Apr		+	-	-	+					-	-	-	-
May	+		-	+	+		+	+	-	-	-	-	-
Jun	-	-		+	-				-	-	-	+	+
Jul	-						-	-	-	-	+	+	+
Aug					+	-	-	-		+	+	-	-
Sep		-	+	+	-	-	-	-	+	+	+	-	-
Oct		+			-	-	-	+	+	+	-	-	+
Nov/Dec	-							-	-	-	+	+	+

Figure 32. Enlistment Month by Delay in Months Matrix

Figure 32 shows the matrix that denotes a plus for a delay that is high for that given month of enlistment, and a minus for a delay that is low.

Looking at the resulting matrix, it appears that applicants enlisting in the first half of the year are less likely to delay more than 8 months. The highlighted diagonal of plusses starting with a three month delay in May to a seven month delay in January corresponds to August. This diagonal is surrounded by neutral cells, and seems to indicate that the summer months of the same year are not unusual for applicants enlisting January through May. The diagonal associated with February has no plusses and only two neutral cells, indicating that February is not a favorite month to start BCT or phase 1 OSUT. There are two more highlighted rows of plusses from October with a seven and eight month delay to June with an eleven and twelve month delay. One of these “months” includes the combined November/December “month”, and thus corresponds to June and July of the following year.

The cells associated with June, July and August collected the most “plusses”, indicating that those months may be the most favorable. Starting with June the year

before, June and July are the main high demand months. In January, the high demand months are July, August, and September. Starting in February, the high demand diagonal is July. The neutral cells corresponding to June are neutral until April, possibly representing that there are training seats with start dates available, but not for all specialties. Further analysis is required to say more with any certainty.

V. DEMOGRAPHIC OVERVIEW

In the demographic overview I will look at several quantitative and qualitative variables for the entire accession population, straight-through and split-option recruits, and three MOSs: 54B (Chemical Operations Specialist), 91S (Preventive Medical Specialist), and 95B (Military Police). These three MOSs were chosen because 95B is a high quota usage MOS, 91S a low quota usage MOS, and 54B an average quota usage MOS.

A. THE QUANTITATIVE VARIABLES

Descriptive statistics for several quantitative demographic variables are shown in Table 13. These are listed for the overall population, as well as separately by the training program (straight-through or split-option training) and the MOSs.

Table 13. Quantitative Descriptive Statistics

Table 13 lists the quantitative descriptive statistics for six populations. The categories are education in years, Armed Forces Qualification Test (AFQT) score, age, and days between enlistment date and BCT/OSUT start date. The statistics include the record counts, means, and standard deviations. The +/- rows provide the 99% confidence half-interval width for the mean.

		Total	Straight -through	Split -option	95B	54B	91S
EDYRS	Count	62361	59189	3172	3375	2056	159
	Mean	12.10	12.09	12.29	12.08	12.15	12.50
	SD	3.87	3.95	2.03	1.81	4.05	1.56
	+/-	0.04	0.04	0.09	0.08	0.23	0.32
AFQT	Count	72506	61343	11163	3779	2452	159
	Mean	59.86	59.40	62.40	63.02	62.26	76.33
	SD	19.11	19.26	18.09	17.15	18.00	14.09
	+/-	0.18	0.20	0.44	0.72	0.94	2.88
AGE	Count	72509	61346	11163	3780	2452	159
	Mean	20.062	20.39	18.27	20.24	19.89	20.48
	SD	3.42	3.51	2.13	3.47	3.26	2.87
	+/-	0.03	0.04	0.05	0.15	0.17	0.59
Days Enlst to Train	Count	72508	61346	11162	3780	2452	159
	Mean	111.136	107.809	129.425	137.27	115.498	122.616
	SD	96.38	100.69	65.01	100.62	89.23	88.98
	+/-	0.92	1.05	1.59	4.22	4.64	18.18

Of the four quantitative variables, I found education in years to be a problem, particularly so for the split-option trainees. There were 7,989 of 11,163 records that had a null or blank value for education in years. The split-option trainees accounted for 80% of these values. As such, I will make no comparisons that reference split-options and education in years. The 91S had, on average, nearly 5 months additional education than the total population. The fact that 91S has an enlistment

requirement of one year of high school algebra/chemistry or equivalent means it has a somewhat higher educational requirement than most specialties.

The AFQT score was interesting in that all the sub-populations other than the straight-through had a higher mean AFQT score than the base population. Although there is no minimum score required for the AFQT, to be a 91S (Preventive Medicine Specialist), an applicant must score a minimum of 105 in the Skilled Technical (ST) section of the Armed Services Vocational Aptitude Battery (ASVAB) (as stated in Department of the Army Pamphlet 611–21, Military Occupational Classification and Structure). 54B (Chemical Operations Specialist) and 95B (Military Policeman) also have a requirement for a minimum ST score, each requiring a score of 95 or better. These minimum scores may be part of the reason for the above average AFQT scores.

There is a large difference in split-option trainees who are, on average, nearly two years younger than those who select the straight-through option. This difference is nothing unexpected, given that the split-option program primarily targets students.

The last quantitative variable I examined was the time in days between enlistment date and BCT/OSUT start date. The split-option and straight-through enlistments differed in the mean number of days, with the split-option program seeing a 21 day longer delay on average than the straight through enlistments. Of the three MOSs, 91S and 95B both have longer average delays. The longer average delay for split-options is not a surprise, as they enlist throughout the year from predominately summer training start dates. The longer delays for the 91S may be a number of things,

one possibly being there are only seven classes conducted during the year. For 95B, with a high average quota usage and an average of 21 classes conducted a year, the delay would indicate that the classes fill up quickly and that an applicant would be willing to delay longer to be a Military Police.

Table 14. MOS Quantitative Descriptive Statistics

Table 14 lists the quantitative descriptive statistics for six populations. The categories are education in years, Armed Forces Qualification Test (AFQT) score, age, and days between enlistment date and BCT/OSUT start date. The statistics include the record counts, means, and standard deviations. The +/- rows provide the 99% confidence half-intervals.

		Straight-through	Split-Option	95B ST	95B SO	54B ST	54B SO
EDYRS	Count	59189	3172	3120	255	1909	147
	Mean	12.09	12.29	12.05	12.41	12.15	12.07
	SD	3.95	2.03	1.84	1.31	4.18	1.38
	+/-	0.04	0.09	0.09	0.21	0.25	0.29
AFQT	Count	61343	11163	3133	646	1934	518
	Mean	59.40	62.40	62.91	63.56	61.73	64.25
	SD	19.26	18.09	17.20	16.90	18.14	17.32
	+/-	0.20	0.44	0.79	1.71	1.06	1.96
AGE	Count	61346	11163	3134	646	1934	518
	Mean	20.388	18.27	20.55	18.70	20.37	18.08
	SD	3.51	2.13	3.57	2.38	3.39	1.81
	+/-	0.04	0.05	0.16	0.24	0.20	0.20
Days Enlist to Train	Count	61346	11162	3134	646	1934	518
	Mean	107.809	129.425	135.95	143.676	112.601	126.315
	SD	100.69	65.01	107.34	57.49	96.23	54.55
	+/-	1.05	1.59	4.94	5.83	5.64	6.17

Since there are differences between the split-option and straight-through trainees, it is hard to make any statements about the specific MOSs without looking the

populations broken down to split-option and straight-through populations. Since 91S (Preventive Medicine Specialist) had only four split-option trainees over the period examined, I have restricted further analysis to 54B and 95B. Table 14 shows descriptive statistics for the overall straight-through and split-option populations, as well as the two MOSs by training program.

The mean delays for 95B (Military Policeman) are clearly higher than the overall means. A 95B enlistee, on average, delays 28 days more than the population average for overall straight-through accessions. The 95B split-options also tend to begin later. Their delay is 14 more days on average. The 54B (Chemical Operations Specialist) accession delays from enlistment to training start are in line with the population averages.

The two MOSs' populations are not significantly different than the norm in terms of age, although the average AFQT scores are slightly higher than the respective overall populations.

B. THE QUALITATIVE VARIABLES

The qualitative variables I will consider are the market segment (a clustering of the population by economic indicators associated with a specific zip code plus , or nine-digit zip code), and the distribution of gender in the accession population.

This market segment is a commercial data product purchased by USAREC for use in their marketing analysis. It is a useful starting point for demographic analysis. A breakdown of the 50 segments, including names for each

segment and their categorization into one of 10 larger groups, is outlined in Appendix 4.

Once again, there is a large amount of missing information. For the overall population, the proportion of records missing a market segment was larger than the proportion shown having any one of the 50 market segments. One in five of the accessions did not have a valid market segment. The segments with 2% or more of the population are 38, 16, 18, 10, 40, 25, 11, 24, 15, 46, 17, 35, 23, and 5, as shown in Figure 33. The names for these market segments are listed in Table 15. Missing values correspond to market segment 99 in Figure 33.

Table 15. Sample Market Segment Names

Table 15 lists the names of the market segments that are used for comparisons with the MOS and the populations.

SEGMENT	SEGMENT NAME
5	PROSPEROUS METRO MIX
10	HOME SWEET HOME
11	FAMILY TIES
15	GREAT BEGINNINGS
16	COUNTRY HOME FAMILY
17	STARS AND STRIPES
18	WHITE PICKET FENCE
23	SETTLED IN
24	CITY TIES
25	BEDROCK AMERICA
32	METRO SINGLES
35	BUY AMERICAN
36	METRO MIX
40	TRYING METRO TIMES
46	DIFFICULT TIMES

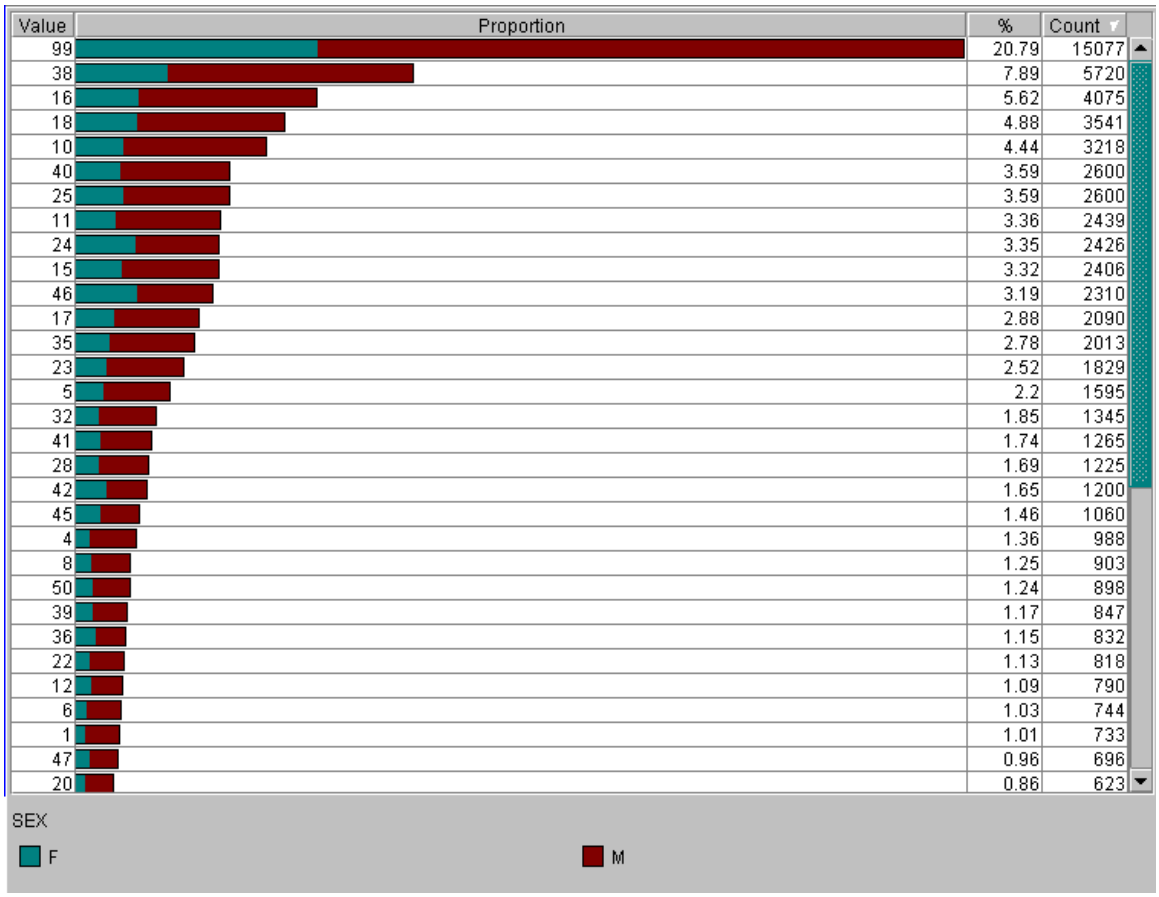


Figure 33. Overall Accession Market Segment

Figure 33 is a distribution graph of the market segments associated with the USAR accessions from 1999 through 2002 listed from top to bottom by proportion of population.

Comparing the distributions of the different MOSs against the overall is difficult with such a significant proportion of “segment-less” accessions. I will only look at the top segments from the overall population against the two training programs and the three MOSs. In building the chart in Figure 34, the 54B and 91S MOSs had two segments that are not in the top overall market segments appear in the top for their specialties, segments 32 and 36.

The bar chart in Figure 34 is based on proportions; so keep in mind that the population for 91S is relatively

small with all but two market segments consisting of fewer than 10 individuals.

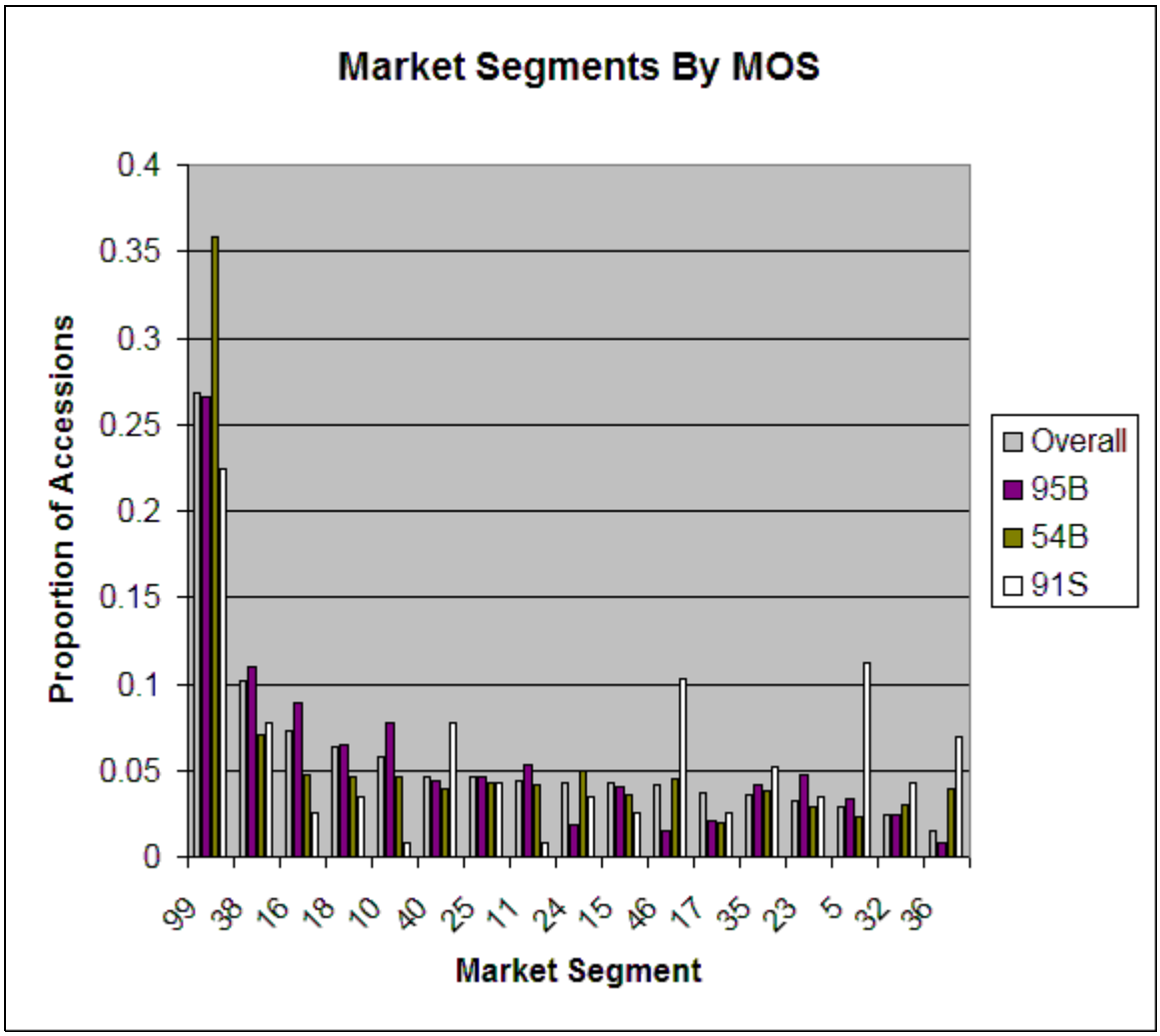


Figure 34. Top Market Segments for Three MOS

Figure 34 lists the proportions of the top market segments for the overall population, and the proportions for 95B, 91S and 54B. The proportions are for a subset of the accessions for just the listed segments, not all segments.

The 91S MOS does appear to differ from the overall population in terms of the market segments associated with its enlistees. Segment 38, the top market segment for the

overall, 95B and 54B, is third behind segments 5 and 46 for the 91S. Three of the top five market segments for 91S (segments 17, 32 and 36) were not in the top nine overall, and 91S also had markedly fewer in the segments 10, 1

The matter of the unassigned segments poses problems for making assessments on most variations. I will use the segment data to point out that, combined with the quantitative variable summaries, it appears that 91S (Preventive Medicine Specialist) is a different population from the overall, 95B, and 54B accession populations. The three market segments which 91S drew from less often (10, 11, 16) represent major segments of the overall population, and are all in the mainstream families group. But looking at the distribution of MOS against the groups, shown in Figure 35, it seems that 91S is the same in terms of the proportion of mainstream families. The interesting groups are called mainstream singles and sustaining singles, which contain the market segments from which 91S draws from more heavily. These are 32, 36, 40 and 46: three of these segments have “metro” in their segment name.

The 95B MOS (Military Policeman), although similar to the overall, seems to have a significantly lower proportion of sustaining families and a higher proportion of mainstream families.

Market Group	Overall	95B	54B	91S	Overall	95B	54B	91S
ACCUMULATED WEALTH	6.8%	8.1%	5.4%	9.4%	8.6%	10.2%	7.6%	11.3%
MAINSTREAM FAMILIES	35.5%	40.7%	27.9%	35.2%	44.8%	51.3%	39.2%	42.1%
YOUNG ACCUMULATORS	5.5%	5.4%	5.1%	2.5%	7.0%	6.8%	7.2%	3.0%
MAINSTREAM SINGLES	12.9%	13.4%	12.3%	16.4%	16.3%	16.8%	17.2%	19.5%
ASSET-BUILDING FAMILIES	0.8%	0.9%	1.0%	1.9%	1.0%	1.1%	1.4%	2.3%
CONSERVATIVE CLASSICS	1.3%	1.5%	1.2%	0.0%	1.7%	1.9%	1.7%	0.0%
CAUTIOUS COUPLES	0.2%	0.4%	0.1%	0.0%	0.3%	0.5%	0.2%	0.0%
SUSTAINING FAMILIES	10.5%	4.4%	10.8%	8.8%	13.3%	5.5%	15.1%	10.5%
SUSTAINING SINGLES	4.2%	3.1%	6.2%	8.2%	5.3%	3.9%	8.7%	9.8%
ANOMALIES	0.1%	0.1%	0.0%	0.6%	0.1%	0.2%	0.1%	0.8%
UNCLASSIFIED	1.2%	1.5%	1.2%	0.6%	1.6%	1.8%	1.7%	0.8%
UNMATCHED	20.8%	20.7%	28.7%	16.4%				

Figure 35. Market Group by MOS

Figure 35 shows the market group proportions for the three MOSs and the overall population.

Once again, with the high level of unknown market groups, it is hard to draw conclusions with any certainty. The large proportion of missing information must be addressed before further analysis is conducted with the demographic data, in case the pattern of missing values is not random. This might be accomplished by using the distribution of market segments and population by five-digit zip code to try to estimate the segment density associated with of the accessions for which no nine-digit zip code market segment match was obtained. By quantifying the unknown segments, then the data may prove to be more useful in making descriptions about the accession population.

The demographic data, when combined with REQUEST enlistment incentives data, may provide insight into relationships between market and incentives. These comparisons would have to be done first by MOS, and contrasted to the overall population. Adding in a geographic element, such as the recruiting battalion area where the applicant enlisted, could provide another

discriminator for analyzing the MOS demographic data. Contrasting the same MOS and incentive package by geographic area, then contrasting with other MOSs and the overall population, could in turn provide some information about regional differences in terms of enlistment patterns, MOS choices, and the effectiveness of incentives. This in turn could assist in making policy decisions such as assignment and composition of enlistment incentives or location of units or detachments.

THIS PAGE INTENTIONALLY LEFT BLANK

VI. CONCLUSIONS AND RECOMMENDATIONS

The first thing I will state is that quality analysis comes from quality data. I spent a great deal of time and effort to get the best quality data possible. My goal was to develop a process that could be repeated by me and others in future analysis with the REQUEST data. Since REQUEST is an accessioning system and not necessarily a decision support system, allowances have to be made for the data drawn from it. The method of extraction of this data is a software package called FOCUS. The data draws that I used were relatively large, and I do not believe that FOCUS is designed for this kind of use. Nonetheless, larger draws will be the norm if analysis is to be done over periods of time that entail a large number of accessions.

Implementing a structured process for cleaning and categorizing accessions data is important for any analysis in this regard.

The REQUEST data provided by the Army Reserve Personnel Command contained 87,958 records. Of these records, 15,443 of the records were duplicates or partial duplicates of some of the 72,156 unique SSNs. Without accounting for blank and invalid field entries, 17.6% of the records representing duplicate SSNs already needed to be reduced.

The process I built screened out all but 19 duplicate SSNs, deleted 2,546 blank and invalid unique SSNs, and deleted 69 other SSNs with duplicate records and data field inconsistencies. All records not included in the dataset for analysis were placed in a separate file with a deletion

code for further analysis as to assist in determining the problem with record by defining the reason it was not included in the data set for analysis.

This process is designed specifically for reuse, so that subsequent USAR analysts can start with a better understanding of the data problems associated with the dataset, and a relatively quick process to generate a quality dataset.

The training data provided by the Department of the Army from ATRRS provides an overview of the flow of enlistees into the system. Binning these data by month by training category and MOS (if applicable) from 1999 to 2002, I was able to look at the data over time. During the overview of the training seat data, I observed that quotas and quota usage patterns vary across training programs (split-option versus straight-through) and MOSs. Usage is particularly low for phase 2 split option quotas, averaging 65% from 1999 to 2002. The split-option IET completion rate, which is the ratio of phase 1 inputs to the following year's phase 2 inputs, is consistently low over the same time frame at 65%. This low rate of 65% matches the split-option phase 2 quota usage over the same time frame. The main problem seems to be the lack of scheduling of phase 2 split-option training. Improvements in the split-option training seat usage need to focus on getting phase 1 enlistees into phase 2, and a good start would be to schedule them for training. Currently, the applicant only schedules phase 1 when he or she enlists, and is supposed to schedule phase 2 after they complete phase 1. The USAR needs to improve the management of phase 1 enlistees to get more inputs into phase 2 the following year. The current

process relies on the individual enlistee and his or her assigned unit to make this happen, and is resulting in only a 65% estimated completion rate.

Analysis of the training seat data indicated a seasonal usage, with February being historically low and June and July being high. During the three-year period from 2000 through 2002, February overall BCT usage was less than 70%, while June and July were over 90%.

I used the REQUEST data, aggregated to month and year of enlistment, MOS, and start date of training, to link recruiting to IET training. These data, which are similar to the ATRRS data, are binned by training start month. I used the aggregated REQUEST data to try to uncover the relationship between the month of enlistment and the date training starts. The results support the seasonal highs and lows noted in the ATRRS summaries, particularly with respect to the high volume for summer months and the low volume for February. USAREC's suggestion for a USAR Seasonal Ship Bonus (monetary enlistment incentive) to encourage new potential applicants to enlist for February start dates seems to be a good way to address this problem. Further analysis into time relationships by MOS may provide other valuable insights into training seat scheduling and quota management issues.

The time of year an applicant enlists can affect both the selection of specialty and the resulting time he or she will start training. I found that the fall quarter enlistments tend to start training in the fall or in the summer of the following year; winter enlistments mostly began training in March or August; spring (April and May) enlistments generally began training in April, May, or

July; and summer enlistments began training the following summer. In the case of 95B (Military Policeman), the much higher delay after enlistment suggests applicants are willing to wait for training in order to become a 95B.

Identifying low IET training seat usage MOSs is the first step towards highlighting potential “problem” MOSs. The second step is to look for factors that might contribute to a lack of accessions for those particular specialties. In some cases, as with 91S, the population recruited to the specialty varies from the general accession population, and most certainly from other specialties. Identifying MOS-specific demographics and characteristics is a starting point for using marketing tools such as market surveys, advertising, and enlistment incentives to target accessions for “problem” MOSs. For example, the 91S (Preventive Medicine Specialist) accessions used only 130 of 276 AIT school quotas from 1999 through 2002. Its enlistees are 54% female, and tend to have higher education levels and AFQT scores. 91S also had a higher proportion of accessions than average in the single market segments but still, as a whole, has not come close to filling the 91S AIT quotas allotted to the USAR. The USAR enlistment incentive for 91S has consistently been the $5,000 Enlistment Bonus (EB) and the $10,000 or 20,000 Student Loan Repayment Program (SLRP), a generous incentives package. What would accessions look like without the incentive package, or possibly with a different incentive package? Understanding these types of effects can assist decision-makers in making policies that positively affect USAR NPS accessions.

Linking IET training with recruiting is important because IET is a fundamental part of the recruiting process. The fact is that monetary enlistment incentives have been, and continue to be, related to the MOS an applicant chooses. If we are to ever get to a point where we analyze the impact of various enlistment incentives with the purpose of assigning them more effectively, we must understand the relationships between incentives, enlistments, and IET training seat usage. The range of training options and training availability need to be accounted for in the analysis of USAR recruiting.

I recommend further development of the data to provide an analysis of all the high density, high usage, and low usage MOSs. Additional data from REQUEST should be added to the analysis, including the recruiting incentives received by the enlistee, the opportunity display (or number of positions looked at before choosing their position or MOS), and the unit of assignment.

I believe that including ATRRS and TAPDB-R data by SSN into this process would further improve data clarity. The analysis could then be expanded to consider the effects of geographic region, demographic effects, and force structure (USAR unit locations and composition of entry level positions) on manpower and recruiting issues.

With regard to the demographic data, I recommend using the zip code aggregate data from USAREC that lists each five-digit zip code, the recruitable population, and the proportion of market segments for the zip code to qualify the blank market segments for the accession data. If we can replace the “black-hole” of unknown market segments

with valid data or a reasonable estimated distribution, then the demographic data can be better used in accessions analysis.

Effective management of both the demographically-based recruitment process and the seasonally-based IET management process is necessary in order to provide the right soldier for the right job at the right time. Until such time as the interrelated processes are more closely lashed together, we will not fully realize efficiencies in the recruitment and training environment.

APPENDICES: LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX 1. MOS DESCRIPTIONS

Below are the three digit codes corresponding to all the Military Occupational Specialties and associated job titles that were in the USAR personnel inventory during the period 1999 through 2002.

MOS	JOB TITLE
00B	DIVER
00D	SPECIAL DUTY ASSIGNMENT
00G	AADEP LOSS
01H	NOW (ASI P9) BIOLOGICAL SPECIALIST
02B	CORNET OR TRUMPET PLAYER
02C	EUPHONIUM PLAYER
02D	FRENCH HORN PLAYER
02E	TROMBONE PLAYER
02F	TUBA PLAYER
02G	FLUTE OR PICCOLO PLAYER
02H	OBOE PLAYER
02J	CLARINET PLAYER
02K	BASSOON PLAYER
02L	SAXOPHONE PLAYER
02M	PERCUSSION PLAYER
02N	KEYBOARD PLAYER
02S	SP BANDSPERSON
02T	GUITAR PLAYER
02U	ELECTRIC BASS GUITAR PLAYER
09B	TRAINEE
09C	TRAINEE (ESL)
09R	SIMULTANEOUS MEMBERSHIP P
09S	COMMISSIONED OFFICER CANDIDATE
09T	RESERVE FORCES RPT CODE
09W	WARRANT OFFICER CANDIDATE
11B	INFANTRYMAN
11C	INDIRECT FIRE INFANTRYMAN
11H	HEAVY ANTIARMOR WEAPON INFANTRYMAN
11M	FIGHTING VEHICLE INFANTRYMAN
11X	INFANTRY RECRUIT
12B	COMBAT ENGINEER
12C	BRIDGE CREWMEMBER
12F	ENGINEER TRACKED VEHICLE
13B	CANNON CREWMEMBER
13C	TACTICAL AUTOMATED FIRE CONTROL SYSTEM SPECIALIST

MOS	JOB TITLE
13D	FIELD ARTILLERY AUTO TACT DATA SYS SPECIALIST
13E	CANNON FIRE DIRECTN SPECIALIST (E7 IN RC ONLY)
13F	FIRE SUPPORT SPECIALIST
13M	MULTIPLE LAUNCH ROCKET SYSTEM(MLRS) CREWMEMBER
13P	MLRS/AUTOMATED DATA SYSTEMS SPECIALIST
13R	FIELD ARTILLERY FIREFINDER RADAR OPERATOR
14D	HAWK MISSILE SYSTEMS CREWMEMBER
14E	PATRIOT FIRE CONTROL ENHANCED OPERATOR
14J	AIR DEFENSE TACTICAL OPERATIONS CENTER OPERATOR
14L	AN/TSQ-73 CCS OP/MNT
14M	MAN PORTABLE AIR DEFENSE SYSTEM CREWMEMBER
14R	BRADLEY LINEBACKER CREWMEMBER
14S	AVENGER CREWMEMBER
14T	PATRIOT LAUNCHING SYSTEM ENHANCED OPER/MNT
16P	CHAPARRAL CREWMEMBER
16R	VULCAN CREWMEMBER
16S	FY 96 (RC ONLY) MAN PORTABLE
16T	NOW (14T1)
18B	SPECIAL FORCES WEAPONS SERGEANT
18C	SPECIAL OPERATIONS ENGINEER
18D	SPECIAL OPERATIONS MEDICAL SERGEANT
18E	SPECIAL FORCES COMMUNICATIONS SERGEANT
18X	SPECIAL FORCES RECRUIT
19D	CAVALRY SCOUT
19E	M48-M60 ARMOR CREWMAN
19K	M1 ARMOR CREWMAN
23R	HAWK MISSILE SYSTEM MECHANIC
24H	HAWK FIRE CONTROL REPAIRER
24K	HAWK CONTINUOUS WAVE RADAR
24M	VULCAN SYSTEM MECHANIC
24N	CHAPARRAL SYSTEM MECHANIC
24T	FY97 CHG TO (14E1)
25L	AN/TSQ-73 ADA COMMAND & CONTROL
25M	MULTIMEDIA ILLUSTRATOR
25R	VISUAL INFORMATION EQUIPMNT OPERATOR/MAINTAINER
25V	COMBAT DOCUMENTATION/PRODUCTION SPECIALIST
27B	NOW (35B1)
27E	LC ELEC MSL SYS REPAIRER
27F	VULCAN REPAIRER
27G	CHAPARRAL/REDEYE REPAIRER
27H	HAWK FIRING SECTION REPAIRER
27J	NOW 8A HAWK FIELD MAINTENANCE
27K	HAWK FIRE CONTROL/CONTINUOUS WAVE
27M	MLRS REPAIRER
27T	AVENGER SYSTEM REPAIRER

MOS	JOB TITLE
27X	PATRIOT SYSTEM REPAIRER
29E	NOW (35E1) RADIO REPAIRER
29F	FIXED COMSEC EQUIPMENT REPAIRER
29J	NOW (35J1)
29N	NOW (35N1)
29S	NOW (35E1)
29Y	SATELLITE COMMUNICATIONS
31C	RADIO OPERATOR/MAINTAINER
31D	NOW (31R1) MOB
31F	NETWORK SWITCHING SYSTEMS OPERATOR
31L	CABLE SYSTEMS INSTALLER/MAINTAINER
31M	NOW (31R1)
31P	MICROWAVE SYSTEMS OPERATOR/MAINTAINER
31R	MULTI-CHANNEL TRANSMISSION SYSTEMS OPERATOR
31S	SATELLITE COMMUNICATIONS SYSTEMS OPER/MAINT
31U	SIGNAL SUPPORT SYSTEMS SPECIALIST
33R	EW/I AVN SYS REPAIRER
33T	EW/I TACTICAL SYSTEMS REPAIRER
33W	MILITARY INTELLIGENCE SYSTEMS MAINT/INTEGRATOR
33Y	STRATEGIC SYSTEMS REPAIRER
35B	LAND COMBAT SUPPORT SYSTEMS TEST SPECIALIST
35C	SURVEILLANCE RADAR REPAIR
35D	AIR TRAFFIC CONTROL EQUIPMENT REPAIRER
35E	RADIO/COMMUNICATIONS SECURITY REPAIRER
35F	SPECIAL ELECTRONIC DEVICES REPAIRER
35G	MEDICAL EQUIPMENT REPAIRER UL
35H	TMDE MAINTENANCE SUPPORT SPECIALIST
35J	COMPUTER/AUTOMATION SYSTEMS REPAIRER
35L	AVIONIC COMMUNICATION EQUIPMENT REPAIRER
35M	RADAR REPAIRER
35N	WIRE SYSTEMS EQUIPMENT REPAIRER
35Q	AVIONIC FLIGHT SYSTEMS REPAIRER
35R	AVIONIC RADAR REPAIRER
35Y	INTEGRATED FAMILY OF TEST EQUIPMENT OPER/MAINT
36L	NOW (31F1)
36M	SWITCHING SYSTEMS OPERATOR
37F	PSYCHOLOGICAL OPERATIONS
38A	CIVIL AFFAIRS SPECIALIST
39B	ATE OPERATOR/MAINTAINER
39C	TGT ACQ/SVL RDR REPAIRER
39D	DECENTRALIZED AUTOMATED SPECIALIST
39E	NOW (35F1)
39G	NOW (74G1)
42C	ORTHOTIC SPECIALIST
42D	NOW (ASI N5) DENTAL LABOR
42E	OPTICAL LAB SPECIALIST

MOS	JOB TITLE
43E	NOW (92R1)
43M	FABRIC REP SPECIALIST
44B	METAL WORKER
44E	MACHINIST
45B	SMALL ARMS/ARTILLERY REPAIRER
45D	SP FA TURRET MECHANIC
45E	M1 ABRAMS TANK TURRET MECHANIC
45G	FIRE CONTROL REPAIRER
45K	ARMAMENT REPAIRER
45N	M60A1/A3 TANK TURRET MECHANIC
45T	BRADLEY FVS TURRET MECHANIC
46Q	JOURNALIST
46R	BROADCAST JOURNALIST
51B	CARPENTRY/MASONRY SPECIALIST
51K	PLUMBER
51M	FIREFIGHTER
51R	INTERIOR ELECTRICIAN
51T	TECHNICAL ENGINEERING SPECIALIST
52C	UTILITIES EQUIPMENT REPAIRER
52D	POWER GENERATION EQUIPMENT REPAIRER
52E	PRIME POWER PROD SPECIALIST
52F	TURBINE ENGINE DRIVEN GENERATOR REPAIRER
52G	TRANSMISSION AND DISTRIBUTION SPECIALIST
54B	CHEMICAL OPERATIONS SPECIALIST
55B	AMMO SPECIALIST
55D	EOD SPECIALIST
56M	CHAPLAIN ASSISTANT
57E	LAUNDRY/BATH SPECIALIST
57F	NOW (92M1) MORTUARY AFFAIRS
62B	CONSTRUCTION EQUIPMENT REPAIRER
62E	HEAVY CONSTRUCTION EQUIPMENT OPERATOR
62F	CRANE OPERATOR
62G	QUARRYING SPECIALIST
62H	CONCRETE/ASPHALT EQUIPMENT OPERATOR
62J	GENERAL CONSTRUCTION EQUIPMENT OPERATOR
63A	M1 ABRAMS TANK SYSTEM MAINTAINER
63B	LIGHT-WHEEL VEHICLE MECHANIC
63D	ARTILLERY MECHANIC
63E	M1 TANK SYSTEMS MECHANIC
63G	FUEL AND ELEC SYS REPAIRER
63H	TRACK VEHICLE REPAIRER
63J	QUARTERMASTER AND CHEMICAL EQUIPMENT REPAIRER
63M	BRADLEY FIGHTING VEHICLE SYSTEM MAINTAINER
63N	M60A1/AE TANK SYSTEMS MECHANIC
63S	HEAVY WHEEL VEHICLE MECHANIC
63T	BRADLEY FIGHTING VEHICLE SYSTEMS MECHANIC

MOS	JOB TITLE
63W	WHEEL VEHICLE REPAIRER
63Y	TRACK VEHICLE MECHANIC
67G	UTILITY AIRPLANE REPAIRER
67N	UH-1 HEL REPAIRER
67R	AH-64 ATTACK HELICOPTER REPAIRER
67S	OH-58D HELICOPTER REPAIRER
67T	UH-60 HELICOPTER REPAIRER
67U	CH-47 HELICOPTER REPAIRER
67V	OBSN/SCOUT HELICOPTER REPAIRER
67X	HEAVY LIFT HELICOPTER REPAIRER
67Y	AH-1 ATTACK HELICOPTER REPAIRER
68B	AIRCRAFT POWERPLANT REPAIRER
68D	AIRCRAFT POWERTRAIN REPAIRER
68F	AIRCRAFT ELECTRICIAN
68G	AIRCRAFT STRUCTURAL REPAIRER
68H	AIRCRAFT PNEUDRAULICS REPAIRER
68J	AIRCRAFT ARMAMENT/MISSILE SYSTEMS REPAIRER
68L	FY 96 CHG TO (35L1)
68N	AVIONIC MECHANIC
68Q	FY 96 CHG TO (35Q1)
68R	FY 96 CHG TO (35R1)
68S	OH-58D ARMAMENT/ELECTRICAL SYSTEMS REPAIRER
68X	AH-64 ARMAMENT/ELECTRICAL SYSTEMS REPAIRER
68Y	AH-64D ARMAMENT/ELECTRICAL SYSTEMS REPAIRER
71C	EXECUTIVE ADMINISTRATIVE
71D	LEGAL SPECIALIST
71G	PATIENT ADMIN SPECIALIST
71L	ADMINISTRATIVE SPECIALIST
71M	CHAPLAIN ASSISTANT
73C	FINANCE SPECIALIST
73D	ACCOUNTING SPECIALIST
74B	INFORMATION SYSTEMS OPERATOR/ANALYST
74C	TELECOMMMUNICATIONS OPERATOR/MAINTAINER
74G	TELECOMMUNICATIONS COMPUTER
75B	PERSONNEL ADMINISTRATION SPECIALIST
75E	PERSONNEL ACTIONS SPECIALIST
75F	PERSONNEL INFORMATION SYSTEM MGMT SPECIALIST
75H	PERSONNEL SERVICES SPECIALIST
76J	MEDICAL SUPPLY SPECIALIST
77F	PETROLEUM SUPPLY SPECIALIST
77L	PETROLEUM LABORATORY SPECIALIST
77W	WATER TREATMENT SPECIALIST
79R	RECRUITER NONCOMMISSIONED OFFICER
81C	CARTOGRAPHER
81L	LITHOGRAPHER
81Q	TERRAIN ANALYST

MOS	JOB TITLE
81T	TOPOGRAPHIC ANALYST
82C	FIELD ARTILLERY SURVEYOR
82D	TOPOGRAPHIC SURVEYOR
88H	CARGO SPECIALIST
88K	WATERCRAFT OPERATOR
88L	WATERCRAFT ENGINEER
88M	MOTOR TRANSPORT OPERATOR
88N	MOTOR TRANSPORTATION COORDINATOR
88P	RAILWAY EQUIPMENT REPAIRER (RC)
88T	RAILWAY SECTION REPAIRER (RC)
88U	RAILWAY OPERATIONS CREWMEMBER
88V	TRAIN CREWMEMBER (USAR ONLY)
91A	MEDICAL EQUIPMENT REPAIRER
91B	MEDICAL SPECIALIST
91C	PRACTICAL NURSE
91D	OPERATING ROOM SPECIALIST
91E	DENTAL SPECIALIST
91F	PSYCHIATRIC SPECIALIST
91G	PATIENT ADMINISTRATION SPECIALIST
91H	OPTICAL LABORATORY SPECIALIST
91J	MEDICAL LOGISTICS SPECIALIST
91K	MEDICAL LABORATORY SPECIALIST
91M	HOSPITAL FOOD SERVICE SPECIALIST
91P	RADIOLOGY SPECIALIST
91Q	PHARMACY SPECIALIST
91R	VETERINARY FOOD INSPECTION SPECIALIST
91S	PREVENTIVE MEDICINE SPECIALIST
91T	ANIMAL CARE SPECIALIST
91V	RESPIRATORY SPECIALIST
91W	HEALTH CARE SPECIALIST
91X	MENTAL HEALTH SPECIALIST
92A	AUTOMATED LOGISTICAL SPECIALIST
92G	FOOD SERVICE OPERATIONS
92M	MORTUARY AFFAIRS SPECIALIST
92R	PARACHUTE RIGGER
92S	LAUNDRY & BATH SPECIALIST
92Y	UNIT SUPPLY SPECIALIST
93C	AIR TRAFFIC CONTROL (ATC)
93F	FA MET CREWMEMBER
93P	AVIATION OPERATIONS SPECIALIST
95B	MILITARY POLICE
95C	INTERNMENT/RESETTLEMENT SPECIALIST
96B	INTELLIGENCE ANALYST
96D	IMAGERY ANALYST
96H	COMMON GROUND STATION OPERATOR
96R	GROUND SURVEILLANCE SYSTEMS OPERATOR

MOS	JOB TITLE
96U	UNMANNED AERIAL VEHICLE OPERATOR
97B	COUNTERINTELLIGENCE AGENT
97E	HUMAN INTELLIGENCE COLLECTOR
97G	MDCI ANALYST
97L	TRANSLATOR/INTERPRETER
98C	SIGNALS INTELLIGENCE ANALYST
98D	EMITTER LOCATOR/IDENTIFIER
98G	CRYPTOLOGIC LINGUIST
98H	COMMUNICATIONS LOCATOR/INTERCEPTOR
98J	ELECTRONIC INTELLIGENCE INTERCEPTER/ANALYST
98K	SIGNAL COLLECTION/IDENTIFICATION ANALYST
98X	EW/SIGINT RECRUIT

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX 2. REQUEST DATA PREPARATION PROCESS

This appendix will list the required inputs, the four data streams that constitute the process, and the process outputs. Each stream is diagrammed and the nodes numbered. The function of each node will be annotated in numbered entries corresponding to each node in the diagram.

The first step is acquisition of REQUEST data, with the following minimum necessary data fields (see appendix 3 for field definitions).

- IND_SSN
- VAC_CTRL_N
- BT_START_D
- TNG_PATH_S
- ALT_TNG_PH
- IND_SHIP_V
- MOS_OR_AOC
- ASG_UIC
- AFQT_PCTL_

The data format I used for these queries is the DBF 4 (dBase IV)(*.dbf). Accommodations can be made to the input nodes if a different format is used for the queries.

Secondly, data locations for the streams and the data should be created ahead of time, for ease of management. For the purposes of outlining the process, the data structure will be used as shown in Figure 2-1. The queries are placed in one directory, the streams in another

directory, and the output files from the process in a third directory.

Figure 2-1 Project Directory Structure. This diagram represents the directory structure for the project.

The third item to keep in mind is the consistency of data between streams and even nodes within a stream. This is controlled through the Data and Type tabs in the input nodes, and the Settings tab of the type node. The input node data types should reflect the types shown in Figure 2-2, and the type nodes should reflect the data types in figure 2-3. Notice they are the same for the common fields, as this is the purpose of the setting the types. The data storage, which is denoted by the symbol on the far left, needs to be set in the input node. It is critical to insure SSN fields have the box “A” representing a string storage, or the leading zero will get omitted and can potentially create additional duplicate records. The Data tab of the input node is where you change these settings, and is shown in Figure 2-4.

Default settings for the sort nodes, which are present in various places throughout the process, is sort ascending by SSN, MOS, ALT_TNG_PH, EnlistmentDate, BCTDate, AITDate, and ShipDate.

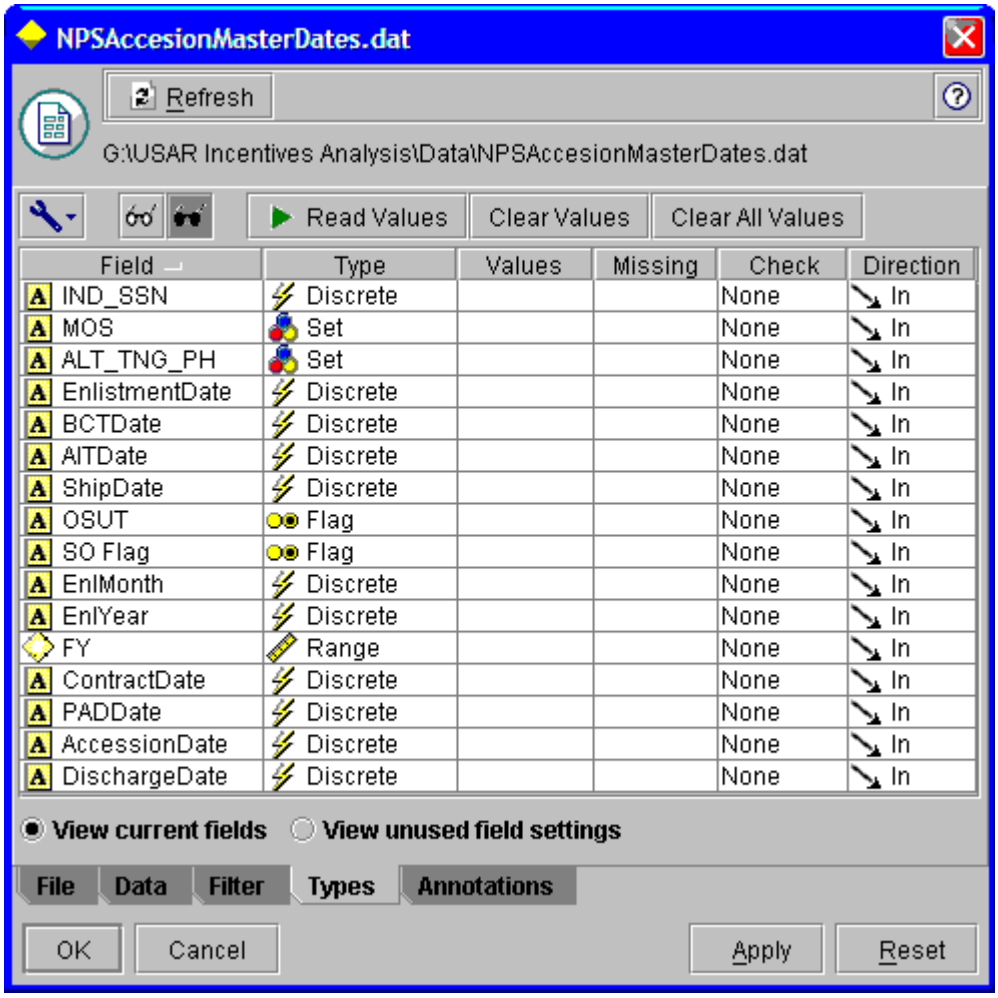


Figure 2-2 Input Data Node Type. The shows the Types tab for an input node. The far left symbol denotes the storage type, a box "A" representing a String and the diamond representing an Integer. The Types with the names are shown in the next column.

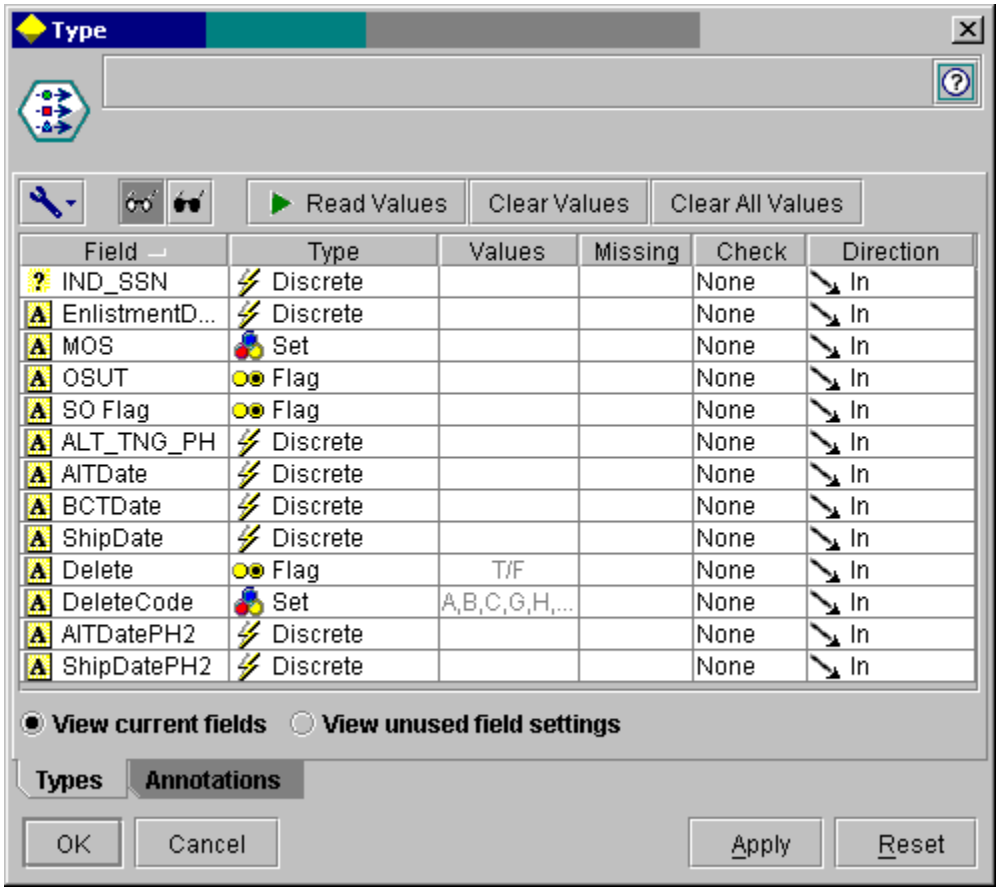


Figure 2-3 Example Type Node Settings. This shows the standard settings for the data types. These are the settings used in a majority of the type nodes throughout the process.

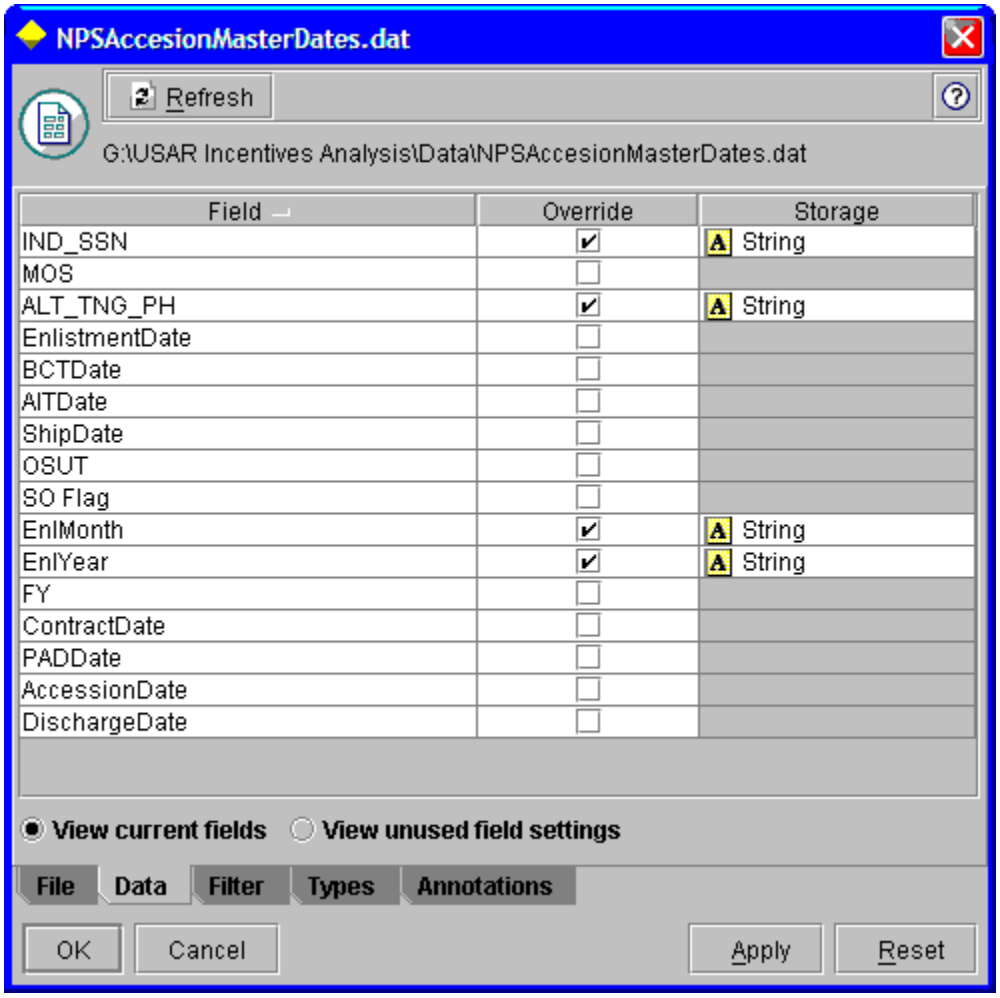


Figure 2-4 Input Node Data Tab. This shows the Data tab for an input node. The Override column is checked in the cases where the default storage value is other than what is desired. In this case, since the IND_SSN field consists of integer numbers, Clementine defaults to Integer storage. The override box for that item is checked and the Storage is set to String.

A. REQUEST DATA MERGE

This stream merges separate queries of data from REQUEST into a single output file called the NPSAcc.txt, and prepares data subsets listing the duplicate records, records without a ship date, and records with identical SSNs and different enlistment dates.

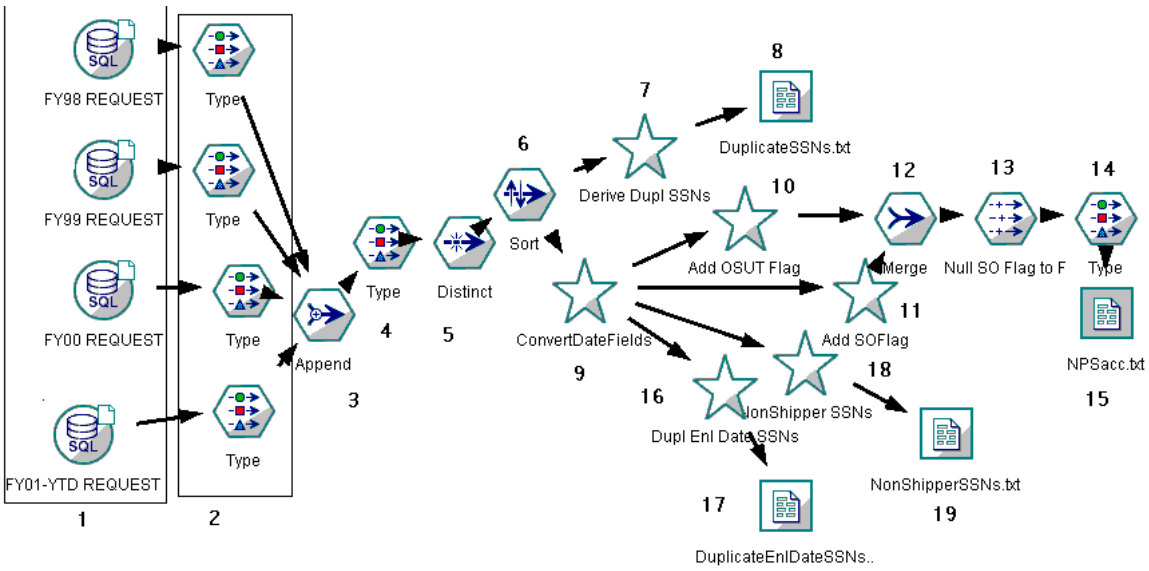


Figure 2-5 REQUEST Data Merge Stream. The stream merges the four years of REQUEST data, converts the date fields, adds the flags for split-option and OSUT accessions, generates the duplicate tables, and creates the accessions table called NPSacc.txt on the right.

Node(s) 1: The input nodes link to the REQUEST queries in DBF format. Ensure that text fields consisting of numeric elements such as vacancy control number, SSN, zip code, and training phase code have the data defaults set to string storage, as they default to integer storage. This is done in the data tab in the node.

Node(s) 2: Set types as shown in Figure 2-3.

Node 3: Append on keys for all the data fields.

Node 4: Types as shown in Figure 2-3.

Node 5: Selects the distinct records on all the input data fields to screen out full duplicates.

Node 6: Sorts data in this order: IND_SSN, MOS_AOC, ENL_VER.

Node 7: **Derive Dupl SSNs** supernode (Figure 2-6).

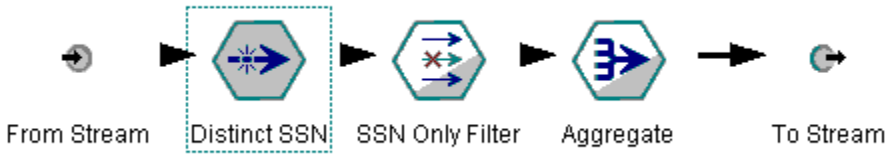


Figure 2-6 Derive Dupl SSNs Supernode. This distinct node discards the first distinct record for a SSN. The subsequent nodes filter all but SSN, and then aggregate to SSN. The result is a list of unique SSNs with a record count of the number of duplicate records.

Node 8: Outputs a list of SSNs with multiple records to a text file in the PrepOutput directory. This is used in later streams to identify SSNs with duplicate records.

Node 9: **ConvertDateFields** supernode (Figure 2-7).

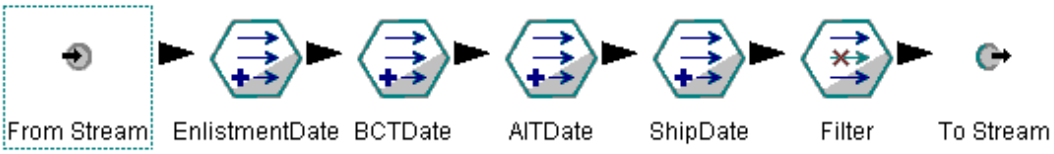


Figure 2-7 ConvertDateFields Supernode. This node adds fields named EnlistmentDate, BCTDate, AITDate, and ShipDate from corresponding fields in the REQUEST date (See Appendix 3 for definitions). Each uses the command to_date() to perform the conversion. For example, the EnlistmentDate is set equal to to_date(ENLST_VER_). Once the conversions are accomplished, the filter node eliminates the unconverted date fields used in the four derive nodes.

Node 10: **Add OSUT Flag** supernode (Figure 2-8).

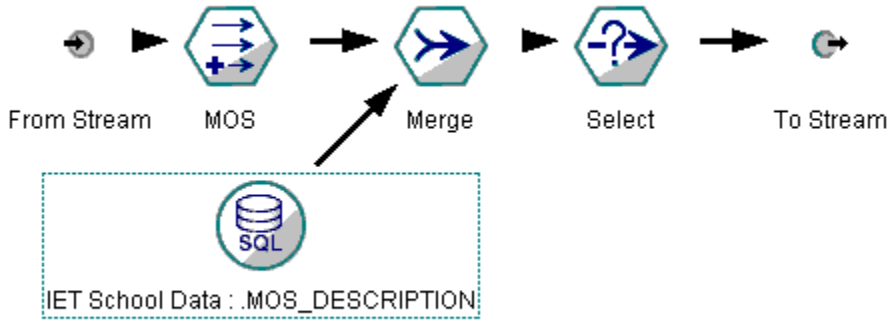


Figure 2-8 "OSUT Flag" Supernode. This supernode first converts the four-digit MOS_OR_AOC to a three-digit MOS field, setting MOS equal to substring(1,3, MOS_OR_AOC). The data input is a listing of MOSs with an "O" or "N" denoting OSUT or non-OSUT, respectively. The file is located in the root directory as a dbf file called MOS_DESCRIPTON.dbf, a file derived from MOS listings in DA PAM 611–21.

Node 11: **Add SO Flag** supernode (Figure 2-9).

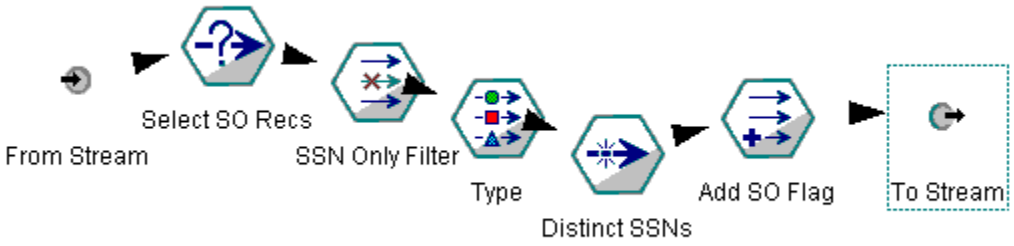


Figure 2-9 "Add SO Flag" Supernode. This supernode selects records with to_integer(ALT_TNG_PH) = 1 or to_integer(ALT_TNG_PH) = 2. It then filters to retain only the SSN. The "Distinct SSNs" node reduces the data to the distinct SSNs, and then the SO Flag is added to the record with the value set to "T" (true) indicating it is a split-option record.

Node 12: Merges records on all fields except the SO Flag and the OSUT Flag. Result is that all records have both flags.

Node 13: For any SO Flag fields that are undefined or null, the value is set equal to “F” (false) indicating the record is not a split-option record.

Node 14: Type node with same setting as Figure 2-3. Values should read O (OSUT) and N (not OSUT) for the OSUT Flag, and T (split-option) and F (not split-option) for the SO Flag.

Node 15: Outputs all records without full duplicates with the date fields now stored as dates, and flags included for the split-option (SO Flag) and OSUT record identification. Outputs are sent to a flat file called NPSacc.txt in the output directory.

Node 16: Duplicate Enlistment Date supernode (Figure 2-10).

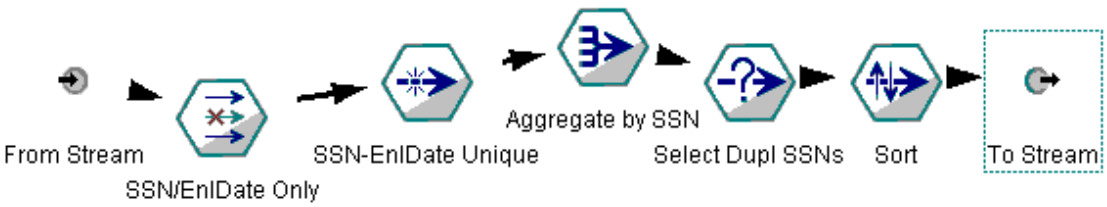


Figure 2-10 “Dupl Enl Dates SSNs” Supernode. This supernode filters the records to SSN and Enlistment date only, then the distinct node reduces the records to the unique vales for SSN and Enlistment Date combination. The aggregation by SSN with a record count provides the number of enlistment dates by SSN. Then only SSNs with a record count greater than one are selected, sorted by SSN and passed back to the stream.

Node 17: Outputs a list of SSNs with multiple enlistment dates, and the corresponding numbers of different enlistment dates, to a flat file called DuplicateEnlDateSSNs.txt in the output directory.

Node 18: **Non-Shipper SSNs** supernode (Figure 2-11).

Figure 2-11 "Non-Shipper SSNs" Supernode. The data are filtered to SSN and Ship Date only, distinct combinations are passed on, and then records with Ship Date null or undefined are selected, sorted by SSN, and passed to the stream.

Node 19: Outputs a list of SSNs with null or undefined ship date fields to a flat file called NonShipperSSNs.txt in the output directory.

B. QUALIFY REQUEST DUPLICATES

This stream examines the data for null and blank fields for BCT and AIT start dates, and examines all the duplicate records for field value inconsistencies. Any record that meets one of the alphabetical delete code criteria (see Section III.B.2) is flagged for deletion and assigned a deletion code. The result is an output file called NPSdeletions.txt containing all the records marked for deletion.

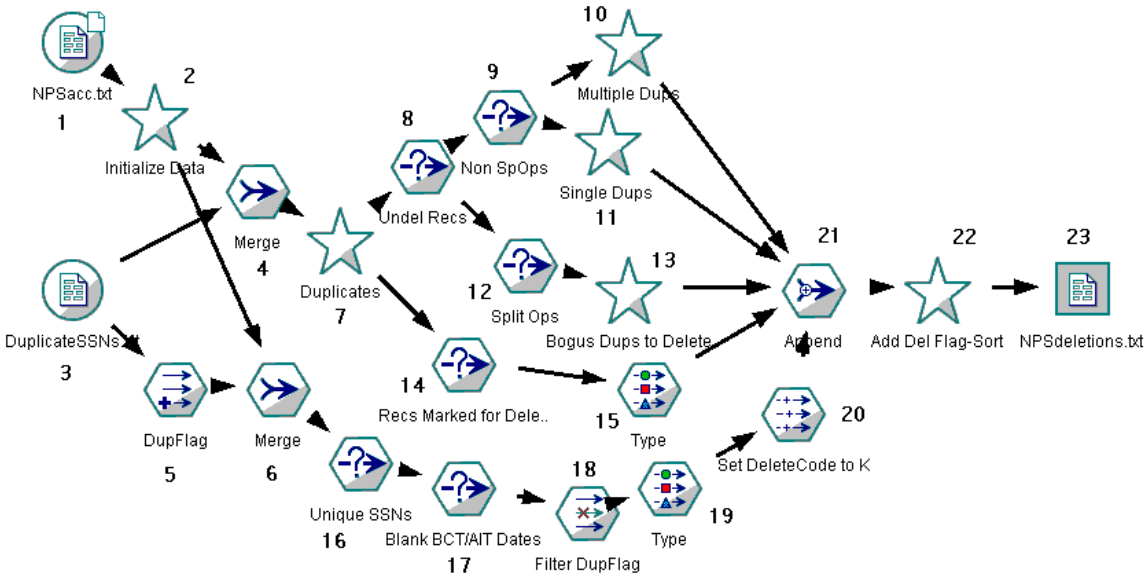


Figure 2-12 Qualify REQUEST Duplicates Stream. This stream takes the merged REQUEST file and duplicates file, and qualifies the records based on the lettered criteria through a series of node operations. The records are flagged for deletion and output to a deletion file that catalogues all records marked for deletion.

Node 1: The input file is the NPSacc.txt file in the output directory. Once again, the analyst should ensure that text fields consisting of numeric elements such as vacancy control number, SSN, zip code, and training phase code have the data defaults set to string storage, as they may default to integer storage.

Node 2: **Initialize Data** supernode (Figure 2-13).

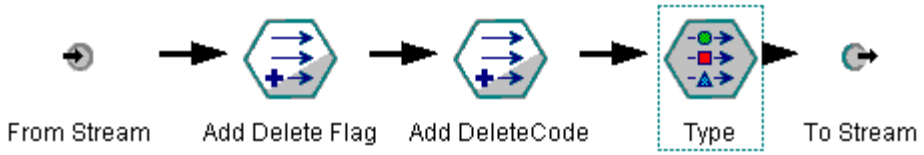


Figure 2-13 "Initialize Data" Supernode. This supernode adds fields for Delete with a default value of "F" (false), and DeleteCode with a default value of "" (null). The type ensures that these new fields are of type Flag, in addition to reflecting the Types as shown in Figure 2-3.

Node 3: Ensures that the SSN field is set to string storage in the data tab.

Node 4: Merges files on SSN key, using an inclusive-join. This means that only the records in both files with data are merged and passed out of the stream. In this case, only the records with duplicate SSNs are passed out of this node.

Node 5: Derives a flag field called **DupFlag** that is set to "T" (true), since the only records entering this node are the duplicate SSNs.

Node 6: Merges input from NPSacc.txt file with the duplicates on SSN key, using an outer-join. This means that all records are merged. All records in one set that do not have a field are automatically given one with an undefined value. The result in this case is the NPSacc records now have a DupFlag field. Records that have duplicates have this field set to "T" and the unique SSNS value undefined.

Node 7: **Duplicates** supernode (Figure 2-14).

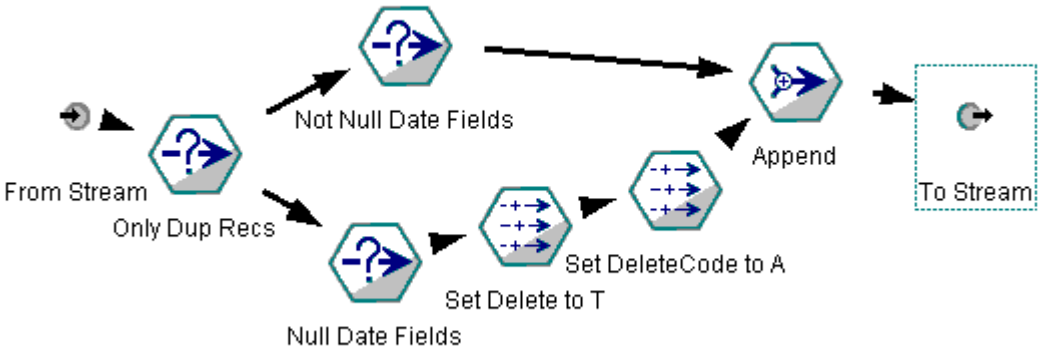


Figure 2-14 "Duplicates" Supernode. The select node selects records with the DupFlag equal to "T" (true). The bottom select node ("Null Date Fields") selects records with both AIT and BCT dates blank or undefined [(BCTDate="" or BCTDate=undef) and (AITDate = "" or AITDate = undef)], marks them for deletion, and codes the records with deletion code "A." The top select node ("Not Null Date Fields") discards records meeting the same criteria. The append node adds the two together, passing back to the stream all the records that were passed in.

Node 8: Select node that selects records with the Delete flag field set to "F."

Node 9: Select node selects records with SO Flag equal to "F" (straight-through records).

Node 10: **Multiple Dups** supernode (Figure 2-15).

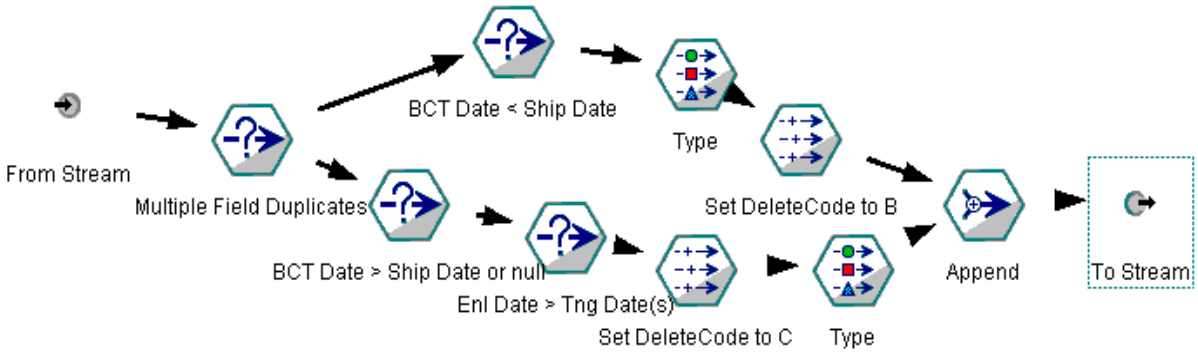


Figure 2-15 "Multiple Dups" Supernode. The select node at the left selects records with a duplicate record count greater than 1, which means there are 3 or more records for the SSN. The upper select node selects records that have a BCT Date prior to the Ship Date, and then these records are set to delete code "B." The lower select node selects records that have a BCT date after the Ship date or a null or blank ShipDate field. The second select node selects records with enlistment dates later than training dates, coded as [(EnlistmentDate>BCTDate and BCTDate /= "") or (EnlistmentDate>AITDate and AITDate /= "")]. Records selected are coded to delete code "C." The "Append" node groups together all the records marked for deletion, and passes them back to the stream.

Node 11: **Single Dups** supernode (Figure 2-16).

Figure 2-16 "Single Dups" Supernode. This supernode starts by selecting all records with a single duplicate (record_count equal to 1). The second select node selects records with blank or null BCT and AIT start dates [(AITDate = "" or AITDate=undef) and (BCTDate = "" or BCTDate=undef)]. The records are reduced to a distinct set unique on all input fields, set to delete code "D" and passed back to the main stream.

Node 12: Select node that selects records with a SO Flag equal to "T" (split-option records).

Node 13: **Bogus Dups to Delete** supernode (Figure 2-17). Two supernodes nested inside this supernode are **Mark Bogus Non-Osut** (Figure 2-18) and **Mark Bogus OSUT** (Figure 2-19)

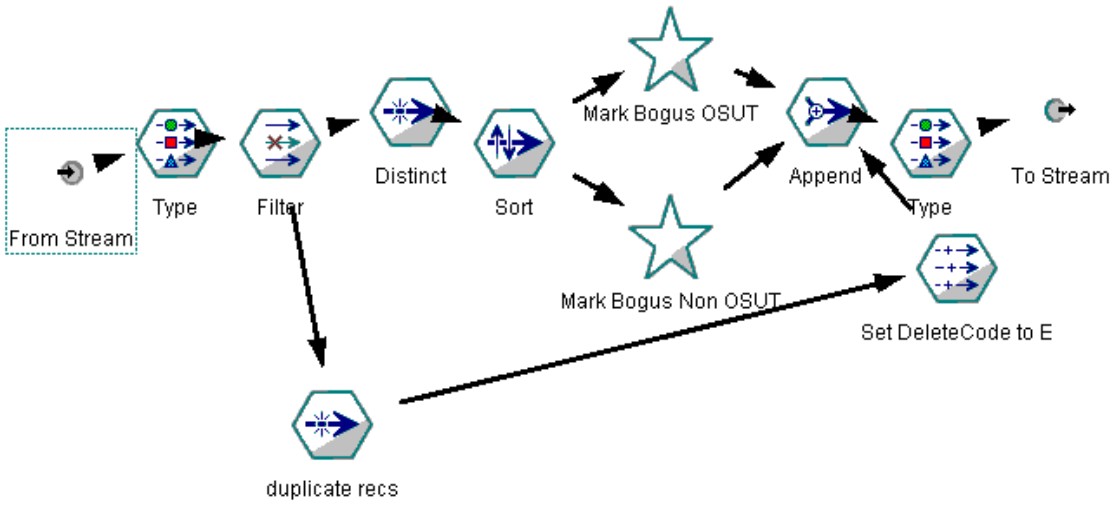


Figure 2-17 “Bogus Dups to Delete” Supernode. The type is set as in Figure 2-3. The distinct node is distinct by all fields input from the NPSacc.txt file. They are then sorted ascending by SSN, MOS, ALT_TNG_PH, EnlistmentDate, BCTDate, AITDate, and ShipDate. The Mark Bogus OSUT and mark Bogus Non-OSUT are discussed in Figures 2-18 and 2-19. The results from these nodes are appended together and sent back to the main stream.

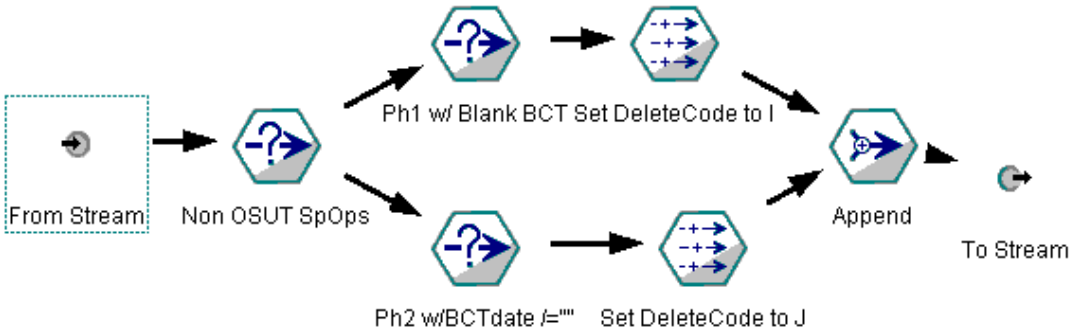


Figure 2-18 “Mark Bogus Non-OSUT” Supernode. This supernode selects SO Flag equal “T” and OSUT equal “N” records. The upper path selects records with ALT_TNG_PH = "1" and BCTDate is null or blank, and sets the DeleteCode to “I”. The lower path selects records with ALT_TNG_PH = "2" and BCTDate not equal to null or blank, and sets the DeleteCode to “J”. This supernode has marked the extraneous split-option duplicates for deletion except for one phase 1 record and 1 phase 2 record.

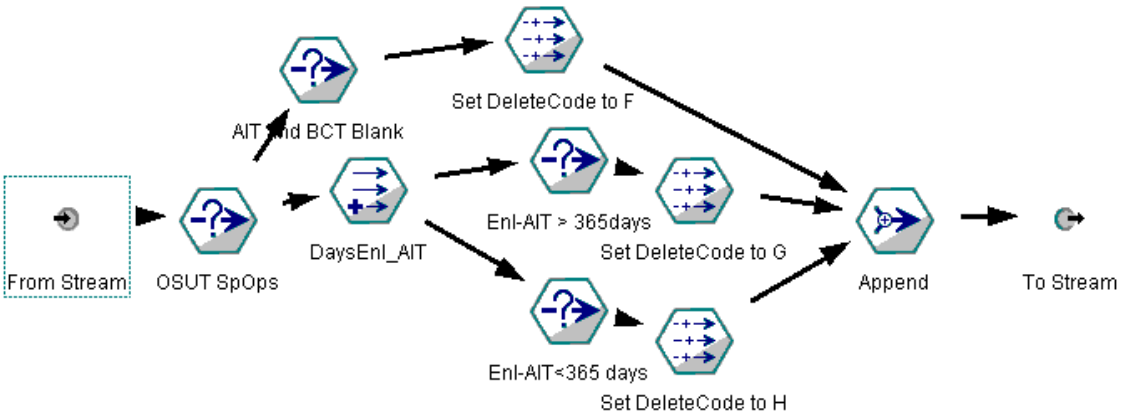


Figure 2-19 “Mark Bogus OSUT” Supernode. This supernode selects SO Flag equal “T” and OSUT equal “O” records. The upper path selects records with AIT and BCT dates that are null or blank [(BCTDate = "" or BCTDate=undef) and (AITDate="" and AITDate=undef)], and sets the DeleteCode to “F”. The lower path derives a fields for days between EnlistmentDate and AITDate (date_days_difference(EnlistmentDate, AITDate)), then evaluates the date in two select nodes. The first one selects records with ALT_TNG_PH = "1" and DaysEnl_AIT > 335, and sets the DeleteCode to “G”. The second selects [(ALT_TNG_PH = "2" or ALT_TNG_PH="" or ALT_TNG_PH=undef) and DaysEnl_AIT < 365] and sets the DeleteCode to “H”. This supernode has marked the extraneous split-option duplicates for deletion except for one phase 1 records and one phase 2 record. They are appended and passed to the Qualify Bogus Split-Option Records supernode.

Node 14: Select node that selects records with Delete flag field set to “T” (records marked for deletion).

Node 15: Type node with settings as shown in Figure 2-3.

Node 16: Select node that selects records with DupsFlag set to “F” (unique SSN records).

Node 17: Select node that selects records with blank or undefined (null) BCT and AIT start date fields.

Node 18: Filter node that eliminates the DupFlag field from the records.

Node 19: Type node with same settings as Figure 2-3.

Node 20: Sets the DeleteCode field equal to "K," which represents unique SSN records that have BCT and AIT fields that are both either blank or undefined.

Node 21: Appends all the records together. This node combines all records marked for deletion that have a deletion code with a value from A to K.

Node 22: **Add DelFlag Sort** supernode (Figure 2-20).

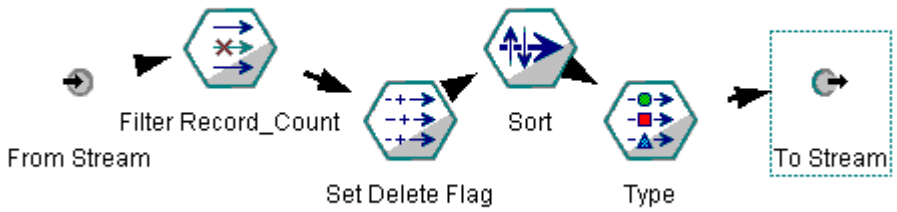


Figure 2-20 "Add DelFlag Sort" Supernode. The first node filters out the RECORD_COUNT field, and the second node ensures the delete flag is set to "T." The sort node sorts the records ascending by SSN, MOS, ALT_TNG_PH, EnlistmentDate, BCTDate, AITDate, and ShipDate. The type node ensures that the fields are as shown in Figure 2-3.

Node 23: Outputs a list of records with all original fields from the NPSacc.txt file in addition to a deletion code field and a DelFlag field. The output is a flat file called NPSdeletions1.txt in the output directory.

C. SPLIT-OPTION MERGE

This stream aggregates the split-option records for phase 1 and phase 2 into a single record with additional fields representing the phase 2 training start date and ship date.

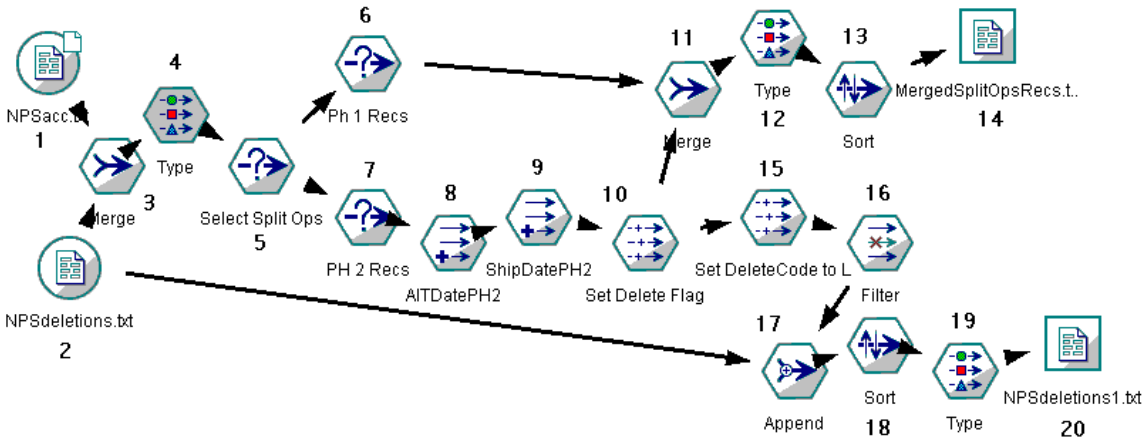


Figure 2-21 Split-Option Merge Stream. This stream takes the two remaining records for each split-option SSN, and merges the information into a single record for insertion later into the master accession file.

Node 1: The input file is the NPSacc.txt file in the output directory. Once again, the analyst should ensure that text fields consisting of numeric elements such as vacancy control number, SSN, zip code, and training phase code have the data defaults set to string storage, as they may default to integer storage.

Node 2: The input file is the NPSdeletions.txt file in the output directory. Once again, the analyst should ensure that text fields consisting of numeric elements such as vacancy control number, SSN, zip code, and training

phase code have the data defaults set to string storage, as they may default to integer storage.

Node 3: Merges on all input fields from the NPSacc.txt file in an outer-join.

Node 4: Sets the data types as shown in Figure 2-3.

Node 5: Selects only records with a SO Flag equal to “T” and Delete undefined (Undeleted split-option records only).

Node 6: Selects records with ALT_TNG_PH equal to “1.”

Node 7: Selects records with ALT_TNG_PH equal to “2.”

Node 8: Derives a new field called AITDate2 with the value of the AIT start date field. This is the phase 2 training start date.

Node 9: Derives a new field with the ShipDate2 value of the ShipDate date field. This is the phase 2 training ship date.

Node 10: Sets DelFlag equal to “T.” This marks all the phase 2 split-option records for deletion. These records are no longer needed as key dates are placed in the newly derived fields AITDate2 and ShipDate2.

Node 11: Merges phase 1 and phase 2 split-option records together on the key fields SSN, EnlistmentDate, and MOS on an inner-join. The result is a SSN unique set of records with the phase 1 and phase 2 training data now located in a single record.

Node 12: Type node with same settings as Figure 2-3.

Node 13: Sorts records ascending by fields SSN, MOS, ALT_TNG_PH, EnlistmentDate, BCTDate, AITDate, ShipDate, and Delete.

Node 14: Outputs merged split-option records to a flat file in the output directory called MergedSplitOpsRecs.txt.

Node 15: Filler node sets DeleteCode to “L,” to represent phase 2 records eliminated during the split-option merge stream.

Node 16: Filter node removes AITDate2 and ShipDate2, as these records are to be deleted and these fields are not in the NPSdeletions1.txt file.

Node 17: Appends records to the contents of NPSdeletions1.txt.

Node 18: Sorts as in node 13.

Node 19: Sets type as in node 4.

Node 20: Outputs a list of records with all original fields from the NPSdeletions.txt file, in addition to the records marked for deletion during the split-option merge process, to a flat file called NPSdeletions1.txt in the output directory.

D. REQUEST DUPLICATE RECONCILE

This stream is the last data preparation stream. It performs additional coding not done earlier, and outputs to file known duplicates that cannot be screened out using the earlier process. This output file can be used for further analysis.

This is the stream that was used to add additional coding processes as the understanding of the data issues

increased, and allowed for better “exception” handling for subsets of the duplicate records.

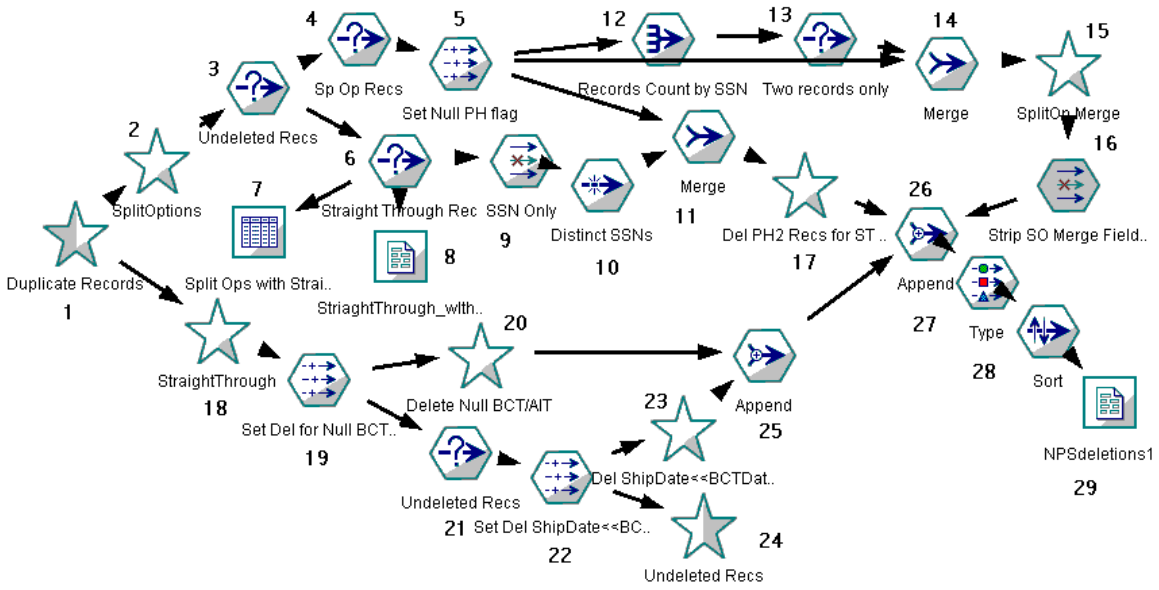


Figure 2-22 “Request Exception Reconcile” Stream. This stream performs additional screening functions to further eliminate duplicates, and also output some subsets of the duplicate population by category for manual reconciliation.

Node 1: **Duplicate Records** supernode (Figure 2-23).

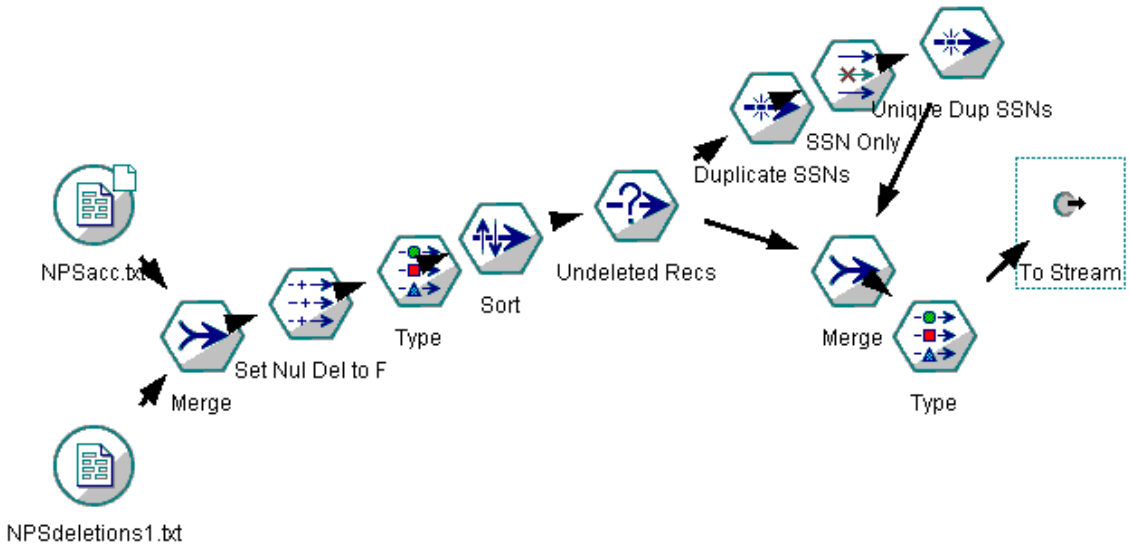


Figure 2-23 "Duplicate Records" Supernode. This supernode has input nodes exactly as in the "Split-Option Merge" stream input nodes. The merge node merges the records on all the fields in the NPSacc.txt on an outer-join. Then the filler node sets the undefined values for the Delete field to "F" (undeleted Records), sets the data type as in Figure 2-2, and sorts by SSN, MOS, ALT_TNG_PH, EnlistmentDate, BCTDate, AITDate, ShipDate and Delete. The select node selects the records with Delete equal to "F." The branch in the upper right discards the unique records by SSN, filters to SSN, and then reduces to for the records to a unique listing of SSNs with duplicate records remaining from the NPSacc.txt input. The last merge is an inner-join on the key field SSN to create a list of remaining duplicate records from the NPSacc.txt input.

Node 2: The **Split-Option** supernode (Figure 2-24)

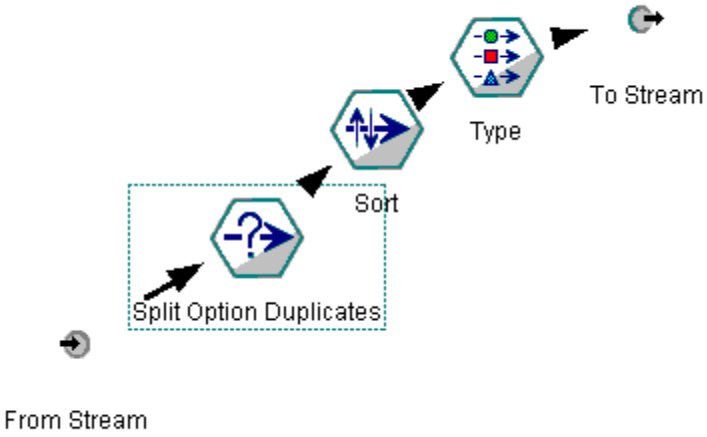


Figure 2-24 Split-Option Supernode. The records are selected for SO Flag equal "T," sorted by SSN, MOS, ALT_TNG_PH, EnlistmentDate, BCTDate, AITDate, and ShipDate, and then the type is set as in Figure 2-3.

Node 3: Selects records with Delete equal to "F."

Node 4: Selects records with SO Flag equal to "T."

Node 5: Filler node sets all undefined records to "2." This represents correcting the phase 2 split-option records with that field null or blank.

Node 6: Selects records with SO Flag equal to "F."

Node 7: Outputs records of SSNs that are associated with both split-option and straight-through records to the screen as a table.

Node 8: Same as above, except the output is sent to a flat file called StraightThrough_with_SplitOp Flagged.txt in the output directory.

Node 9: Filters all but the SSN.

Node 10: Reduces the records to a set of distinct SSNs.

Node 11: Merges on an inner-join by SSN. This produces records with straight-through and split-option records.

Node 12: Aggregates records to SSN with a record count.

Node 13: Selects only SSNs with RECORD_COUNT = 2.

Node 14: Merges on SSN using an inner-join. This creates a group of split-option records with phase 1 and phase 2 records for the same SSN with different enlistment dates. They are ready for merging into a single record.

Node 15: This supernode is essentially a duplicate of the split-option merge stream shown in section C of this appendix. The only differences are the that deletion code is “L” and there is no enlistment date in any merge in this supernode.

Node 16: This filter node strips out the AITDate2, ShipDate, and RECORD_COUNT fields.

Node 17: The **Del PH2 Recs for ST** supernode (Figure 2-25).

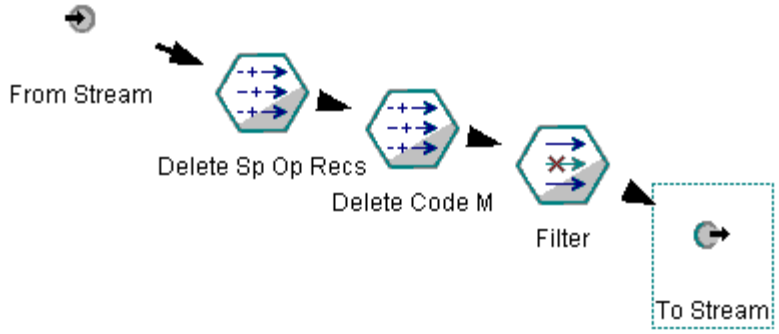


Figure 2-25 "Del PH2 Recs for ST" Supernode. It marks the split-option records for the SSNs with both split-option and straight-through records for deletion and assigns them deletes code M.

Node 18: **Straight-Through** supernode (Figure 2-26).

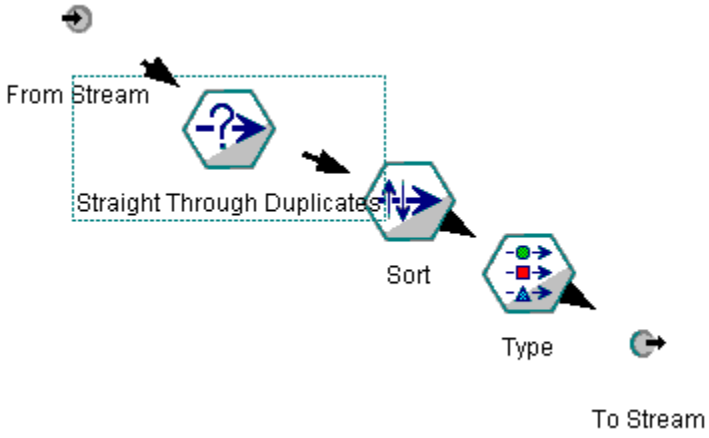


Figure 2-26 "Straight-Through" Supernode. The records are selected for SO Flag equal "F," sorted by SSN, MOS, ALT_TNG_PH, EnlistmentDate, BCTDate, AITDate, and ShipDate, and then the types are set as in Figure 2-3.

Node 19: This filler node sets Delete to "T" for records with a blank or null value for BCTDate or EnlistmentDate (BCTDAte BCTDate = "" or BCTDate=undef or EnlistmentDate= "" or EnlistmentDate = undef).

Node 20: The **Delete Null BCT/AIT** supernode contains two nodes: a select node that selects records with Delete equal to “T,” and then a filler node that sets the DeleteCode to “N.”

Node 21: Selects undeleted records (Delete equal to “F”).

Node 22: This filler node sets the Delete field to “T” for records that have a Shipdate at least five weeks earlier than the BCTDate (date_weeks_difference(ShipDate,BCTDate)>5).

Node 23: The **Del ShipDate<<BCTDate** supernode contains two nodes: a select node that selects records with Delete equal to “T,” and then a filler node that sets the DeleteCode to “O.”

Node 24: **Undeleted Recs** supernode (Figure 2-27).

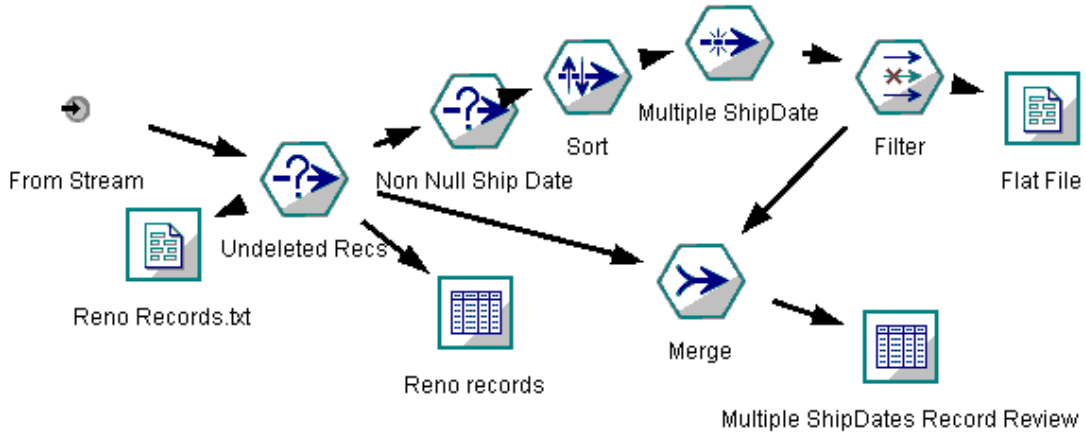


Figure 2-27 "Undeleted Recs" Supernode. This supernode outputs any remaining undeleted records with identified problems without an identified "fix" to a combination of flat file and screen outputs. The first node selects records with Delete equal to "F." These records represent the remaining duplicates that have not been deleted. The upper path selects the records that have ShipDates that are not null or blank, then sorts by SSN and ShipDate. The filter node screens out all the fields except SSN, ShipDate and DeleteCode. These records are then output to a flat file in the output directory called "Multiple Ship Date SSNs.txt." The merge node uses an inner-join by SSN. The output node displays the records on the screen in a table called "Multiple ShipDates Record Review."

Node 25: Combines the records from nodes 20 and 23.

Node 26: Appends all the records that have been marked for deletion together.

Node 27: Sets data types as in Figure 2-3, with the addition of Delete set to data type "Flag," and DeleteCode set to data type "Set."

Node 28: Sorts ascending by SSN, MOS, ALT_TNG_PH, EnlistmentDate, BCTDate, AITDate, ShipDate and Delete.

Node 29: Appends records to an existing flat file in output directory named NPSdeletions1.txt.

E. REQUEST DATA PREPARATION SUMMARY

This final stream provides summary data on the REQUEST data that was input into the process, the results of the screening of the duplicates and blank and null fields in terms of a single record summary, and a distribution chart of the deletion codes used.

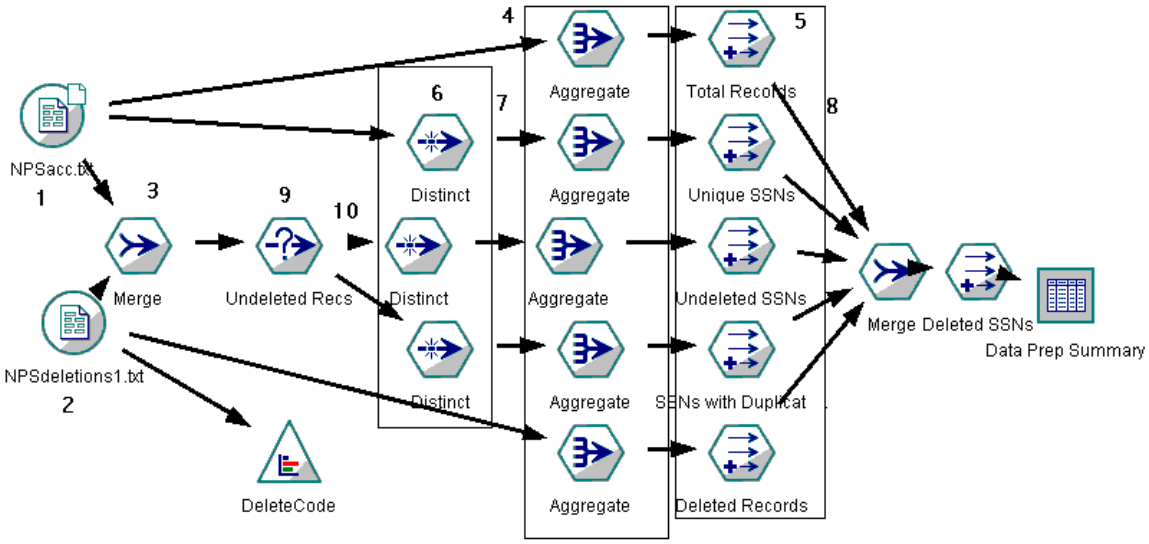


Figure 2-28 REQUEST Data Prep Summary Stream. This stream generates a single record summary of the records, the deletions, and the remaining duplicates. It generates a proportion graph of the deletion codes as well.

This stream uses the same inputs as previous streams, merges using an outer-join on all records from NPSacc.txt at node 3, and at node 9 selects the undeleted nodes (Delete equal undef). The distinct nodes at node group 6 all reduce the records to unique SSNs, and the nodes at

node group 4 aggregate records by SSN with a record count. They are then merged using an outer-join without a key, and the derive node calculates the deleted SSNs by subtracting the aggregate for unique SSNs minus the aggregate for SSNs undeleted. The result is output to the screen in a single record table.

The “Distribution graph” node, the triangular node labeled “DeleteCode” in the lower left of Figure 2-24, generates a graph of the DeleteCode by record count.

APPENDIX 3. REQUEST DATA DICTIONARY

The fields included in the data preparation and used in the resulting analysis were generated through queries of the REQUEST system. The source file names and definitions are listed below.

REQUEST Field Definitions

VAC_CTRL_N - Unique seven-digit number referencing to the vacant position in REQUEST.

BT_START_D - Date string for date scheduled to start Basic Combat Training (BCT start date).

TNG_PATH_S - Date string for date scheduled to start One Station Unit training or Advanced Individual Training (AIT start date).

ALT_TNG_PH - Single digit number, 1 or 2, representing phase of training if a split-option trainer, otherwise null.

IND_SSN - Individual SSN for accessing individual.

IND_SHIP_V - Date string for date individual shipped to Initial Entry training (Ship date).

MOS_OR_AOC - Four-digit code representing Military Occupational Specialty and grade (e.g. 95B1 for a skill level one Military Police).

ASG_UIC - Unit Identification Code for the unit with the vacant position.

AFQT_PCTL - Armed Forces Qualification Test Percentile for the accessing individual.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX 4. MARKET SEGMENTATION DATA

These market segments are provided by USAREC, which they obtained from a commercial source. Each accession in the Reserve Enhanced Applicant File should be coded with a two-digit number corresponding to their particular market segment. There are actually 48 market segments, with additional segments reflecting anomalies and unclassified. These segments are grouped into 9 different groups, with two additional groups for the anomalies and unclassified segments. The last two represent less than 0.2% of the population.

The data fields include information such as overall percent of the base population, percent veterans, percent white collar and blue collar, percent by ethnicity, median income, age range, and so on.

Also included is a summary for the population labeled as segment 0 representing the entire United States.

Field definitions were not available from USAREC, so they are not included. Given that, I will note that in the FORCEPCT field there are a couple of anomalous entries. Without better information, I cannot clarify the accuracy of these entries, or any of the others. This information is appended for supplemental reference only.

SEGMENT	SNAME	GROUP	GNAME
0	US BASE DEMOGRAPHICS		US BASE DEMOGRAPHICS
1	UPPER CRUST	01	ACCUMULATED WEALTH
2	LAP OF LUXURY	01	ACCUMULATED WEALTH
3	ESTABLISHED WEALTH	01	ACCUMULATED WEALTH
4	MID-LIFE SUCCESS	01	ACCUMULATED WEALTH
5	PROSPEROUS METRO MIX	01	ACCUMULATED WEALTH
6	GOOD FAMILY LIFE	01	ACCUMULATED WEALTH
7	COMFORTABLE TIMES	06	CONSERVATIVE CLASSICS
8	MOVERS AND SHAKERS	04	MAINSTREAM SINGLES
9	BUILDING A HOME LIFE	03	YOUNG ACCUMULATORS
10	HOME SWEET HOME	02	MAINSTREAM FAMILIES
11	FAMILY TIES	02	MAINSTREAM FAMILIES
12	A GOOD STEP FORWARD	04	MAINSTREAM SINGLES
13	SUCCESSFUL SINGLES	09	SUSTAINING SINGLES
14	MIDDLE YEARS	01	ACCUMULATED WEALTH
15	GREAT BEGINNINGS	04	MAINSTREAM SINGLES
16	COUNTRY HOME FAMILY	02	MAINSTREAM FAMILIES
17	STARS AND STRIPES	02	MAINSTREAM FAMILIES
18	WHITE PICKET FENCE	02	MAINSTREAM FAMILIES
19	YOUNG AND CAREFREE	03	YOUNG ACCUMULATORS
20	SECURE ADULTS	06	CONSERVATIVE CLASSICS
21	AMERICAN CLASSICS	06	CONSERVATIVE CLASSICS
22	TRADITIONAL TIMES	02	MAINSTREAM FAMILIES
23	SETTLED IN	02	MAINSTREAM FAMILIES
24	CITY TIES	08	SUSTAINING FAMILIES
25	BEDROCK AMERICA	03	YOUNG ACCUMULATORS

SEGMENT	SNAME	GROUP	GNAME
26	THE MATURE YEARS	07	CAUTIOUS COUPLES
27	MIDDLE OF THE ROAD	05	ASSET-BUILDING FAMILIES
28	BUILDING A FAMILY	03	YOUNG ACCUMULATORS
29	ESTABLISHING ROOTS	05	ASSET-BUILDING FAMILIES
30	DOMESTIC DUOS	06	CONSERVATIVE CLASSICS
31	COUNTRY CLASSICS	06	CONSERVATIVE CLASSICS
32	METRO SINGLES	04	MAINSTREAM SINGLES
33	LIVING OFF THE LAND	07	CAUTIOUS COUPLES
34	BOOKS AND NEW RECRUITS	04	MAINSTREAM SINGLES
35	BUY AMERICAN	02	MAINSTREAM FAMILIES
36	METRO MIX	09	SUSTAINING SINGLES
37	URBAN UP AND COMES	09	SUSTAINING SINGLES
38	RUSTIC HOMESTEADERS	02	MAINSTREAM FAMILIES
39	ON THEIR OWN	04	MAINSTREAM SINGLES
40	TRYING METRO TIMES	04	MAINSTREAM SINGLES
41	CLOSE KNIT FAMILIES	08	SUSTAINING FAMILIES
42	TRYING RURAL TIMES	08	SUSTAINING FAMILIES
43	MANUFACTURING USA	08	SUSTAINING FAMILIES
44	HARD YEARS	08	SUSTAINING FAMILIES
45	STRUGGLING METRO MIX	09	SUSTAINING SINGLES
46	DIFFICULT TIMES	08	SUSTAINING FAMILIES
47	UNIVERSITY AMERICA	09	SUSTAINING SINGLES
48	URBAN SINGLES	09	SUSTAINING SINGLES
49	ANOMALIES	10	ANOMALIES
50	UNCLASSIFIED	11	UNCLASSIFIED

SEGMENT	BASE	VETERAN	FORCEPCT	PERCAPIT	INCOME	LOCATION
0	100.0	14.33	0.89	21272	40824	
1	1.4	15.87	0.17	58704	119981	SUBURBAN
2	1.3	14.47	0.71	33698	77425	SUBURBAN
3	2.1	15.63	0.47	33557	66562	SUBURBAN
4	3.0	15.36	0.43	36893	68788	SUBURBAN
5	2.6	14.62	1.32	25718	61311	SUBURBAN
6	2.0	15.88	0.41	26286	57588	RURAL
7	0.7	17.38	0.39	29601	57282	SUBURBAN
8	2.8	14.42	0.37	38334	59792	SUBURBAN
9	0.1	16.02	0.76	26039	54189	RURAL
10	6.0	16.47	0.47	25791	52309	SUBURBAN
11	3.6	16.05	0.77	20027	48642	SUBURBAN
12	3.2	12.52	0.53	37575	45950	URBAN
13	0.6	9.39	0.14	61880	64140	URBAN
14	0.4	14.72	0.49	42755	76920	RURAL
15	4.4	13.63	0.90	25109	44238	URBAN
16	6.1	16.10	0.38	18788	40806	RURAL
17	2.5	12.73	6.71	15340	39970	URBAN
18	4.7	16.15	0.75	18227	37857	SUBURBAN
19	0.1	14.96	0.63	25851	41040	SUBURBAN
20	1.9	16.94	0.39	20418	36346	SUBURBAN
21	0.4	15.87	0.41	22519	36798	SUBURBAN
22	2.2	16.83	0.40	17659	34203	SUBURBAN
23	4.8	17.21	0.31	20937	36084	SUBURBAN
24	2.2	13.39	0.42	15986	36922	URBAN
25	3.5	15.55	1.07	16428	32993	RURAL

SEGMENT	BASE	VETERAN	FORCEPCT	PERCAPIT	INCOME	LOCATION
26	0.2	15.36	0.63	15784	30470	SUBURBAN
27	0.4	14.06	0.62	16440	31697	RURAL
28	1.7	13.99	1.12	15497	30405	RURAL
29	0.5	13.89	0.92	15034	29185	RURAL
30	1.1	20.05	0.28	23593	33970	SUBURBAN
31	0.6	16.29	0.29	15339	29944	RURAL
32	2.1	11.33	0.37	17794	33872	URBAN
33	0.3	15.44	0.25	14575	29175	RURAL
34	0.5	6.83	19.30	17100	30874	SUBURBAN
35	2.9	15.38	0.22	14661	27508	SUBURBAN
36	1.4	7.44	0.09	18133	33074	URBAN
37	0.5	9.99	0.41	33140	36502	URBAN
38	8.0	15.11	0.25	13950	27601	RURAL
39	3.5	15.04	0.85	21736	30279	SUBURBAN
40	4.3	13.91	0.71	13902	24286	SUBURBAN
41	1.7	6.89	0.23	9432	24927	URBAN
42	1.3	11.70	0.29	11751	23203	RURAL
43	0.5	11.21	0.23	11212	18675	SUBURBAN
44	0.1	12.20	0.88	14722	23133	URBAN
45	1.5	10.49	0.50	17347	27650	URBAN
46	2.5	9.34	0.20	10904	19981	URBAN
47	0.7	3.86	1.16	14119	20748	URBAN
48	0.9	13.22	0.24	20020	19630	URBAN
49	0.1	13.83	0.70	19099	38323	
50	0.1	8.50	37.24	14157	36740	

SEGMENT	WHITE	BLACK	ASIAN	HISPANIC	EDUC
0	80.00	12.00	3.00	9.00	
1	92.78	1.97	4.62	2.60	Bachelors Degree
2	90.65	2.38	5.71	3.78	Bachelors Degree
3	93.09	3.12	2.78	2.90	Bachelors Degree
4	88.69	2.68	6.30	5.67	Associate Degree
5	78.40	6.51	11.19	8.01	Associate Degree
6	95.56	2.41	0.98	2.13	Associate Degree
7	93.37	2.91	2.29	3.48	Some College
8	90.93	4.22	3.37	3.88	Bachelors Degree
9	92.28	2.99	2.40	4.25	Associate Degree
10	91.99	3.55	2.53	4.75	Some College
11	91.66	4.08	1.70	6.05	Associate Degree
12	86.66	6.34	4.30	6.67	Bachelors Degree
13	86.05	6.17	5.16	7.81	Post Graduate Degree
14	84.90	4.11	6.80	8.64	Associate Degree
15	82.95	7.47	5.09	9.80	Associate Degree
16	94.92	2.98	0.50	2.40	HSDG
17	68.01	9.23	7.90	27.61	Some College
18	90.47	4.89	1.42	6.78	HSDG
19	90.23	4.48	2.69	5.25	Associate Degree
20	91.45	4.60	1.36	5.07	HSDG
21	88.21	7.02	1.62	5.72	HSDG
22	91.71	4.53	1.01	5.47	HSDG
23	94.49	2.99	1.00	3.11	HSDG
24	20.29	75.42	1.22	5.52	Some HS
25	86.95	8.10	1.15	6.45	HSDG

SEGMENT	WHITE	BLACK	ASIAN	HISPANIC	EDUC
26	87.40	6.09	1.29	7.97	HSDG
27	77.65	14.78	1.68	8.46	Some HS
28	74.73	17.48	1.62	9.69	Some HS
29	73.79	18.86	1.40	8.69	Some HS
30	94.24	3.00	1.10	3.67	HSDG
31	92.01	4.74	0.49	4.09	HSDG
32	77.73	8.10	4.58	21.58	Some HS
33	93.35	3.17	0.49	3.37	HSDG
34	80.95	11.40	4.16	5.54	Some College
35	90.31	6.33	0.44	4.37	HSDG
36	46.43	26.98	10.20	32.65	Some HS
37	69.70	17.63	8.44	8.50	Bachelors Degree
38	92.52	5.19	0.27	2.65	HSDG
39	88.82	6.36	1.88	5.75	Some College
40	77.96	12.69	1.63	14.04	Some HS
41	48.05	7.94	3.84	68.36	Some HS
42	52.98	41.30	0.33	4.63	Some HS
43	21.64	72.54	0.74	7.89	Some HS
44	71.76	13.14	3.36	20.75	Some HS
45	31.62	47.18	11.48	16.01	Some HS
46	12.96	77.48	1.09	13.19	Some HS
47	82.59	8.21	6.58	4.50	Bachelors Degree
48	76.93	15.73	2.58	10.39	Some HS
49	73.91	18.47	2.26	8.74	
50	68.77	22.14	2.67	11.00	

SEGMENT	PCTWHITE	PCTBLUE	RENTPAID	HOUSE	PROPERTY
0	58.14	41.86	374		79098
1	87.70	12.30	786	HOME OWNER	324899
2	81.93	18.07	783	HOME OWNER	192592
3	80.53	19.48	573	HOME OWNER	149073
4	76.38	23.62	671	HOME OWNER	245155
5	71.67	28.33	716	HOME OWNER	165768
6	66.95	33.05	455	HOME OWNER	132996
7	72.60	27.40	515	HOME OWNER	133859
8	81.63	18.37	555	OWN/RENT	163390
9	66.20	33.80	497	HOME OWNER	138367
10	68.43	31.57	512	HOME OWNER	123589
11	60.48	39.52	465	HOME OWNER	91691
12	78.56	21.44	551	RENT	177666
13	89.10	10.90	687	RENT	380053
14	73.61	26.39	611	HOME OWNER	324322
15	66.02	33.98	518	OWN/RENT	130593
16	50.13	49.87	324	HOME OWNER	81301
17	50.32	49.68	499	HOME OWNER	106735
18	53.35	46.65	379	HOME OWNER	71720
19	67.06	32.94	456	OWN/RENT	124702
20	57.65	42.35	364	HOME OWNER	80858
21	58.68	41.32	418	HOME OWNER	95664
22	52.83	47.17	324	HOME OWNER	64177
23	59.97	40.03	355	HOME OWNER	74787
24	51.96	48.04	379	HOME OWNER	68386
25	48.70	51.30	322	HOME OWNER	63897

SEGMENT	PCTWHITE	PCTBLUE	RENTPAID	HOUSE	PROPERTY
26	47.92	52.08	298	HOME OWNER	60624
27	47.26	52.74	323	HOME OWNER	73487
28	46.70	53.30	316	HOME OWNER	62739
29	45.31	54.69	298	HOME OWNER	59775
30	62.37	37.63	397	HOME OWNER	95030
31	43.95	56.05	252	HOME OWNER	56835
32	51.68	48.32	429	RENT	112754
33	40.51	59.49	231	HOME OWNER	52154
34	67.14	32.86	383	RENT	90800
35	44.96	55.04	236	HOME OWNER	45959
36	57.83	42.17	427	RENT	208036
37	77.74	22.26	501	RENT	215890
38	39.49	60.51	214	HOME OWNER	47217
39	61.39	38.61	376	OWN/RENT	80913
40	43.23	56.77	286	OWN/RENT	47022
41	32.83	67.17	356	OWN/RENT	64667
42	38.15	61.85	175	HOME OWNER	41729
43	37.20	62.80	208	OWN/RENT	37053
44	44.96	55.04	327	RENT	60669
45	54.55	45.45	379	RENT	103439
46	39.63	60.37	263	RENT	42010
47	67.30	32.70	379	RENT	80934
48	57.75	42.25	294	RENT	73664
49	49.51	50.49	328		109307
50	57.47	42.53	470		87167

SEGMENT	MARITALS	SFEMALE	SMALE	HOUSHOLD	HMARRIED	AGERANGE
0	54.79	12.17	14.76	83.67	55.15	35.6
1	67.11	10.62	11.97	91.24	76.07	45-49
2	70.74	9.79	11.59	95.48	82.41	35-49
3	66.37	10.13	11.79	90.94	71.74	40-54
4	62.04	10.77	13.15	87.04	65.57	40-54
5	65.04	10.29	12.85	92.71	73.70	30-44
6	67.90	9.17	11.84	91.73	74.72	35-49
7	65.04	9.63	11.72	89.58	69.53	45-59
8	55.20	12.98	14.34	77.55	52.35	35-49
9	63.25	10.14	12.81	88.77	67.80	35-49
10	61.52	10.50	12.70	88.20	64.88	50-65
11	65.11	9.85	12.35	92.89	72.41	35-49
12	40.25	17.27	19.08	58.58	31.73	22-34
13	34.95	23.58	23.28	48.37	25.07	30-44
14	57.10	11.98	14.63	81.70	58.35	45-59
15	50.09	13.79	16.37	77.84	46.40	25-34
16	65.05	8.79	12.07	89.76	69.22	40-54
17	58.02	11.19	15.97	90.67	65.64	25-34
18	58.35	10.35	12.91	87.18	59.21	25-34
19	50.41	14.02	15.69	72.81	48.41	21-24
20	58.51	9.69	11.74	83.47	57.03	55-84
21	55.69	10.15	11.74	78.69	52.76	55-84
22	58.99	9.72	12.04	86.15	58.99	50-69
23	56.82	10.28	11.90	81.35	55.01	55-69
24	43.10	17.02	17.26	89.31	45.90	40-59
25	58.97	9.68	12.79	86.65	59.68	50-64

SEGMENT	MARITALS	SFEMALE	SMALE	HOUSHOLD	HMARRIED	AGERANGE
26	57.92	9.69	12.79	84.75	57.73	55-84
27	55.09	11.12	14.53	84.66	55.32	25-55
28	54.16	11.47	14.45	85.40	54.14	25-55
29	54.05	11.50	14.44	85.31	53.88	18-21
						55-74
30	58.97	8.35	9.40	76.21	53.17	60-84
31	62.87	8.25	11.53	87.32	63.61	45-59
32	44.63	14.90	18.85	78.84	41.99	21-34
33	63.83	8.12	11.99	87.70	65.16	45-59
34	34.12	22.78	34.50	49.18	48.04	18-24
35	58.33	9.24	12.00	85.10	57.75	45-59
36	40.32	17.91	19.94	79.29	35.61	25-34
37	24.90	24.60	29.98	39.95	18.07	25-34
38	62.98	8.18	11.76	87.90	64.30	45-64
39	44.36	13.70	15.64	66.85	37.85	18-34
40	47.12	12.41	14.97	81.35	43.40	21-29
41	49.13	14.91	19.60	92.79	55.12	18-29
42	52.02	12.59	14.70	87.88	53.72	18-20
						55-84
43	34.31	18.61	17.90	84.76	31.99	18-20
44	40.32	14.61	19.03	73.25	35.30	18-20
45	33.32	18.94	22.14	71.39	27.90	18-34
46	29.65	22.57	20.45	86.61	27.48	18-24
47	13.43	38.60	41.50	21.91	19.75	18-24
48	25.71	15.81	21.04	42.15	17.87	18-29
						65-84
49	53.10	11.89	15.57	83.19	53.94	
50	32.09	11.01	45.59	32.77	62.54	

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Interview with MSG Patrick Sarley, REQUEST NCO, Army Reserve Personnel Command, St Louis, MO, and the author 14 November 2002.

Telephone conversation between MSG Patrick Sarley, REQUEST NCO, Army Reserve Personnel Command, St Louis, MO, and the author 19 February 2003.

Telephone conversation between MSG Patrick Sarley, REQUEST NCO, Army Reserve Personnel Command, St Louis, MO, and the author 29 April 2003.

Telephone conversation between MSG Patrick Sarley, REQUEST NCO, Army Reserve Personnel Command, St Louis, MO, and the author 12 May 2003.

Telephone conversation between Mr. Dennis Sullivan, Senior Personnel Analyst for AT&T Government Solutions, contracted to U.S. Army Reserve Command G-1, and the author, 14 May 2003.

Telephone conversation between Mr. Charles Dalbec, Senior Personnel Analyst - Resource Consultants Inc, contracted to U.S. Army Reserve Command G-1, and the author 4 June 2003.

Department of the Army Pamphlet 611–21, Military Occupational Classification and Structure, Headquarters Department of the Army, Washington, DC 31 March 1999.

Department of the Army Pamphlet 601–5–14, Personnel Procurement: REQUEST MANAGER'S HANDBOOK, Headquarters, Department of the Army, Washington, DC 1 December 1984.

Devore, Jay L., Probability and Statistics for Engineering and the Sciences, fifth edition, Duxbury, 2000.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Major Ward Litzenberg
   Programs Analysis and Evaluation
   Office of the Commander of the Army Reserve
   Arlington, Virginia

4. Joyce Dyer
   Programs Analysis and Evaluation
   Office of the Commander of the Army Reserve
   Arlington, Virginia

5. Major Michael Kamei
   Programs Analysis and Evaluation
   United States Army Recruiting Command
   Fort Knox, Kentucky

6. Dennis Sullivan
   Deputy Chief of Staff for Personnel, G-1
   U.S. Army Reserve Command
   Fort McPherson, Georgia

7. Master Sergeant Patrick Sarley
   Deputy Chief of Staff for Personnel, G-1
   U.S. Army Reserve Command
   Fort McPherson, Georgia

8. Charles Dalbec
   Deputy Chief of Staff for Personnel, G-1
   U.S. Army Reserve Command
   Fort McPherson, Georgia

9. Alan Craig
   Manpower, Personnel, and Training
   Deputy Chief of Staff for Personnel
   Department of the Army
   Arlington, Virginia

10. Mary Wood
    Personnel Proponency Office
    Army Reserve Personnel Command
    St Louis, Missouri

11. Professor Susan Sanchez
    Naval Postgraduate School
    Monterey, California

12. Professor David Olwell
    Naval Postgraduate School
    Monterey, California