NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

FORECASTING MARINE CORPS ENLISTED ATTRITION THROUGH PARAMETRIC MODELING

by

Jeremy T. Hall

March 2009

Thesis Advisor:	Samuel E. Buttrey
Second Reader:	Jeremy Arkes

Approved for public release, distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** March 2009	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE Forecasting Marine Corps Enlisted Attrition Through Parametric Modeling		**5. FUNDING NUMBERS**	
6. AUTHOR(S) Jeremy T. Hall			
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) N/A		**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**	

13. ABSTRACT (maximum 200 words)

The Marine Corps, as with any organization with a large workforce, must accurately monitor and more importantly predict the transition rates among personnel entering and exiting the enlisted and officer ranks. This emphasis is even more appropriate given that the Marine Corps has been authorized to increase the current authorized end strength by 13,000 personnel from Fiscal Year 2008 to Fiscal Year 2010. The purpose of this thesis is to apply parametric modeling (specifically survival analysis) to historical data sets of enlisted personnel in order develop a more efficient forecasting tool for military planners. It is the intent to include in the model those characteristics that significantly influence attrition behavior, and aggregate these findings to an efficient, yet effective forecasting model. Therefore, this thesis will analyze the interaction of time, individual characteristics, and those causal attributes that determine whether a Marine completes his or her contracted service. The current forecasting method used by the Marine Corps forecasts enlisted attrition annually. This study forecasts enlisted attrition monthly within occupational field. Hence, the data was structured to provide this depth of analysis. In comparison to the current forecasting method of exponential smoothing this study found that the use of survival analysis could be beneficial to not only forecast attrition, but also provide a descriptive assessment of attrition rates amongst occupation fields without loss of information due to averaging or weighting probabilities.

14. SUBJECT TERMS Forecasting, Attrition, Marine Corps NEAS losses, Gompertz Model, Survival Analysis			**15. NUMBER OF PAGES** 85
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500 Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release, distribution is unlimited

FORECASTING MARINE CORPS ENLISTED ATTRITION THROUGH PARAMETRIC MODELING

Jeremy T. Hall
Captain, United States Marine Corps
B.A., Ohio State University, 2002

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
March 2009

Author: Jeremy T. Hall

Approved by: Samuel E. Buttrey
Thesis Advisor

Jeremy Arkes
Second Reader

Bill Gates
Dean, Graduate School of Business and Public Policy

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The Marine Corps, as with any organization with a large workforce, must accurately monitor and more importantly predict the transition rates among personnel entering and exiting the enlisted and officer ranks. This emphasis is even more appropriate given that the Marine Corps has been authorized to increase the current authorized end strength by 13,000 personnel from Fiscal Year 2008 to Fiscal Year 2010. The purpose of this thesis is to apply parametric modeling (specifically survival analysis) to historical data sets of enlisted personnel in order develop a more efficient forecasting tool for military planners. It is the intent to include in the model those characteristics that significantly influence attrition behavior, and aggregate these findings to an efficient, yet effective forecasting model. Therefore, this thesis will analyze the interaction of time, individual characteristics, and those causal attributes that determine whether a Marine completes his or her contracted service. The current forecasting method used by the Marine Corps forecasts enlisted attrition annually. This study forecasts enlisted attrition monthly within occupational field. Hence, the data was structured to provide this depth of analysis. In comparison to the current forecasting method of exponential smoothing this study found that the use of survival analysis could be beneficial to not only forecast attrition, but also provide a descriptive assessment of attrition rates amongst occupation fields without loss of information due to averaging or weighting probabilities.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

My sincerest thanks and admiration to the faculty and staff of the Naval Postgraduate School for their professional and personal dedication in the advancement of higher learning, which enabled me to complete this research.

I also like to recognize the efforts of my family in the support of the long hours dedicated in the completion of this study. Dad is finally “done with his homework.”

Lastly, to the soul-wrenching and inspirational music of John Lee Hooker that motivated me through long hours and times of great pondering.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. BACKGROUND

The Marine Corps, as with any organization with a large workforce, must accurately monitor and more importantly predict the accessions and losses for the enlisted and officer ranks. This emphasis is even more appropriate given that the Marine Corps has been authorized to increase the current authorized end strength from 189,000 to 194,000 Marines in fiscal year 2009 and by an additional 8,000 Marines for fiscal year 2010 (The National Defense Authorization Act for Fiscal Year 2008, Public Law 110–181). Thus, in a three-year period the Marine Corps will have grown by 13,000 personnel, consequently increasing manpower costs.

The manpower costs of the Marine Corps comprise over 60% of the total fiscal year budget. The annual costs associated with maintaining an all-volunteer force were $9.5 billion for fiscal year 2008 (The National Defense Authorization Act for Fiscal Year 2008, Public Law 110–181), and they will only continue to rise as the force grows larger. Appendix A provides a complete listing of Marine Corps personnel end strength for fiscal year 2008. Therefore, accurately and efficiently managing the force and forecasting attrition rates is crucial. Recent endeavors to accomplish this requirement have not been successful. An over-estimation of the end of fiscal year-end strength for 2001-2002 cost the Marine Corps $200 million in Operation and Maintenance Funds (Hattiangadi, Kimble, Lambert, Quester, CNA, 2005). Such reductions in the O&M funds can reduce operational and material readiness. The tightrope walked in forecasting year-end strengths is a precarious one. If the Corps under-estimates enlisted losses, then new accessions will not be sufficient to replace personnel required and mission readiness could suffer across the Marine Corps. On the other hand, if enlisted losses are over-

predicted, and new accession quantities are not adjusted, then the Corps will overspend the personnel budget. There is an art and science to managing labor force transition rates; the art comprises the ability to "see into the future" of personnel strengths.

Marine Corps personnel end strength is calculated at the end of each fiscal year as follows:

Endstrength = Fiscal Year (FY) beginning strength minus losses + gains

The U.S. Congress mandates the end-strength. Title X allows for an overage of overall personnel not to exceed 2-3 percent. The Secretary of the Navy must authorize an overage of 2%, while the Secretary of Defense must approve a 3% overage. The end-strength may not exceed 103% of the end-strength authorized in the current year's National Defense Authorization Act. Table of Organization (T/O) requirements and manpower policies determine the required FY beginning strength. The Enlisted Strength Planners (MPP-20) in concert with the Officer Inventory Planner (OIC) (MPP-30) construct the plans for end-strength requirements. The plan, by pay grade per month, is for the current budget year and for six years into the future (Hattiangadi, Kimble, Lambert, Quester, CNA, 2005).

The T/O is a personnel requirements roster that is broken down for any unit within the Marine Corps. It specifies the required rank, Military Occupational Specialty (MOS), component code, and personnel quantities for that specific unit to operate the designated mission. In unison with the T/O, manpower policies determine the number of reenlistment contracts available for each MOS and boat spaces are allotted on a first-come-first-served basis for qualified Marines within each MOS. All of this is governed by the Budget Authority (BA), which dictates the available funds for personnel costs for the current fiscal year. Therefore, forecasting future end of active service (EAS), retirement, or non-end of active service (NEAS) losses directly affects the required accessions for each year. The focus of this thesis is to contribute to the forecasting of NEAS losses, specifically attritions: those Marines who do not complete their contracted active service.

B. PURPOSE OF THIS STUDY

The purpose of this thesis is to apply parametric modeling (specifically survival analysis) to historical data sets of enlisted personnel in order develop a more efficient forecasting tool for military planners. It is the intent to include in the model those characteristics that significantly influence attrition behavior, and aggregate these findings to an efficient, yet effective forecasting model. Therefore, this thesis will analyze the interaction of time, individual characteristics, and those causal attributes that determine whether a Marine completes his or her contracted service. The following two primary questions will drive this analysis:

1. What causal factors and individual characteristics contribute to attrition behavior?

2. Can a more efficient and effective forecasting model be developed to either replace or complement current forecasting methods for NEAS losses?

C. SCOPE AND METHODOLOGY

This study will rely on survival analysis to assess those factors that are associated with attrition. Event history analysis, duration analysis, or life-to-death-analysis, are other common names for this methodology, but the fundamental approach is the same. A *subject* is observed, in an *origin state*, for a *duration or episode* until that subject leaves the origin state through an *event,* or is *censored* and cannot be further observed. The duration of the origin state or episode and those causal factors that may have caused the event are analyzed. An event could be death, as in medical studies, or generator failure, as in mechanical studies. Survival analysis has been used in the medical community to study the effects of a drug on cancer patient survival, and the effects of a new surgery on heart patients. The engineering community has used event history analysis to study the effects of new engine components or synthetic oils on the life expectancy of diesel engines. Social scientists have increasingly used survival analysis to forecast drug-use among teenagers, labor force transition rates for large organizations, and expected duration for peacekeeping missions responding to civil and international crisis. The

applications of survival analysis can describe the influences explanatory variables have on the probability of an event occurring in the future. This descriptive ability is achieved through the temporal ordering of a cause (change in a time-dependent variable) to an effect (departure from an initial state) and the analysis of the temporal interval of the time between the cause and the effect (Blossfeld, Golsch, Rohwer, 2007).

In all systems, there is a temporal order to causes and their effects. There is also a temporal interval between cause and effect. In very rare cases does a change in value of a time-dependent covariate result in an instantaneous occurrence of an effect. In most cases, there is a lag between a cause and an event. Of specific importance to applying survival analysis to manpower studies is the comparison of this temporal interval (or lag) amongst a sample of a population that experiences the same change in a time-dependent variable (marriage, number of combat deployments, promotion, etc.). Therefore, survival analysis can model the importance time has on the probability that a cause will produce an effect (NEAS loss in this study). The Human Capital Theory states that as employee-specific investments decline (On-the-job-training, task-specific training, etc), exits of experienced longer tenured employees from the organization will also decrease due to the accumulation of job-specific experience and skills that may not be transferable to another organization. Survival analysis can measure this accumulation of job-specific experience and apply probabilities of future failures based on the individual characters within the sample and the time already spent in the organization.

This study examines enlisted Marines who enter the origin state upon initial enlistment until failure (NEAS loss) or until they exit the analysis by choosing not to reenlist. Then personal characteristics such as race, sex, and marital status, are analyzed to determine if these attributes can be used to forecast future attrition behavior. Lastly, a forecasting model is introduced to calculate the hazard rate (or probability of failure) for a specific time of interest given the covariate estimates used in the study.

D. ORGANIZATION OF THE STUDY

Six chapters comprise this study. Chapter II is a literary review of previous attrition studies that utilized survival analysis to forecast future attritions. Chapter III is a basic description of the mathematical formulas and terms used for survival analysis. Chapter IV defines the covariates and provides descriptive statistics of the data used in the parametric model. Chapter V defines the model and discusses the results. Chapter VI summarizes the results and concludes with the author's recommendations for follow-on research in forecasting enlisted attrition.

THIS PAGE INTENTIONALLY LEFT BLANK

II. LITERATURE REVIEW

A. PREVIOUS ATTRITION AND LOSS STUDIES

The Marine Corps, as with any organization with a large labor force, is keenly aware of the need for efficient monitoring of labor force transition rates. Labor is the most expensive resource to maintain, develop, and replace. Accounting for the personnel required to complete an organization's mission requires diligence, continual re-examination of personnel policies, and a forecasting model developed by sound theory and explanatory variables. The techniques to study and analyze personnel behavior are as varied as the personnel within an organization. Therefore, predicting who leaves an organization, and when, will continue to be a critical topic of interest for any large organization. This may be even more true for the military, which must effectively and efficiently maintain material and personnel readiness in order to service the nation's interests through the threat of or the use of force. Unaccounted budgetary expenditures in manpower overages due to poor or inaccurate end-strength forecasts diminish available funds to maintain material readiness and can affect the military's ability to conduct combat operations. The following studies were conducted to understand better and more specifically, improve the predictive capabilities of personnel attrition rates from military service.

The first study reviewed utilized a binary logit model in order to predict enlisted Marine Corps End of Active Service (EAS) and Non-End of Active Service (NEAS) losses from the period 1997-2007. The second study utilized the Weibull and exponential models to obtain the survival functions of individual characteristics of enlisted Coast Guard personnel for the fiscal years 1983 to 1990. Then the author constructed a logit model to predict future attritions using past accession and attrition information. The third study reviewed for this thesis is a CNA report that describes in great detail the current methodology the Marine Corps utilizes to forecast personnel endstrength.

1. Forecasting Marine Corps Enlisted Losses (Orrick, 2008)

A Naval Postgraduate School master's thesis, written by Captain Sanford C. Orrick in March 2008, examined the current methodology of forecasting enlisted loss rates in the Marine Corps and then attempted to develop a more efficient forecasting model using logit regression. The focus of his study was to compare those attributes and characteristics of NEAS losses to EAS losses for the period 1997–2007. In order to develop a more accurate forecasting tool for Marine Corps personnel planners, his research attempted to identify factors contributing to enlisted Marines who leave the service prior to their EAS. Essentially, he sought the use of recordable attributes that significantly affect the probability of attrition.

Captain Orrick's data was obtained from the Total Force Data Warehouse (TFDW), and included three different sets of data captured by fiscal year. The first data set is enlisted Marine accessions from 1997 to 2007. The second data set is Marine enlisted losses, either EAS or NEAS, from 1997 to April 2007. The last data set is a snapshot of Marine enlisted endstrength for fiscal year 1997.

The methodology for this research was to compare the attributes of those Marines in the data set who completed their obligated service (classified as EAS) to those of the Marines who did not complete their obligated service (classified as NEAS). The snapshot end-strength data set for fiscal year 2007 was used to capture attributes for those enlisted losses that may not have been captured in the accession data.

The independent variables of the model were estimated for all NEAS losses across all fiscal years in the study. Next, the model was computed again, using only data from fiscal years 1998 through 2004 in order to predict NEAS losses for 2005. The model was continued in this manner including the predicted years' observations to predict the next fiscal years, concluding in fiscal year 2007. The model results were encouraging. According to his results, his model predicted NEAS losses accurately 76.2% of the time.

Unfortunately, several questions of bias and data structure taint this study's findings. The merge of the three data sets yielded 587,154 observations. However, only 167,269 observations were used due to missing data. It is difficult to take the model results as representative of the probability of attrition given the selection bias of being forced to omit potentially viable observations due to missing data.

The data structures used for the logit model suffer from inherent issues in the concept of causality employed for predictive modeling. The first issue is the use of panel data that was used for predicting the next fiscal year's attritions from an accumulation of the previous fiscal years'. Time plays an important role in moderating causality on a dependent variable. Specifically, not only is there a temporal order (cause precedes effect), but also a temporal interval. There is some time that elapses between an event occurring and the impact on the dependent variable. A restrictive assumption of panel analysis is that, the cause and effect happen at the same time. Consequently, the lag time between a cause, and to an event is irrelevant. The larger the discrepancy is between the true lag and observed lag in the data, the less likely panel analysis will uncover the true causal process. The second issue in the data structure is the use of cross-sectional data to compare the NEAS and EAS Marines. Cross-sectional data can over-predict change and overestimate the significance of explanatory variables (Blossfeld, Golsch, Rohwer, 2007). The explanatory variables can only explain the outcome at the specific point in time the data was collected. Thus, changes in time-dependent explanatory variables are not captured over multiple occurrences in the same duration. The last occurrence of a change in a variable is the value used for estimating the probability of a potential outcome. An important analytical aspect lacking from this form of analysis is that predictions might have been different if the previous changes to the time-varying variable had also been included in the model. A Marine just recently divorced may be more likely to attrite than a Marine who has been divorced for ten years and has presumably adjusted to single life. In a cross-sectional data structure, this would not be evident; all divorced Marines would share an equal probability of becoming an NEAS loss, *ceteris paribus*.

The main contribution of Orrick (2008) is that it highlights the importance of time on causality and the subsequent use of forecasting. The ability to capture duration between the cause and effect in cross-sectional and panel studies is limited and therefore limits the effectiveness of the logit model to capture causality over long durations, such as the length of a Marine's career.

2. An Analysis of the Coast Guard Enlisted Attrition (Rubiano, 1993)

A Naval Postgraduate School master's thesis, written by Laureano Enrique Onate Rubiano, analyzed attrition behavior with survival analysis and then attempted to develop a forecasting model in order to predict monthly attritions of enlisted United States Coast Guard (USCG) personnel.

The first goal of developing survival functions for USCG personnel was accomplished by defining personnel characteristics for all observations that would be used to categorize attrition behavior. The data set consisted of USCG enlisted personnel from fiscal year 1983 to fiscal year 1990. The study included pay grades from E-1 to E-9, Military Occupational Skill (MOS), gender, race, minority designation, marital status, and dates of entry and exit from the Coast Guard. Overall, there were 50,036 people in the data set with 29,405 of them exiting the analysis due to discharge from active service.

The author constructed the number of months on active duty, as an integer, by calculating the duration from date of entry to date of exit from the USCG. Survival functions were then generated by pay grade, sex, race, marital status, and rating. The study plotted the estimated survival against time, the negative natural log of the estimated survival function against time, and the natural log of the negative natural log of the estimated survival function against time. The second and third plots were used to check the validity of using the Weibull and exponential survival models. In either model, there was not enough empirical evidence to justify their implementation. Yet, the author continued to use these models to graph the survival functions. The empirical test plots one used to compare *pseudoresiduals,* presumably from the nonparametric Kaplan-Meier estimate (though not stated by the author), against the predicted residuals apparently from

the Cox-Snell model (though again this was not stated by the author). If the data fits with the model then the log survival plot should have assumed a linear pattern through the origin. As stated before, nether of these graphs exhibited this characteristic. The author should have attempted another parametric survival model such as the log-logistic, log-normal, gamma, or Gompertz, to plot the survival curves. As published, the survival functions cannot be relied on to accurately report the historical survival probabilities in the USCG.

It is worth further mentioning that the author did not differentiate between personnel exiting the USCG for retirement, non-reenlistments, or administrative reasons. This is evident in the sharp drops in survival curves in months 48 and 240 for each characteristic. These are the times most first-term enlistees choose not to reenlist and the time of standard retirement after twenty years of service. In order to accurately plot attrition behavior, the study should have calculated the survivor function for each departure event separately.

The thesis constructed a multiple regression model to predict monthly attritions. The independent variable was monthly attritions and the explanatory variables were monthly attrition in the previous months, the number of accessions for the previous four and twenty years, and monthly unemployment rates. In order to measure the performance of the model the mean squared error (MSE) and mean relative error (MRE) were used for 96 observations (one for each month in eight fiscal years) and 33 observations (October 1990 to June 1993) were utilized to validate the model. The author chose to use the four-year and twenty-year attrition number as explanatory variables in the model, citing the drastic change in survival probabilities for these time periods. As mentioned above, an over-emphasis on times that are considered normal, such as choosing not to reenlist or to retire, can bias predictions towards these times (Cleves, Gould, Gutierrez, Marchenko, 2008). These biases can skew potential informative survival probabilities of early attrition behavior toward these two times (48 and 240 months) and precluding the author from determining those characteristics that comprise personnel discharging before their

respective contract date. Nonetheless, the author claimed his model performed better than current USCG forecasting methods. However, the study did not provide a direct comparison.

The main contribution of Rubiano's study is that it advanced the use of survival analysis in an attempt to quantify trends of attrition behavior according to individual characteristics. Though the survival probabilities are probably not nearly as representative of empirical attritions than a better selection of parametric models (log-logistic, lognormal, gamma, Gompertz) would have been, it does demonstrate the ability of survival analysis to link causality to an effect within a duration. The temporal order and the temporal interval were explained in the data.

3. Endstrength: Forecasting Marine Corps Losses Final Report (Hattiangadi, Kimble, Lambert, Quester, CNA, 2005)

This CNA report, from 2005 is a comprehensive and detailed report on the manpower systems, techniques, and procedures the Marine Corps employs to forecast end strength gains and losses. The recognition of the severe consequences of incorrect estimates was the motivation this study. The first approach was to assess the existing loss forecasting processes. The next step was to make the processes systematic for all military personnel planners. Improvements and additions were made to the existing forecasting models and the whole process was documented for continuity amongst personnel planners. Several issues were identified with missing or incorrect data fields that made forecasting calculations less robust. The NEAS Loss Model and active-duty strength planning chapters are the focus of this review.

Forecasting enlisted endstrength entails dividing the enlisted force into three categories; first-term, intermediate, and career Marines. The first-term enlistees are those non-prior service members serving their initial enlistment contract. The personnel planners use the first-term alignment plan (FTAP) to calculate reenlistment rates. The FTAP directly contributes the projected end-strength at the end of the current fiscal year by estimating the number of qualified Marines who are likely to reenlist. These

projections are applied for future fiscal years in order to estimate the number of new accessions required per fiscal year to maintain personnel requirements for a specific MOS. The model is a steady-state model with planner-influenced adjustments that averages reenlistment rates for the past three years. The intent of Hattiangadi, Kimble, Lambert, Quester (2005) is to add survival analysis to the planner's arsenal of tools. Steady-state models, though easy to calculate, cannot readily capture changes in behavior variables that would directly affect a Marine's decision to reenlist, such as an armed conflict or changes in the economy. Moving average models tend to be reactive to changes in the historical data, whereas the use of survival functions calculated from proven covariates of reenlistment influencers can be more descriptive of changes in time-varying variables.

The Marines who have reenlisted after the first-term of enlistment are categorized as Intermediate-term and Career Marines for end-strength forecasting. Intermediate-term Marines are those who have reenlisted and have between three and fourteen years of service. Career Marines, those with fourteen or more years of service, are forecasted similarly. The purpose of the Intermediate-term forecasting model is to forecast the number of first-term Marines that will remain in the Corps after their first term. The model is using an unweighted average over three years of historical data on the reenlistment and attrition behavior of intermediate-term Marines. The CNA report is very detailed on this process and it should be reviewed for greater comprehension of the current Marine Corps' methodology of forecasting end-strength levels. The CNA authors discovered that the strength planners' continuation rates have been under-estimating the empirical EAS continuation rates that is, more Marines are exiting from active than the model is predicting. The authors believe that the under-estimations are caused by the use of unweighted averages and that survival functions could better align forecasting with true force continuation rates.

The NEAS Loss model employed is to predict those Marines who will either retire or fail to complete their contractual obligation. The NEAS Loss model has three components: recruit losses, retirements, and category losses. The review of my study

focuses only on category losses analyzed in the CNA study; recruit losses as they pertain to the data set and are not discussed separately as in the CNA study. Category losses (or attritions) account for 28% of all NEAS losses (Hattiangadi, Kimble, Lambert, Quester, CNA, 2005). Forecasting personnel leaving the service as a category loss is critical. The authors found that the Marine Corps accounts for the six categories of the category losses collectively, a method used in this study as well. One forecasting model calculates a weighted average of the past three years of category losses; the other uses Monte Carlo simulation. The personnel strength planners may decide to use one method rather than the other depending on the accuracy of the forecasts from previous periods to the actual attrition rates. Again, talent and experience of the planners is used in this decision and within the weighting of the averages for future forecasts. It is the premise of this thesis that moving or weighted averages cannot amply explain developing or shifting trends in attrition behavior as responsively as survival analysis. This study intends to demonstrate the power of utilizing historical hazard and survival rates for future forecasting.

The studies reviewed for this thesis all have the same goal: to improve the forecasting of attrition from the active duty forces. There are strengths and weaknesses but the intent is paramount. In a resource-scarce environment, managing the enlisted transition rates by efficiently predicting losses and establishing recruiting efforts to replace these losses can positively impact service personnel readiness.

III. SURVIVAL ANALYSIS

A. INTRODUCTION

The survival analysis in this study is utilized to develop a model in order to more accurately predict enlisted attrition rates. Therefore, it is appropriate to offer a brief introduction to the terminology and equations that are employed to facilitate this type of modeling.

Event history analysis (EHA), duration analysis, and time-to-failure analysis are other common terms used to model the time a subject under study enters a risk set and subsequently fails or leaves the analysis. At its core, EHA measures transitions from discrete states or durations from entry to and exit of the state under observation known as the survival times. The basic analytical structure of event history analysis is the state space and some defined time axis (Blossfeld, Golsch, Rohwer, 2007). Throughout this study, the state space is enlistment and the continuous time axis is 'months enlisted' on active duty. However, there are several ways a Marine can exit or leave the state space of being "enlisted." Essentially, a Marine enters a single state, enlistment, but has multiple destinations. A Marine could be discharged prior to completing his or her contractual obligation, complete his or her obligation and leave the service, or continue his or her enlistment until retirement.

This chapter will describe the basic functions of survival analysis. Following that, there will be an introduction of the parametric modeling technique and log-logistic function used in this study. The chapter will conclude with a brief discussion on censoring and truncation.

B. BASIC MATHEMATICAL COMPONENTS OF SURVIVAL ANALYSIS

Defining T as a positive random variable denoting survival time or time to a failure event is a logical starting point for introduction. The study assumes T is continuous and the actual survival time of a unit is t.

1. Cumulative Distribution and Probability Density Function

The probability distribution of *T* is defined by the probability density function, *f(t)*, and the cumulative distribution function, *F(t)*.

a. Cumulative Distribution Function

$$F(t) = \int_0^t f(u)d(u) = \Pr(T \le t) \tag{3.1}$$

This equation denotes the probability that a survival time *T* is less than or equal to some value *t* in the future. All points that are differentiable in *F(t)* can be used to define *f(t)*.

b. Probability Density Function

The density $f(t)$ is defined by

$$f(t) = \frac{dF(t)}{d(t)} = F'(t) \tag{3.2}$$

Implying

$$f(t) = \lim_{\Delta t \to 0} \frac{F(t+\Delta t) - F(t)}{\Delta t}. \tag{3.3}$$

The probability density function is the unconditional failure rate of the events occurring in a smaller and smaller time unit. The function can be expressed as the instantaneous probability an event (failure) occurs within the specified state space bounded by *t* and $t + \Delta t$,

$$f(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \le T \le t + \Delta t)}{\Delta t} \tag{3.4}$$

2. Survivor Function

The survivor function is given by

$$S(t) = 1 - F(t) = \Pr(T \geq t) . \tag{3.5}$$

The survivor function gives the probability of surviving beyond time *t*. That is *S(t)* is the proportion of units surviving beyond *t.* $S(0) = 1$ at the origin time and monotonically decreases as *t* increases. Thus, at some value of $S(t)$, the probability that one unit has not failed will be zero.

The probability density function gives the unconditional failure while the survivor function provides the proportion of units that will not have failed at time *t*. The important link between these two functions is the hazard function.

3. Hazard Function

The probabilities of failure and survivor functions are linked accordingly:

$$h(t) = \frac{f(t)}{S(t)} . \tag{3.6}$$

The hazard function is the conditional failure rate that denotes the rate of unit failure (or duration ends) by *t* given that a unit survived until *t*. Equation (3.6) can be written as

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T \leq t + \Delta t \mid T \geq t)}{\Delta t} . \tag{3.7}$$

The rate can be increasing, decreasing, or a combination of either increasing then decreasing, or decreasing then increasing as time elapses. In essence, the failure event is conditional on the history (Blossfeld, Golsch, Rohwer, 2007). The conditional aspect of time on the probability of failure can be expanded to include time-constant and time-varying covariates with the following function:

$$h(t \mid x) = \lim_{\Delta t \to 0} \frac{\Pr(t \le T \le t + \Delta t \mid T \ge t, x)}{\Delta t}. \quad (3.8)$$

Therefore, the effects of time and of covariates on a unit's probability of survival to time *t* can be measured with the changes in the hazard function. The interpretation of the effects of covariates to the hazard function in survival analysis is in terms of risk. (Blossfeld, Golsch, Rohwer, 2007)

C. PARAMETRIC MODELS

Parametric models differ from nonparametric and semi-parametric models in one specific, and important, way. Nonparametric and semi-parametric models compare units when events happen to occur. Time is therefore treated as a nuisance and not dependent on an event occurring. Therefore, a covariate within these models that changes value when a failure event does not occur, is not considered in the respective hazard function. Parametric modeling considers the entire duration of a unit given what was known during time (x_j). Thus for each observation in the data for the duration, (t_{0j}, t_j) parametric schemes assigned probabilities utilizing covariate values at (x_j). In addition to accounting for changes in covariates throughout a duration, parametric models allow researchers to assume the shape of the hazard rate whereas nonparametric models allow the "date to speak for itself." The difference is a matter of efficiency and allows for a more precise estimation of the effects covariates have on the hazard rate (Blossfeld, Golsch, Rohwer, 2007).

1. Parametric Proportional Hazards Models

The proportional hazard models begin with the basis equation,

$$h(t \mid x_j) = h_0(t) \exp(x_j \beta_x). \quad (3.9)$$

The equation stipulates that the baseline hazard rate is a product of the covariate value (*x*) and the estimated value (β_x) (the log relative hazard) (Cleves, Gould,

Gutierrez, Marchenko, 2008). A standard interpretation is that $\exp(\beta_i)$ is the hazard ratio for the *i*th coefficient. Parametric models differ from the semi-parametric models in this assumption of the baseline (Blossfeld, Golsch, Rohwer, 2007). Semi-parametric models do not parameterize the baseline hazard $h_0(t)$.

Parametric models specify the shape of the baseline in order to gain a more efficient estimate of the covariates. For example, the Gompertz Model specifies the functional form as,

$$h_0(t) = \exp(a)\exp(\gamma t). \tag{3.10}$$

2. Gompertz Models

The Gompertz Model is one example of a proportional hazards model and it assumes a monotonically and exponentially increasing or decreasing hazard rate. The Gompertz distribution based on the "Gompertz's Law" that proposes that transition rates decline monotonically as duration increases (Blossfeld, Golsch, Rohwer, 2007). The expression for the transition is

$$r(t) = b\exp(ct). \quad b \geq 0 \tag{3.11}$$

Where $r(t)$ is the transition rate. The hazard, cumulative hazard, and survival functions are given below.

a. Gompertz Hazard Function

$$\begin{aligned} h(t \mid x_j) &= h_0(t)\exp(x_j\beta_x). \\ &= \exp(\gamma t)\exp(\beta_0 + x_j\beta_x) \end{aligned} \tag{3.12}$$

b. Gompertz Cumulative Hazard Function

$$H(t \mid x_j) = \gamma^{-1}\exp(\beta_0 + x_j\beta_x)\{\exp(\gamma t) - 1\}, \tag{3.13}$$

c. Gompertz Survival Function

$$S(t \mid x_j) = \exp[-\gamma^{-1} \exp(\beta_0 + x_j \beta_x)\{\exp(\gamma t) - 1\}] \tag{3.14}$$

In STATA $\hat{c} = \hat{\gamma}_0 = gamma$, $\hat{b} = \exp(\hat{\beta}_0)$ for analysis purposes (Cleves, Gould, Gutierrez, Marchenko, 2008).

D. CENSORING AND TRUNCATION

Censoring and truncation occur in nearly all real data-analysis situations. Censoring occurs when a subject is not under observation and a failure event occurs (Cleves, Gould, Gutierrez, Marchenko, 2008). Truncation is slightly different in that there is ignorance of the information that the researcher does not have for a given observation. The strength of survival analysis over Ordinary Least Squares (OLS) logistic or regression is that these censored or truncated observations are included in the analysis as long as a portion of the duration falls within the analysis time. OLS and logistic regressions typically exclude these observations. The following paragraphs explain censoring and truncation more in depth.

1. Censoring

Censoring, in addition to as described above, occurs when an observation's full event history is not observed. Observations can either be right censored or interval censored. Right censoring occurs when an observation is under study for some time then either exits the study or the study concludes and a failure event was not observed. Typically, right censoring occurs when the analysis time ends due to data collection limitations or some other factor that causes to researcher to end the analysis. It becomes unknown when a right censored observation fails, only that the observation survived until the end of the study.

Interval censoring occurs when an observation fails between two points in time of observation, but the exact time of failure is unknown. This type of censoring is usually experienced in medical or interview studies when observations are assessed at discrete times (Cleves, Gould, Gutierrez, Marchenko, 2008).

2. Truncation

Truncation is when a period of an observation's history is unobserved and, therefore, cannot be included in the analysis. Observations can either be left truncated or interval truncated. Left truncation occurs when an observation enters into the risk set prior to the analysis time. In this situation, an observation will be at risk longer than other observations that entered the risk set on or after the analysis time.

Interval truncation (or gaps) occur when a unit under study is not observable for a portion of the observation time. Essentially, the subject disappears for a time; then returns for observation. The obvious disadvantage to this form of truncation is that the time a change to a time-varying covariate occurs is not observed.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. DATA AND METHODOLOGY

A. INTRODUCTION

The analysis begins with the assumption that the hazard rate, or risk of attrition, decreases the longer a Marine serves on active duty. This assumption rests within the Human Capital Theory, which theorizes that within an imperfect market and with imperfect information, an employee will continue to "invest" through continued labor, as long as the perceived benefits are greater than the perceived costs. Therefore, a Marine who views promotion opportunities, educational benefits, pay, health care, etc; as a positive return for deployments, changes in duty station, regimented lifestyle of the military, and general sacrifices away from the family as positive; they will choose to be committed to their service obligations. Once a Marine perceives that the costs are greater than the benefits for continued service, he or she may decide to increase his or her net benefit through the application of his or her talents, knowledge, and abilities outside the military.

A hypothesis of this study is that those who attrite are examples of people for whom the perceived benefits of exiting the service outweigh the consequences of failing to complete the contracted service obligation. The immediate benefits to a person's initial investment of their time is first one to two years of a Marine's enlistment is the duration he or she receives the most regimented of training. This period includes recruit training, combat training, and basic education within the Military Occupational Skill (MOS). Once an enlisted Marine completes these entry-level schools, he or she is assigned to an active duty unit.

A Marine's active duty unit provides further training to round out the "schoolhouse" skills with the techniques and procedures used in daily operations. As Marines develop from basic trained personnel, further education becomes primarily the responsibility of those individuals. It is theorized that at this time, some individuals no longer feeling guided in their development, will weigh the "costs" of continued

commitment to their unit, MOS, and Marine Corps against the perceived benefits of opportunities outside military service. Consequently, these individuals may become less productive on duty, may be less motivated to put forth the effort to assimilate into military structure, and may develop behavioral deficiencies on off-duty hours. In essence, their inaction and resistance to assimilate become the catalyst for early discharge for administrative, disciplinary, or convenience of the government discharges. These are the majority of NEAS Losses.

As the catalyst for this and many other studies of attrition, is the question, "How can an organization identify these types of individuals?" The simplest answer is that an organization cannot. Unless all employees could either be continually surveyed for job satisfaction or managers become mind readers, organizations cannot identify who would leave. Continual surveys are inefficient and mind reading is impossible. Therefore, we can only look to historical trends, characteristics and attributes in attrition behavior, and apply relevant theories to predict those most likely to attrite. Previous studies in attrition have compared and analyzed similar characteristics of service members who became NEAS losses and applied statistical methodologies to probabilities of attrition based on the average of these characteristics. This study follows this formula as well, but also attempts to model time to attrition behavior. The premise that the probability of attrition diminishes with time is central to the Marine's perceived future value of continued service. Time, especially in the military, is a determinant of service requirements. Enlistment contracts, deployment lengths, time in service and time in grade requirements for promotion to the next grade all form the perceived costs of continued service. Therefore, time is not treated as a nuisance in this study, as it is in non-parametric analysis. Time is the decision factor that all service members must consider when choosing to "re-up" or "get out", regardless of the means of "getting out."

There is ample evidence that certain specific attributes contribute to attrition behavior. Education attainment, race, gender, and age are all substantiated indicators of likely attrition behavior. However, these cannot define the entire likelihood of attrition through statistical averaging in an assumed "steady-state" as in exponential smoothing or

moving-average techniques. Rather, viewing these and other characteristics as investment criteria can shed light on attrition behavior. A Marine with a family may perceive more benefit from extended deployments than a single Marine. Though deployed and away from the spouse and children, that family is receiving benefits from the Marine's pay, benefits, and service provided healthcare, whereas, a single Marine, typically, has only himself or herself to care for. These two Marines will react in different ways to increases in perceived costs: for example cost from extended or back-to-back deployments, longer service time for promotion, or stresses of their MOS. The married Marine may be more willing to "pay" in time and personal application to continue to receive the benefit of providing for the family, while single Marines may be more apt to seek higher benefits outside the Corps and cease to invest personal abilities and talents to service obligation; perhaps creating the conditions for early separation. Another theory that this thesis applies is the concept of causality.

The concept of causality is employed for this predictive modeling. The concept states that each unit of a population must be exposable to any of the various levels of a cause. There are Occupation Fields (OccFlds) in the Marine Corps that are not open to every Marine. Infantry, Artillery, and Tank/Assault Vehicle are male-only. However, these male-only OccFlds do not violate the concept of causality in the study as these are limitation based on associated attributes (gender) and are restrictive to all female Marines. This study attempts to apply the concept of causality with the available data and the resources available to military planners in mind. The optimum modeling strategy would employ all the known variables of a population and that the information would be accurate. In either case, this was not possible given the limited scope of this study. The goal of this study is to develop a forecasting model for military personnel planners. The more complex methodologies employed in this study would likely translate into a complex and time-consuming model for planners. Military planners are concerned with the aggregate when forecasting period attrition rates and not computing the various combinations of attributes in order to make predictions. Therefore, care has been taken in

the organization of this data to construct a simple and efficient model that forecasts the aggregate probabilities of attrition utilizing the statistical findings of numerous, yet descriptive, variables.

This chapter provides some descriptive statistics of the data set and concludes with the parametric model selection process.

B. DATA COLLECTION

The data employed in this study is from the Marine Corps' Total Force Data Warehouse (TFDW). The master data set is the combination of twenty-five individual data sets. The master data set used in the parametric model contains data on all enlisted Marines who enlisted in the Marine Corps between January 1, 1996 and October 31, 2008. The master data set does not contain Officers and enlisted Marines who accessed prior to January 1996. Twelve of the data sets are yearly information containing monthly "snapshots" per enlisted Marine, per fiscal year, beginning January 1996 and concluding October 2008. These twelve data sets primarily contained the accession date for Marines that joined in that data set's fiscal year. The purpose of these data sets is to capture all new accessions and to verify the continued service of enlisted Marines that accessed in previous fiscal years. A "Personal Statistic" data set for each fiscal year accompanied the accession data sets. The utility of these data sets is to capture changes in time-varying variables for each month per observation. The "Personal Statistic" data sets contain individual information such as education level, rank, marital status, etc. for analysis. Lastly, a "Separation" data set captured the separations per fiscal year as recorded by the "Type Change Code" and "Action Date". The data sets are merged into one master file. STATA/10C for Windows is the statistical software used in the analysis. Monthly observations per fiscal year were collapsed to a one duration per observation. For example, a Marine who enlisted in January 1996 and who is still on active duty as of October 2008, began with twenty-four observations. Those twenty-four observations were then consolidated to one observation detailing the duration of the Marine's service.

The fiscal years 1996 to 2008, and the limitation of analyzing only those enlistees who assessed after January 1996, were specifically determined by the author to capture the duration of greatest volatility among attrition rates of enlisted Marines. Previous studies in attrition, (Hattiangadi et al., 2005; Rubiano, 1993) found that attrition rates drastically decline after twelve years of service. Therefore, this study seeks to capture attrition behavior of enlisted Marines from their initial entry to the twelfth year of service.

C. DATA SUMMARY

The initial master data set contained 419,893 individual observations. However, 39,562 observations were dropped due to data abnormalities. The active duty statuses of 39,559 observations could not be obtained from the individual data sets. These observations did not have entries indicating separation from active duty, but contained less and less information in the following periods of data. In some cases, many variables were blank. Therefore, a separation date could not be calculated for these observations and continued service was not verifiable. Three observations were dropped due to erroneous gender codes. The adjusted master data set used in the analysis contains 376,710 observations and 373,647 individual subjects. The additional 3063 observations are residue left over from the coding of the data. Of the 3,063 multiple observations, 2,216 have a "Deserter" status. There was an initial attempt to include deserters as a failure event but this disrupted the model because if a deserter returns from desertion their PEBD (origin) is adjusted to reflect the time lost. In the master data set for these observations are two durations; 1) from time of entry to desertion and 2) return from desertion to separation. The study counts the duration from the date of accession to the date of separation. The remaining 847 duplicate observations of the 3,063 are administrative corrections. The first duration for these observations is from the date of accession to the date of separation. The second duration is from the day after the date of separation to an arbitrary date entered in the record at TFDW to remove the record from the master data file. Essentially, these arbitrary entries are administrative corrections. These 3,063 duplicate records in the master data set are not included in the analysis.

The initial intent of the study was to use the “Separation Code” to identify observations exiting the active duty, but irregularities in the data required abandonment of this strategy and “Type Action Codes” were used instead. The separation code is a four-character code that describes the nature of the discharge from active duty. Missing or apparent typographical errors were resident in the original data. Thus, ascertaining the type of discharge was not possible within reasonable fidelity for analysis. Instead, the more general “Type Change Code” and “Action Date” was used. The Type Change Code is a two-character code that describes an enlisted loss to the active duty end strength. The Action Date is simply the date at which the Type Change Code occurred. The specific Type Change codes used are listed in Table 1.

Table 1. Type Change Codes

R1	Discharge
R3	Transfer to the IRR
RZ	Implied Loss

The code R1 and RZ signifies those Marines who did not complete their active duty service requirements and are counted as NEAS Losses in the study. Code R3 is assigned for a Marine who was transferred to the Individual Ready Reserve (IRR), signifying a satisfactory completion of service obligations. Extensive sampling of these codes, specifically their definition in the data set, show sufficiently high accuracy for further analysis. Over 90% of the Marines who were assigned Type change Code R3 were discharged on or about their EAS. Approximately 92% of the observations assigned the R1 or RZ code were discharged prior to their respective EAS, and are assumed to be NEAS Losses in the data set.

The master data set was coded for duration analysis within STATA. Specifically, and in accordance with the information contained in Chapter III, the data was structured for survival analysis. Enlisted Marines enter the origin state upon their respective Pay Entry Base Date (PEBD), where they become "at risk" in the analysis. Their analysis time (duration and episode) continues until they experience a failure or exit the origin state (enlistment). A failure occurs when a Marine is discharged from active duty (Code

R1 or RZ) for the purpose of this study, these are NEAS Losses. A Marine exits the origin state upon transfer to the IRR (Code R3) or on the date October 31, 2008. It is important to note that transfers to the IRR are not considered failures in the analysis. The assumption is that these Marines completed their required service and chose not to reenlist. These are considered EAS Losses in the analysis. In order to capture those attributes of NEAS losses (failures), EAS Losses are excluded, because too much emphasis would be placed on the periods of 48, 96, 134 months of service; potentially over estimating the effects of these times in the analysis. These are the times that four-year contracts expire and when the majority of Marines who do reenlist exit the service. The date October 31, 2008 signifies the end of the duration for those still on active duty, because it is the last date for which data is available. Therefore, these observations are right-censored. The analysis did not observe a failure on these observations, but can still use the fact that they did not fail in application to the population under study.

In summary, the duration time (analysis time) for each observation in the master data set, begins on the respective PEBD and concludes when either, the Marine is discharged, transferred to the IRR or the end of the study on October 31, 2008. There was an initial attempt in the study to include desertions in the study. Deserter status is for any Marine who is on an "Unauthorized Absence" status for a minimum of thirty days. However, these records could not be formed into the proper duration time-frame for analysis. In total, 2,916 deserter records were dropped from the master data set.

D. DESCRIPTIVE STATISTICS

The master data set contains 88 variables, which are listed in Table 2. Not all variables were used in the model; some variables are retained only for further analysis of specific duration of events. The variables beginning with "occfld…" refer to the Occupational Fields utilized by the Marine Corps to classify an enlistee's job description. Resident within the Occupational Fields (OccFld), are the Military Occupational

Specialties (MOS). For the purpose of the analysis, the Occupational Fields are used for forecasting attrition. Three combat OccFlds are restricted to males only OccFld03 Infantry, OccFld08 Artillery and OccFld18 Tanks and Assault Amphibious Vehicles.

Table 2. Variable Description

Variable Label		Frequency	Percentage
Unique Identifier per Marine		376,710	100
Observation relevant to analysis		375,070	-
Type Change Code		*See Table 3	*
Female		26,970	7.16
Ranks/Pay Grade			
Private/E1		114,354	30.36
Private First Class/E2		75,696	20.09
Lance Corporal/E3		107,165	28.45
Corporal/E4		65,383	17.36
Sergeant/E5		12,728	3.38
Staff Sergeant/E6		1,384	0.37
Citizenship			
Nationalized U.S. citizen		7,069	1.88
U.S. resident		670	0.18
U.S. Alien		11,967	3.18
U.S. citizen		356,932	94.75
Contract Terms			
Open Contract		53,100	14.10
Race			
American/Alaskan Indian		4,530	1.20
Asian		7,857	2.09
Black/African American		39,114	10.38
Hawaiian/Pacific Islander		2,150	0.57
White		290,532	77.12
Declined to comment on race		32,527	8.63
Education Level			
Less than 12 years education		7,346	1.95
Equal to 12 years education		338,485	89.85
Equal to 13 years education		4,699	1.25
Equal to 14 years education		4,341	1.15
Equal to 15 years education		1,309	0.35
Equal to 16 years education		3,447	0.92
Equal to 17 to 19 years education		17,083	4.53
Marital Status			
Married		118,982	31.58
Occupational Field			
Occfld01	Personnel & Administration	16,107	4.28
Occfld02	Intelligence	3,493	0.93
Occfld03	Infantry	77,717	20.63
Occfld04	Logistics	7,048	1.87
Occfld05	MAGTF Plans	514	0.14
Occfld06	Communications	23,848	6.33

Table 2 continued

Variable Label		Frequency	Percentage
Occfld08	Artillery	8,825	2.34
Occfld11	Utilities	6,440	1.71
Occfld13	Engineer, Equipment & Shore Party	17,591	4.67
Occfld18	Tank and Assault Amphibious Vehicle	5,697	1.51
Occfld21	Ground Ordnance Maintenance	8,668	2.30
Occfld23	Ammunition & Explosive Ordnance Disposal	3,648	0.97
Occfld25	Operational Communications	1,614	0.43
Occfld26	Signal Intelligence/Electronic Warfare	4,314	1.15
Occfld28	Data/Communications Maintenance	7,666	2.03
Occfld30	Supply Administration and Operations	14,518	3.85
Occfld31	Traffic Management	1,250	0.33
Occfld33	Food Service	5,620	1.49
Occfld34	Financial Management	2,585	0.69
Occfld35	Motor Transport	29,680	7.88
Occfld40	Data Systems	1,098	0.29
Occfld41	Morale, Welfare & Recreation	109	0.03
Occfld43	Public Affairs	746	0.20
Occfld44	Legal Services	965	0.26
Occfld46	Combat Camera	1,063	0.28
Occfld57	Chem, Bio, Radio & Nuclear Defense	1,850	0.49
Occfld58	Military Police and Corrections	8,255	2.19
Occfld59	Electronics Maintenance	2,801	0.74
Occfld606162	Aircraft Maintenance	25,724	6.83
Occfld6364	Avionics	11,528	3.06
Occfld65	Aviation Ordnance	5,265	1.40
Occfld66	Aviation Logistics	9,028	2.40
Occfld68	Meteorological & Oceanographic	542	0.14
Occfld70	Airfield Services	4,706	1.25
Occfld72	Air Spt Anti-air Warfare/Air Trfc Cntrl	3,763	1.00
Occfld73	Enlisted Flight Crew	477	0.13
Occfld8490	Enlisted B-Billet	550	0.15
Occfld99	General Marine	39,596	10.51

Source: created by author from master data set

The frequency and percentage of the total observations for the Type Change Codes are contained in the next table.

Table 3. Summary Statistic per Type Change Codes

R1	Discharge	96,601	43.46%
R3	Tr IRR	121,963	54.87%
RZ	Implied loss	3,719	1.67%
	Total	222,283	

E. METHODOLOGY

There are several graphical and statistical methods to examine the "fit" of the data to the five parametric models. These methods are not the sole determinant of model selection, but can provide the basis for matching the data and sound theory to survival analysis. This section of the study will briefly introduce the five parametric models, discuss the graphical and statistical methods employed for the model used in this study and conclude with the model selection.

1. Five Parametric Models

Five parametric models that can be used in survival analysis are the exponential, Weibull, Gompertz, log-logistic, and log-normal. The exponential model assumes the baseline hazard (or risk of failure) is constant for all observations. Hence, failure rates are independent of process or “lacks memory” of past durations. The Weibull Model is an extension of the exponential model that allows the hazard function to monotonically increase, decrease, or remain constant. It is most suitable for data that displays monotone hazard rates (Cleves, Gutierrez, Gould, Marchenko, 2008). The exponential and Weibull are unique amongst the parametric models in that both models can be fitted with either the Proportional Hazard (PH) or Accelerated Failure Time (AFT) metric. The Gompertz model is suitable for exponentially increasing or decreasing hazard rates. The model only has the PH interpretation available. The Log-Logistic and Log-Normal models are similar in computation to the LOGIT and PROBIT models and assume log-logistic distribution implying a nonmonotonic relationship between the transition rate and episode duration. (Cleves, Gutierrez, Gould, Marchenko, 2008). The models do not have a PH interpretation, but allow for changes in the direction of the hazard rate. A logical beginning step to parametric model selection is an examination of the product limit estimator (Kaplan-Meier).

2. Kaplan-Meier Survival Estimate

The Kaplan-Meier survival estimate is a nonparametric calculation of estimated cumulative survival function of all observations in the data. The product-limit method is derived by calculating a risk set at every interval an event occurred. In this study, a failure event is defined as an NEAS loss (Type Change Code R1 or RZ). The graph depicts an initial decrease in the cumulative survival rates at approximately $t = 5$, signifying an initial increased cumulative hazard rate. Then the survival rate declines at a slower rate until approximately $t = 90$ when another drastic drop in survival rate (increased cumulative hazard rate) is experienced. Nonetheless, this graph depicts a monotonic decreasing survival rate indicating a monotonically decreasing cumulative survival rate.

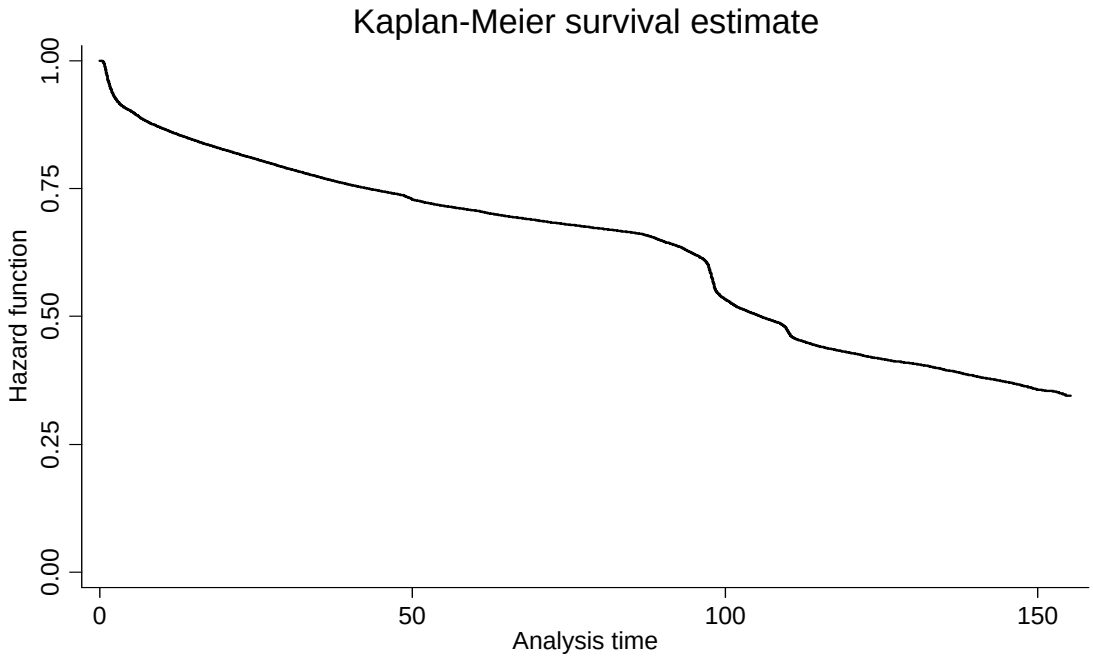

Figure 1. Kaplan Meier Survival Estimate

3. Pseudoresidual Graph of Model Suitability

The development of the graph involves specifying the Cox-Snell residuals as the variable for time against the cumulative hazard function as the log of the Kaplan-Meier

estimates (Cleves, Gutierrez, Gould, Marchenko, 2008). If the model "fits" the data, the Cox-Snell residuals will have an exponential distribution and the set of *pseudoresiduals* will cluster near a straight line passing through the origin with a slope of one. Figure 2 is obtained by first estimating the Cox-Snell residuals and then the integrated hazard rate is estimated utilizing the Kaplan-Meier estimates. The y-axis is the log of Kaplan-Meier estimates and the x-axis is the computed Cox-Snell residuals for each parametric model assessed in the figure.

Examining the below graph, it is easy to determine that the exponential and Weibull models are not appropriately suited for the data because the estimated Kaplan-Meier (or pseudoresiduals) estimates of the integrated hazard function do not follow an exponential distribution. The closer the pseudoresiduals follow along the Cox-Snell residual, the better the data "fits" the specified model. The psuedoresiduls for the Weibull and expontneial models are plotted far from the linear line indicating a weak data "fit." The Log-Logistic and Log-Normal perform somewhat better. The Gompertz Model seems to be best suited amongst the models for the data. The departure of the estimated residuals from the linear line is a normal occurrence. This "flaring" off is primarily due to fewer observed failures towards the end of the analysis time as fewer observations "survive." In the Gompertz Model the departure seems to be less drastic.

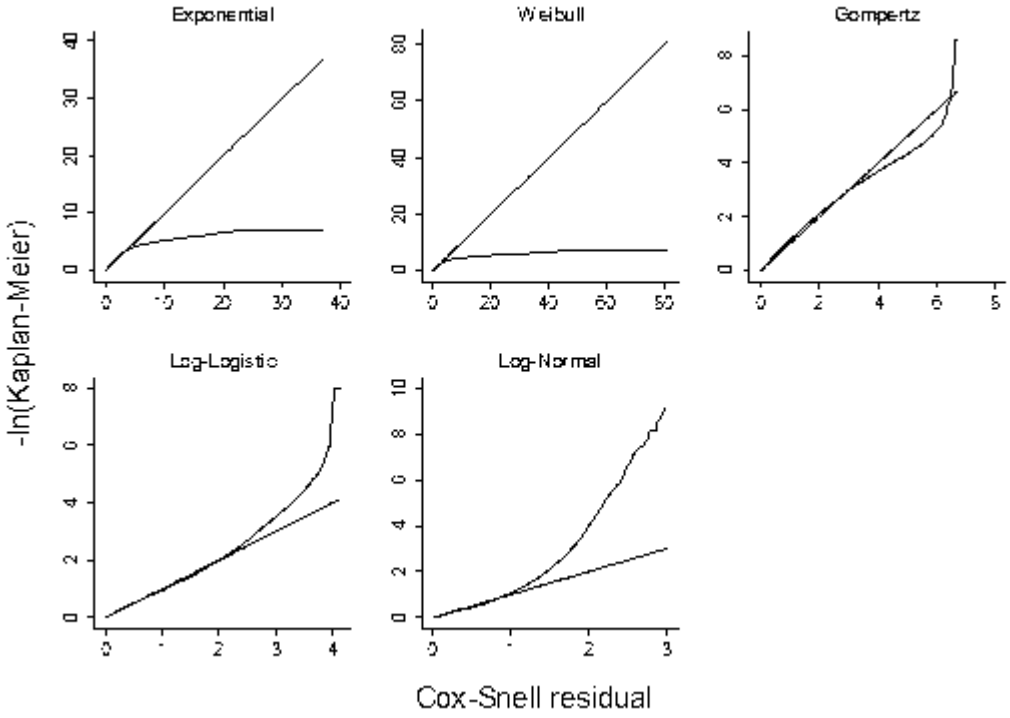

Figure 2. Graphical Representation of Pseudoresiduals

4. Akaike Information Criterion (AIC)

The aim of the AIC is to penalize each parametric model's log likelihood function for each covariate estimated. The AIC criterion is given by

$$AIC = -2(\log L) + 2(c + p + 1),$$

where c is the number of covariates in the model and p is the number of structural parameters in the model. In Table 4, each model's log likelihood and AIC are estimated. Based on selecting the model with the lowest AIC, the preferred model for this data is the Gompertz Model.

Table 4. AIC values for Parametric Model Selection

Model	Log likelihood(null)	Log likelihood (model)	AIC
Exponential	-232756.6	-231084.3	462296.5
Weibull	-232565	-230428.1	460986.2
Gompertz	*-230404.9*	*-227057.3*	*454244.5*
Log-Logistic	-241655.9	-240518.9	481167.9
Log-Normal	-259442.2	-258113.7	516357.3

F. DISCUSSION

The Gompertz Model seems to be the most appropriate model for the data set as demonstrated through a graphical and statistical test of model "fitness." However, it must be emphasized that these tests are not a "goodness of fit" test, but rather a model selection process that evaluates a model's assumptions on the distribution of hazard rates. As theorized in this study, hazard rates decrease as time elapses, which is an appropriate assumption for the Gompertz Model. The theory to support the use of this distribution assumption rests in the Human Capital Theory. Specifically, as an employee incurs more experience within an organization, additional personal developmental investments decline, and thus job-exits decrease. The determination of the author not to include EAS losses as failures in the model allows for a direct estimation of NEAS loss rates without over-estimating the transition rates of enlisted Marines who choose not to reenlist. Typically, these times would be every 48 months.

The next chapter will provide an analysis of the data estimated with the Gompertz Model. The chapter will begin with an estimation of the data without covariates, expand to a model with covariates and then test specific influences of the individual covariates on the hazard rate. The chapter will conclude with a description of survival and hazard rates of specific covariate values in preparation for the development of a forecasting model for NEAS losses.

V. PARAMETRIC MODEL RESULTS

The Gompertz Model was chosen for the analysis of Marine Corps enlisted attrition rates for the period 1996 to 2008 by means of the methodology described in the previous chapter. The analysis will begin by estimating and interpreting a model without covariates. Further expansion of the model will be provided in the subsequent sections of this chapter.

A. GOMPERTZ MODEL WITHOUT COVARIATES

The hypothesis for the study is that transition rates (hazard rates) will decline at a monotonic rate as time increases.

Table 5. Gompertz Model without Covariates

t	Coefficient	Std. Err.	z	P>z	[95% Conf.	Interval]
constant	-4.553	.0048	-940	0.000	-4.563	-4.544
gamma	-.0160	.0002	-99	0.000	-.0163	-.0157

Source: generated by author in STATA

As expected, the transition rate as estimated in STATA by the gamma coefficient is negative and significant. Thus, the transition rate for the observations is decreasing as enlistment time increases. Comparisons can be made between varying times in service levels. For example, the estimated parameters are $\hat{c} = \hat{\gamma}_0 = -.0160$, $\hat{b} = \exp(-4.553) = .0105$. An enlistee's initial transition rate $(\hat{r}(0) = .0105)$ compared to that of an enlisted Marine with one year of service (or 12 months) $(\hat{r}(12) = .0105\exp(-.0160*12) = .009)$ demonstrates a 14% decrease in an expected hazard rate as time in service increases by 12 months for enlisted Marines. The survivor function

$$G(t) = \exp\left\{\frac{-b}{c}(\exp(c^t - 1)\right\}, \tag{5.1}$$

defined by $G(M) = .5$, is

$$G(\widehat{M}) = \exp\left\{\frac{-0.011}{-.0160}(\exp(-.0160^{\widehat{M}} - 1)\right\}$$

The probability a service member is still enlisted, say, at four years (48 months) can be calculated as,

$$G(48) = \exp\left\{\frac{-0.011}{-.0160}(\exp(-.0160^{48} - 1)\right\}$$

However, the coefficients in Table 5 do not include other explanatory factors that influence transition rates. A model without covariates assumes that there is no heterogeneity amongst individuals (Blossfield, Golsch, Rohwer). The assumption that transition rates decrease with time because of the accumulation of MOS-specific skills and returns to investment in the form of promotions, higher pay, family benefits, etc. could be misleading without the inclusion of substantiated covariates. Therefore, a second model is estimated utilizing the covariates outlined in Chapter IV.

B. GOMPERTZ MODEL WITH COVARIATES

The model's covariates (Table 6) are linked to the *b* parameter. Furthermore, the model takes as a baseline for white males, at the rank of Private, with the Occupational Field 9900, designated as a United States citizen, with an education level equal to 12 years, who are serving on a guaranteed contract.

The value of the log likelihood for this model is -201720.91 with 56 parameters compared to the log likelihood of the model without covariates -374750.53. Therefore, the model with covariates provides a better description of the hazard rate.

All variables, with the exception of (races American Indian, black, and race declined comment, education level equal to 13 years, and Occupational Field 6500 (Aviation Ordnance)) are significant at the 95% confidence level. A negative coefficient signifies a decreasing effect on the hazard rate (i.e. lower probability of attrition), while a positive coefficient reflects an increasing effect on the hazard function. For example, the coefficient for female is positive and significant at the 5% level. Therefore, females attrite at a higher rate than males. The hazard function for a specific set of covariates can be calculated per (3.12) in Chapter III.

Table 6. Gompertz Model with Covariates linked to the *b* parameter

Variable	Coefficient	S.E.	z	P-value
Female	.294	.011	26.34	0.00
PFC\E2	-1.475	.001	-151.65	0.00
LCpl\E3	-3.285	.014	-233.68	0.00
Cpl\E4	-3.263	.014	-236.87	0.00
Sgt\E5	-4.259	.025	-167.87	0.00
SSgt\E6	-7.043	.172	-40.87	0.00
Enlist Age^2	.001	.000	43.67	0.00
U.S. Nationalized	- .214	.030	-7.18	0.00
U.S. Resident	- .250	.081	-3.08	0.02
U.S. Alien	- .213	.021	-10.23	0.00
Open Contract	.057	.009	6.23	0.00
American/Alaskan Indian	- .278	.031	-9.10	0.00
Asian	- .159	.027	-5.89	0.00
Black/African American	- .003	.010	-0.25	0.80
Hawaiian/Pacific Islander	- .510	.061	-8.34	0.00
Declined race	- .012	.012	-1.03	0.30
Less than 12 years education	.236	.020	12.00	0.00
Equal to 13 years education	- .058	.032	-1.84	0.07
Equal to 14 years education	- .142	.037	-3.87	0.00
Equal to 15 years education	.586	.063	9.23	0.00
Equal to 16 years education	.180	.033	5.52	0.00
Married	.113	.010	11.61	0.00
Number of Dependents	- .260	.001	-45.46	0.00
Occfld01	-2.710	.020	-134.33	0.00
Occfld02	-2.658	.048	-55.38	0.00
Occfld03	-2.463	.010	-238.07	0.00
Occfld04	-2.721	.030	-89.70	0.00

Table 6 continued

Variable	Coefficient	S.E.	z	P-value
Occfld05	-2.427	.111	-21.79	0.00
Occfld06	-2.740	.019	-147.66	0.00
Occfld08	-2.600	.026	-99.42	0.00

Occfld11	-2.682	.030	-89.39	0.00
Occfld13	-2.699	.020	-132.90	0.00
Occfld18	-2.461	.031	-80.31	0.00
Occfld21	-2.665	.027	-97.18	0.00
Occfld23	-2.647	.040	-65.52	0.00
Occfld25	-1.837	.035	-52.38	0.00
Occfld26	-2.524	.039	-64.97	0.00
Occfld28	-2.547	.027	-94.48	0.00
Occfld30	-2.712	.021	-129.25	0.00
Occfld31	-2.867	.069	-41.79	0.00
Occfld33	-2.624	.029	-89.34	0.00
Occfld34	-2.631	.044	-59.47	0.00
Occfld35	-2.665	.015	-172.68	0.00
Occfld40	-2.001	.054	-37.28	0.00
Occfld41	-2.524	.209	-12.08	0.00
Occfld43	-2.705	.090	-30.07	0.00
Occfld44	-2.780	.079	-35.01	0.00
Occfld46	-2.786	.079	-35.19	0.00
Occfld57	-2.675	.062	-43.03	0.00
Occfld58	-2.771	.030	-93.54	0.00
Occfld59	-2.587	.046	-57.79	0.00
Occfld606162	-2.768	.017	-159.31	0.00
Occfld6364	-2.756	.024	-113.77	0.00
Occfld65	.059	.055	1.06	0.29
Occfld66	-2.839	.042	-67.69	0.00
Occfld68	-2.828	.102	-27.64	0.00
Occfld70	-2.699	.036	-74.07	0.00
Occfld72	-2.499	.038	-66.50	0.00
Occfld73	-2.567	.118	-21.30	0.00
Occfld8490	-3.022	.165	-18.37	0.00
Intercept	-2.142	.015	-147.39	0.00
Gamma	.028	.000	172.83	0.00

Source: created by author from master data set in STATA

1. Sample Hazard Function Calculation

In order to demonstrate the probability a failure event will occur given a specific time of service (using Table 6 estimates) an example of the hazard function (3.12) is computed for a particular set of covariates.

$$h(36 \mid x_j) = \exp(.0282 * 36)\exp(-2.142 + (1 * \beta_{female}) + (1 * \beta_{occfld\,01})$$
$$= \exp(.0282 * 36)\exp(-2.142 + (1 * (.294)) + (1 * (-2.707))$$
$$= .0290$$

The hazard function depicts a .029 probability that a white female, in the Occfld 0100 (Administration), will become a NEAS loss in the 36th month of service, given she

survived to 36 months of service. This compares to the estimate for a white male, with the same covariate valuess and time of service, of .021, a as shown below. Therefore, these females have a 26% higher hazard rate than males with identical covariates.

$$h(36 \mid x_j) = \exp(.0282 * 36)\exp(-2.142 + (1 * \beta_{male}) + (1 * \beta_{occfld\,01})$$
$$= \exp(.0282 * 36)\exp(-2.142 + (1 * (0)) + (1 * (-2.707))$$
$$= .0216$$

The effects of the coefficients on the hazard function for a combination of estimated covariant values is proportional to changes to the hazard function and the frequency of covariant values that experienced an event (Cleves, Gould, Gutierrez, Marchenko, 2008). In the above demonstration, the estimated value for the coefficient *female,* was positive and significant and increased the probability a "failure event" will occur when t = 36 when compared to a male with identical combination of covariates. The example of calculating the hazard function for any set or combination can easily be expanded to include all variables estimated in Table 6.

2. Positive *c*-Parameter

It is the premise of this study that the hazard rate would decrease monotonically as time passes; the Marine Corps would experience fewer NEAS losses as enlistees accumulated time in service. A change in the hazard with covariates linked to the *b* parameter from the model estimated without covariates is the sign of the *gamma* (γ) coefficient (or *c*-parameter). This appears to be in violation of the hypothesized declining rate of transitions as the shape parameter is now positive, .0281, indicating an increasing hazard rate. However, a positive *c* parameter indicates an increasing hazard function. The apparent discrepancy between the theory of a declining hazard rate and the contrary positive *gamma* coefficient can be explained by the influence of different sub-populations within the data.

The data is comprised of a multitude of sub-populations. The various Occupational Fields, gender, and race are examples. The combinations of these attributes

further divide the data into more sub-populations. An individual hazard rate can be calculated for each sub-population. In addition, within this data, the proportions of observations within these sub-populations are constantly changing as observations, fail, exit, or enter the analysis representing the daily accessions and exits of enlisted Marines to and from active duty. The enlisted end-strength is in a constant and daily flux. Therefore, the *gamma* (γ) coefficient is a combination of the hazard rates of the various sub-populations. This does not mean the *gamma* (γ) coefficient can be dismissed as false, but it does emphasize the need to construct hazard rates by the sub-populations of interest.

The estimated effects of the covariates on the hazard rate remains the same for the population and the sub-populations, but as sub-populations are formed from different combinations of covariates, the multiplicative effect of those covariates will change the hazard and survival rate for each sub-population. (This concept was demonstrated in Section B.1 of Chapter V.)

The next few sections of this chapter will demonstrate the varying hazard rates per sub-populations.

3. Gompertz Model with Covariates Linked to the *b* and *c* Parameter

The Gompertz Model can be expanded to estimate the effects individual time-constant coefficients have on the hazard rate as duration time increases. A negative coefficient signifies a decreasing effect of the shape of the hazard function, while a positive coefficient has an increasing effect of the hazard function. The purpose of such a model is to determine if those covariates that serve as initial predictors of attrition behavior actually decline in significance as enlistment duration increases (Cleves, Gould, Gutierrez, Marchenko, 2008).

The coefficient estimates listed under the "gamma" section in Table 7 differ from those listed in the top half the table in that they estimate the effects of the covariate on the hazard rate over the duration of enlistment. For example, the “occfld01” coefficient is negative signifying a reduction in the hazard rate compared to the baseline. However, the

"occfld01" coefficient listed under the "gamma" section is positive and significant at the 95% confidence level. This means that the covariate has an increasing positive effect on the hazard rate as duration time increases. The coefficients U.S. Resident, American Indian, Asian and Hawaiian/Pacific Islander are not statistically significant at the 95% confidence level. The interpretation of the "gamma" coefficients is difficult to apply to forecasting and requires calculation beyond the scope of this study. The point of this description is to demonstrate how individual covariates influence the hazard rate over time and that coefficients that are originally negative can eventually have a positive effect on the hazard rate over time. Thus, a positive *c* parameter $(\gamma = .0282)$ that indicates an increasing hazard rate is a result of the cumulative hazard rates of sub-populations within the study and not necessarily in opposition to the study's theory of decreasing hazard rates. The aggregated grouped effects of sub-population's hazard rates are causing the shape of the hazard function to be position. Thus, some sub-populations are experiencing increasing attrition rates which may be occurring to events not explained by the Human Capital Theory.

Table 7. Gompertz Model with Covariates linked to the *b* and *c* parameter

Variable	Coefficient	S.E.	z	P-value
Female	.185	.013	13.87	0.00
PFC\E2	-1.421	.010	-143.64	0.00
LCpl\E3	-3.269	.014	-229.82	0.00
Cpl\E4	-3.422	.014	-237.39	0.00
Sgt\E5	-4.742	.027	-178.50	0.00
SSgt\E6	-7.572	.173	-43.69	0.00
Enlist Age^2	.002	.000	41.14	0.00
U.S. Nationalized	- .160	.040	-3.96	0.00
U.S. Resident	- .300	.120	-2.50	0.01

Table 7 continued

Variable	Coefficient	S.E.	z	P-value
U.S. Alien	- .161	.027	-6.03	0.00
Open Contract	.082	.012	7.05	0.00
American/Alaskan Indian	- .254	.042	-6.09	0.00
Asian	- .109	.035	-3.14	0.00
Black/African American	.025	.013	1.87	0.06
Hawaiian/Pacific Islander	- .429	.081	-5.31	0.00
Declined race	.038	.015	2.51	0.01
Less than 12 years education	.301	.020	15.29	0.00
Equal to 13 years education	- .023	.032	-0.71	0.48
Equal to 14 years education	- .217	.037	-5.87	0.00
Equal to 15 years education	.646	.064	10.17	0.00

Equal to 16 years education	.433	.033	13.21	0.00
Married	.089	.001	9.19	0.00
Occfld01	-3.339	.031	-108.03	0.00
Occfld02	-3.781	.093	-40.52	0.00
Occfld03	-3.032	.014	-208.14	0.00
Occfld04	-3.508	.049	-70.99	0.00
Occfld05	-3.394	.194	-17.47	0.00
Occfld06	-3.689	.030	-122.57	0.00
Occfld08	-3.670	.041	-81.67	0.00
Occfld11	-3.258	.047	-69.94	0.00
Occfld13	-3.313	.031	-107.76	0.00
Occfld18	-3.110	.047	-66.61	0.00
Occfld21	-3.431	.043	-79.87	0.00
Occfld23	-3.485	.067	-52.07	0.00
Occfld25	-2.238	.058	-38.40	0.00
Occfld26	-3.458	.067	-51.40	0.00
Occfld28	-3.261	.045	-73.00	0.00
Occfld30	-3.223	.032	-101.96	0.00
Occfld31	-3.329	.110	-30.14	0.00
Occfld33	-3.117	.046	-68.38	0.00
Occfld34	-3.099	.068	-45.30	0.00
Occfld35	-3.266	.023	-140.60	0.00
Occfld40	-2.804	.105	-26.79	0.00
Occfld41	-4.500	.776	-6.44	0.00
Occfld43	-3.285	.150	-21.94	0.00
Occfld44	-3.427	.127	-26.99	0.00
Occfld46	-3.651	.134	-27.21	0.00
Occfld57	-3.492	.099	-35.39	0.00
Occfld58	-3.438	.045	-75.83	0.00
Occfld59	-3.129	.073	-42.82	0.00
Occfld606162	-3.424	.027	-126.95	0.00
Occfld6364	-3.395	.040	-84.06	0.00
Occfld65	- .394	.090	-4.37	0.00
Occfld66	-3.392	.066	-51.33	0.00
Occfld68	-3.571	.184	-19.45	0.00
Occfld70	-3.536	.059	-59.91	0.00
Occfld72	-3.177	.061	-52.50	0.00
Occfld73	-3.328	.207	-16.10	0.00
Occfld8490	-4.073	.244	-16.71	0.00
Intercept	-1.848	.017	-106.27	0.00

Table 7 continued

Variable	**Coefficient**	**S.E.**	**z**	**P-value**
Gamma				
Female	.005	.000	12.30	0.00
Enlist Age^2	- .000	.000	-11.63	0.00
U.S. Nationalized	- .002	.000	-2.49	0.00
U.S. Resident	.002	.002	0.81	0.42
U.S. Alien	- .002	.001	-2.78	0.01
Open Contract	- .003	.000	-8.83	0.00
American/Alaskan Indian	.001	.001	0.59	0.55
Asian	- .002	.000	-1.81	0.70
Black/African American	- .002	.000	-6.52	0.00
Hawaiian/Pacific Islander	.000	.002	0.24	0.81

Declined race	- .002	.000	-4.69	0.00
Occfld01	.070	.001	67.06	0.00
Occfld02	.078	.002	51.56	0.00
Occfld03	.069	.001	75.54	0.00
Occfld04	.074	.001	60.21	0.00
Occfld05	.077	.003	25.26	0.00
Occfld06	.077	.001	78.67	0.00
Occfld08	.075	.001	62.11	0.00
Occfld11	.069	.001	51.35	0.00
Occfld13	.070	.001	65.68	0.00
Occfld18	.072	.001	54.26	0.00
Occfld21	.074	.001	61.54	0.00
Occfld23	.075	.001	50.91	0.00
Occfld25	.064	.002	29.14	0.00
Occfld26	.077	.001	53.06	0.00
Occfld28	.072	.001	62.85	0.00
Occfld30	.067	.001	61.77	0.00
Occfld31	.066	.003	23.65	0.00
Occfld33	.067	.001	50.16	0.00
Occfld34	.065	.002	41.31	0.00
Occfld35	.070	.001	70.59	0.00
Occfld40	.078	.003	24.53	0.00
Occfld41	.088	.007	11.99	0.00
Occfld43	.067	.002	27.43	0.00
Occfld44	.070	.003	26.67	0.00
Occfld46	.076	.003	28.54	0.00
Occfld57	.075	.002	37.86	0.00
Occfld58	.071	.001	60.09	0.00
Occfld59	.068	.002	37.86	0.00
Occfld606162	.070	.001	71.52	0.00
Occfld6364	.070	.001	60.87	0.00
Occfld65	.011	.002	6.80	0.00
Occfld66	.068	.002	45.24	0.00
Occfld68	.071	.003	22.56	0.00
Occfld70	.076	.001	53.40	0.00
Occfld72	.071	.001	50.80	0.00
Occfld73	.073	.003	21.14	0.00
Occfld8490	.085	.003	25.43	0.00
Intercept	- .028	.001	-28.06	0.00

Source: created by author from master data set in STATA

4. Gompertz Hazard Rate

The estimated hazard rate is depicted in Figure 3. As expected with the positive gamma coefficient ($\gamma = .0282$), the rate is increasing monotonically as time increases.

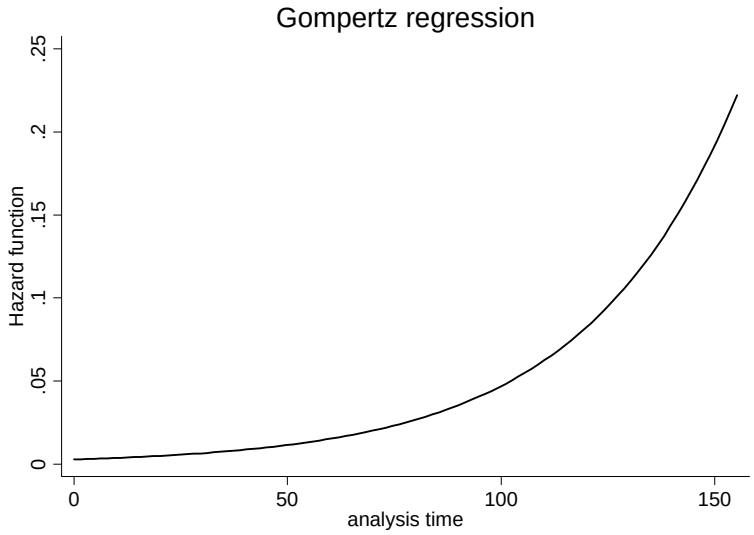

Figure 3. Gompertz Hazard Rate

5. Hazard Function by Rank

Figure 4 depicts the different hazard rates by rank. The rank of Private First Class has the highest probability of experiencing a failure event as the duration of enlistment increases. This is not a surprising result as most PFC's are promoted to the rank of Lance Corporal within their first year of active service. The Marines at the rank of PFC beyond the first year are typically in that rank as a result of poor performance or conduct and have been reduced from a higher pay grade as a result. The rank of Staff Sergeant, on average, is achieved at approximately the eighth year of active service. These Marines have demonstrated competence and proficiency within their MOS and thus the probability of becoming an NEAS loss for discipline or performance issues are reduced. Also, the hazard rate for the rank of Staff Sergeant declines as duration increases. Notably, the ranks of Sergeant and Corporal experience increase hazard rates as enlistment duration increases. This may be caused by poor or sub-average performance, resulting in those Marines being “passed over and not being promoted to the next rank in unison with their respective accession cohort. The increasing hazard rates experienced by

these ranks signify attrition behavior that is the result of reduction of personal investments to continued service and the decision to exit the military rather than incurring additional "costs" of enlisted service.

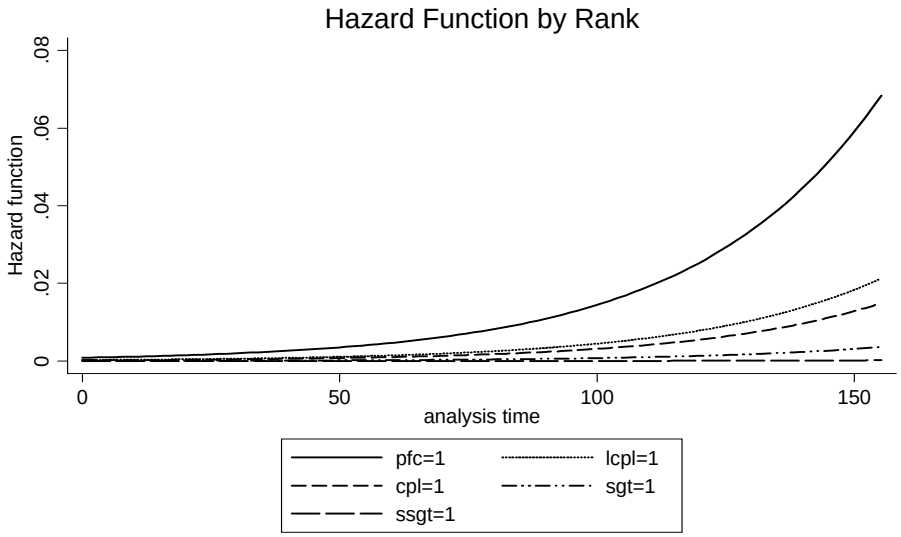

Figure 4. Hazard Functions by Rank

6. The Use of the Gompertz Model with Covariates Linked to the *b* Parameter as a Forecasting Model

Developing a model to forecast attrition from the data presented in Table 6 is dependent on the researcher's ability to construct sub-populations from the data. As discussed, each sub-population within a larger population will have its own hazard rate. That sub-population's specific hazard rate is influenced by the covariates that are in the model. If large numbers of sub-populations were constructed, the hazard and survival rates would still be sensitive to the proportion of observations for each covariate, the quantity of observations entering and exiting the sub-population, and the frequency of "events" observed with the duration under study. However, a simpler model can be developed that would be within the resources a military planner will have in order to build future forecasting models. The next section will construct a simple model including only the covariates of the various ranks. The model will employ interaction terms for the

individual ranks within the 0300 Infantry Occupational Field. The purpose of this simplified model is to demonstrate that restricting the hazard function to a few covariates drastically diminishes the models capability to estimate the probability of failure for a specified time *t*.

C. OCCUPATIONAL FIELD 0300 ATTRITION FORECAST MODEL

The data in the forecasting model is from the master data set used throughout the study. The interaction terms were created to study the attrition behavior of each rank within the 0300 Occupational Field. The control group is the rank of Private. The model estimates and corresponding graphs of the hazard and survivor functions are provided.

Table 8. Gompertz Model results for Occupational Field 0300

Variable	Coefficient	Standard Error	z	P-value
PFC/E2	- .510	.016	-31.03	0.00
LCpl/E3	-1.912	.027	-71.15	0.00
Cpl/E4	-1.406	.027	-51.69	0.00
Sgt/E5	-1.341	.066	-20.46	0.00
SSgt/E6	-3.921	.707	- 5.55	0.00
Intercept	-4.401	.005	-896.95	0.00
Gamma	- .016	.000	-98.86	0.00

Source: created by the author in STATA

1. Descriptive Statistics

As depicted in Table 8, all covariates are significant at the 95% confidence level and have an estimated negative effect on the hazard rate. The negative gamma coefficient signifies a decreasing hazard rate, which supports this study's assumption of declining attrition rates with time. Figure 5 graphically represents the associated hazard rate. The graph depicts Marines in the 0300 Occupational Field have the highest attrition rate in the first 48 months of service. At approximately 50 months of service the probability of attrition is ,004, which typically is for a Marine that is now classified as an Intermediate-term Marine.

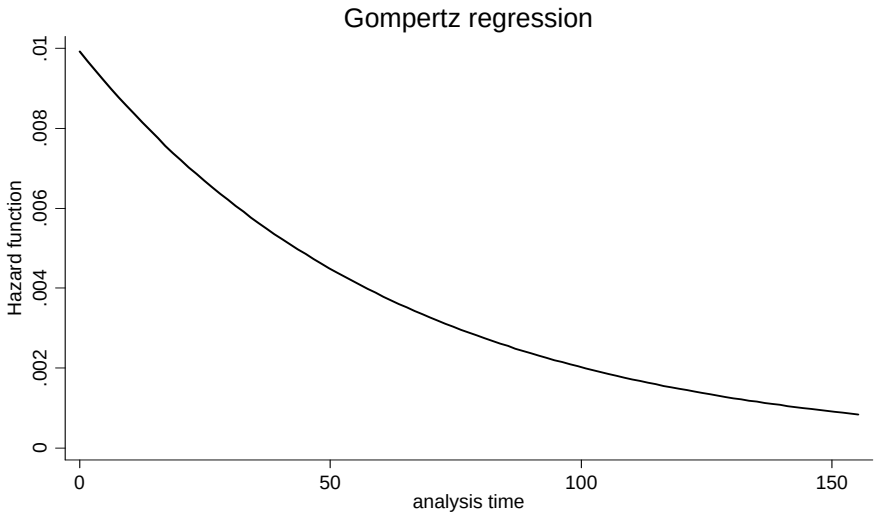

Figure 5. Graph of Occupational Field 0300 Overall Hazard Rate

2. Occupational Field 0300 Hazard Rates by Rank

The graphed hazard rates by rank within the 0300 Occfld differ from the hazard rates from the larger covariate model in Table 6. As expected, the rank of Private First Class experiences the highest hazard rate early in the analysis time. The subsequent hazard rates of the other ranks, diminishes as time elapses, which is synonymous with promotions to next rank as the duration increases. The ranks of Corporal and Sergeant have nearly identical hazard rates.

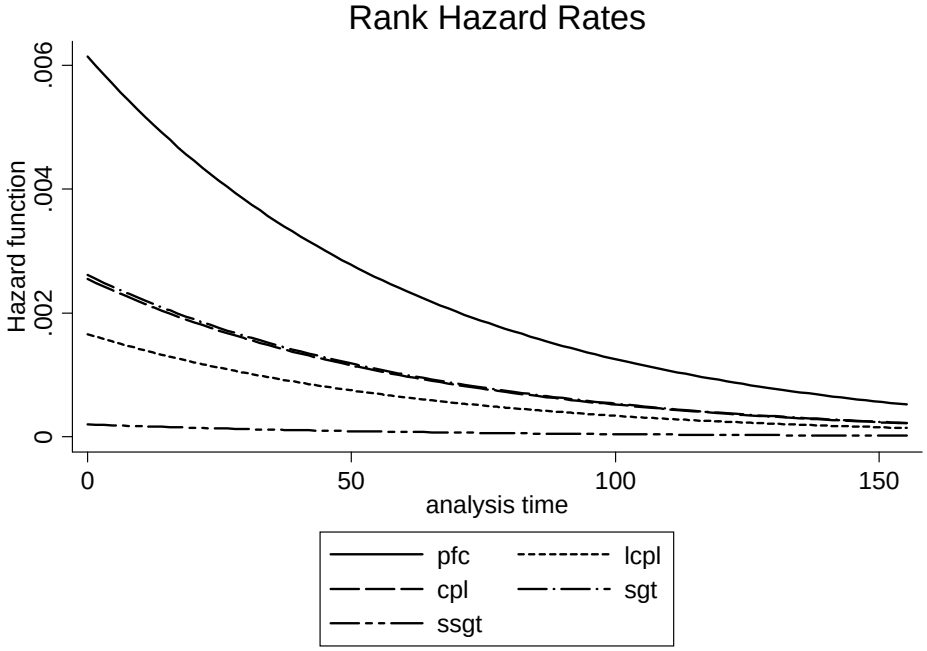

Figure 6. Graph of OccFld 0300 Hazard Rates by Rank

3. Why the Differences in Shape of the Hazard Functions?

The differences in the slope of the estimated hazard functions between the model represented in Table 6 and the simplified model in Table 8 lie in the number of covariates used in the estimations. Transition rate models are sensitive to the set of covariates used to evaluate the model and a change in the values of a set of covariates used in the model can change the shape of the transition rate. This dependency is due to the function of the residuals estimated in the model (Blossfeld, Golsch, Rohwer, 2007). In the model depicted in Table 6, the estimated effect of Corporal is the estimated effect the rank of Corporal has on the entire population. The residual is calculated by combining all the Corporals throughout the sample who had failed and measuring the duration time for the failure event. What is not captured is the varying probabilities of failure for the rank of Corporal within each occupational field.

The increasing hazard function, depicted in Figure 3, is estimated using the mean values of 56 covariates. However, the simplified model that depicts a decreasing hazard rate utilizes only five interaction terms to estimate the hazard function. The exclusion of other explanatory variables reduces the descriptive ability of modeling attrition behavior and demonstrates that a Marine's rank and Occfld alone are not adequate to forecast the probability of attrition. There are other factors, which affect the probability of attrition besides those factors currently used by the Marine Corps, and each set of these factors will affect the attrition rates differently within sub-populations. Forecasting models that utilize only rank, occupational field and service duration can be misleading and non-responsive to changes within sub-population attrition rates. Consequently, the requirement to weight the historical data becomes necessary to compensate for the inefficiency of an averaging technique. The application of survival analysis for each occupational field within the Marine Corps with all possible combinations of covariates significant to characterizing the probability of becoming a NEAS loss will improve the efficiency of an attrition forecast model.

The table in Appendix C provides a frequency distribution of Type Change Codes per fiscal year from the master data set. An examination of this table reveals varying failure rates (Code R1 and RZ) for each occupation field. Furthermore, the failure rates steadily increase for each successive fiscal year differently within each occupational field. The number of failures drastically increases for the majority of the fields from the FY 2003 to FY 2006. The data presented in Appendix C demonstrates that each occupational field has different frequencies of failure and that these frequencies do not change evenly across all fields. Therefore, the assumption of a steady-state modeling technique is inadequate.

THIS PAGE INTENTIONALLY LEFT BLANK

VI. SUMMARY AND RECOMMENDATIONS

The effects a set of covariates have on the probability of attrition, as determined by the hazard rate, can change when the set of covariates used in the model are altered. The number of covariates in a model of transition rates can influence the shape of the hazard function and the estimated coefficient values due to unobserved heterogeneity. The effect a set of covariates will have on the sample transition is more than likely different for each sub-population. For example, not all Corporals in the Marine Corps attrite at the same rate. Within each occupational field attrition, rates can vary due to different performance, educational, and ability requirements in order to be successful. Marines perceive the costs for perceived future benefits differently within each occupational field. Therefore, each occupational field has a different transition rate. Given the varying hazard rates per occupational field with gender, race, citizenship, and any other set of covariates, a transition rate model that estimates the hazard rate for an entire population may suffer because it does not consider these effects. Exponential smoothing models suffer from this inefficiency even more so. Future forecasts of attrition are dependent on previous attrition rates. Yet in a dynamic environment such as the enlisted force, Marines are not influenced by the attrition rates of fellow Marines. They are influenced by the constant weighing of the perceived costs to the perceived benefits of military service. The important distinction a set of covariates have on different sub-populations would be lost in a sample averaging or weighted-average technique that attempts to aggregate effect over a whole population. Eventually, the model would become inefficient in capturing changes in attrition rates and weighted averages would likely be employed to correct forecasting errors. Modeling covariates by sub-population and estimating the effect variables have within each population will provide a better estimate of the hazard rate per sub-group and variables that are contributing to attrition behavior. For example, in Table 7, the gamma estimates for each occupational field are different and significant. This indicates that each occupational affects the attrition rate differently over time. Therefore, the Marines within those

occupational fields will have varying probabilities of becoming an NEAS loss in the future. These differing attrition rates (or hazard functions) become compounded when additional covariates are added to the model.

A. PROPOSED ENLISTED ATTRITION FORECASTING MODEL

The hazard rate introduced in formula (3.12) should be employed for each occupational field in the Marine Corps. Within each occupational field, all possible combinations of covariates listed in Table 6 should be calculated utilizing this formula. For example, the computation of the hazard rate for Occfld 3000 Supply, would be

$$h(t \mid x_j) = \exp(.0282 * t)\exp(\beta_0 + x_j\beta_{Gender} + x_j\beta_{rank} + x_j\beta_{enlistage} + x_j\beta_{citizenship} + x_j\beta_{opencontract} + x_j\beta_{race} + x_j\beta_{education} + x_j\beta_{Occfld\,30} + x_j\beta_{maritalstatus}),$$

where time *t* is defined by the planner and x_j takes on the value of the covariate. The formula is for each gender, rank, citizenship, race and education level within an occupation field. The survivor rate for occupational field is calculated with the formula (3.14).

B. SUMMARY

The study attempted to answer two primary research questions through the application of survival analysis.

1. What Causal Factors and Individual Characteristics Attribute to Attrition Behavior?

The covariates included in the survival analysis were chosen because previous attrition studies have substantiated their relevance to modeling attrition behavior. This study verified the significance of these covariates. However, this study found that the covariates have varying effects amongst different sub-populations. In order to accurately model the probability of an NEAS loss occurring, the effects of the covariate estimates should be modeled per sub-population rather than for the entire sample of the population.

The study did not attempt to discover additional variables that could be used to forecast enlisted attrition as the main intent was to apply survival analysis to an already reliable set of covariates.

2. Can a More Efficient and Effective Forecasting Model be Developed to either Replace or Complement Current Forecasting Methods for NEAS Losses?

The study could not compare the results of the survival analysis to the current method of forecasting NEAS Losses. The only data available from the current forecasting method was the actual and forecasted aggregated amounts for fiscal years 2003 to 2009. There are two key differences between the aggregated amounts available and the data used in the study. The first difference is the data in this study only contained enlisted Marines with a maximum of 12.5 years of service. The study only described those Marines who entered into the service on or after January 1, 1996 and concluded the analysis on October 31, 2008. The characteristics of the enlisted Marines that comprise the aggregated amounts per fiscal year are unknown. The second key difference is the goal of this study to forecast the probability of attrition by occupation field per month. Hence, the data was structured to provide this depth of analysis. The aggregated data available could not be separated by occupational field nor by month in which attrition occurred. Therefore, direct comparison was impossible. However, in comparison to the current forecasting method of exponential smoothing (Hattiangadi, Kimble, Lambert, Quester, CNA, 2005), this study found that the use of survival analysis could be beneficial to not only forecast attrition, but also to provide a descriptive assessment of attrition rates amongst occupation fields without loss of information due to averaging or weighting probabilities.

C. RECOMMENDATIONS

The following recommendations are provided in order to further enhance the survival analysis model used in this study and to provide more tools for military planners in forecasting NEAS losses.

1. Use Separation Category Codes

This study attempted to use Separation Category codes to define when an enlisted Marine failed and became a NEAS loss. Unfortunately, the Separation category codes resident in the TFDW database were not reliable. The Type Change Codes were used as an alternate means to identify attritions. These codes do not describe the nature a Marine's discharge as descriptively as the Separation Category codes. It is possible that the use of these less descriptive Type Change codes, may have erroneously determined a Marine to be a failure. A thorough review of the Separation Category codes should be conducted. When these category codes are determined to be accurate, the model in this study should be re-estimated utilizing Separation Category codes as the event of failure.

2. Forecasting by Military Occupational Specialty

The occupational field was used as a covariate to model the hazard rates of attrition. It is likely that further analysis of attrition rates by MOS will provide even greater clarity in modeling the probability of attrition within an occupational field. There are MOSs within an occupational field that are rank-specific. For example, 0369 within the 0300 Occfld is only for the rank of Staff Sergeant and above; 0193 in the 0100 Occfld is also only for the rank of Staff Sergeant and above. Modeling the hazard rates for specific MOSs will reduce the aggregated hazard rates experienced in modeling the entire occupational field.

3. Current Events Variables

The inclusion of variables that contain data on current operations the Marine Corps is conducting can provide greater modeling of attrition rates. Including in the model information on the number and duration of deployments in support of the Global War on Terrorism can provide estimates on how attrition rates are affected by successive deployments.

APPENDIX A: FY 2008 MARINE CORPS END-STRENGTH

This is personnel end-strength for the Marine Corps in Fiscal Year 2008.

• Personnel (AD)	180,000
• Personnel (FTS)	2,261
• Personnel (SELRES)	37,339
• Uniformed Personnel	219,600
• Civilian Personnel	18,322
• Total Personnel	237,922

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B: FREQUENCY OF RANK BY OCCUPATIONAL FIELD

	Frequency of Rank by Occupational Field													
	0100	*0200*	***0300***	*0400*	*0500*	*0600*	***0800***	*1100*	*1300*	***1800***	*2100*	*2300*	*2500*	*2600*
RANK														
E0	1,196	128	**8,106**	518	28	1,612	**822**	602	1,383	**603**	771	254	280	163
E1	1,841	200	**10,080**	767	43	2,462	**1,036**	798	1,964	**714**	1,094	359	289	445
E2	4,052	454	**22,139**	1,456	101	4,248	**1,993**	1,716	3,746	**1,138**	1,927	732	321	860
E3	5,829	893	**23,844**	2,615	181	8,427	**2,920**	2,136	6,816	**1,873**	2,900	1,018	491	1,200
E4	2,391	1,132	**11,997**	1,247	116	5,836	**1,691**	1,031	3,124	**1,102**	1,654	917	233	1,317
E5	812	586	**1,404**	391	40	1,120	**327**	142	515	**241**	295	311		272
E6	104	100	**147**	54	5	143	**36**	15	43	**26**	27	57		57
Total	16,107	3,493	77,717	7,048	514	23,848	8,825	6,440	17,591	5,697	8,668	3,648	1,614	4,314
	4.28%	0.93%	20.63%	1.87%	0.14%	6.33%	2.34%	1.71%	4.67%	1.51%	2.30%	0.97%	0.43%	1.15%

	Frequency of Rank by Occupational Field													
	2800	*3000*	*3100*	*3300*	*3400*	*3500*	*4000*	*4100*	*4300*	*4400*	*4600*	*5500*	*5700*	*5800*
RANK														
E0	589	1,272	100	595	210	2,816	98		36	66	48	26	140	494
E1	849	1,728	135	741	308	3,700	104		63	100	111	93	177	954
E2	1,260	3,213	358	1,182	591	6,439	174		168	289	307	254	448	2,080
E3	2,271	5,634	463	2,023	925	10,989	442	6	251	311	378	517	659	3,166
E4	2,143	2,110	153	934	398	5,024	278	65	203	148	169	584	331	1,235
E5	505	524	39	142	132	680	2	34	23	49	44	178	87	306
E6	49	37	2	3	21	32		4	2	2	6	48	8	20
Total	7,666	14,518	1,250	5,620	2,585	29,680	1,098	109	746	965	1,063	1,609	1,850	8,255
	2.03%	3.85%	0.33%	1.49%	0.69%	7.88%	0.29%	0.03%	0.20%	0.26%	0.28%	0.43%	0.49%	2.19%

	Frequency of Rank by Occupational Field									
	5900	*60/61/62*	*63/64*	*6500*	*6600*	*6800*	*7000*	*7300*	*80/95*	*9900*
RANK										
E0	344	1,992	588	278	625	38	370	12	125	23,035
E1	383	2,683	890	454	854	46	522	34	214	12,683
E2	528	5,014	2,546	1,258	1,994	82	1,052	64	87	1970
E3	682	7,331	3,518	1,944	3,241	183	1,692	171	9	45
E4	711	7,120	3,312	1,041	1,790	140	844	152	30	1,717
E5	133	1,470	620	270	474	45	211	43	37	251
E6	20	114	54	20	50	8	15	1	48	
Total	2,801	25,724	11,528	5,265	9,028	542	4,706	477	550	39,701
	0.74%	6.83%	3.06%	1.40%	2.40%	0.14%	1.25%	0.13%	0.15%	10.54%

RANK	***Total***	**%**
E0	114,354	30.36%
E1	54,160	14.38%
E2	76,328	20.26%
E3	107,324	28.49%
E4	65,766	17.46%
E5	12,799	3.40%
E6	1,389	0.37%
	376,710	
	100%	

APPENDIX C: FREQUENCY OF TYPE CHANGE CODE BY OCCUPATIONAL FIELD

	Frequency of Type Change Code by Occupational Field											
OccFld	0100			0200			0300			0400		
Sep code	R1	R3	RZ	R1	R3	RZ	R1	R3	RZ	R1	R3	RZ
FY												
1996	12						56	1				2
1997	61	1		7			453	6				23
1998	180	3		12	1		840	3	4	54		
1999	169	5	14	14		1	1109	28	113	66	1	5
2000	202	635	25	23	72	2	1,218	3,523	225	75	256	11
2001	231	726	58	31	143	5	1025	3233	352	63	238	22
2002	219	746	1	20	115		937	2768	1	96	241	
2003	210	697		19	128		645	3026	3	85	266	1
2004	231	694		46	38		1196	3261	5	114	354	
2005	262	705		70	149		1795	3092	2	126	341	1
2006	337	698	3	82	105		2278	3789	1	148	321	
2007	308	682		58	110		1781	3487	4	137	347	1
2008	273	509	19	57	63	1	1675	2723	33	114	230	2
Total	2,695	6,101	120	439	924	9	15,008	28,940	743	1,078	2,595	68
	8,916			1,372			44,691			3,741		

	Frequency of Type Change Code by Occupational Field											
OccFld	1300			1800			2100			2300		
Sep code	R1	R3	RZ	R1	R3	RZ	R1	R3	RZ	R1	R3	RZ
FY												
1996	7	1		1			3			3		
1997	66	1		33	1		40	1		19		
1998	156	1		48	2	1	57	1	2	31	1	1
1999	161	6	12	61		4	74	1	5	26	1	2
2000	224	796	21	84	240	20	120	293	16	37	136	9
2001	178	796	62	77	269	23	122	351	25	42	116	11
2002	229	719		111	196		110	331		45	119	
2003	187	833		64	262		92	372	1	35	93	
2004	308	886		98	293	1	98	386		59	132	
2005	308	947	2	127	269		163	428		77	152	
2006	262	1117	2	135	272	1	172	468		84	174	
2007	283	935	1	112	203		171	383		78	152	
2008	231	618	11	109	169	3	135	287	7	69	108	2
Total	2,600	7,656	111	1,060	2,176	53	1,357	3,302	56	605	1,184	25
	10,367			3,289			4,715			1,814		

Frequency of Type Change Code by Occupational Field												
OccFld	0500			0600			0800			1100		
Sep code	R1	R3	RZ	R1	R3	RZ	R1	R3	RZ	R1	R3	RZ
FY												
1996							4			5		
1997							44			41		
1998							64			76		
1999	3		1				100	2	10	66	3	8
2000	6	10		33	147	5	96	331	25	93	261	12
2001	5	16		217	886	72	150	331	40	82	275	32
2002	7	16		261	806		129	267	1	127	287	1
2003	7	16		261	1226	1	116	426	1	91	314	
2004	3	17		374	1202	2	148	423		126	304	1
2005	10	20		527	1289	1	157	411	2	108	327	
2006	13	38		611	1236	1	175	412		114	368	
2007	19	10		535	1397	1	157	446		101	352	2
2008	7	8		452	739	19	136	300	7	83	217	2
Total	80	151	1	3,271	8,928	102	1,476	3,349	86	1,113	2,708	58
	232			12,301			4,911			3,879		

Frequency of Type Change Code by Occupational Field												
OccFld	2500			2600			2800			3000		
Sep code	R1	R3	RZ	R1	R3	RZ	R1	R3	RZ	R1	R3	RZ
FY												
1996	3	1								5		2
1997	93			14	3		38	1		81	1	
1998	185	2		38	1		60			135	1	
1999	220	3	13	53	2	8	91	3	4	146	6	10
2000	193	726	24	57	66	2	112	150	14	177	598	31
2001	21	1	12	56	151	5	107	278	40	176	811	71
2002	57	1		50	128		191	324		267	618	
2003	7			58	196		111	436		199	697	
2004	7			51	179		109	327	1	207	613	1
2005	2			103	144		126	399		262	717	1
2006	1			76	216		163	346		248	642	1
2007				63	155		177	221	1	260	687	1
2008	4			53	130		144	160	3	264	383	13
Total	793	734	49	672	1,371	15	1,429	2,645	63	2,427	5,774	131
	1,576			2,058			4,137			8,332		

Frequency of Type Change Code by Occupational Field												
OccFld	3100			3300			3400			3500		
Sep code	R1	R3	RZ	R1	R3	RZ	R1	R3	RZ	R1	R3	RZ
FY												
1996				2						12		
1997	6			43			11			163	2	
1998	7			73	2		21	1		270	3	2
1999	12		2	109		6	37	2	1	279	4	21
2000	16	68	1	108	307	24	38	76	14	363	1,389	68
2001	16	38	4	118	293	39	36	137	12	427	1346	101
2002	19	62		105	303		56	91		491	1265	
2003	14	57		101	408		44	110		387	1373	
2004	23	60		96	212		53	97		456	1352	2
2005	22	57		94	240		52	113		473	1455	1
2006	28	46		106	242		50	85		619	1566	1
2007	18	65	1	103	189		55	129		488	1604	1
2008	22	36	4	101	100	5	41	65	4	448	832	15
Total	203	489	12	1,159	2,296	74	494	906	31	4,876	12,191	212
	704			3,529			1,431			17,279		

Frequency of Type Change Code by Occupational Field												
OccFld	4600			5700			5800			5900		
Sep code	R1	R3	RZ	R1	R3	RZ	R1	R3	RZ	R1	R3	RZ
FY												
1996	1						3			4		
1997	6			7			28			14		
1998	9			8			52	2		32		
1999	9		1	12	1		74	1	6	35	1	
2000	9	50	1	12	59	1	81	383	7	30	9	5
2001	22	64	4	17	51	2	85	357	32	46	108	11
2002	14	48		15	61		105	325		32	106	
2003	12	61		18	72		92	404		33	148	1
2004	13	60		27	70		103	417		23	141	
2005	12	49		29	83		125	408		44	162	
2006	16	62		32	98		130	377		55	125	
2007	22	38		45	104		140	371		70	95	
2008	10	13		37	83		144	99	2	75	85	1
Total	155	445	6	259	682	3	1,162	3,144	47	493	980	18
	606			944			4,353			1,491		

Frequency of Type Change Code by Occupational Field												
OccFld	4000			4100			4300			4400		
Sep code	R1	R3	RZ	R1	R3	RZ	R1	R3	RZ	R1	R3	RZ
FY												
1996								1				
1997	5						1			3		
1998	28						4			8	1	
1999	51		3				8			10	1	1
2000	30	155	5				5	22	1	16	47	2
2001	83	228	11				10	19	1	12	30	2
2002	106	282					2	23		10	41	
2003	7	31			1		4	16		14	43	
2004	9	45		3			13	24		14	38	
2005	11	5		5	5		12	24		12	43	
2006				3	1		21	35		21	41	
2007				7	2		21	38		14	35	
2008				5	1		22	22		20	25	1
Total	330	746	19	23	10	0	123	224	2	154	345	6
	1,095			33			349			505		

Frequency of Type Change Code by Occupational Field												
OccFld	60/61/62			63/64			6500			6600		
Sep code	R1	R3	RZ	R1	R3	RZ	R1	R3	RZ	R1	R3	RZ
FY												
1996	14			8			2			5		
1997	82			64			22			44		
1998	189	3		91	3		52	2		75	3	
1999	289	5	9	148		3	42	2	4	71	3	4
2000	303	284	37	138	76	9	57	144	7	90	320	15
2001	323	862	96	171	412	46	49	286	17	91	385	31
2002	316	1061		133	600		43	223		91	356	
2003	266	1223		125	557	2	46	242		82	394	1
2004	270	1200	1	146	577	1	64	242		111	410	
2005	413	1039	2	177	613		84	320		155	480	
2006	475	1151	1	239	643		90	275		159	416	
2007	411	951	1	188	460		105	208		181	350	2
2008	406	809	7	175	387	3	78	186	1	137	278	3
Total	3,757	8,588	154	1,803	4,328	64	734	2,130	29	1,292	3,395	56
	12,499			6,195			2,893			4,743		

	Frequency of Type Change Code by Occupational Field											
OccFld	**6800**			**7000**			**7300**			**80/95**		
Sep code	R1	R3	RZ	R1	R3	RZ	R1	R3	RZ	R1	R3	RZ
FY												
1996												
1997	4			21								
1998	3			27		1	4			1		
1999	3			42	1	6	6		1			
2000	4	19		61	205	13	5	4	2			
2001	10	25	1	61	203	26	8	13				
2002	9	7		61	185		7	8			1	
2003	7	23		44	247	1	5	34		2		
2004	11	25		69	231	1	6	16		3		
2005	16	28		82	206		10	9		4		
2006	6	28		103	228		8	32		6	2	
2007	12	22		83	269		10	25		6	1	
2008	9	18	1	78	139			13		15	2	
Total	94	195	2	732	1,914	48	69	154	3	37	6	0
	291			2,694			226			43		

OccFld	**9900**		
Sep code	R1	R3	RZ
FY			
1996	1,682	2	
1997	3,110	6	
1998	5,107	4	3
1999	3,550	2	217
2000	4,111	9	120
2001	3,755	3	249
2002	3,590	4	7
2003	3,510	1	17
2004	3,133	25	255
2005	2,889	28	119
2006	2,572	19	10
2007	364	12	2
2008	15		
Total	37,388	115	999
	38,502		

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Blossfeld, H.P., Golsch, K., Rohwer, G., Event history analysis with Stata. Lawrence ErlBaum Associates. Mahwah, NJ. (2007).

Cleves, M.A., Gould, W.W., Gutierrez, R.G., Marchenko, Y.U., An introduction to survival analysis using Stata. Second Edition. A Stata Press Publication. StataCorp LP. College Station, TX. (2008).

Hattiangadi, A.U., Kimble, T.H., Lambert, W.B. Quester, A.O. (2005). Center for Naval Analysis. *Endstrength: Forecasting Marine Corps losses final report.* Alexandria, VA: CAN.

Hawes, E.A., (1990). *An application of survival analysis methods to the study of Marine Corps enlisted attrition.* United States Postgraduate School.

Orrick, S.C., (2008). *Forecasting Marine Corps enlisted losses.* United States Postgraduate School.

Read, R.R., (1997). *The use of survival analysis in the prediction of attrition in large scale personnel flow models*. United States Postgraduate School.

Rubiano, L.E.O., (1993). *An analysis of the Coast Guard enlisted attrition*. United States Postgraduate School.

Total Force Data Warehouse. https://tfdw-web.manpower.usmc.mil. (Accessed November 2008).

Wenger, J.W., Hodari, A.K., (2004). Center for Naval Analysis. *Predictors of attrition: attitudes, behaviors, and educational characteristics.* Alexandria, VA: CAN.

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. Marine Corps Representative
   Naval Postgraduate School
   Monterey, California

4. Director, Training and Education
   MCCDC, Code C46
   Quantico, Virginia

5. Director, Marine Corps Research Center
   MCCDC, Code C40RC
   Quantico, Virginia

6. Marine Corps Tactical Systems Support Activity
   (Attn: Operations Officer)
   Camp Pendleton, California

7. Samuel Buttrey
   Naval Postgraduate School
   Monterey, California

8. Jeremy Arkes
   Naval Postgraduate School
   Monterey, California

9. Jeremy Hall
   Marine Corps Recruiting Command
   Quantico, Virginia