THE EFFECTS OF EQUIPMENT AGE ON MISSION-CRITICAL FAILURE RATES

A Study of M1 Tanks

ERIC PELTZ

LISA COLABELLA

BRIAN WILLIAMS

PATRICIA M. BOREN

Prepared for the
United States Army

RAND
ARROYO CENTER

Approved for public release; distribution unlimited

The research described in this report was sponsored by the United States Army under Contract No. DASW01-01-C-0003.

Library of Congress Cataloging-in-Publication Data

The effects of equipment age on mission-critical failure rates : a study of M1 tanks / Eric Peltz ... [et al.].
p. cm.
"MR-1789."
Includes bibliographical references.
ISBN 0-8330-3493-6 (pbk.)
1. M1 (Tank)—Maintenance and repair. 2. United States—Armed Forces—Operational readiness. I. Peltz, Eric, 1968–

UG446.5.E35 2004
623.7'4752—dc22

2004010090

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

RAND® is a registered trademark.

Photo Courtesy of U.S. Army by Sgt. Derek Gaines.
Cover design by Peter Soriano

© Copyright 2004 RAND Corporation

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from RAND.

Published 2004 by the RAND Corporation
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
201 North Craig Street, Suite 202, Pittsburgh, PA 15213-1516
RAND URL: http://www.rand.org/
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

PREFACE

Due to budget limits, the service lives of many Army weapon systems are being extended. There is a widespread belief that the resulting increases in fleet ages are—or will be—creating readiness and cost problems. The Army has therefore launched a program to rebuild and selectively upgrade fielded systems, many of which currently exceed fleet age targets. This program is known as recapitalization (RECAP).

However, initial recapitalization plans combined with investments in new equipment have strained the Army budget, and complete RECAP of current aged fleets has been found unaffordable. Thus, the Office of the Deputy Chief of Staff, G-8 (Programs), the Office of the Deputy Chief of Staff, G-3 (Operations and Plans), the Office of the Deputy Chief of Staff, G-4 (Logistics), the Office of the Assistant Secretary of the Army for Acquisition, Logistics, and Technology (OASA[ALT]), and the Army Materiel Command (AMC) have been examining which systems (both type and portion of the fleet) should be recapitalized and defining what that renewal process should involve (the extent of work for each "overhaul"). Accordingly, OASA(ALT) is sponsoring RAND Arroyo Center research on how equipment age affects readiness and resource requirements, to aid analyses in support of RECAP decisions.

This report describes one component of this study: an assessment of the relationship between tank age and the mission-critical failure rate for the M1 Abrams tank. Findings should be of interest to resource planners, logistics analysts, and weapon system analysts.

This research has been conducted in the Military Logistics Program of RAND Arroyo Center, a federally funded research and development center sponsored by the United States Army.

For more information on RAND Arroyo Center, contact the Director of Operations (telephone 310-393-0411, extension 6419; FAX 310-451-6952; e-mail Marcy_Agmon@rand.org), or visit the Arroyo Center's Web site at *http://www.rand.org/ard/*.

CONTENTS

Preface iii

Figures vii

Tables xi

Summary xiii

Acknowledgments xxi

Glossary xxiii

Chapter One
INTRODUCTION 1

Chapter Two
METHODOLOGY 9
Data Sources 9
Sample Characteristics 9
Measures 11
Tank Study Variables 11
System Failures 11
Age 14
Accumulated Usage During the Study Period 17
Updays 17
Location 18
Subsystem Study Variables 18
Data Refinement Techniques 19
Exclusion of Observations 19
Imputation 19
Analyses 22

Tank Study Analysis 22
Subsystem Study Analysis 24

Chapter Three
RESULTS 27
Tank Study Results 27
Subsystem Study Results 30
Interpretation of Subsystem Results 40
Rebuild Versus Upgrade Candidates 47
The Link Between Age-Failure Relationships and Part Prices 48
Sensitivity Analysis Results 52
Alternative Imputation Approach 52
Additional Control Variable for Odometer Resets 58
Alternative Regression Techniques in the Tank Study . . 59
Alternative Regression Techniques in the Subsystem Study 60

Chapter Four
IMPLICATIONS 69

Appendix

A. GENERAL DESCRIPTIONS OF STATISTICS USED 73

B. DISTRIBUTION OF FAILURE DATA 77

C. CROSS-VALIDATION OF TANK STUDY MODEL 83

D. PLOTS OF SUBSYSTEMS' PREDICTED MEAN FAILURES BY AGE AND USAGE 87

Bibliography 97

FIGURES

1.1. Hazard Functions with Pronounced Wear-out Regions . . . 3
1.2. Hazard Functions Without Pronounced Wear-out Regions . . . 4
2.1. Number of Months of Usage Data per Tank by Location . . . 12
2.2. Distribution of Tank Age by Location . . . 12
2.3. M1A1 Age Histogram . . . 13
2.4. M1A2 Age Histogram . . . 13
2.5. Distribution of Tank Usage by Location . . . 14
2.6. Distribution of Initial M1A1 Odometer Readings by Age . . . 16
2.7. Distribution of Initial M1A2 Odometer Readings by Age . . . 16
3.1. Predicted Mean Failures (over 180 days) by Tank Age . 29
3.2. Predicted Mean Failures by Age at Location 1, with 95 percent Confidence Bars (180 days, usage = 375 km) . . . 29
3.3. Predicted Mean Failures (over 180 days) by Tank Usage . . . 30
3.4. Predicted Mean Failures of Second-Tier Subsystems by Age (Location 1, 180 days) . . . 41
3.5. Predicted Mean Fire Control Failures by Age for the M1A1s, M1A2s, and Combination of M1A1s and M1A2s (Location 1, 180 days) . . . 43
3.6. Predicted Mean Failures of Second-tier Subsystems by Usage (Location 1, 180 days) . . . 45

3.7. Total Parts Demand (during Study Period) per Subsystem by Age 46
3.8. Parts Demand per Part Type by Age 47
3.9. Predicted Mean Part Failures Versus Tank Age (Location 1, 180 days) 52
3.10. Predicted Mean Failures by Age for Hydraulic and Power Train Subsystems, Based on Multiple Imputation Models (Location 1, 180 days) 57
3.11. Predicted Mean Failures by Usage for Hydraulic and Power Train Subsystems, Based on Multiple Imputation Models (Location 1, 180 days) 57
3.12. Confidence Interval Width by Age for Multiple Imputation and Mean Imputation Overall Tank Study Model 59
3.13. GAM Predicted Mean Failures of Chassis, Fire Control, Hardware, and Power Train Subsystems by Age (Location 1, 180 days) 63
3.14. 95 Percent Confidence Bands for Power Train GAM Curve 63
3.15. 95 Percent Confidence Bands for Chassis GAM Curve 64
3.16. 95 Percent Confidence Bands for Fire Control GAM Curve 64
3.17. 95 Percent Confidence Bands for Power Train GAM Curve, with Extrapolation Past Age 15 65
3.18. 95 Percent Confidence Bands for Chassis GAM Curve, with Extrapolation Past Age 1 65
3.19. 95 Percent Confidence Bands for Fire Control GAM Curve, with Extrapolation Past Age 15 66
3.20. Alternate Plot of Predicted Mean Failures of Second-tier Subsystems by Age (Location 1, 180 days) 66
B.1. Illustration of Failure Data Overdispersion 77
B.2. Comparison of Battalion Failure Distributions and Poisson Distribution in 1st Cavalry Division 79
B.3. Comparison of Battalion Failure Distributions and Poisson Distribution in 4th Infantry Division 79
B.4. Comparison of Battalion Failure Distributions and Poisson Distribution in 1st Infantry and 1st Armor Divisions: Fort Riley 80

B.5. Comparison of Battalion Failure Distributions and Poisson Distribution in 2nd Infantry Division 80
B.6. Comparison of Battalion Failure Distributions and Poisson Distribution in 3rd Infantry Division 81
B.7. Comparison of Battalion Failure Distributions and Poisson Distribution in 1st Infantry and 1st Armor Divisions: Europe 82
D.1. Predicted Mean Hull Failures by Tank Age 88
D.2. Predicted Mean Hull Failures by Tank Usage 88
D.3. Predicted Mean Chassis Failures by Tank Age 89
D.4. Predicted Mean Chassis Failures by Tank Usage 89
D.5. Predicted Mean Power Train Failures by Tank Age . . . 90
D.6. Predicted Mean Power Train Failures by Tank Usage . 90
D.7. Predicted Mean Turret Failures by Tank Age 91
D.8. Predicted Mean Turret Failures by Tank Usage 91
D.9. Predicted Mean Gun Failures by Tank Age 92
D.10. Predicted Mean Gun Failures by Tank Usage 92
D.11. Predicted Mean Fire Control Failures by Tank Age . . . 93
D.12. Predicted Mean Fire Control Failures by Tank Usage . 93
D.13. Predicted Mean Electrical Failures by Tank Age 94
D.14. Predicted Mean Electrical Failures by Tank Usage . . . 94
D.15. Predicted Mean Hardware Failures by Tank Age 95
D.16. Predicted Mean Hardware Failures by Tank Usage . . . 95
D.17. Predicted Mean Hydraulic Failures by Tank Age 96
D.18. Predicted Mean Hydraulic Failures by Tank Usage . . . 96

TABLES

2.1. Number of M1 Tanks in Sample by Location and Division 10
3.1. Negative Binomial Regression of Tank Failures on Age, Usage, and Location Variables (N = 1,567) 28
3.2. Summary of Subsystem Age and Usage Effects (Terms in Final Model) 31
3.3. Negative Binomial Regression of Hull Failures on Age, Usage, and Location Variables (N = 1,480) 32
3.4. Negative Binomial Regression of Turret Failures on Age, Usage, and Location Variables (N = 1,480) 33
3.5. Negative Binomial Regression of Chassis Failures on Age, Usage, and Location Variables (N = 1,480) 34
3.6. Negative Binomial Regression of Electrical Failures on Age, Usage, and Location Variables (N = 1,480) 35
3.7. Negative Binomial Regression of Fire Control Failures on Age, Usage, and Location Variables (N = 1,480) . . . 36
3.8. Negative Binomial Regression of Hardware Failures on Age, Usage, and Location Variables (N = 1,480) . . . 37
3.9. Negative Binomial Regression of Power Train Failures on Age, Usage, and Location Variables (N = 1,480) . . . 38
3.10. Negative Binomial Regression of Hydraulic Failures on Age, Usage, and Location Variables (N = 1,480) . . . 39
3.11. Negative Binomial Regression of Gun Failures on Age, Usage, and Location Variables (N = 1,480) 40
3.12. Negative Binomial Regression of Low-Priced Part Failures on Age, Usage, and Location Variables (N = 1,480) 48

3.13. Negative Binomial Regression of Medium-Priced Part Failures on Age, Usage, and Location Variables (N = 1,480) 49
3.14. Negative Binomial Regression of High-Priced Part Failures on Age, Usage, and Location Variables (N = 1,480) 50
3.15. Negative Binomial Regression of Very-High-Priced Part Failures on Age, Usage, and Location Variables (N = 1,480) 51
3.16. Negative Binomial Regression of Hull Failures on Age, Usage, and Location Variables (N = 1,480), with Multiple Imputation Approach 54
3.17. Negative Binomial Regression of Power Train Failures on Age, Usage, and Location Variables (N = 1,480), with Multiple Imputation Approach 55
3.18. Negative Binomial Regression of Hydraulic Failures on Age, Usage, and Location Variables (N = 1,480), with Multiple Imputation Approach 56
C.1. PRESS Statistics for Models in Cross-Validation Study 85

SUMMARY

> Without a significant effort to increase resources devoted to recapitalization of weapon systems, the force structure will not only continue to age but, perhaps more significantly, become operationally and technologically obsolete.
>
> *Quadrennial Defense Review Report*, 2001, p. 47

Aging equipment has become a key concern of Army leaders striving to maintain high operational readiness. Army leaders anticipate that equipment age will pose a continually increasing challenge over the lengthy period in which current equipment is expected to remain in the Army's fleet, anticipated until about 2030 in some cases, as it develops and fully fields its next generation of forces termed the future force. In response, the Army has initiated a recapitalization (RECAP) program to rebuild and/or upgrade selected systems, such that combat capabilities are maintained and maintenance costs are kept affordable.[1] To date, the Army plans to rebuild or upgrade 17 systems—including the M1 Abrams, M2 Bradley Fighting Vehicle, M88 Recovery Vehicle, and other systems that are expected to remain in the inventory for the next 15 to 20 years (Brownlee and Keane, 2002; Army Recapitalization Management, 2003). These modernization plans continue to evolve, however. To help determine the scale of

[1]Rebuilding consists of efforts to restore a system to like-new condition. Upgrading is adding components (or replacing old components with new ones) that increase a system's warfighting capability (Gourley, 2001).

RECAP required to maintain the desired level of operational readiness capability, and to facilitate RECAP program design, statistical analyses of the relationship between age and Army equipment failures are needed.

This report describes a RAND Arroyo Center study, sponsored by the Office of the Assistant Secretary of the Army for Acquisition, Logistics, and Technology (OASA[ALT]), on the impact of age on the M1 Abrams mission-critical failure rate. The M1 Abrams is of particular interest because it is often considered the centerpiece of the Army's heavy ground forces, because it has a high average fleet age that will continue to advance, and because it is scheduled to remain a key part of the force for as many as 30 more years. Consequently, it has been one of the key systems being targeted by the RECAP program.

RESEARCH QUESTIONS

The four research questions in this study are as follows:

1. What is the relationship between age and the M1 Abrams mission-critical failure rate?[2]
2. How is the M1 failure rate related to other factors, such as usage and location-specific factors?
3. If there is a significant relationship between age and the M1 Abrams mission-critical failure rate, which of the various M1 subsystems and individual parts generate this relationship, and to what degree do they do so?
4. How can statistical models of such relationships inform RECAP decisions and planning?

Subsequent studies will address the same questions for other critical Army ground systems.

---

[2]A mission-critical failure is defined in this study as one that makes an item not mission capable, as indicated by the item's technical manual and subsequently reported by its owning unit. Mission-critical failures are also called *deadlining events.*

STUDY DESIGN

To address the research questions, we conducted two "substudies" at the individual tank level of analysis. In substudy 1 (the *Tank Study*) we assessed the impact of age, location, and usage on individual tank failures. In substudy 2 (the *Subsystem Study*) we assessed the impact of tank age, location, and usage on tank subsystem failures. Subsystems included actual subsystems, such as fire control, as well as part technology groups, such as basic hardware. As an additional segment of the Subsystem Study, we assessed the impact of tank age, location, and usage on tank *part* failures, where parts (subsystem components such as transmissions and pumps) were placed into price categories ranging from low to very high. The samples for the two substudies included 1,567 tanks and 1,480 tanks, respectively,[3] which includes the tanks in the Army's six active heavy divisions distributed across what we categorized as six different geographic areas: Germany, Georgia, Korea, Kansas, Colorado, and Texas.

The age, location, usage, and failure data came from Army maintenance database extracts from April 1999 through January 2001.[4] Our primary analytical techniques included imputation of missing data and negative binomial regression. It should be noted that data on the maintenance history of each tank prior to the beginning of the study period were not available. Hence, only the ages of the tanks themselves, and not their components, were known.

RESULTS

The study provides preliminary support for the hypothesis that age is a significant predictor of M1 failures, as are usage and location. The models suggest that M1 age has a positive log-linear effect that is consistent with a 5 ± 2 percent increase in tank failures per year of age. For a given location, usage, and time period, this equates to a 14-

[3]The sample in the Subsystem Study included fewer tanks because we lacked complete data on 4th Infantry Division M1A2 subsystem failures.

[4]Failure data came from Standard Army Maintenance System-2 (SAMS-2) aho01i and aho02i files archived in the Integrated Logistics Analysis Program (ILAP), and age, location, and usage data come from The Army Maintenance Management System (TAMMS) Equipment Database (TEDB). Unit price data for tank parts came from Federal Logistics (FedLog) database extracts for January 2003.

year-old tank having about double the expected failures of a new tank. This conclusion only applies to the first 14 years of a tank's life, since most tanks in the study were 14 years old or younger at the time of the study. (Only two tanks in the dataset were 15 years old.) The conclusion may or may not hold beyond that point; this can be determined as the Army's tank fleet continues to age. In the meantime, it is risky to assume that this compound annual growth rate in failures applies beyond the age range of our dataset.

Usage appears to have a log-quadratic effect on the mean failures of tanks; this implies that as tank usage during a year increases, the expected failures increase, but the rate of increase continually slows as usage increases (in the range of peacetime, home-station usage). Again, this conclusion is only valid within the range of the data—up to approximately 3,000 kilometers in peacetime operations. At some point the usage effect may become linear, with each one-kilometer increase in usage producing the same increase in expected failures.

The magnitude and shape of the observed effects—particularly the relationship between age and failures—differ across tank subsystems. The electrical, hardware, hydraulic, and main gun subsystems experienced larger *absolute* failure rate increases due to aging than the chassis, power train, and fire control subsystems. The chassis, hardware, hydraulic, and main gun subsystems experienced the greatest *relative* increases due to aging. Because the electrical subsystem had a high initial (age-0) failure rate, the relative increase in its failure rate was low, despite a high absolute increase. Because the chassis subsystem had a low initial failure rate, the relative increase in its failure rate was high, despite a low absolute increase. Also, for some subsystems the effect of age diminished or disappeared after tanks reached a certain age, which is probably an indication that the age was beyond the normal wear point for the subsystem's components. The point at which failures no longer increase with age for a subsystem (or part) or actually start to decrease reflects that point at which the peak wearout age region has been passed and sufficient fleet renewal for the subsystem (or part) has occurred to reduce the effective age of the fleet with respect to that subsystem (or part).

For the fire control subsystem, our data suggest an aging effect but also a possible effect with respect to tank variant. (Fully isolating

these two effects was not possible, since age and tank variant are confounded.) M1A2s, which are younger than M1A1s, have different types of fire control components than M1A1s—in particular, digital electronic line replaceable units (LRUs), rather than analog LRUs. The data suggest that the like-new failure rate of M1A2 fire control components is higher than that of fire control components in relatively young M1A1s.

Supplementary analyses of subsystem part failures and the unit prices of those parts provided additional information about the drivers of aging effects. Specifically, aging effects tended to be stronger for low-priced parts than for high-priced parts.

Although not a focus of this study, the effect of location is noteworthy. Some locations had significantly more tank failures than did others, after controlling for usage and age. This could be due to different maintenance practices, climate, terrain, training plans, and failure-reporting practices.

IMPLICATIONS

Consistent with private industry fleet management principles, Army leaders have long believed that older tanks have higher failure rates than newer ones, which increases maintenance demands and stresses operational readiness. However, supporting statistical evidence has been lacking. This study provides such evidence, demonstrating that increasing age, after accounting for usage and location effects, tends to raise M1 failure rates (given the current Army maintenance regime). Although the study is cross-sectional (incorporating one year of data from tanks), its findings—and the results of sensitivity analyses involving additional data and tests—provide initial quantitative support for several conclusions. Specifically, it is reasonable to conclude that, without modernization, time (or age) will pose a threat to operational readiness and increase the demand on resources.

Another important finding is that age is harder on some subsystems than on others. Moreover, within subsystems, age has different effects on different components. Knowledge of these patterns may help RECAP planners determine which subsystems and components should be rebuilt and which should receive higher priority in such ef-

forts. Further, the study indicates which subsystems and components are likely to drive the failure rate of new tanks—specifically, fire control, electrical, and power train; whether new or old, these components constitute reliability "problems." This information suggests where upgrade initiatives such as engineering redesign might have the biggest impact.

Further exploration of the source of age effects on the Abrams failure rate yields valuable insights into the aging problem. Much of the age effect tends to result from what are, in the Abrams, relatively low-cost components, so the age effect on operations and maintenance cost (the budget account used to pay for spare parts) is likely to be less than its effect on readiness and workload. These components are typically simple parts that have dominant failure modes associated with wear-and-tear. The expensive parts, in contrast, tend to be more complex, with many different failure modes. Increased component failures increase the maintenance workload burden. Since Army maintainers are not paid according to the amount of maintenance they perform and do not receive overtime, this does not affect the Army's cost structure. Rather, it can affect maintainer quality of life when the workload necessary to maintain operational readiness increases substantially.

Additionally, there are potential implications for force structure and future operational readiness. Once tank age reaches a certain point, the maintenance system may no longer be able to supply a satisfactory level of operational readiness—even with workarounds such as controlled exchange, necessitating replacement or substantial rebuild or acceptance of lower readiness possibly combined with increased maintenance capacity. There is some indication that a portion of the active Army's tank fleet has already reached this point, causing isolated M1A1 operational readiness problems. For example, Fort Riley units, with the oldest tanks in the Army's active inventory, are the only active units that consistently struggle to meet the Army's operational readiness rate goal for tanks.[5] At the Army's National Training Center (NTC), tank battalions employing relatively old M1A1s (both NTC-owned and from home stations) averaged just 74

[5]From 1999 to 2001, Fort Riley M1A1 operational readiness averaged 88.05 percent, while the active force M1A1 average was 90.75 percent, based on monthly readiness reports extracted from the Logistics Information Database.

percent operational readiness over the course of rotational training events from fiscal years 1999 through 2001; 4 of the 22 battalions for which data are available achieved less than 70 percent, a figure often considered the breakpoint for combat effectiveness.[6] This contrasts with an average of 83 percent for battalions with relatively new M1A2s. Repair time for the two groups was similar, with a difference in failure rates accounting for the difference in operational readiness rate. Thus, for the Abrams fleet, age most likely produces gradual workload increases, possibly resulting in decreasing soldier quality of life and declining operational readiness, and it generates a buildup of deferred financial cost that emerges in the form of programs such as RECAP.

---

[6]The NTC metrics are based on manually collected data provided by NTC observer-controllers (OC) to one of the authors. Each day, OCs collocated with tank platoons report the operational readiness status and failure information to the Forward Support Battalion Support Operations Officer OC, who records the information.

ACKNOWLEDGMENTS

We thank the Honorable Paul J. Hoeper, then Assistant Secretary of the Army for Acquisition, Logistics, and Technology (ASA[ALT]), and his staff for sponsoring this research. Within the Office of the ASA(ALT), Dr. Walter Morrison, Director for Research and Laboratory Management, championed the project, and his action officers, initially Suzanne Kirchoff and then Joseph Flesch, have assisted in coordinating with Army organizations.

The sponsorship and strong support of another RAND Arroyo Center project, "Diagnosing Equipment Serviceability," by MG (ret.) Charles Cannon, as the Army's acting Deputy Chief of Staff for Logistics, and LTG Charles Mahan, first as the Chief of Staff of the Army Materiel Command and later as the Army's Deputy Chief of Staff, G-4, made this research possible. The equipment serviceability project led to the ability to archive individual tank failures, which was the key missing element for enabling this type of research for the Army. Tom Edwards, Deputy to the Commanding General at the Army's Combined Arms Support Command (CASCOM), has also provided strong support for this work. Within the office of the Army G-4, Donna Shands, Associate Director for Sustainment, Kathleen Schulin, Chief of the Retail Supply Policy Division, MAJ Diane Del Rosso, MAJ John Collie, and MAJ Michael Kerzie played key roles in moving the equipment serviceability research forward, as did Jan Smith and CW4 Robert Vachon of CASCOM, CW5 Jonathon Keech and CPT Doug Pietrowski of the Ordnance Center and School, and CW3 David Cardon of the 1st Cavalry Division.

We are grateful to Sharon Gilbert, Karen Weston, and Donita Wright at the Army Materiel Command Logistics Support Agency for providing database extracts of tank year-of-manufacture and usage. We thank Theresa Ho and Mike Hilsinger at CALIBRE Systems for providing Standard Army Maintenance System-2 archives from which we extracted tank failure information.

At RAND, the contributions of John Dumond, Rick Eden, Ron Fricker, Bonnie Ghosh-Dastidar, Sally Morton, Tim Ramey, and Marc Robbins greatly facilitated this study and its documentation. Dan Relles and Ray Pyles provided thorough technical reviews that helped us improve the quality of the research. We also very much appreciate the editorial comments of Nikki Shacklett and the assistance of Pamela Thompson and Joan Myers in preparing the document.

GLOSSARY

1AD	1st Armor Division
1CAV	1st Cavalry Division
2ID	2nd Infantry Division
3ID	3rd Infantry Division
4ID	4th Infantry Division
AGREE	Advisory Group on Reliability of Electronic Equipment
AMC	Army Materiel Command
AMSAA	Army Materiel Systems Analysis Activity
ANOVA	Analysis of Variance
ASA(ALT)	Assistant Secretary of the Army for Acquisition, Logistics, and Technology
CASCOM	Combined Arms Support Command
DoD	Department of Defense
EDA	Equipment Downtime Analyzer
FedLog	Federal Logistics
GAM	Generalized Additive Models
LRUs	Line Replaceable Units
NTC	National Training Center

OASA(ALT)	Office of the Assistant Secretary of the Army for Acquisition, Logistics, and Technology
OLS	Ordinary Least Squares
OR	Operational Readiness
PRESS	Predicted Residual Sum of Squares
RAM	Reliability, Availability, and Maintainability
RCM	Reliability-Centered Maintenance
RECAP	Recapitalization
SAMS-2	Standard Army Maintenance System-2
TAMMS	The Army Maintenance Management System
TEDB	TAMMS Equipment Database
YOM	Year of Manufacture

Chapter One

INTRODUCTION

Equipment reliability has become a high priority of managers in both the private and public sectors. The term has multiple definitions, but the most widespread one is the probability of performing an intended function, for a given interval, under prescribed conditions (Hillier and Lieberman, 1986; Omdahl, 1988; Stevenson, 1993; Morris et al., 1995). Consequences of poor reliability, manifested as high failure rates, can range from minor inconvenience to catastrophe. They include financial costs, essential function or mission-capability losses, and safety consequences. In the Armed Forces, where weapon systems are technology-intensive and used under life-threatening conditions, equipment failure can have particularly severe penalties (Alexander, 1988). Many believe that the age of equipment contributes to failures (Gansler, 1999; United States General Accounting Office, 2001), and with budget constraints forcing longer equipment life cycles, Army officials suspect that aging systems are impairing readiness and increasing financial costs. However, the effects of age on Army equipment have not been quantified and are therefore poorly understood. Accordingly, this study begins an investigation, conducted by RAND Arroyo Center, to assess the impact of age on weapon system failure rates and the resulting consequences. The focal weapon systems are U.S. Army ground equipment.

Interest in the age-reliability relationship has grown steadily over the past century. Prior to World War II, the simplicity of equipment made repairs straightforward and inexpensive (Moubray, 1997). Historical records suggest that, in addition, the military did not keep

vehicles for long periods.[1] As a result, reliability and the effects of age received little attention. With the exception of a few material fatigue studies (e.g., Weibull, 1939), approaches to the subject were "largely intuitive, subjective, and qualitative" (Blischke and Murthy, 2000:19).During World War II, however, the labor shortage combined with productivity demands led to more dependence on complex technology and systems (Burrill and Ledolter, 1999; Moubray, 1997). Overall weapon-system-level reliability suffered, leading to higher weapon system failure rates; consequently, keeping military systems operational began to consume resources at a higher rate and raised concerns about cost (Barringer, 1998:4). Additionally, downtime became a significant issue. Prompted by these difficulties, Army, Navy, and Air Force officials appointed committees to address reliability. To coordinate the efforts of these committees, the Department of Defense (DoD) established the Advisory Group on Reliability of Electronic Equipment (AGREE) in 1952 (Kapur and Lamberson, 1977). The AGREE report in 1957 led the DoD to establish standards for such activities as reliability testing, program management, and prediction, and the field of reliability engineering emerged (Kales, 1997; O'Connor, 1998).

In the first decade following the AGREE report, empirical papers and texts (e.g., Gosling, 1962; Krohn, 1969; Machol, Tanner, and Alexander, 1965) advanced the notion that age-failure relationships were best described by one of the two curves in Figure 1.1. The first curve displays a constant or slowly increasing failure probability, followed by a wear-out region with a rapidly increasing failure probability. The second is commonly known as the "bathtub curve," which depicts a "burn-in" or infant mortality period, a constant failure probability, and then a wear-out region (Moubray, 1997; Nowlan and Heap, 1978:46).

Preventive or scheduled maintenance programs were, for many years, designed with one of those two conditional probability curves

[1]Interview with Timothy Ramey, July 2001. Ramey's research suggests that the retirement age of Air Force aircraft has increased steadily between 1932 and 1995. He cautions that some of the early data may be missing. But as he also points out, "Even allowing for the possibility that records prior to 1946 may be missing, the trend in design service lives is clear: the Air Force has been operating its *oldest* designs for roughly 6 months longer each year."

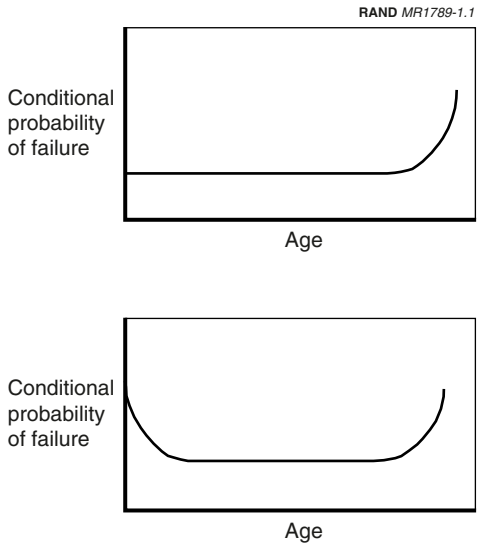


Figure 1.1—Hazard Functions with Pronounced Wear-out Regions

(or *hazard functions*) in mind (Harrington, 2000). Such programs would take equipment out of service for maintenance at regular intervals or for overhaul, even if it did not show signs of wear (Robinson, Anderson, and Meiers, 2003).

In the late 1960s, an analysis of United Airlines failure data challenged the notion that most equipment could be characterized by the curves in Figure 1.1, with their pronounced wear-out regions. Analysts found that the majority of aircraft parts had hazard functions represented by the curves in Figure 1.2 (Moubray, 1997; Nowlan and Heap, 1978:46). This was especially true for *complex items*, those subject to many types or *modes* of failure (Nowlan and Heap, 1978:37). Unless they had a dominant failure mode, complex items generally were found to lack wear-out characteristics (p. 48).[2]

[2]A *dominant failure mode* is one that accounts for a large percentage of an item's failures (Nowlan and Heap, 1978:38). Like most simple items, complex items with dominant failure modes do tend to reach a point at which their failure probability increases rapidly with age. Most complex items, however, experience widely distributed failure modes; thus, they often do not reach a wear-out region. Many types

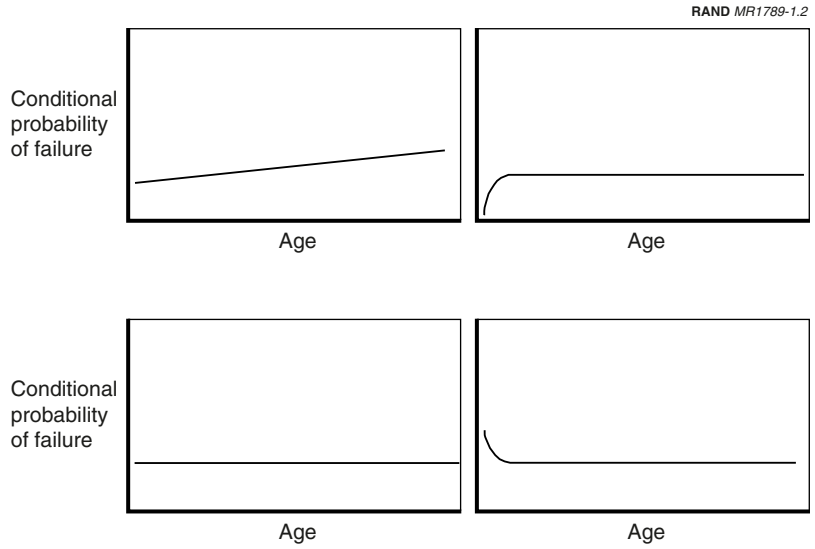


Figure 1.2—Hazard Functions Without Pronounced Wear-out Regions

These results first appeared only in civil aviation reports, but a decade later they reached a broader audience via a seminal publication by Nowlan and Heap (1978).

The United Airlines findings led to the development of *Reliability-Centered Maintenance* (RCM), the idea that a maintenance regime should be based on the specific failure characteristics (e.g., patterns, causes, modes, criticality, detectability) associated with a system/component under review (Moubray, 1997; Nowlan and Heap, 1978; Robinson, Anderson, and Meiers, 2003).[3] Recognizing that a variety

of "overstress" conditions, other than those related to wear, can cause failures at random points in an item's life. Alternatively, the various failure modes could experience different wear-out regions, none of which is dominant. Thus, the hazard rate curve for complex items often reflects the convolution of many different hazard rate curves for different types of failure modes.

[3]Formally defined, Reliability-Centered Maintenance is "a process used to determine the maintenance requirements of any physical asset in its operating context" (Moubray, 1997:7). The RCM process involves answering a series of questions about an item (Moubray, 1997): What are the item's purpose and performance standards? How does it fail? What causes its failures? What are the failure effects? What is the

of failure patterns and causes are possible, researchers in the 1980s and 1990s continued to analyze empirical data to estimate the hazard functions of different systems and components (e.g., Mudholkar, 1995). Commercial and military organizations encouraged and supported such efforts. For example, the U.S. Army Materiel Systems Analysis Activity (AMSAA) recommended a methodology, including field data collection, for developing a replacement strategy for Army Tactical Wheeled Vehicles (Streilein, 1984).

Still, much more can be done. Political and economic changes over the past decade have heightened the need for more refined models of equipment age and failure rates—particularly in the U.S. Army. With less funding for procurement, military services are using weapon systems for more years than originally intended (Kitfield, 1997). General Paul Kern (2001:5) recently noted that

> the average age of critical systems such as the Abrams tank, AH-64 Apache, UH-60 BLACK HAWK, CH-47 Chinook, and Bradley Infantry Fighting Vehicle will exceed their 20-year expected service lives by 2010. The potential exists for the Army to move into the second decade of this century with a significant portion of its forces incapable of meeting a world-class threat.

As General Kern's statement indicates, Army leaders intuitively believe that age eventually impairs the functioning of equipment, harming readiness or requiring substantially more resources to maintain readiness. Hence, they have embarked on a program of recapitalization (RECAP), which "involves rebuilding and selectively upgrading currently fielded systems to ensure they are operationally ready, 'zero-time/zero-mile' systems" (Orsini and Harrold, 2001:2). The Army is currently deciding which systems should be rebuilt (i.e., restored to like-new condition) and which should be upgraded (i.e., given new capabilities) based on the age, cost to maintain, fleet readiness, and importance of equipment.

Some evidence of stresses on the ability to maintain desired operational readiness (OR) is already present. Fort Riley's two heavy

---

significance of the failures? How can its failures be prevented or predicted? What actions can be taken if prevention and prediction are not possible? RCM is designed to help items achieve maximum reliability at minimum cost.

brigades, which have the oldest tanks in the active Army, are the only two brigades that consistently have trouble meeting the Army's 90 percent peacetime OR goal for tanks.[4] More significantly, some units with the older M1A1s struggle to maintain even 70 percent OR during high-intensity training events at the National Training Center (NTC). During fiscal years 1999 to 2001, tank battalions with M1A1s averaged just 74 percent OR, versus 83 percent for those with newer M1A2s. (Average downtimes per failure were similar, with M1A2 times slightly longer.) Even with the relatively robust supply support infrastructure at NTC, four of 22 M1A1-equipped battalions failed to achieve a 70 percent average, with a low of 63 percent. The primary reason for the difference in the OR rates was a difference in failure rates: About 12.4 percent of available M1A1 tanks failed each day, versus 7.6 percent of M1A2s.[5]

Additional quantitative evidence, however, is needed to characterize age-failure relationships for systems under consideration for rebuild or upgrade. Further knowledge about system failure patterns would facilitate and improve RECAP decisions—providing better justification for funding, where merited. The present study aims to provide such information by addressing the following research questions:

1. What is the relationship between age and the M1 Abrams mission-critical failure rate?[6]
2. How is the M1 failure rate related to other factors, such as usage and location-specific factors?
3. If there is a significant relationship between age and the M1 Abrams mission-critical failure rate, which of the various M1

[4]From 1999 to 2001, Fort Riley operational readiness averaged 88.05 percent, while the active force average was 90.75 percent, based on monthly readiness reports extracted from the Logistics Information Database.

[5]The NTC metrics are based on manually collected data provided by NTC observer-controllers (OC) to one of the authors. Each day, tank platoon OCs collocated with the platoons report the OR status and failure information to the Forward Support Battalion Support Operations officer OC, who records the information.

[6]A mission-critical failure is defined in this study as one that makes an item not mission capable, as indicated by the item's technical manual and subsequently reported by its owning unit. Mission-critical failures are also called *deadlining events.*

subsystems and individual parts generate this relationship, and to what degree do they do so?

4. How can statistical models of such relationships inform RECAP decisions and planning?

Our analysis focused initially on the M1 Abrams, for several reasons. First, it is one of the key systems in Army equipment readiness reporting. Second, it plays a central role in armored combat, often being considered the centerpiece of the Army's heavy ground forces. Third, the already-aging M1 fleet is projected to continue serving the Army for quite some time—perhaps 30 years or more (Konwinski and Wilson, 2000). M1 maintenance is expensive, in terms of both parts cost and maintenance personnel cost. Fourth, the availability of M1 data prompted us to begin with the Abrams. Subsequent studies will focus on other critical Army ground systems that cover a broad range of technologies, complexity, and missions.[7]

As mentioned earlier, the consequences of equipment failure tend to fall into three categories: financial costs, function/mission losses, and safety. This research focuses on mission-critical failures, not cost or safety. It does, however, provide insights with regard to the financial cost implications, and it lays some of the groundwork (conceptual and data preparation) for related cost and safety studies.

In Chapter Two we describe the study methodology. Chapter Three then summarizes our findings, and Chapter Four discusses implications of those findings.

---

[7]The currently planned analyses are limited to ground systems because the requisite data for the methods applied in this study are not widely available for aviation and missile systems.

Chapter Two

METHODOLOGY

DATA SOURCES

Data for this study came from several sources. Tank type, year-of-manufacture, odometer readings, and site information—i.e., the tank's division, location, battalion, and company—were from a database called TEDB. TEDB refers to the TAMMS Equipment Database. (TAMMS is The Army Maintenance Management System.)[1] Missing year-of-manufacture data was supplemented by fielding dates reported by selected units. Tank failure records came from the Equipment Downtime Analyzer (EDA) (Peltz et al., 2002),[2] which incorporates data from the Standard Army Maintenance System-2 (SAMS-2) daily deadline reports (026 prints). Data on tank part prices came from the Federal Logistics (FedLog) database.

SAMPLE CHARACTERISTICS

The study sample included tanks from the Army's six active divisions categorized into six locations defined by geographic regions. After data refinement (see the subsection "Data Refinement Techniques" below), our sample size was 1,567. Approximately 1,162 tanks were M1A1 variants, and 405 were newer M1A2 variants. Table 2.1 shows the number and types of tanks by location.

---

[1]The Logistics Support Agency (LOGSA) extracted and forwarded the TEDB information we needed.

[2]The EDA is a new information system decision support tool, developed by RAND Arroyo Center and now implemented by the Army in the Global Combat Support System-Army, that facilitates the diagnosis of equipment downtime.

We used odometer readings collected between April 16, 1999, and January 15, 2001, to compute the monthly and annual usage of tanks. Odometer/usage data were available for all tanks in the sample, but some tanks had more monthly data than others. Monthly data availability was affected by the length of time a tank belonged to a unit (e.g., new equipment fielding resulted in M1A2s replacing M1A1s in some units during the study period), data quality issues, and overlap with EDA data collection.[3] Following data refinement, most tanks in the sample had at least 9 months of usage data, as Figure 2.1 indicates. We included up to one year of usage data for each tank. For example, if a tank had 16 months of usage data available between April 1999 and January 2001, we included only the first 12 months. Of the 1,567 tanks in the sample, 627 (40 percent) had a full year of usage data.

Table 2.1

Number of M1 Tanks in Sample by Location and Division

Location Code	Location	Division(s)	Number of M1A1 Tanks	Number of M1A2 Tanks
1	Fort Hood, TX	1st Cavalry (1CAV) 4th Infantry (4ID)	184	405
2	Fort Carson, CO	4ID	58	—
3	Korea	2nd Infantry (2ID)	141	—
4	Europe	1st Infantry (1ID) 1st Armor (1AD)	355	—
5	Fort Riley, KS	1ID 1AD	175	—
6	GA	3rd Infantry (3ID)	249	—

NOTE: 3ID is stationed at Fort Stewart and Fort Benning, both of which are in Georgia.

[3]EDA data collection began in April 1999 for the 1st Cavalry Division, October 1999 for the 4th Infantry (Mechanized) Division, November 1999 for the 3rd Infantry (Mechanized) Division, February 2000 for the 1st Armor Division and 1st Infantry (Mechanized) Division, and April 2000 for the 2nd Infantry (Mechanized) Division.

For some tanks, the first year of data—hereafter referred to as the *study period*—began in 1999, and for others it began in 2000. We took the study period into account when computing tank age (see the subsection "Tank Study Variables" below).

Figures 2.2 through 2.4 show the distribution of age by location. Figure 2.5 shows the distribution of accumulated usage by location. Accumulated usage was the number of kilometers traveled by a tank during its study period. As Figure 2.5 indicates, usage varied greatly among locations.

In summary, tanks had many months of data, spanned a range of ages, and came from multiple settings with distinct usage patterns. However, the ages of M1A1s and M1A2s did not overlap, preventing the isolation of tank variant effects from other effects.

MEASURES

Two "substudies," each at the individual tank level of analysis, comprised the overall study:

1. an assessment of factors affecting M1 failures, and
2. an assessment of factors affecting M1 subsystem failures.

In substudy 1, hereafter called the *Tank Study*, we assessed the impact of age, location, and usage on individual tank failures. In substudy 2, hereafter called the *Subsystem Study*, we assessed the impact of tank age, location, and usage on tank subsystem failures. Below we describe the key variables in these substudies.

TANK STUDY VARIABLES

System Failures

In the Tank Study, the outcome variable was a tank's total number of mission-critical failures during the study period. Repair records showed each date on which the tank became inoperable. A simple count of those dates yielded the number of deadlining failures.

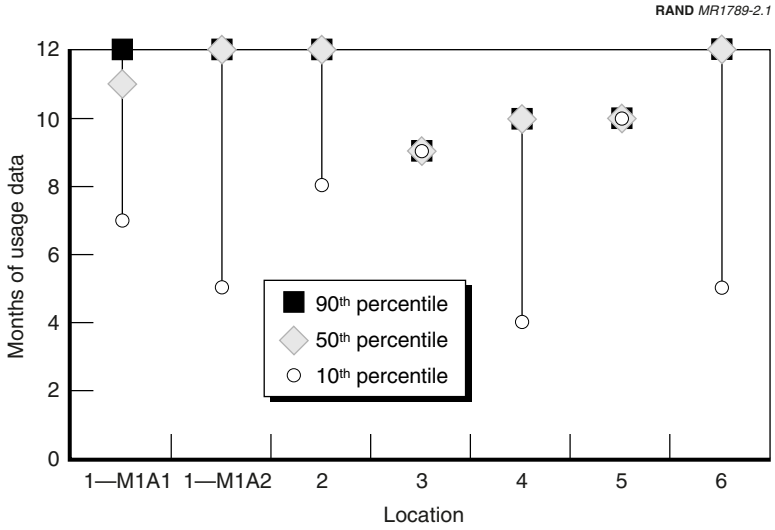


Figure 2.1—Number of Months of Usage Data per Tank by Location

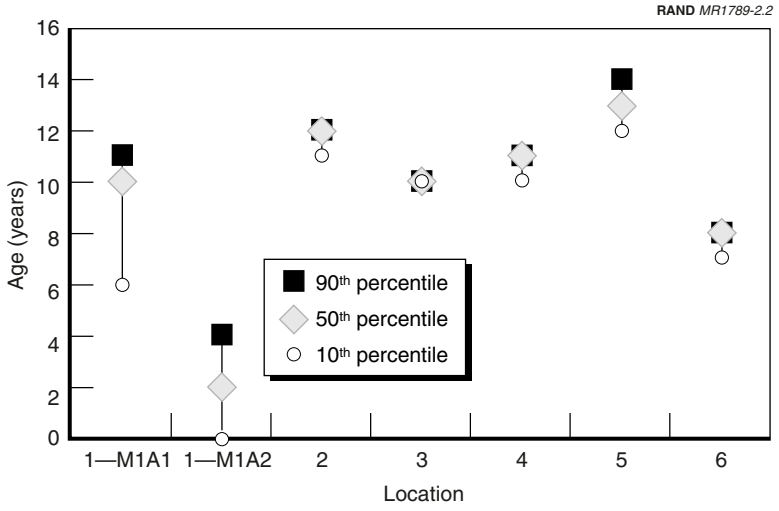


Figure 2.2—Distribution of Tank Age by Location

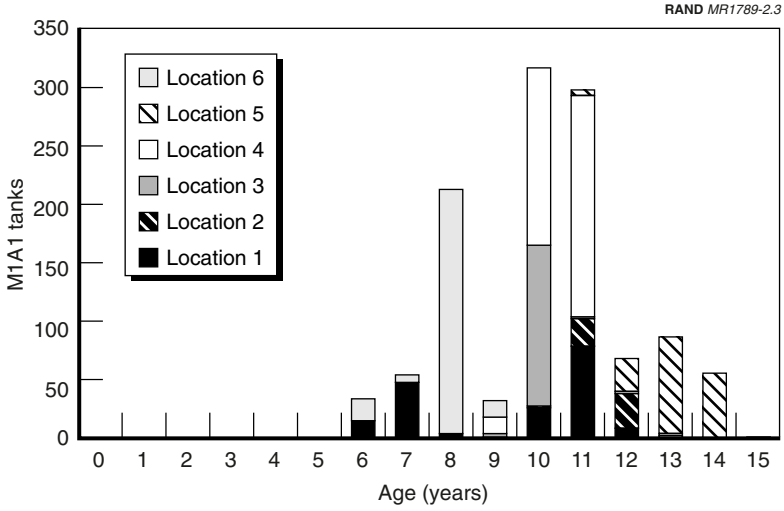


Figure 2.3—M1A1 Age Histogram

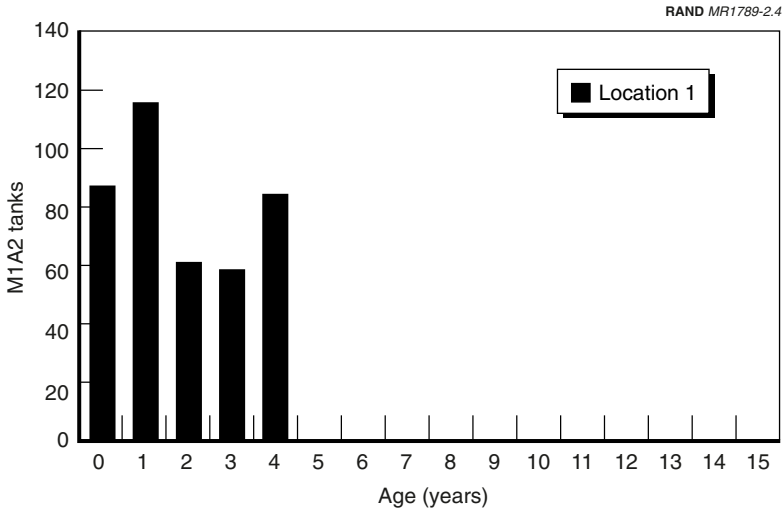


Figure 2.4—M1A2 Age Histogram

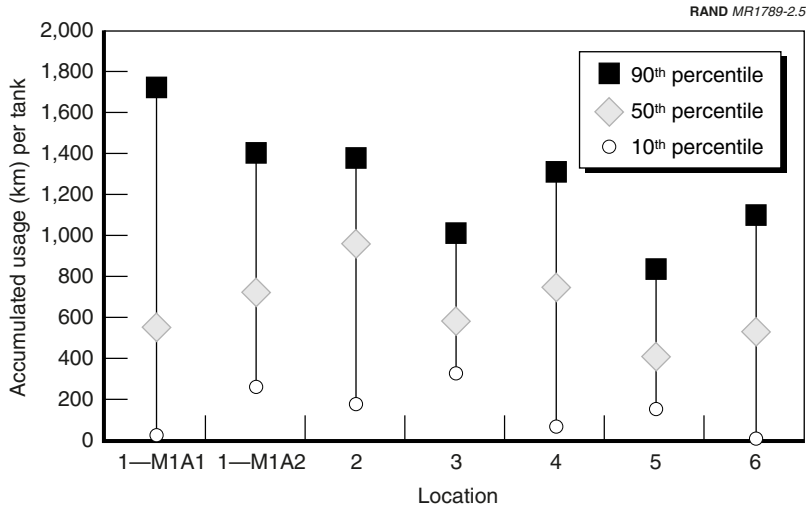


Figure 2.5—Distribution of Tank Usage by Location

Age

Tank age was, in most cases, computed using year of manufacture (YOM).[4] Because YOM data were not available in the TEDB for M1A2s in the 1st Cavalry and 4th Infantry divisions, we used fielding dates instead. The formula for tank age is shown below:

$$\text{Age} = \text{study year} - \text{YOM or fielding date},$$

where the study year was either 1999 or 2000, depending on the tank. Because we defined age in terms of YOM of the entire tank, the age variable does not necessarily reflect the age of tank components. Many tank components may have been replaced or refurbished during a tank's lifetime. However, data on the replacement history or ages of individual tanks' components are not available. Thus, any age

[4]All M1A2 tanks were younger than M1A1 tanks, so tank type was confounded with age. The correlation between tank type and age was $r = .90$ ($p < .0001$). Such a high correlation precluded controlling for tank type in our analyses.

effects we observe are those that appear despite the component renewal histories of the tanks over their entire service lives prior to the study period.

We mean-centered the age variable to reduce multicollinearity problems that occur when first-order and higher-order terms (e.g., age and age-squared) are included in the same regression (Aiken and West, 1991). This step involved transforming the age variable by subtracting the mean tank age.

In addition to serving as a predictor in our models, the age data provided a bit of guidance in the selection of other model variables. Originally, we planned to include initial odometer reading, i.e., the first reading during a tank's study period, as a predictor that would serve as another type of "age" indicator.[5] However, plots of initial odometer readings versus age revealed a data-quality issue: Possibly due to odometer resets, the expected relationship between initial odometer reading and age was not apparent. Figures 2.6 and 2.7 illustrate this data problem, which was a greater issue for M1A1s than M1A2s. The patterns in the graphs are consistent with a situation in which, as time progresses, more and more tanks have reset odometers from maintenance actions. The percentiles on the graph indicate the percentage of tanks by age with an odometer reading less than or equal to the point on the y-axis. Up to 12 years of age, Figure 2.6 shows a fairly linear year-to-year increase at the 90th and 95th percentiles of usage. The 75th percentile time series is fairly linear until age 9, the 50th to age 8 or 9, the 25th to 8, and the 10th to 7. This suggests that by age 13, most tanks have had their odometers reset at least once, 75 percent have their odometers reset by age 10, between 25 and 50 percent by age 9, and so forth.

This problem prevented us from including initial odometer reading in the model. Still, changes in odometer readings served a purpose in our study: They allowed us to compute tank usage during the study period.

---

[5]If the first reading was greater than 50,000 km, we checked subsequent monthly readings until we found a valid one to use as the initial value. The Army funds tank usage at 800 miles (1,290 km) per year. At 14 years of age (the maximum age in the study), this implies an accumulated usage of 11,200 miles (18,065 km). Readings exceeding 50,000 km were therefore considered infeasible.

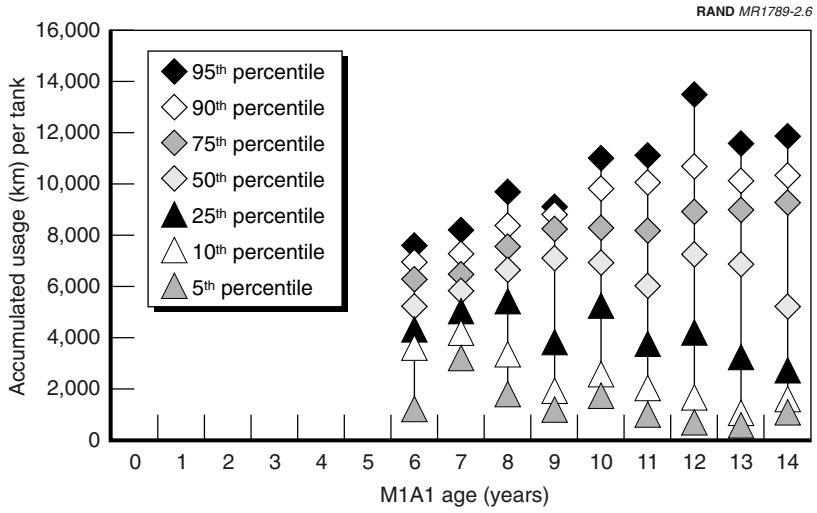


Figure 2.6—Distribution of Initial M1A1 Odometer Readings by Age

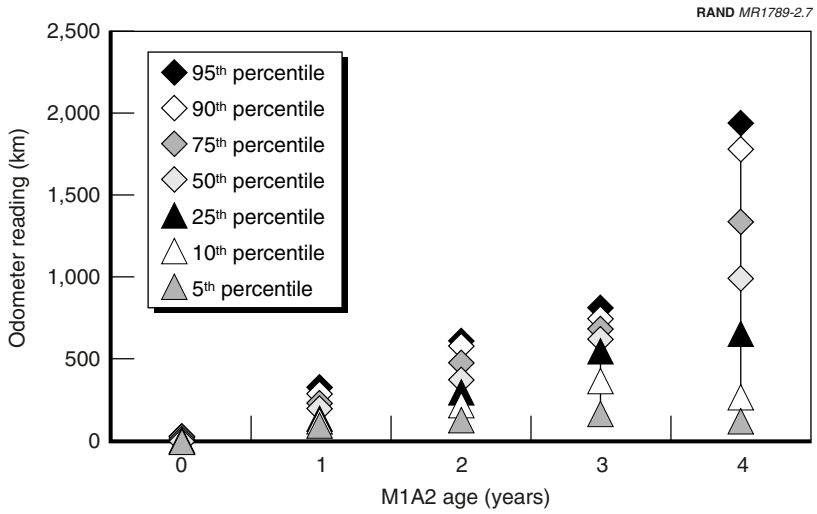


Figure 2.7—Distribution of Initial M1A2 Odometer Readings by Age

Accumulated Usage During the Study Period

For each tank, accumulated usage was the total distance traveled by the tank during its study period. Through a data-filtering technique (and imputation techniques that will be described later), we were able to derive accumulated usage (in kilometers) from monthly odometer readings[6] despite reset odometers and data-quality problems with odometer readings. When a tank's odometer reading from month $n + 1$ was smaller than its odometer reading from month n, we deleted the month $n + 1$ reading and treated it as a missing data point. Similarly, when the month $n + 1$ reading exceeded the month n reading by more than 1,000 km, we deleted the month $n + 1$ reading. Upon completing this filtering process, we computed monthly usage as follows:

$$\begin{aligned}\text{Usage during month } n = {} & \text{odometer reading for month } n+1 \\ & - \text{odometer reading for month } n.\end{aligned}$$

To get a tank's accumulated usage during the study, we summed its monthly usage values. Just as we transformed the age variable, we transformed the accumulated usage variable via mean centering.

Updays

This variable captured the total number of days a tank was available to fail—i.e., those study period days in which it was "up" or operative.

$$\begin{aligned}\text{Updays} = {} & \text{number of days in tank's study period} \\ & - \text{number of days tank was down during study period.}\end{aligned}$$

Tanks that are down most of the time may have low failure counts simply because they have less opportunity to fail. Thus, it was important to control for this factor.

---

[6] Army units report odometer readings as of the 15th of each month.

Location

To control for the combined effects of different environmental conditions, training schedules, maintenance practices, and command policies, we included location as a predictor. (Data were not available to enable the isolation of each of these factors.) As Table 2.1 indicates, there were six possible locations; hence, we used five dummy variables to capture the location of each tank in the sample.

SUBSYSTEM STUDY VARIABLES

Subsystem Study variables were identical to Tank Study variables except that the key outcome was subsystem failures, rather than system failures. In a secondary analysis delving further into subsystem failures, the outcome variable was part failures, where parts were subsystem components classified into price categories.

Subsystem failures. Based on the parts ordered for repair, each M1 system failure was classified into two tiers of categories: (1) hull, turret, or either, and (2) chassis, power train, fire control, main gun, electrical, hardware, hydraulics, and miscellaneous.[7] Categories in the second tier are subsets of categories in the first tier. Specifically, chassis and power train are types of hull failures; fire control and main gun are types of turret failures; and electrical, hardware, hydraulics, and miscellaneous can be either hull or turret failures. If the parts ordered for a repair fell into multiple subsystem categories for a given tier, then the system failure counted as more than one subsystem failure. For example, suppose a repair required two turret parts: a fire control part and a main gun part. In that case, we counted the repair as a turret failure at the first tier and as fire control and main gun failures at the second tier.

For each tank, we computed the total number of failures in each category during the study period. This process yielded 11 subsystem

[7]This set of categories was based on Federal Supply Class (FSC) codes associated with parts ordered for the repairs. The FSC consists of the first four digits of a part's National Stock Number (NSN). FSCs were mapped to the listed categories, and maintenance personnel in the 1st Cavalry and 4th Infantry divisions reviewed the mappings. Originally there was one additional second-tier category, communications, but we found no part orders in that category during the study period.

failure variables. Our analyses focused on 9 of these, excluding failures labeled "miscellaneous" or "either."

Part failures. We also classified M1 part orders into four price categories. Low-priced parts were those having a unit price of $100 or less; medium-priced parts were those having a unit price greater than $100 but less than or equal to $1,000; high-priced parts were those having a unit price greater than $1,000 but less than or equal to $10,000; and very-high-priced parts were those having a unit price above $10,000.

For each tank, we computed the total number of part failures in each price category during the study period. This process yielded four part failure variables.

DATA REFINEMENT TECHNIQUES

Due to data-quality issues, multiple refinement techniques were necessary. A previous section mentioned several data-filtering processes that discarded faulty odometer readings. In this section we describe additional techniques employed.

Exclusion of Observations

Tanks that were down (inoperative) during most of their study period were problematic for our analyses. Such extended downtime significantly reduced the tanks' usage and failure opportunities, potentially distorting our findings. Our final sample of 1,567 tanks therefore excluded 26 tanks that were down for more than 50 percent of their study period. Many of these were probably serving as the tank equivalents of "hangar queens," with their parts used to provide parts for other tanks.

Imputation

Imputation of usage data. Many tanks in the sample were missing some monthly usage data.[8] About 42 percent of the tanks lacked us-

[8]It is important to clarify our interpretation of missing usage data. Recall that some tanks had 12 months of usage data while others had fewer. A tank with less than 12

age data for at least one month during their study period. On average, tanks were missing 2.2 months of usage data. Such gaps diminished the accuracy of the accumulated usage measure. Thus, we used *imputation*, a technique that compensates for missing data by assigning values that "fill in the blanks."

Our primary imputation approach was mean substitution. This common technique entails using an auxiliary variable to divide a sample into a set of classes, and "assigning the class mean . . . to all nonrespondents in each class" (Brick and Kalton, 1996:228).[9] In the present study, the auxiliary variable was the tank's company.[10] When a tank was missing usage data from a particular month, we gave it the company mean for that month—i.e., the average usage of other tanks in the same company.

An imputed mean should, ideally, come from a group of tanks with similar usage patterns. Anticipating that usage among tanks would be most consistent within a platoon, we initially considered using platoon means rather than company means for imputation. However, we ultimately rejected the use of platoon means because(1) platoon information was incomplete[11] and (2) nested Analysis of Variance (ANOVA) suggested that, in general, tanks in the same company tend to have similar usage. (Given the lack of complete pla-

---

months of usage data was not necessarily missing any information. The tank may simply have been brand new—or may have been replaced. For a tank to have one month of missing usage data, the following had to occur: The tank's serial number, NSN, and unit identifier code (UIC) appeared in two consecutive monthly odometer data files (e.g., March 2000 and April 2000), but one of those odometer readings was absent or discarded as being unreasonable (e.g., had an extra digit due to a typographical error). Hence, the tank was missing one month of usage information (March 16 through April 15).

[9]Although widely utilized, mean substitution does not preserve the distribution of the variable on which it is performed.

[10]Tank companies have 14 tanks: 4 tanks in each of 3 platoons and 2 headquarters section tanks.

[11]The second digit of a tank's *bumper number* identifies its platoon (or if it is in the headquarters section). Because bumper numbers came from SAMS-2 repair records, only tanks that failed had platoon identifiers. The TEDB database only identifies tanks down to the company level.

toon information, the ANOVA found the company to be a better discriminator than either the platoon or battalion.)[12]

Occasionally, an entire company of tanks was missing data for a particular month. When this occurred, mean substitution was not possible, so we took an alternative route. Specifically, we subtracted the tank's minimum odometer reading from its maximum odometer reading during the study. If this figure was less than 2,000 km, it became the tank's accumulated usage.[13] For example, if a tank had an initial odometer reading of 10,000 km in September 1999 and a final odometer reading of 10,800 km in August 2000, its accumulated usage was 800 km.

Imputation of age data. Originally, before data refinement, there were 1,636 tanks in our sample. Of those, 83 lacked age data. To reduce the amount of missing age data, we used deductive imputation, the process of deducing the value of a variable based on the values of other variables. As Brick and Kalton (1996:226) note,

> Deductive imputation is applicable when a missing response can be deduced from responses to other items. For example, a person under 16 years may be imputed to be single and a university teacher may be imputed to have a college degree. Deductive imputation is often considered to be editing rather than imputation.

Because tank serial numbers were assigned in order of production, we were able to deduce the YOM of some tanks from their serial numbers. When a tank's serial number fell in the middle of a sequence, we inferred its YOM from that of other tanks in the se-

---

[12]We ran 21 nested ANOVAs, one for each month between May 1999 and January 2001. Tank usage during a particular month was the dependent variable, and the independent variables were the platoon, company, and battalion corresponding to each tank. We used nested ANOVA because platoon was nested within company, and company was nested within battalion. In each ANOVA the company variable was significant at $p < .0001$.

[13]Tanks rarely travel more than 2,000 km per year. Indeed, of the 1,214 tanks in our sample that did not require the second type of usage imputation, only 11 had an accumulated usage exceeding 2,000 km.

Ordinarily, odometer resets might affect the validity of this second type of imputation; the 2,000 km filter minimizes the impact of such resets, however.

quence. Once we followed this procedure, the number of tanks with missing age data fell to 69—hence the final sample size of 1,567.

ANALYSES

Because of the nature of the dependent variable in the Tank and Subsystem Studies, our primary analytical technique was negative binomial regression; however, we used additional techniques in our sensitivity analyses. Conducting multiple analyses allowed us to elicit different insights and helped ensure the validity of results. Below we describe these analyses in detail.

Tank Study Analysis

The Tank Study's outcome measure, System Failures, was a count variable. Typically, such data follow a Poisson distribution, in which the variance across all observations equals the mean of those observations. In some datasets, however, the variance is greater than the mean. This feature, called *overdispersion*, suggests that a negative binomial distribution is more applicable. Appendix A provides general descriptions of each distribution.

Before determining which distribution applied to the failure data in this study, we had to consider whether tanks' mean failures changed during the study. Typically, longitudinal failure data exhibit nonconstant means; however, within a one-year period (the timeframe of our study), any change in the failure mean is likely to be minimal. We therefore assumed each tank's failure rate did not significantly vary during its study period. Essentially we are conducting a cross-sectional examination of the data instead of a longitudinal one. In effect, this breaks the time series into discrete periods within which the change in the failure rate is low. Comparing the one-year periods at different points in tank life cycles enabled us to examine how the failure rate changes over time.

Next, we began to fit a distribution to the failure data. When the degree of overdispersion is small, the choice between the Poisson and negative binomial distributions may not be clear-cut. Such was the case in this study: The M1 failure data were only slightly overdis-

persed. Appendix B illustrates and provides a detailed discussion of the nature of that overdispersion.

Given the slight overdispersion, our primary analysis was negative binomial regression, but we ran a Poisson regression as a sensitivity analysis. Both regressions treated the natural log of mean failures as a function of the predictor variables.[14] That is, we applied a "log link," specifying a full quadratic model in the following manner:

$$\ln \mu = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \beta_{11} x_1^2 + \cdots + \beta_{pp} x_p^2 + \beta_{12} x_1 x_2 + \cdots + \beta_{1p} x_1 x_p + \cdots + \beta_{p-1} x_{p-1} x_p,$$

where μ is the mean of Y, system failures during the study period, and p is the number of explanatory variables. (For simplification, higher-order terms were left out of the equation above.) However, in one case we assumed that Y had a negative binomial distribution, and in the other case we assumed that Y had a Poisson distribution.

The technique of backward elimination helped us select terms to serve as predictors in the model. We began with a model that included higher-order and interaction terms; then, in a step-by-step process, we removed those terms that failed to satisfy our criterion for remaining in the model.[15] Likelihood ratio tests[16] were utilized throughout this process. For the Tank Study, the full model was as follows:

---

[14]It is possible to treat other transformations of μ, rather than the natural log, as functions of the predictor variables in Poisson and negative binomial regressions. However, McCullagh and Nelder (1989) have noted that using the natural log of the mean leads to Poisson regression models with desirable statistical properties. The natural log is also commonly used in negative binomial regressions.

[15]In the backward elimination process we began with a cubic model and then reduced it, to the extent possible. The decision to limit the order of terms was based on preliminary data exploration and the need to preserve model interpretability.

[16]As described by Maddala (1988:84), the likelihood ratio test requires computing $-2\ln(\lambda)$, where λ = [max $L(\theta)$ of restricted model / max $L(\theta)$ of unrestricted model], and $L(\theta)$ is the likelihood function. (The restricted model has fewer terms than the unrestricted model. For example, if one is comparing a model with a cubic term to a model without that term, the latter is the restricted model.) The statistic $-2\ln(\lambda)$ is treated as a χ^2 with k degrees of freedom, where k is the number of restrictions in the model.

ln (mean tank failures during study period) =

$$\beta_0 + \ln(\text{updays}) + \beta_1(\text{location 2}) + \beta_2(\text{location 3}) + \beta_3(\text{location 4}) + \beta_4(\text{location 5}) + \beta_5(\text{location 6}) + \beta_6(\text{usage}) + \beta_7(\text{usage}^2) + \beta_8(\text{usage}^3) + \beta_9(\text{age}) + \beta_{10}(\text{age}^2) + \beta_{11}(\text{age}^3) + \beta_{12}(\text{usage} \times \text{age}) + \beta_{13}(\text{usage} \times \text{age}^2) + \beta_{14}(\text{usage}^2 \times \text{age})$$

Note that ln(updays) is an offset variable, so it is assigned a coefficient of 1. Also, age and usage are mean-centered in this equation.

To supplement the regressions and facilitate interpretation, we plotted predicted mean failures by age and by usage. When generating plots of predicted mean failures by age, we assigned updays a value of 180, and we assigned the usage variable a value of 375 km, the approximate median tank usage during 180 days. Then, for each location, we substituted a set of age values into the regression equation to get predicted mean failures.[17] The set of age values was kept within the actual age range at each location. For example, because all tanks at Location 6 were between six and nine years old, for that location we plotted the predicted mean failures corresponding to ages six through nine only.

To generate plots of predicted mean failures by usage, we set the age variable equal to the approximate median at each location, and we assigned updays a value of 180. We then substituted a range of usage values into the regression equation to get predicted mean failures.

Diagnostic plots assessed the goodness of fit of the Tank Study model, and a cross-validation procedure (Appendix C) assessed the model's predictive performance.

Subsystem Study Analysis

Like the Tank Study, the Subsystem Study relied on negative binomial regressions. For each subsystem, the full model was as follows:

---

[17] As the regression equation shows, this method first yielded the natural log of predicted mean failures (since the dependent variable in the regression equation is ln(mean failures)), which we then converted to predicted mean failures.

$$\ln(\text{mean subsystem failures during study period}) =$$

$$\beta_0 + \ln(\text{updays}) + \beta_1(\text{location 2}) + \beta_2(\text{location 3}) + \beta_3(\text{location 4}) + \beta_4(\text{location 5}) + \beta_5(\text{location 6}) + \beta_6(\text{usage}) + \beta_7(\text{usage}^2) + \beta_8(\text{usage}^3) + \beta_9(\text{age}) + \beta_{10}(\text{age}^2) + \beta_{11}(\text{age}^3) + \beta_{12}(\text{usage} \times \text{age}) + \beta_{13}(\text{usage} \times \text{age}^2) + \beta_{14}(\text{usage}^2 \times \text{age})$$

Recall that, in addition to nine subsystem failure variables, we had four part failure variables (low-, medium-, high-, and very-high-priced part failures during the study period) that constituted dependent variables in our Subsystem Study. In each of these 13 regressions, the full model was, except for the dependent variable, identical to the full model for the Tank Study. Once again, we used backward elimination to find the most appropriate model for each subsystem or part group. This procedure allowed different subsystems and part groups to have distinct regression models.[18] We supplemented the regressions with plots of predicted mean subsystem or part failures by age and by usage.

---

[18]For example, depending on their respective likelihood ratio tests, one subsystem might warrant a quadratic age term, while another subsystem might not.

Chapter Three

RESULTS

TANK STUDY RESULTS

Negative binomial regression results for the Tank Study appear in Table 3.1. Parameter estimates (regression coefficients) appear in the column labeled "β," and corresponding standard errors appear in the column labeled "s.e." The "t" column contains Wald t-statistics.

Both tank age and tank usage were found to have statistically significant effects: Age had a log-linear relationship with mean failures, and usage had a log-quadratic relationship with mean failures. We describe these effects as "log-linear" and "log-quadratic," rather than simply "linear" and "quadratic," because the regression coefficients reflect the impact of predictors on the natural log of mean failures.

Plots of the observed effects facilitate interpretation. Figures 3.1 through 3.3 illustrate how age and usage were related to predicted mean failures. Figure 3.1 shows that predicted mean failures increase with age at a compound annual growth rate of about 5 percent (with ± 2 percent at 95 percent confidence). Figure 3.2 shows 95 percent confidence bars for the Location 1 age-failure curve; the bars widen at the tail of the curve, as the sample contained fewer tanks at the end of the age range. This widening means that predictions based on earlier portions of the curve are likely to be more accurate than predictions based on the tail region. Figure 3.3 confirms that predicted mean failures increase with usage (kilometers driven is a good predictor of failures); also, it suggests that as usage increases, failures increase at a decreasing rate, which is indicated by the concave form of the curves.

Table 3.1

Negative Binomial Regression of Tank Failures on Age, Usage, and Location Variables (N = 1,567)

	Tank Failures		
Predictor Variables	β	s.e.	t
Age	.04848000	.00915600	5.30***
Accumulated usage	.00062840	.00006174	10.18***
(Accumulated usage)^2	–.00000046	.00000008	–5.35***
Location 2	–.38010000	.13090000	–2.90**
Location 3	–.14110000	.09628000	–1.47
Location 4	–.49200000	.08355000	–5.89***
Location 5	–.27890000	.10840000	–2.57*
Location 6	–.01741000	.07029000	–0.24
Null deviance: 1,893.6 on 1,566 degrees of freedom. Residual deviance: 1,753.0 on 1,558 degrees of freedom. Dispersion parameter: 5.1833.			

NOTE: Continuous predictor variables were mean-centered. Also, ln(updays) was treated as an offset variable in the regression; this treatment is equivalent to making ln(updays) a predictor variable with $\beta = 1$.

*$p < .05$.

**$p < .01$.

***$p < .001$.

Although not the focus of this study, it is noteworthy that some locations, controlling for age and usage, had more M1 failures than others. The different intercepts of the curves in Figure 3.1 capture distinctions among locations. Also note that the intercepts in Figure 3.3 are greater than zero, meaning that kilometers driven explains only a portion of a tank's failures. Consistent with the fact that significant numbers of failures were recorded during months with zero usage, the model indicates that failures are likely to occur (or at least be recorded) during periods of nonuse. Because the only usage-based explanatory variable available was kilometers driven, it is likely that the positive intercepts also stem from other tank activities, such as firing rounds or idling (e.g., running the tank engine while the tank is not moving in a defensive position). Such "hidden" explanatory variables can generate failures but no increase in kilometers. (These variables are termed hidden because they were not available for the study.)

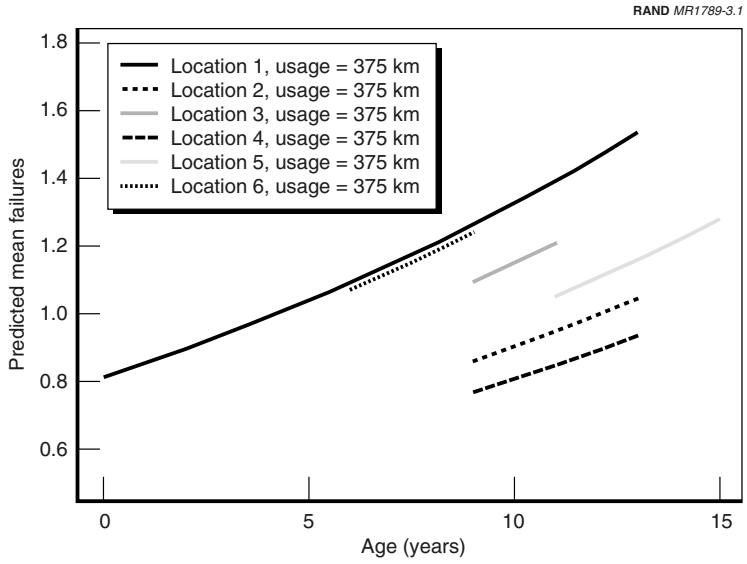


Figure 3.1—Predicted Mean Failures (over 180 days) by Tank Age

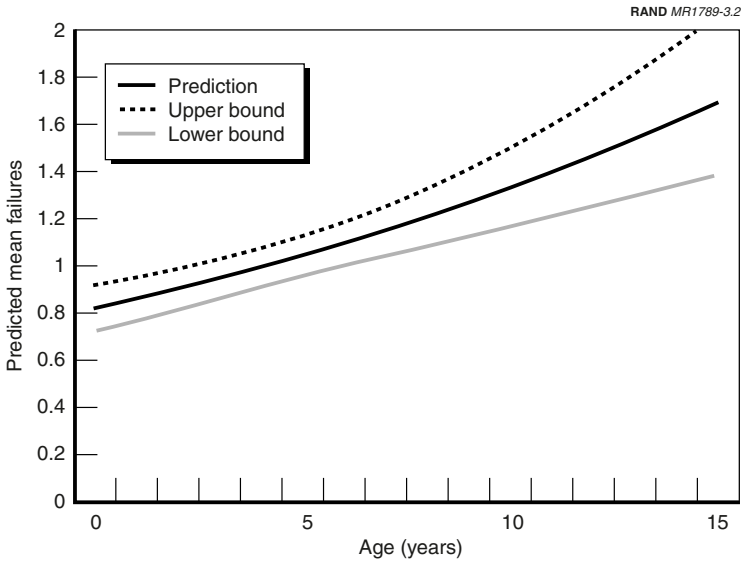


Figure 3.2—Predicted Mean Failures by Age at Location 1, with 95 percent Confidence Bars (180 days, usage = 375 km)

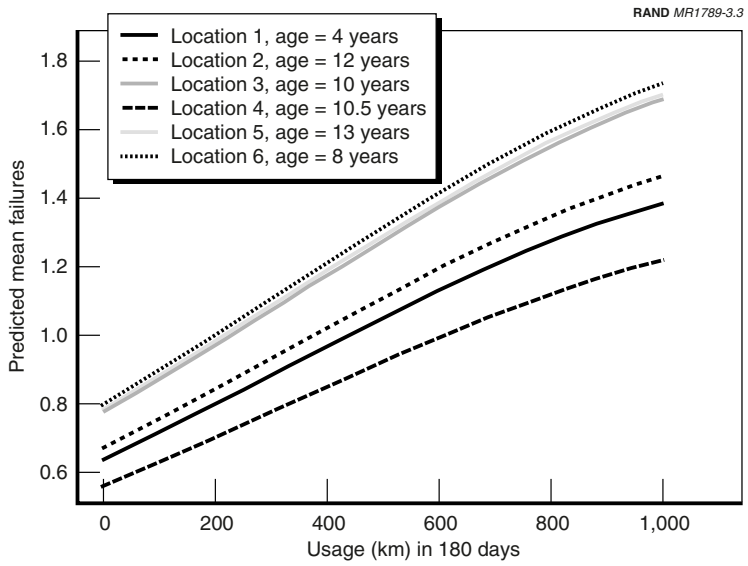


Figure 3.3—Predicted Mean Failures (over 180 days) by Tank Usage

Diagnostic plots—including plots of deviance residuals versus usage, age, and fitted values (natural log of predicted mean failures), a normal probability plot, and a Cook's distance plot—suggest that the data in this study satisfy the model assumptions. The model has a slight tendency to overestimate failures, but the amount of overestimation is negligible.

Appendix C displays the results of a cross-validation procedure, which suggests the model is likely to yield valid predictions with other, similar datasets.

SUBSYSTEM STUDY RESULTS

The backward elimination procedure yielded different negative binomial regression models for different subsystems. Table 3.2 summarizes age and usage effects on all subsystems. Checkmarks indicate all terms that appeared in the final model. For example, the Hull had log-quadratic age and usage effects on failures. Although only Age^2, Usage, and $Usage^2$ had significant effects, we have also checked the Age column because that term appears in the final model.

Table 3.2

Summary of Subsystem Age and Usage Effects (Terms in Final Model)

	Age	Age^2	Age^3	Usage	$Usage^2$	Age × Usage	Age^2 × Usage	Age × $Usage^2$
Hull	✓	✓		✓	✓			
Turret	✓			✓	✓	✓		
Chassis	✓	✓		✓	✓	✓	✓	✓
Electrical	✓	✓	✓	✓	✓			
Fire control	✓	✓	✓*	✓	✓	✓		
Hardware	✓	✓		✓	✓	✓	✓	
Power train	✓	✓		✓	✓			
Hydraulic	✓			✓				
Gun	✓			✓	✓			

*As will be discussed later, it is possible that the cubic effect seen for fire control could be an artifact of combining M1A1 and M1A2 data, given that the two variants have distinct fire control components with different failure rate characteristics.

Tables 3.3 through 3.11 present regression results corresponding to each subsystem.[1] Table 3.3 shows the log-quadratic age and usage effects on hull failures. Table 3.4 indicates that age had a positive, log-linear effect and usage had a log-quadratic effect on turret failures. There was also a significant interaction effect of age and usage on turret failures.

Table 3.5 shows that age and usage each had log-quadratic effects on chassis failures, and they had interaction effects as well. In the case of electrical failures, age had a log-cubic effect and usage had a log-quadratic effect (Table 3.6).

[1]The Cook's distances for each subsystem were small in magnitude, indicating that none of the tanks in the dataset imparted excessive influence on estimation of the subsystem model coefficients. Thus, early-life tanks possibly experiencing "break-in" or "infantile failure" phenomena are not inducing distortions in the estimated models.

In the case of fire control failures, age had a log-cubic effect, while usage had a log-quadratic effect (Table 3.7). Age and usage also had an interaction effect on fire control failures. As we will discuss later (see the subsection "Interpretation of Subsystem Results"), results for the fire control subsystem are more likely to be confounded by differences in tank type, which we could not include as a control variable in this study, than are results for other subsystems.

Age and usage had separate log-quadratic effects as well as interaction effects on hardware failures (Table 3.8). Similarly, both had log-quadratic effects on power train failures (Table 3.9). Age and usage had positive log-linear effects on hydraulic failures (Table 3.10). In the case of gun failures, age had a positive log-linear effect while usage had a log-quadratic effect (Table 3.11).

Table 3.3

Negative Binomial Regression of Hull Failures on Age, Usage, and Location Variables (N = 1,480)

	Hull Failures		
Predictor Variables	β	s.e.	t
Age	.01358000	.02524000	0.54
$(\text{Age})^2$	–.01428000	.00488300	–2.92**
Accumulated usage	.00055160	.00010300	5.35***
$(\text{Accumulated usage})^2$	–.00000027	.00000011	–2.40*
Location 2	.12370000	.18550000	0.67
Location 3	–.27710000	.14970000	–1.85†
Location 4	–.64990000	.13200000	–4.92***
Location 5	.10440000	.21390000	0.49
Location 6	–.14950000	.11950000	–1.25
Null deviance: 1,611.3 on 1,479 degrees of freedom. Residual deviance: 1,504.4 on 1,470 degrees of freedom. Dispersion parameter: 8.6032.			

NOTE: Continuous predictor variables were mean-centered. Also, ln(updays) was treated as an offset variable in the regression; this treatment is equivalent to making ln(updays) a predictor variable with $\beta = 1$.

†$p < .10$.

*$p < .05$.

**$p < .01$.

***$p < .001$.

Table 3.4

Negative Binomial Regression of Turret Failures on Age, Usage, and Location Variables (N = 1,480)

	Turret Failures		
Predictor Variables	β	s.e.	t
Age	.04762000	.01670000	2.85**
Accumulated usage	.00059670	.00011820	5.05***
(Accumulated usage)^2	–.00000063	.00000017	–3.76***
Age × Accumulated usage	.00007645	.00002951	2.59**
Location 2	–.61180000	.02393000	–2.56*
Location 3	–.69870000	.18840000	–3.71***
Location 4	–.80350000	.15400000	–5.22***
Location 5	–.42950000	.19080000	–2.25*
Location 6	–.29700000	.12540000	–2.37*
Null deviance: 1,418.0 on 1,479 degrees of freedom. Residual deviance: 1,364.0 on 1,470 degrees of freedom. Dispersion parameter: 3.4654.			

NOTE: Continuous predictor variables were mean-centered. Also, ln(updays) was treated as an offset variable in the regression; this treatment is equivalent to making ln(updays) a predictor variable with $\beta = 1$.

*$p < .05$.

**$p < .01$.

***$p < .001$.

Table 3.5

Negative Binomial Regression of Chassis Failures on Age, Usage, and Location Variables (N = 1,480)

	Chassis Failures		
Predictor Variables	β	s.e.	t
Age	.05929000	.05027000	1.18
$(\text{Age})^2$	–.03487000	.00998500	–3.49***
Accumulated usage	.00090160	.00026780	3.37***
$(\text{Accumulated usage})^2$	–.00000064	.00000030	–2.13*
Age × Accumulated usage	.00003235	.00007145	0.45
$(\text{Age})^2 \times$ Accumulated usage	–.00003759	.00001657	–2.27*
Age × $(\text{Accumulated usage})^2$	–.00000024	.00000011	–2.30*
Location 2	–.53860000	.03967000	–1.36
Location 3	–1.24300000	.33590000	–3.70***
Location 4	–1.07600000	.24860000	–4.33***
Location 5	.10360000	.38520000	0.27
Location 6	–.38690000	.21710000	–1.78†
Null deviance: 933.51 on 1,479 degrees of freedom. Residual deviance: 849.72 on 1,467 degrees of freedom. Dispersion parameter: 6.1212.			

NOTE: Continuous predictor variables were mean-centered. Also, ln(updays) was treated as an offset variable in the regression; this treatment is equivalent to making ln(updays) a predictor variable with $\beta = 1$.

†$p < .10$.

*$p < .05$.

**$p < .01$.

***$p < .001$.

Table 3.6

Negative Binomial Regression of Electrical Failures on Age, Usage, and Location Variables (N = 1,480)

	Electrical Failures		
Predictor Variables	β	s.e.	t
Age	–.00230100	.03887000	–0.06
(Age)^2	.02078000	.00083020	2.50*
(Age)^3	.00379300	.00121000	3.14**
Accumulated usage	.00033780	.00014250	2.37*
(Accumulated usage)^2	–.00000069	.00000020	–3.37***
Location 2	–.80680000	.29530000	–2.73†
Location 3	–.39890000	.20510000	–1.95**
Location 4	–.80330000	.17810000	–4.51***
Location 5	–1.03400000	.32370000	–3.19**
Location 6	–.46150000	.17030000	–2.71**
Null deviance: 1,281.3 on 1,479 degrees of freedom. Residual deviance: 1,227.1 on 1,469 degrees of freedom. Dispersion parameter: 2.621.			

NOTE: Continuous predictor variables were mean-centered. Also, ln(updays) was treated as an offset variable in the regression; this treatment is equivalent to making ln(updays) a predictor variable with $\beta = 1$.

†$p < .10$.

*$p < .05$.

**$p < .01$.

***$p < .001$.

Table 3.7

Negative Binomial Regression of Fire Control Failures on Age, Usage, and Location Variables (N = 1,480)

	Fire Control Failures		
Predictor Variables	β	s.e.	t
Age	–.089440000	.05316000	–1.68†
$(\text{Age})^2$	.03304000	.01120000	2.95**
$(\text{Age})^3$	.00495300	.00161000	3.08**
Accumulated usage	.0007997	.00020730	3.86***
$(\text{Accumulated usage})^2$	–.00000101	.00000028	–3.54***
Age × Accumulated usage	.00012000	.00004521	2.66**
Location 2	–.17680000	.36940000	–0.48
Location 3	–.23640000	.30110000	–0.79
Location 4	–.89800000	.28070000	–3.20***
Location 5	–.76310000	.46680000	–1.64
Location 6	–.55500000	.24520000	–2.26*
Null deviance: 1,016.81 on 1,479 degrees of freedom. Residual deviance: 916.06 on 1,468 degrees of freedom. Dispersion parameter: 2.302.			

NOTE: Continuous predictor variables were mean-centered. Also, ln(updays) was treated as an offset variable in the regression; this treatment is equivalent to making ln(updays) a predictor variable with $\beta = 1$.

†$p < .10$.

*$p < .05$

**$p < .01$

***$p < .001$

Table 3.8

Negative Binomial Regression of Hardware Failures on Age, Usage, and Location Variables (N = 1,480)

	Hardware Failures		
Predictor Variables	β	s.e.	t
Age	.11020000	.03332000	3.31***
$(\text{Age})^2$	–.01773000	.00727700	–2.44*
Accumulated usage	.00069010	.00017430	3.96***
$(\text{Accumulated usage})^2$	–.00000041	.00000019	–2.09**
Age × Accumulated usage	–.00002486	.00004556	–0.06
$(\text{Age})^2 \times$ Accumulated usage	–.00002350	.00001141	–2.06*
Location 2	–.85450000	.30720000	–2.78**
Location 3	–.54990000	.20530000	–2.68**
Location 4	–.63510000	.16650000	–3.81***
Location 5	–.12170000	.27230000	–0.45
Location 6	.18490000	.16160000	1.14
Null deviance: 1,398.4 on 1,479 degrees of freedom. Residual deviance: 1,260.3 on 1,468 degrees of freedom. Dispersion parameter: 4.3714.			

NOTE: Continuous predictor variables were mean-centered. Also, ln(updays) was treated as an offset variable in the regression; this treatment is equivalent to making ln(updays) a predictor variable with $\beta = 1$.

*$p < .05$.

**$p < .01$.

***$p < .001$.

Table 3.9

Negative Binomial Regression of Power Train Failures on Age, Usage, and Location Variables (N = 1,480)

	Power Train Failures		
Predictor Variables	β	s.e.	t
Age	.02112000	.02753000	0.77
$(\text{Age})^2$	–.01113000	.00530700	–2.10*
Accumulated usage	.00051920	.00012100	4.63***
$(\text{Accumulated usage})^2$	–.00000028	.00000012	–2.31*
Location 2	.28180000	.01953000	1.44
Location 3	–.11970000	.15810000	–0.76
Location 4	–.61030000	.14420000	–4.23***
Location 5	.01399000	.23480000	0.06
Location 6	–.14170000	.13210000	–1.07
Null deviance: 1,528.9 on 1,479 degrees of freedom. Residual deviance: 1,445.1 on 1,470 degrees of freedom. Dispersion parameter: 9.84.			

NOTE: Continuous predictor variables were mean-centered. Also, ln(updays) was treated as an offset variable in the regression; this treatment is equivalent to making ln(updays) a predictor variable with $\beta = 1$.

*$p < .05$.

**$p < .01$.

***$p < .001$.

Table 3.10

Negative Binomial Regression of Hydraulic Failures on Age, Usage, and Location Variables (N = 1,480)

	Hydraulic Failures		
Predictor Variables	β	s.e.	t
Age	.14040000	.02028000	6.92***
Accumulated usage	.00048190	.00009638	5.00***
Location 2	–1.05900000	.28000000	–3.79***
Location 3	–.42770000	.18860000	–2.27*
Location 4	–.65970000	.15320000	–4.31***
Location 5	–.59500000	.19830000	–3.00**
Location 6	.05966000	.12910000	0.46
Null deviance: 1,351.9 on 1,479 degrees of freedom. Residual deviance: 1,274.0 on 1.472 degrees of freedom. Dispersion parameter: 11.0931.			

NOTE: Continuous predictor variables were mean-centered. Also, ln(updays) was treated as an offset variable in the regression; this treatment is equivalent to making ln(updays) a predictor variable with $\beta = 1$.

*$p < .05$.

**$p < .01$.

***$p < .001$.

Table 3.11

Negative Binomial Regression of Gun Failures on Age, Usage, and Location Variables (N = 1,480)

	Gun Failures		
Predictor Variables	β	s.e.	t
Age	.17880000	.02744000	6.52***
Accumulated usage	.00068350	.00016620	4.11***
(Accumulated usage)^2	–.00000059	.00000024	–2.51*
Location 2	–.93860000	.33910000	–2.77**
Location 3	–.99870000	.28440000	–3.51***
Location 4	–.59860000	.19680000	–3.04**
Location 5	–.92620000	.26680000	–3.47***
Location 6	.32090000	.16280000	1.97*
Null deviance: 1,039.82 on 1,479 degrees of freedom. Residual deviance: 961.84 on 1,471 degrees of freedom. Dispersion parameter: 1.6232.			

NOTE: Continuous predictor variables were mean-centered. Also, ln(updays) was treated as an offset variable in the regression; this treatment is equivalent to making ln(updays) a predictor variable with $\beta = 1$.

*$p < .05$.

**$p < .01$.

***$p < .001$.

Interpretation of Subsystem Results

Figure 3.4 displays predicted mean failures versus age for all second-tier subsystems at Location 1. When interpreting the figure, it is helpful to consider both the absolute and relative change in subsystem failures with age. In particular, the absolute change represents the opportunity for improving the failure rate that might be possible for tanks of various ages by systematically replacing selected components with new ones. The difference between the peak failure rate and the age-0 (new tank) failure rate captures the absolute change in failures with age. (For such examinations we have elected to use the first 12 years of data because we have less confidence in the shapes of the curves' right tails, which are based on fewer data points.) The ratio of the peak rate to the age-0 rate captures the relative change—i.e., how much the subsystem degraded relative to its condition at 0 years of age. For example, chassis has a failure peak of 0.177, which

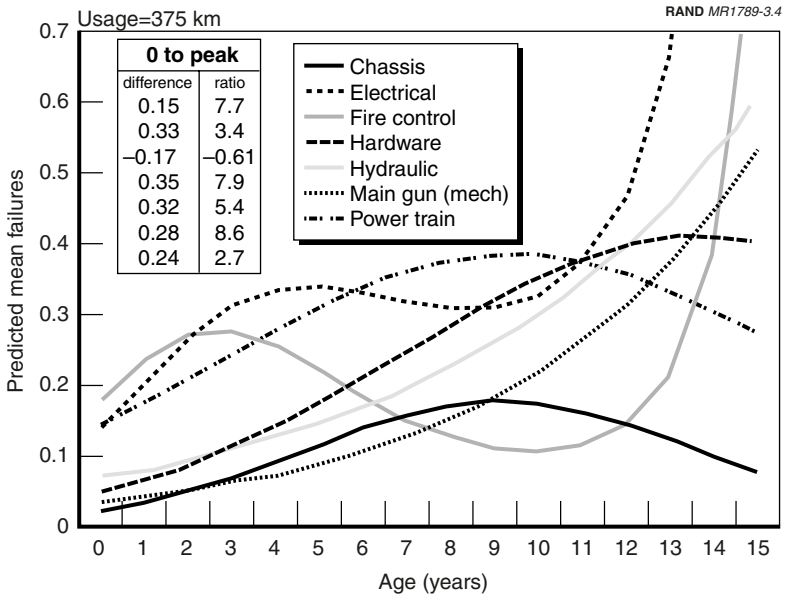


Figure 3.4—Predicted Mean Failures of Second-Tier Subsystems by Age (Location 1, 180 days)

occurs at 9 years of age. Its failure minimum is about 0.023, which occurs at 0 years of age. The difference between the peak and 0-age rates is 0.177 – 0.023, or 0.154. The ratio is 0.177/0.023, or 7.70. By illustrating such absolute and relative changes in failures, Figure 3.4 facilitates comparison of age-failure relationships across subsystems. The graph also shows which subsystems drive the failure rate of new tanks—specifically, fire control, electrical, and power train—and provides information about the contribution of each subsystem to the overall failure rate; this information suggests where engineering redesign might have the biggest impact.

The preceding regression tables and corresponding plots demonstrate that age is a stronger failure predictor for some subsystems than for others. The electrical, hardware, hydraulic, and main gun subsystems experienced the greatest *absolute* failure rate increases due to aging. While the model shows a steep increase at the tail for fire control, we have less confidence in this portion of the curve be-

cause of the limited data discussed earlier. Further study as tanks continue to age will be necessary to determine if there really is a steep increase at about the 14-year point. (The shape of the curve will also be discussed further in the section "Sensitivity Analysis Results" below.)

The chassis, hardware, hydraulic, and main gun subsystems experienced the greatest *relative* increases due to aging. Because the electrical subsystem had a high initial (age-0) failure rate, the relative increase in its failure rate was low, despite a high absolute increase. Because the chassis subsystem had a low initial failure rate, the relative increase in its failure rate was high, despite a low absolute increase. The chassis' low initial failure rate mutes the overall impact of the relative increase in its failure rate.

Although the power train subsystem had a high initial failure rate, it had low absolute and relative failure rate increases due to aging. Even this relatively high initial failure rate, though, may be understated as a result of part-ordering practices. Because of the way many units manage tank engines, they often do not order engines against non-mission-capable (NMC) work orders; the work order appears on the deadline report without any part orders placed against it. Since failures are classified based on the parts ordered against NMC work orders, power train failures (specifically, those associated with engine replacements) are probably undercounted in this analysis. This is likely to be the case for tanks of all ages.

The effect of age on fire control failures may, in part, reflect differences between the M1A1 and M1A2 tanks. Although we could not control for tank variant in our regressions (due to its high correlation with age), we were able to run separate M1A1 and M1A2 analyses. These subsample analyses suggest that the fire control curve represents a convolution of two distinct failure patterns. The data suggest that age generally increases both M1A2 and M1A1 failures; however, the data also suggest that the digital fire control system of like-new M1A2s has a higher failure rate than the analog fire control system of like-new M1A1s. Because the M1A2s in our sample were all younger than the M1A1s, the M1A2s influenced the first part of the combined fire control curve, while the M1A1s influenced the second part. An explanation for the cubic shape of the combined curve is that the curve starts high to account for M1A2 fire control failures and then

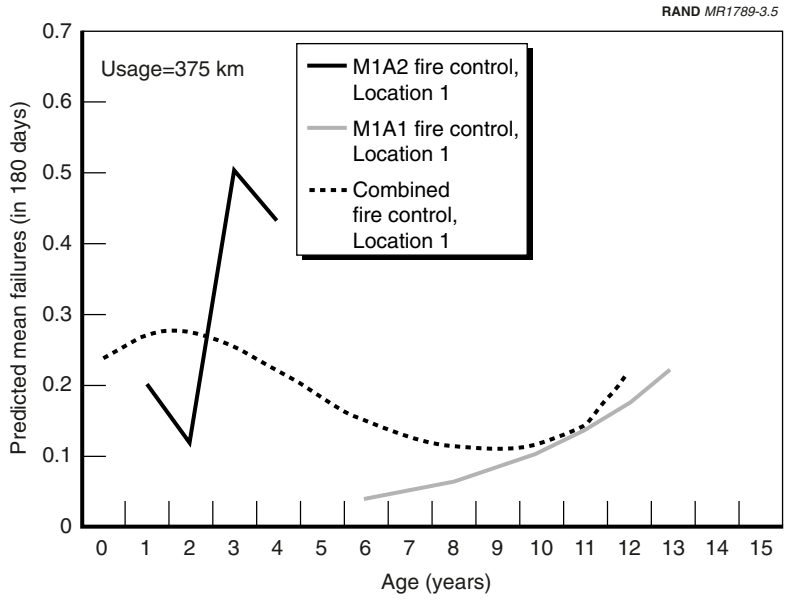


Figure 3.5—Predicted Mean Fire Control Failures by Age for the M1A1s, M1A2s, and Combination of M1A1s and M1A2s (Location 1, 180 days)

turns down to account for the lower fire control failure rate in young M1A1s before turning back up to account for the aging effect.[2] Figure 3.5 displays the M1A2, M1A1, and combined fire control curves.

Note that even though age had a log-linear effect on tank failures, it had higher-order effects on some subsystem failures. A quadratic or cubic age effect means that the relationship between age and the change in failure rate (i.e., the slope of the age-failure curve) changes over time. A sharp increase in the number of components reaching their wear-out regions, and the subsequent renewal (i.e., replacement) of those components, may explain such changes. Once items begin to fail, renewal begins. For a component exhibiting a visible wear pattern, as a significant portion of the fleet experiences a failure

[2]While the data do suggest a cubic model for M1A2s (which could help explain the cubic shape of the combined curve), confidence in the shape of the M1A2 curve is low because of the limited M1A2 age range.

of the component and has the component replaced, the wear pattern begins anew. If the entire fleet were to experience a failure for a given component in a short period of time, the fleet failure rate would quickly increase during this wear-out region and then drop significantly. The wider the wear-out region, the more the age of a given component will be distributed across a fleet. In this case the fleet failure rate for the component may gradually increase and at some point begin a gradual decrease. Ultimately, with renewal, a component's age distribution may become uniform across a "cohort" of tanks—say one particular year of manufacture. At this point, the failure rate with respect to that component will then stay the same as the cohort continues to age. As more and more parts reach this point through renewal for a cohort of tanks, the aging effect should tail off.[3]

It is possible for a subsystem to experience several such wear-out/renewal cycles; cubic age-failure curves (for the electrical and fire control subsystems) may reflect this pattern.[4] The steepness of these curves' tail regions must be interpreted with caution, though, as they are based on fewer data points than other portions of the curves. In practice, most fleets will not remain in the total Army fleet for a lengthy enough period to see many or even multiple wear-out/renewal cycles. Thus, when the age-failure curve for a subsystem shows a plateau—or even a decrease—in the failure rate, it is probably an indication that sufficient fleet renewal for the subsystem has occurred to limit any further increase in the fleet failure rate. The renewal phenomenon should also be seen at the tank level. The model results suggest that at the tank level, it does not occur within the first 14 years of tank age, the range of our data.

Figure 3.6 displays predicted mean failures versus usage for all second-tier subsystems at Location 1.

---

[3]For many complex parts, the Army actually uses rebuilt parts for which there is uncertainty about whether they are "like-new." This could affect the relationship we see between failures and age, as it would contribute to higher failure rates of tanks having such replacement parts and would lessen the benefit of renewal.

[4]The electrical subsystem curve could reflect several sequential renewal cycles, with the sparse data at the tail region exaggerating the effect. Alternatively, the curve could reflect two separate wear-out regions corresponding to different sets of electrical components.

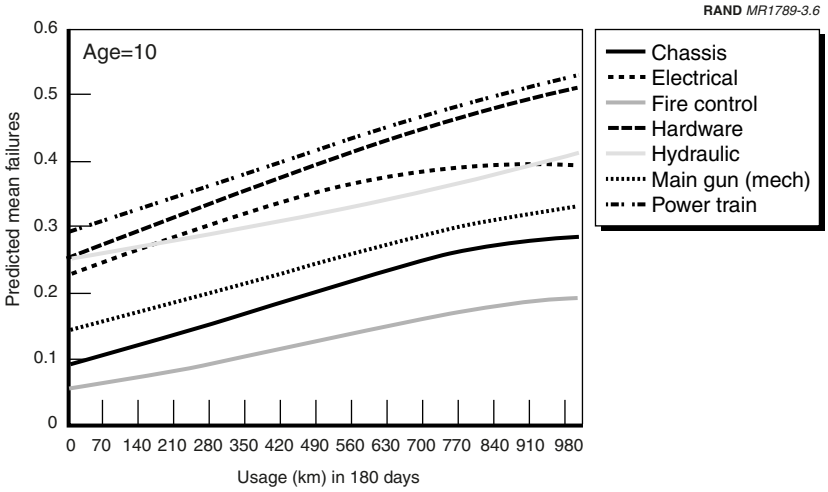


Figure 3.6—Predicted Mean Failures of Second-tier Subsystems by Usage (Location 1, 180 days)

Some subtle differences in usage effects across subsystems were seen as well. Usage had log-linear effects in some cases but log-quadratic effects in others. In the former cases, it is likely that usage—more specifically, kilometers traveled—was the primary explanatory variable; thus the linear effect. In contrast, the quadratic effects may reflect not only kilometers traveled, but also other, unmeasured types of usage. For example, the observed effect of usage on electrical failures is most likely being influenced by rounds fired, idling of the tank engine, operation of electrical systems while the engine is not running, or perhaps the *frequency* (rather than the absolute amount of usage) or cycles of usage. While firing rounds and idling place little or no demand on the chassis, they place demands on the electrical subsystem. Had we been able to control for these alternative usage types, we might have observed that kilometers traveled had a log-linear effect (or perhaps no effect!), rather than a log-quadratic effect.[5] Furthermore, we might have seen a second failure distribution based on rounds fired and a third based on idle time.

[5]To the extent that there is correlation among the usage types, kilometers driven may simply be serving as a proxy for other usage types when it explains failures for

It is noteworthy that, as in the Tank Study, location was a significant predictor in the Subsystem Study. In each subsystem plot of predicted mean failures versus age (see Appendix D for separate subsystem plots), the multiple intercepts reflect the effects of different locations.

When we group part orders by subsystem and plot deadlining part orders by age (Figure 3.7), the plot suggests multiple shifts in aging effects for some subsystems, consistent with the nonlinearity seen in earlier subsystem curves. We see these shifts even without controlling for location or absolute usage, which affect the failure rate.[6]

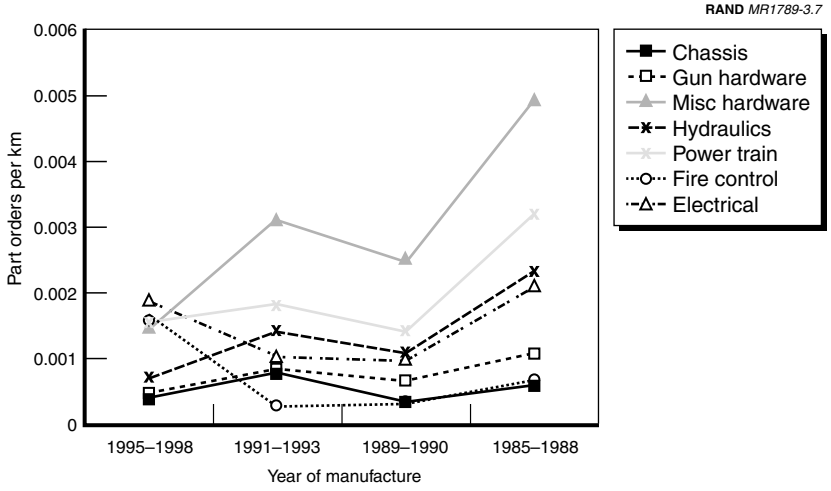


Figure 3.7—Total Parts Demand (during Study Period) per Subsystem by Age

components or subsystems that are generally not associated with movement. Hence the apparent log-quadratic effect of kilometers driven might disappear if other usage types could be included in the models.

[6]Total part orders for a given subsystem and age group were divided by total kilometers traveled (during the study period) by tanks in the corresponding age group.

Rebuild Versus Upgrade Candidates

Simple plots of deadlining part orders for the various subsystems may point out which components *within* subsystems should be rebuilt, upgraded, or left unchanged. As an example, we constructed a plot of deadlining part orders (per million kilometers of usage) versus M1 age for a sample of parts. The plot (Figure 3.8) shows that orders for some parts clearly increased with age; such parts are good candidates for the rebuild portion of the Abrams RECAP program. In other words, simply replacing these components in old tanks is likely to improve the overall failure rate for some period of time. In contrast, parts with relatively high-order levels *regardless of tank age*, which are readiness drivers, or expensive parts with moderate failure levels *regardless of tank age*, which are cost drivers, are good upgrade candidates. By "upgrade" we mean an engineering or repair-process redesign that reduces the failure rate of a "new" part.[7] Still other parts require neither rebuild nor upgrade; demand for them was very low, making age effects, if any, less important.

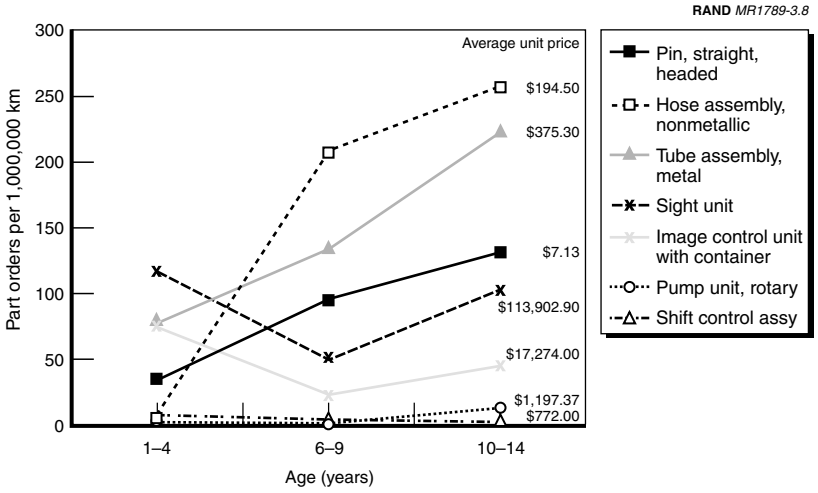


Figure 3.8—Parts Demand per Part Type by Age

[7] In this context a new part is either brand new or rebuilt.

The Link Between Age-Failure Relationships and Part Prices

Another portion of our subsystem analysis entailed four regressions of low-priced, medium-priced, high-priced, and very-high-priced part orders (i.e., orders for parts within subsystems) on predictor variables. Tables 3.12 through 3.15 display the results of those regressions (with the backward elimination procedure), and Figure 3.9 displays corresponding part failure versus age curves. As the results indicate, the age-failure relationship is generally strongest for lower-priced parts, which tend to be simpler parts with wear-related dominant failure modes. Parts in the two middle-price categories appear to have a moderate age-failure relationship, and the most expensive parts have little or no age-failure relationship.

Table 3.12

Negative Binomial Regression of Low-Priced Part Failures on Age, Usage, and Location Variables (N = 1,480)

	Low-Priced Part Failures		
Predictor Variables	β	s.e.	t
Age	.22156	.02777	7.98***
Location 2	–1.37102	.37547	–3.65***
Location 3	–.55070	.26179	–2.10*
Location 4	–.60451	.22055	–2.74**
Location 5	–.85547	.29972	–2.85**
Location 6	.22408	.19642	1.14
Null deviance: 1,180.8 on 1,479 degrees of freedom. Residual deviance: 1,101.1 on 1,473 degrees of freedom. Dispersion parameter: 0.2008.			

NOTE: Continuous predictor variables were mean-centered. Also, ln(updays) was treated as an offset variable in the regression; this treatment is equivalent to making ln(updays) a predictor variable with $\beta = 1$.

*$p < .05$.

**$p < .01$.

***$p < .001$.

Table 3.13

Negative Binomial Regression of Medium-Priced Part Failures on Age, Usage, and Location Variables (N = 1,480)

	Medium-Priced Part Failures		
Predictor Variables	β	s.e.	t
Age	.041921	.041886	1.00
$(\text{Age})^2$	.006583	.008657	0.76
$(\text{Age})^3$	.003669	.001361	2.70**
Location 2	–.431849	.284764	–1.52
Location 3	–.324781	.208311	–1.56
Location 4	–.533127	.175348	–3.04**
Location 5	–.677012	.322640	–2.10*
Location 6	.140302	.169610	0.83
Null deviance: 1,343.2 on 1,479 degrees of freedom. Residual deviance: 1,265.8 on 1,471 degrees of freedom. Dispersion parameter: 0.4245.			

NOTE: Continuous predictor variables were mean-centered. Also, ln(updays) was treated as an offset variable in the regression; this treatment is equivalent to making ln(updays) a predictor variable with $\beta = 1$.

*$p < .05$.

**$p < .01$.

***$p < .001$.

Table 3.14

Negative Binomial Regression of High-Priced Part Failures on Age, Usage, and Location Variables (N = 1,480)

	High-Priced Part Failures		
Predictor Variables	β	s.e.	t
Age	–.00202500	.04096000	–.049
$(\text{Age})^2$	.00168900	.00767600	0.22
$(\text{Age})^3$	.00442900	.00129700	3.42***
Accumulated usage	.00035430	.00012230	2.90**
$(\text{Accumulated usage})^2$	–.00000038	.00000015	–2.59**
Age × Accumulated usage	–.00008233	.00003531	2.33*
Location 2	–.51030000	.25520000	–2.00*
Location 3	–.63170000	.18740000	–3.37***
Location 4	–1.07900000	.16180000	–6.67***
Location 5	–.69900000	.28460000	–2.46*
Location 6	–.44350000	.16120000	–2.75**
Null deviance: 1,475.2 on 1,479 degrees of freedom. Residual deviance: 1,361.7 on 1,468 degrees of freedom. Dispersion parameter: 0.6015.			

NOTE: Continuous predictor variables were mean-centered. Also, ln(updays) was treated as an offset variable in the regression; this treatment is equivalent to making ln(updays) a predictor variable with $\beta = 1$.

*$p < .05$.

**$p < .01$.

***$p < .001$.

Table 3.15

Negative Binomial Regression of Very-High-Priced Part Failures on Age, Usage, and Location Variables (N = 1,480)

	Very-High-Priced Part Failures		
Predictor Variables	β	s.e.	t
Age	–.16800000	.04562000	–3.68***
$(\text{Age})^2$	.00418000	.00941800	0.44
$(\text{Age})^3$	.00426600	.00144900	2.95**
Accumulated usage	.00041950	.00014580	2.88**
$(\text{Accumulated usage})^2$	–.00000045	.00000019	–2.76**
Age × Accumulated usage	.00013180	.00003732	3.53***
Location 2	.87930000	.28150000	3.12**
Location 3	.38620000	.21180000	1.82†
Location 4	–.71610000	.20340000	–3.52***
Location 5	–.23910000	.35810000	–0.67
Location 6	–.81430000	.19170000	–4.25***
Null deviance: 1,314.0 on 1,479 degrees of freedom. Residual deviance: 1,173.4 on 1,468 degrees of freedom. Dispersion parameter: 0.5837.			

NOTE: Continuous predictor variables were mean-centered. Also, ln(updays) was treated as an offset variable in the regression; this treatment is equivalent to making ln(updays) a predictor variable with $\beta = 1$.

†$p < .10$.

*$p < .05$.

**$p < .01$.

***$p < .001$.

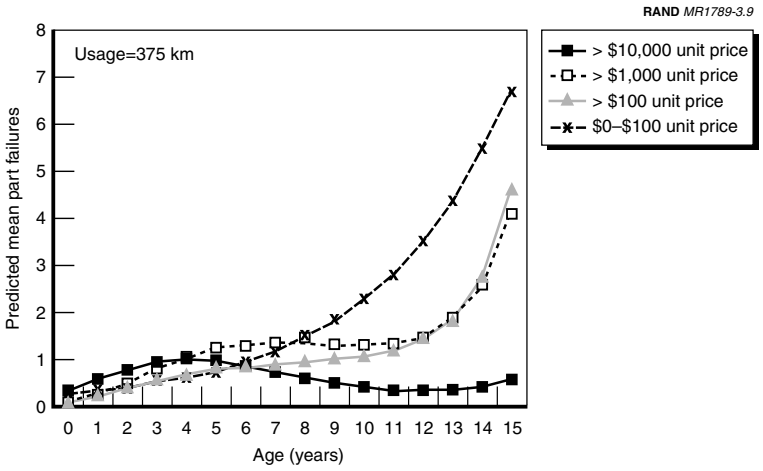


Figure 3.9—Predicted Mean Part Failures Versus Tank Age (Location 1, 180 days)

SENSITIVITY ANALYSIS RESULTS

Alternative Imputation Approach

Our first sensitivity analysis involved an alternative approach to coping with missing usage data. As described earlier, our primary imputation approach was mean substitution. While mean substitution (also known as mean imputation) is widely utilized, other more sophisticated imputation techniques are also available. Thus, we examined how the use of multiple imputation, rather than mean imputation, affected our findings.

We began the process by replacing each missing monthly usage reading via random draws from the same population used in the mean substitution method. In other words, when a tank was missing a usage reading for a particular month, we assigned it a reading that was randomly drawn from those tanks in its company that had complete data for that month. We did 10 such random draws, creating 10 datasets.

Next, for each of the datasets, we ran a negative binomial regression of failures on our predictor variables using the backward elimination technique described in the "Tank Study Analysis" subsection above. We then used the 10 regression results—specifically, the significance of their coefficients—as a guideline for proposing a final model structure (i.e., which terms should be included in the final model).[8] Finally, we employed three techniques (Schafer, 1997) to draw upon results of the 10 regressions and test the proposed final model against the full cubic model.[9] All three techniques supported the proposed final model.

The overall M1 model yielded by multiple imputation had the same structure as the model yielded by mean imputation. The subsystem models yielded by multiple imputation had the same structure as the mean imputation models in all but three cases: hull, power train, and hydraulic. The hull and power train multiple imputation models each had two additional terms: a cubic age term and an age × usage interaction term. The hydraulic multiple imputation model had four additional terms: a quadratic usage term, a quadratic age term, an age × usage interaction term, and an age-squared × usage interaction term. To facilitate comparison of multiple imputation and mean imputation results for those subsystems, we fit the multiple imputation model structure to the mean imputation dataset. Tables 3.16 through 3.18 show the results of those regressions. Figures 3.10 and 3.11 show the corresponding failure versus age and failure versus usage curves at Location 1 for the second-tier (power train and hydraulic) subsys-

---

[8]This step involved a judgment call. For example, if most of the 10 regressions had a particular coefficient that was significant, then our proposed final model included that term. On the other hand, if very few or none had that significant term, we excluded it from the final model. Note that this step focused on identifying those terms (e.g., age, activity, age × activity) that should be included in the proposed model, *not* what their coefficients should be.

[9]The first technique (Li, Raghunathan, and Rubin, 1991) combined covariance matrices and point estimates of coefficients from the 10 regressions and used those combined values to compute a Wald statistic for variables (terms) in the model. The second technique (Li, Meng, Raghunathan, and Rubin, 1991) combined Wald statistics from the 10 regressions to compute an F-test statistic for terms in the model. The third technique (Meng and Rubin, 1992) combined likelihood-ratio statistics from the 10 regressions. For each technique, an F-test based on the corresponding test statistic (Wald, combined Wald, or likelihood ratio) determined which variables could be removed from the model, allowing comparison of a full model to a proposed reduced model.

tems, and how they compare to those resulting from mean imputation. (The hull curves were very similar to the power train curves, as one might expect given that the power train is a key hull component.)

Table 3.16

Negative Binomial Regression of Hull Failures on Age, Usage, and Location Variables (N = 1,480), with Multiple Imputation Approach

	Hull Failures		
Predictor Variables	β	s.e.	t
Age	.14040000	.03461000	–1.04
$(\text{Age})^2$	–.00779900	.00646900	–1.21
$(\text{Age})^3$	.00221500	.00110100	2.01*
Accumulated usage	.00051380	.00010790	4.76***
$(\text{Accumulated usage})^2$	–.00000025	.00000011	–2.31*
Age × Accumulated usage	.00005284	.00002894	1.83†
Location 2	.07919000	.19870000	0.40
Location 3	–.23960000	.16110000	–1.49
Location 4	–.62930000	.14380000	–4.38***
Location 5	–.02531000	.24290000	–0.10
Location 6	–.19790000	.01282000	–1.54
Null deviance: 1,614.7 on 1,479 degrees of freedom. Residual deviance: 1,501.7 on 1,472 degrees of freedom. Dispersion parameter: 8.8973.			

NOTE: Continuous predictor variables were mean-centered. Also, ln(updays) was treated as an offset variable in the regression; this treatment is equivalent to making ln(updays) a predictor variable with $\beta = 1$.

†$p < .10$.

*$p < .05$.

**$p < .01$.

***$p < .001$.

Table 3.17

Negative Binomial Regression of Power Train Failures on Age, Usage, and Location Variables (N = 1,480), with Multiple Imputation Approach

	Power Train Failures		
Predictor Variables	β	s.e.	t
Age	–.02761000	.03791000	–0.73
$(\text{Age})^2$	–.00415000	.00709000	–0.59
$(\text{Age})^3$	.00223100	.00120200	1.86†
Accumulated usage	.00047910	.00011790	4.06***
$(\text{Accumulated usage})^2$	–.00000027	.00000012	–2.22*
Age × Accumulated usage	.00004407	.00003159	1.40
Location 2	.25240000	.20980000	1.20
Location 3	–.07361000	.17120000	–0.43
Location 4	–.57880000	.15760000	–3.67***
Location 5	–.12980000	.26840000	–0.48
Location 6	–.17890000	.14220000	–1.26
Null deviance: 1,530.8 on 1,479 degrees of freedom. Residual deviance: 1,442.6 on 1,468 degrees of freedom. Dispersion parameter: 10.0994.			

NOTE: Continuous predictor variables were mean-centered. Also, ln(updays) was treated as an offset variable in the regression; this treatment is equivalent to making ln(updays) a predictor variable with $\beta = 1$.

†$p < .10$.

*$p < .05$.

**$p < .01$.

***$p < .001$.

Table 3.18

Negative Binomial Regression of Hydraulic Failures on Age, Usage, and Location Variables (N = 1,480), with Multiple Imputation Approach

	Hydraulic Failures		
Predictor Variables	β	s.e.	t
Age	.11370000	.03398000	3.35***
$(\text{Age})^2$	–.00948500	.00726200	–1.31
Accumulated usage	.00079800	.00018930	4.22***
$(\text{Accumulated usage})^2$	–.00000040	.00000020	–2.05*
Age × Accumulated usage	.00001789	.00004630	–0.39
$(\text{Age})^2$ × Accumulated usage	–.00002018	.00001134	–1.78†
Location 2	–.93830000	.31050000	–3.02**
Location 3	–.53590000	.20600000	–2.60**
Location 4	–.71040000	.16860000	–4.21***
Location 5	–.45840000	.28570000	–1.61
Location 6	–.09955000	.16860000	–0.59
Null deviance: 1,368.0 on 1,479 degrees of freedom. Residual deviance: 1,280.0 on 1,468 degrees of freedom. Dispersion parameter: 17.2907.			

NOTE: Continuous predictor variables were mean-centered. Also, ln(updays) was treated as an offset variable in the regression; this treatment is equivalent to making ln(updays) a predictor variable with $\beta = 1$.

†$p < .10$.

*$p < .05$.

**$p < .01$.

***$p < .001$.

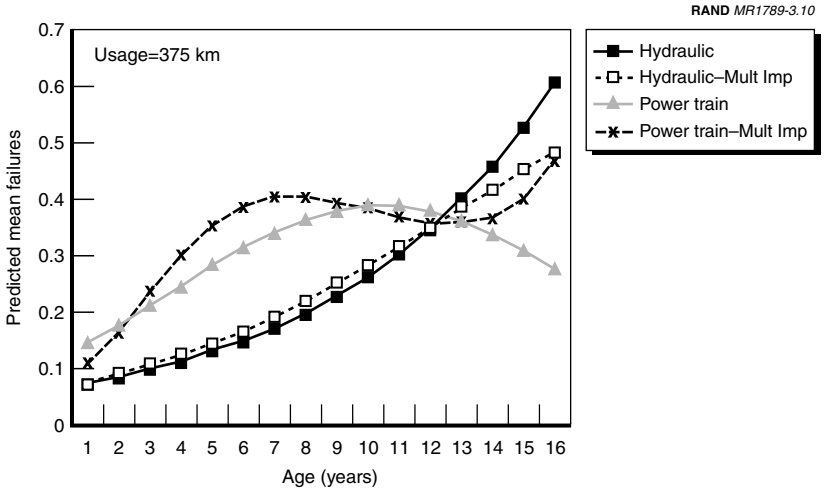


Figure 3.10—Predicted Mean Failures by Age for Hydraulic and Power Train Subsystems, Based on Multiple Imputation Models (Location 1, 180 days)

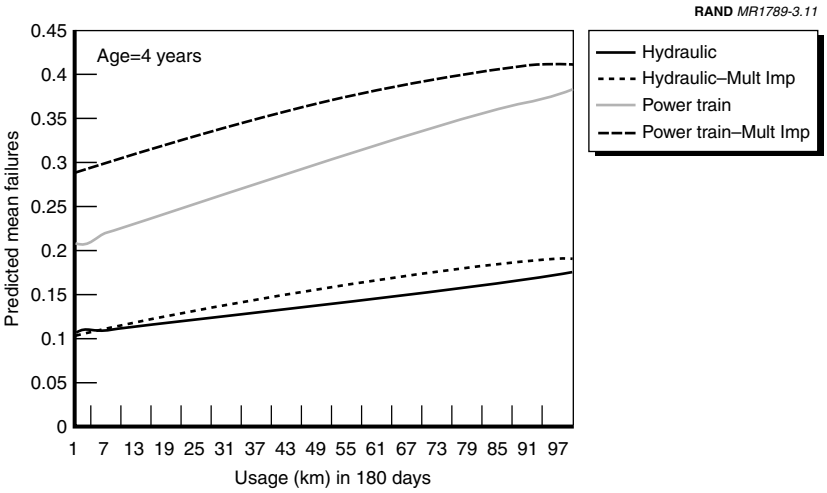


Figure 3.11—Predicted Mean Failures by Usage for Hydraulic and Power Train Subsystems, Based on Multiple Imputation Models (Location 1, 180 days)

As Figures 3.10 and 3.11 indicate, the hydraulic subsystem curves yielded by mean imputation and multiple imputation had considerable overlap. The power train models were more distinct, however; multiple imputation yielded a power train age–failure curve that is perhaps more plausible, without a sharp downturn in its tail region. In general, the similarity of M1 system and subsystem models resulting from multiple imputation and mean imputation suggests that mean imputation does not distort results significantly and may be a more practical, less cumbersome approach when one is analyzing data on a large number of systems. Nevertheless, these power train findings point to the importance of further analyses in some cases—in particular, when unexpected patterns appear.

It is also important to compare the confidence intervals associated with the multiple imputation and mean imputation models, as multiple imputation tends to capture true variability of estimates more accurately than mean imputation. For this step, we used Rubin's (1987) method for combining estimates from the 10 regressions based on the multiple imputation datasets.[10] Figure 3.12 shows confidence interval widths for multiple imputation estimates and mean imputation estimates for the overall M1 model. The small difference between the two curves suggests that there was little difference in confidence intervals resulting from the two techniques; this finding further suggests that mean imputation was sufficient for this study.

Additional Control Variable for Odometer Resets

We also considered the possibility that reset odometers from major maintenance events or even overhauls affected our findings. As a check, we added a control variable called "reset" to the negative binomial regressions in both the Tank Study and Subsystem Study to see if the low-odometer tanks differed from the higher-odometer tanks of a given age. In other words, do the low readings represent something other than simple resets? The value of this dummy variable was "Y" if a tank had at least one odometer reading that fell

[10]Rubin's (1987) method entailed (a) averaging the 10 sets of estimated mean failures to get overall estimates of mean failures versus age, and (b) combining within-dataset and between-dataset estimation variability to obtain confidence intervals for mean failures.

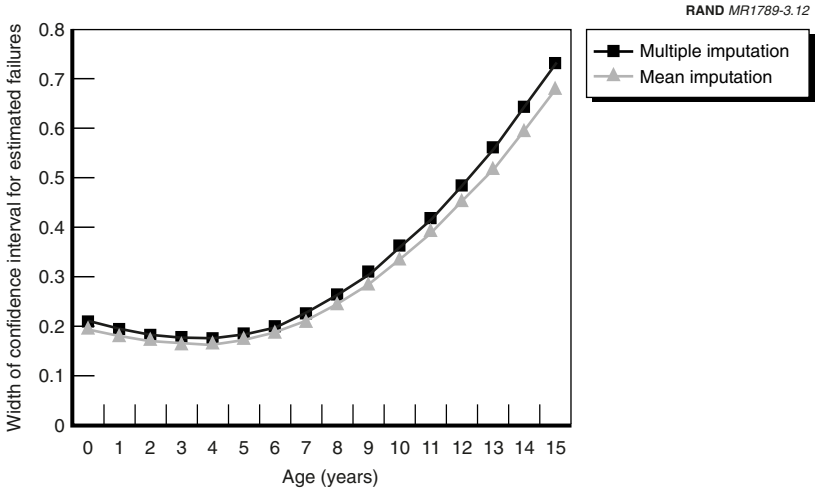


Figure 3.12—Confidence Interval Width by Age for Multiple Imputation and Mean Imputation Overall Tank Study Model

below the 25th percentile for its age group, according to the distribution shown in Figure 2.6; otherwise the value was "N." The variable did not have a significant impact on failures, and the backward elimination regression procedure yielded the same final model as that shown in Table 3.1. As cleaner overhaul and tank maintenance histories (including every part of replacement) become more available, more sophisticated tests of the impact of major maintenance events will be possible. Such tests may determine with greater certainty the degree to which maintenance events help account for the higher-order age effects we found.

Alternative Regression Techniques in the Tank Study

For our third sensitivity analysis, we performed the Tank Study and Subsystem Study analyses using Poisson regression, rather than negative binomial regressions. While the parameter estimates generated by the two techniques were virtually identical, the Poisson regressions yielded larger residual deviances. This discrepancy suggests that the negative binomial regressions addressed overdispersion in the data.

To further assess the robustness of our final Tank Study model, we performed ordinary least squares (OLS) regression on the data. When a Poisson-distributed variable (such as the number of failures per day) is measured for a long-enough period (i.e., a large number of days), the square root of that variable is approximately normal with constant variance.[11] Thus, we used the square root of tank failures as the dependent variable in the OLS regression. Stepwise OLS regression yielded a final model with the same structure and significant terms as our final negative binomial regression model. The standard error of the OLS regression was .68. Since OLS standard errors greater than .5 are consistent with overdispersion, this provides further support for our use of the negative binomial model.

Alternative Regression Techniques in the Subsystem Study

Another sensitivity analysis addressed the subsystem age effects shown in Figure 3.4. Several of the age-failure curves had downturns (power train and chassis) or sharp upturns (electrical and fire control) in their tail regions. Those end regions were based on relatively few data points, as only a small percentage of tanks in our sample exceeded 12 years of age. Consequently, the curves may have less validity for older tanks. Moreover, there is little theoretical rationale for the tails of the curves. Although it is reasonable to expect a slight, temporary reversal in the aging effect (due to renewal as parts are replaced), it is unlikely that increasing age will reduce tank failures for a lengthy period of time and do so at an increasing rate, as the quadratic curves suggest. Much of the quadratic curve is quite plausible, but the tail region advances the unlikely notion that very old tanks will have no failures at all.[12] Similarly, it is unlikely that failures rise as sharply as depicted in the tails of the cubic curves. In other words, the overall curves fit well, but the end regions are character-

---

[11] Let λ_i represent the number of failures per day. Suppose $X_i \sim \text{Poisson}(n_i * \lambda_i)$, where n_i is the number of updays for tank i and X_i is the number of failures in n_i updays. (That is, X_i has a Poisson distribution with mean $n_i * \lambda_i$.) Then, when n_i is large, $2*(X_i)^{1/2}$ is approximately normally distributed with mean = $2*(n_i * \lambda_i)^{1/2}$ and variance = 1. Thus, $(X_i)^{1/2}$ is approximately normally distributed with mean = $(n_i * \lambda_i)^{1/2}$ and variance = 1/4.

[12] The shape at the ends of the quadratic curves may have been distorted by the late appearance of the peak failure rate (relative to the total age range), since a disproportionately small number of tanks were available to fit the change in the trend.

ized by more uncertainty. For example, a cubic model may be necessary to produce a second inflection point toward the high end of the age range, but the cubic term could then cause an unreasonably sharp escalation in predicted mean failures at the curve's tail.

Thus, we fit several other models to the power train, chassis, electrical, and fire control data to further explore the data and models. First, we applied generalized additive models (GAM), using the R statistical software package (Ihaka and Gentleman, 1996)[13] and assuming a negative binomial failure distribution with the log link function. GAM utilizes penalized regression splines with smoothing parameters chosen by generalized cross-validation (Gu, 2002). The penalized likelihood approach balances the overall fit of the model with its complexity. We began with the following full model:

$$\begin{aligned}\ln(\text{Mean Failures}) = {} & b_0 + b_1(\text{Location}) + s_1(\text{Accumulated Usage}) \\ & + s_2(\text{Age}) + s_3(\text{Accumulated Usage} \times \text{Age}) \\ & + \text{offset}(\log(\text{Updays}))\end{aligned}$$

The s(·) functions represent thin plate regression splines. To limit overfitting, each spline function was restricted to a maximum basis dimension of 4, corresponding to 3 degrees of freedom. This constrains the model degrees of freedom to be at most that of a full cubic model in the continuous covariates.

An approximate χ^2 test (at the 5 percent level) was conducted to measure the statistical significance of the bivariate spline. When this test indicated it was appropriate to do so, we based predictions on the following additive reduced model:

$$\begin{aligned}\ln(\text{Mean Failures}) = {} & b_0 + b_1(\text{Location}) + s_1(\text{Accumulated Usage}) \\ & + s_2(\text{Age}) + \text{offset}(\log(\text{Updays}))\end{aligned}$$

GAM models, described more generally in Appendix A, serve as a useful check on the final parametric models because of their non-

[13] R is a computing language and run-time environment designed for statistical computation and graphics (Hornik, 2003). It is an implementation of the S language (Becker, Chambers, and Wilks, 1988), which was developed at Bell Laboratories.

parametric properties: specific parametric models in the continuous covariates are avoided in favor of a data-adaptive approach to modeling that prevents overfitting—both by penalizing more complex models and by limiting the richness of the spline family from which function estimates are obtained.

Figure 3.13 shows the age-failure curves that resulted from the fitted GAM model. For the power train and chassis subsystems, the GAM generally yielded curves with more plausible tail regions than the parametric models. For the power train, the aging effect tapered off, rather than fully reversing itself. For the chassis, the GAM model suggested a second inflection point with a local minimum beyond 14 years of age. In both cases, the results are more consistent with the earlier theoretical discussion of renewal and aging. The GAM curve for the fire control subsystem had a more plausible tail region as well, with an upturn that was not as sharp as that of the parametric curve. For the electrical subsystem, however, the GAM curves and parametric curves were similar.

Although the GAM yielded more reasonable curves—from a theoretical standpoint—for the power train, chassis, and fire control subsystems, confidence bands for those curves were farther apart at the curves' tails, indicating that predictions are less accurate in that vicinity. Figures 3.14 through 3.16 display confidence bands for the GAM curves.

Figures 3.17 through 3.19 extrapolate the GAM curves and confidence bands beyond 15 years (the maximum age of tanks in our dataset); consistent with our previous discussion, the considerable distance between confidence bands illustrates the risks of such extrapolation. Thus, we cannot draw definitive conclusions about the curves' tail regions until more data on older tanks are available. These curves demonstrate that extrapolating beyond the range of the data is probably meaningless.

Nevertheless, since it is possible that GAM curves offer a more accurate depiction of power train, chassis, and fire control age-failure relationships (at least within the age range of tanks in our dataset), we conclude this section with an alternative version of Figure 3.4. In Figure 3.20, parametric curves have been replaced with GAM curves for those three subsystems.

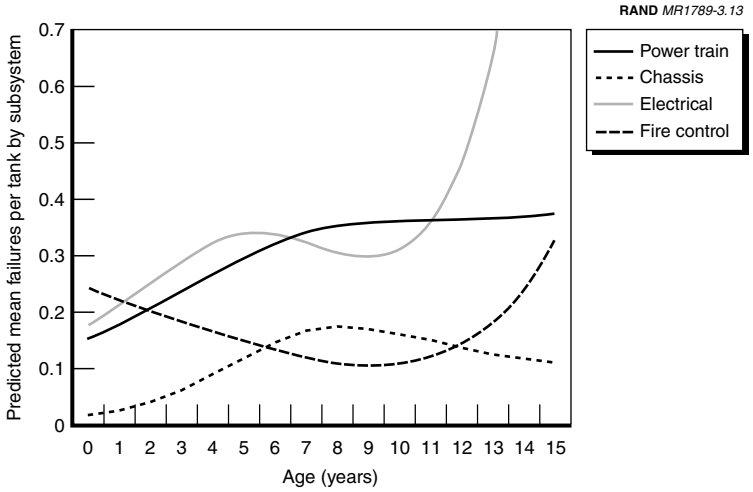


Figure 3.13—GAM Predicted Mean Failures of Chassis, Fire Control, Hardware, and Power Train Subsystems by Age (Location 1, 180 days)

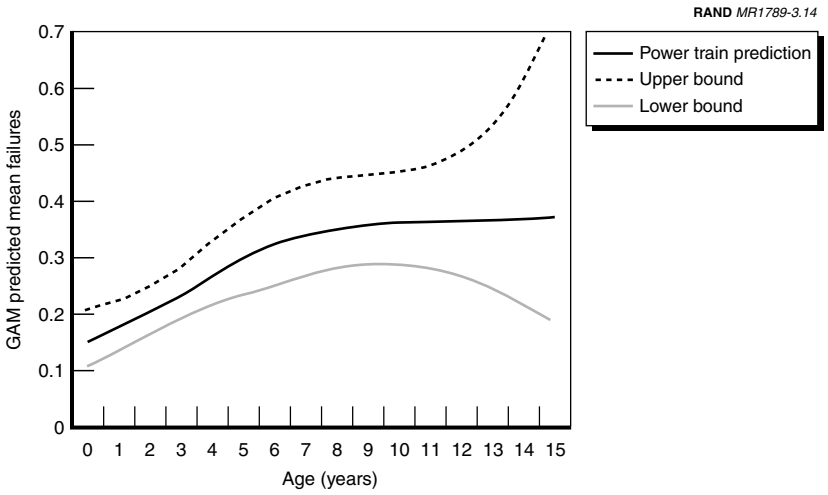


Figure 3.14—95 Percent Confidence Bands for Power Train GAM Curve

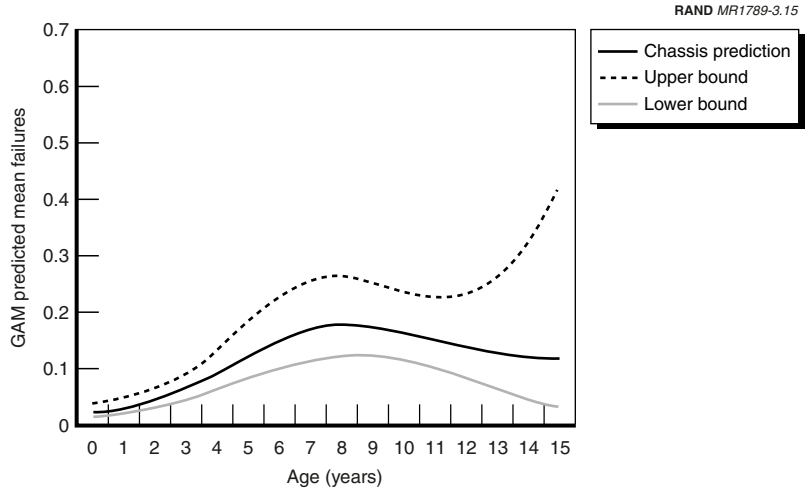


Figure 3.15—95 Percent Confidence Bands for Chassis GAM Curve

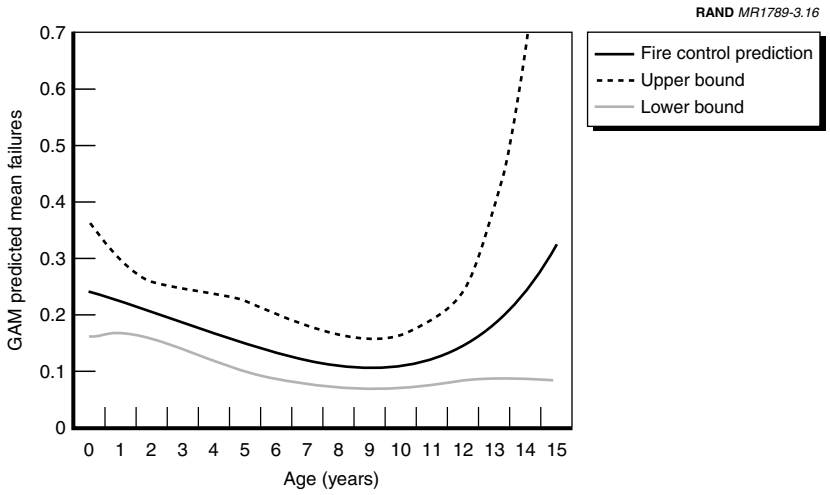


Figure 3.16—95 Percent Confidence Bands for Fire Control GAM Curve

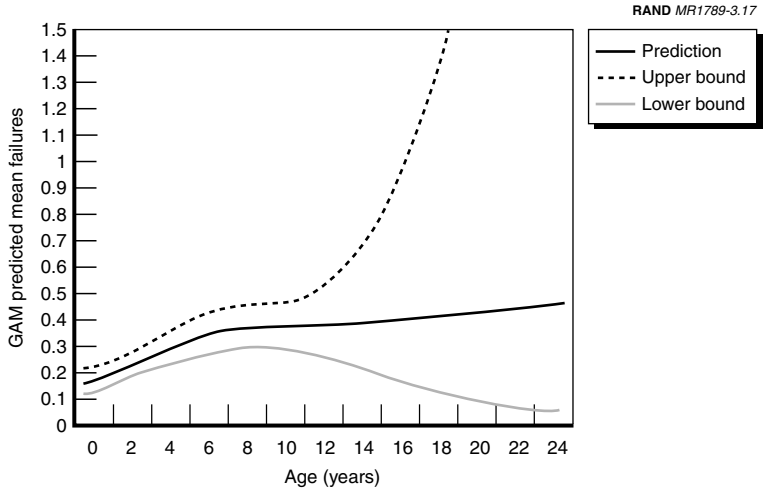


Figure 3.17—95 Percent Confidence Bands for Power Train GAM Curve, with Extrapolation Past Age 15

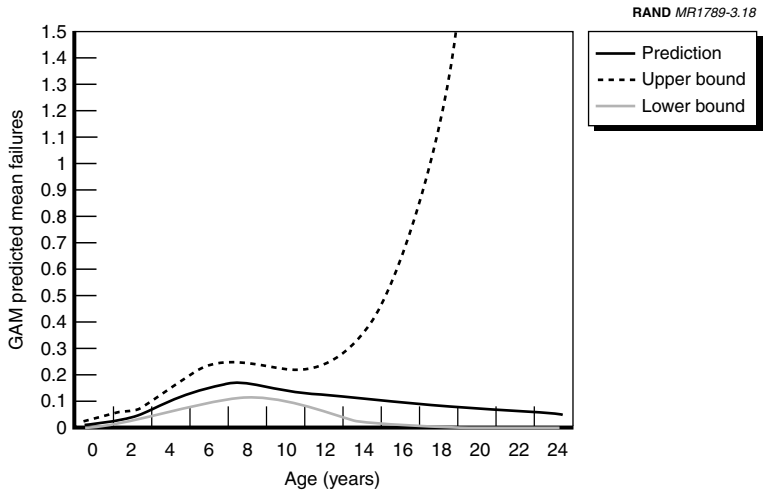


Figure 3.18—95 Percent Confidence Bands for Chassis GAM Curve, with Extrapolation Past Age 1

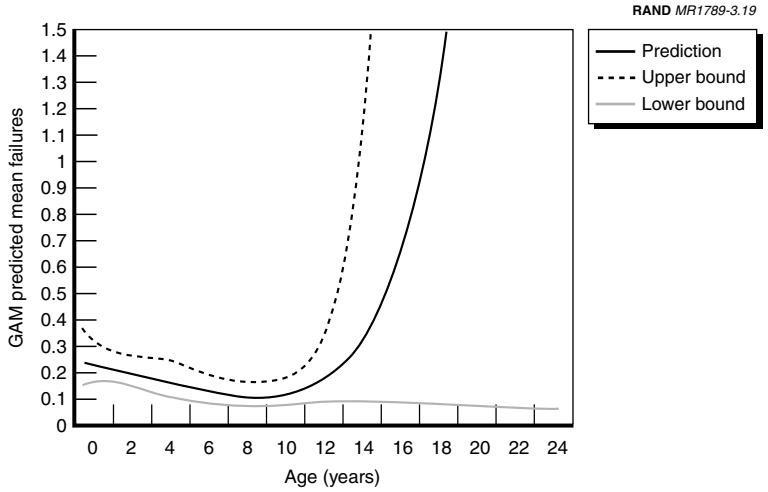


Figure 3.19—95 Percent Confidence Bands for Fire Control GAM Curve, with Extrapolation Past Age 15

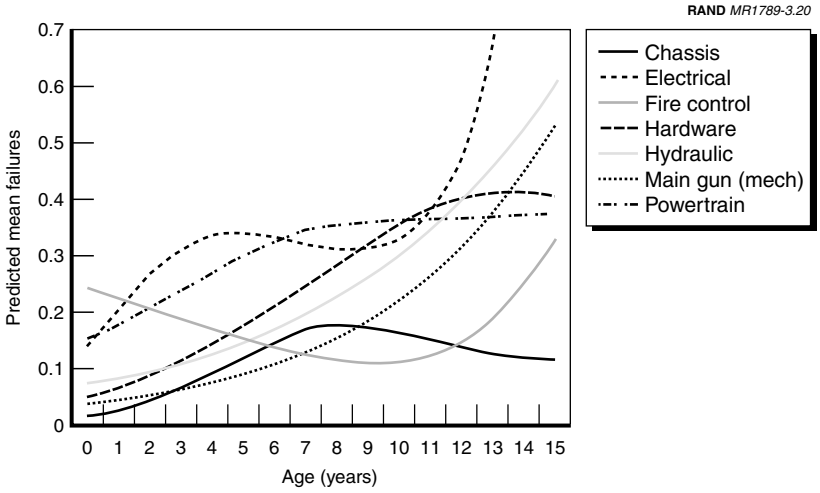


Figure 3.20—Alternate Plot of Predicted Mean Failures of Second-tier Subsystems by Age (Location 1, 180 days)

Although using the GAM curves caused the zero-to-peak differences and ratios to change, the changes were modest. As before, the zero-to-peak differences and ratios indicate that electrical, hardware, hydraulic, and main gun subsystems experienced the greatest absolute failure increases due to aging, and the chassis, hardware, hydraulic, and main gun subsystems experienced the greatest relative increases due to aging.

Chapter Four

IMPLICATIONS

Studies of failure characteristics have become widespread, reflecting efforts to improve design reliability and to better tailor preventive maintenance and scheduled service/overhaul programs to specific systems. Such assessments of the rate and nature of failures are especially important to the U.S. Army. The Army wants to ensure that it can sustain the warfighting capability of its current forces until they are fully replaced by the future force sometime between 2025 and 2030. Hence, this M1 Abrams study is one of a series investigating patterns and causes of Army equipment failures.

Our analysis of cross-sectional data provides preliminary support for the hypothesis that older tanks have higher failure rates than newer ones. Although longitudinal analyses offer a more rigorous test, these results suggest that mission-critical failures increase at a compound annual rate of 5 ± 2 percent during the first 14 years of life. This rate of increase means that a 14-year-old tank will have approximately double the expected number of failures of a brand new tank, for a given location, usage, and time period. (Additional data on older tanks are needed to draw conclusions about the failure rates of tanks beyond 14 years of age.) This result supports the notion that rebuilds of the M1 Abrams will improve readiness, if targeted at those components responsible for the age effect on reliability.

The analysis also showed that, after accounting for age, different Army locations had dramatically different failure rates during the study period. Additional study to understand the source of the differences could reveal further opportunity to reduce failure rates. Finally, the analysis suggested that, for a given age and location,

higher usage corresponded to a lower failure rate (that is, failures per kilometer). This may simply be an artifact of peacetime usage patterns characterized by low and sporadic usage.

Exploring the source of Abrams age effects yields valuable insights into the aging problem. Much of the mission-critical failure age effect appears to be produced by lower-cost "wear-and-tear" type components, so the resulting effect on operation and maintenance cost (the budget account used to pay for spare parts) may be minimal. However, the associated workload necessary to maintain operational readiness can increase substantially. This has no direct financial impact, because it does not affect maintainer pay, but it does have implications for quality of life, force structure needs, and future operational readiness. Once tank age reaches a certain point, the maintenance system may no longer be able to provide for a satisfactory level of operational readiness, even through the use of "workarounds" such as controlled exchange, given the number of maintainers in a unit. Such a condition necessitates replacement or substantial rebuild to maintain an acceptable level of operational readiness or the acceptance of lower operational readiness. As the Fort Riley and NTC data suggest, there is some indication that a portion of the active Army's tank fleet has already reached this point, leading to isolated M1A1 operational readiness problems. Thus, for the Abrams fleet, increasing fleet age most likely generates gradual workload increases, which result in quality-of-life issues, declining operational readiness, and a buildup of deferred financial cost that emerges in the form of programs such as RECAP.

While increasing age could also potentially lead to manpower—and thus maintenance—cost increases, other factors seem to prevent this today. Army maintenance manpower is based on annual maintenance man-hours required by a system. For a variety of reasons beyond the scope of this report, these data are infrequently and poorly updated. Thus, aging generally has no effect on manpower. However, if frequent studies accurately captured increases in workload, the Army's force structure process would automatically increase the maintainer requirement. Note, however, that personnel requirements are not always "resourced" or made part of the actual force structure. With budget and personnel constraints, increases in maintenance requirements would have to be traded off against other resource requirements to become part of the force structure.

Subsystem and part-level analyses offer insight for designing recapitalization programs to produce the greatest benefit. The age-failure relationship was stronger for some M1 subsystems. These distinctions suggest that certain subsystems (chassis, electrical, hardware, hydraulic, and main gun) are, in general, better candidates for rebuild efforts based on equipment age. These findings may also be instructive in the design of overhaul regimes or "phased" recapitalization schemes. Using this information to identify the wear-out patterns and regions more precisely for different types of components will help determine the most beneficial point, balancing cost and readiness, for scheduled component replacement (or suggest that *no* point is beneficial, if the wear-out region is too broad for economical scheduled replacement). The age-failure relationships by subsystem also revealed which subsystems (fire control, electrical, and power train) generally drive the failure rate of new tanks; this information suggests where engineering redesign might have the biggest impact.

In addition to the age and usage effects described above, it is of interest that tank location predicted failures. That is, some locations had more tank failures than did others. This finding could reflect distinct maintenance practices or personnel skill sets, the different operating environments, different reporting practices, different training regimes, or a combination of all these factors.

In summary, this report provides preliminary evidence that increasing age is likely to contribute to failure rate problems for the M1 Abrams series of tanks, and it shows that usage and location are influential as well. We have also highlighted how age and usage effects differ among M1 subsystems. Moreover, we have shown how an analysis of part orders can identify components that should be replaced in a rebuild program and those that would be good candidates for upgrade initiatives. These findings suggest that the Abrams RECAP program could have substantial benefit, and they offer insights that could potentially be used to enhance the program's value. We are currently assessing the robustness of these findings via additional tests, including regressions with longitudinal data.

Further extensions of the current study are also possible. For example, comparisons of maintenance practices, personnel skill sets, terrain, climate, and training practices across locations may help explain the location effects observed in this study. Moreover, they

may highlight opportunities for beneficial procedural, technological, and training changes at certain posts.

Appendix A

GENERAL DESCRIPTIONS OF STATISTICS USED

Two discrete distributions that are used to describe count data are the Poisson distribution and the negative binomial distribution. The first two portions of this appendix briefly describe the nature of each distribution. These descriptions were drawn from Hillier and Lieberman (1986) and from SAS Institute Incorporated (1999).

The final portion of this appendix describes the Generalized Additive Model (GAM), a technique used in our sensitivity analysis. The GAM description was largely based on Xiang (2001).

POISSON DISTRIBUTION

Suppose the distribution of Y, the number of occurrences of an event (e.g., equipment failures), is Poisson. Then the probability that $Y = a$ may be obtained from the following formula:

$$P(Y = a) = \frac{\mu^a e^{-\mu}}{a!}$$

where μ is the mean number of events, or $E(Y)$. Both the mean and the variance of the Poisson distribution are equal to μ. Let μ_i be the mean number of events for tank i:

$$\mu_i = c_i \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \beta_{11} x_1^2 + \cdots + \beta_{pp} x_p^2 + \beta_{1p} x_1 x_p + \cdots + \beta_{p-1} x_{p-1} x_p),$$

where c_i is the exposure time for tank i—that is, the number of days in the study period of tank i.

NEGATIVE BINOMIAL DISTRIBUTION

When data follow a negative binomial distribution (also known as the Pascal distribution), the probability that $Y = a$ is as follows:

$$P(Y = a) = \frac{\Gamma(a + 1/k) \quad (k\mu)^a}{a!\Gamma(1/k) \ (1 + k\mu)^{a+1/k}},$$

where $\Gamma(\cdot)$ is the gamma function, k is a dispersion parameter, and μ is the mean number of events. While the mean of the negative binomial distribution is μ, the variance is $\mu + k\mu^2$.

GENERALIZED ADDITIVE MODEL

The GAM is an extension of the traditional linear regression model. In its most basic form, the GAM defines the expected value of Y as follows:

$$E(Y) = \mu = f(X_1,\ldots,X_p) = s_0 + s_1(X_1) + \ldots + s_p(X_p),$$

where $s_i(X_i)$, $i = 1,\ldots,p$ are smooth functions estimated nonparametrically. For example, X_1 may be the predictor variable age, and s_1 may be a spline smoother.

Some GAMs have a more complex expected value of Y. Specifically, they have a link function, such as the natural log, defining the relationship between μ and $f(X_1,\ldots,X_p)$. Thus, when using a GAM, one does not have to assume that the dependent variable is normally distributed. If the GAM is applied to a dependent variable having a Poisson distribution, then one can set up the model so that

$$\ln(\mu) = f(X_1,\ldots,X_p) = s_0 + s_1(X_1) + \ldots + s_p(X_p).$$

As in the Poisson and negative binomial regressions described previously, μ incorporates a factor that accounts for tank exposure time.

The shape of a GAM curve can help determine whether a parametric model should include a cubic or other higher-order term.

Recent implementations of the R statistical software (Ihaka and Gentleman, 1996) allow GAM models to include nonparametric two-factor interaction terms as well. Thus, the above GAM model can be modified as follows:

$$\ln(\mu) = f(X_1,\ldots,X_p) = s_0 + s_1(X_1) + \ldots + s_p(X_p) + s_{12}(X_{1,}X_2) + \ldots + s_{p-1,p}(X_{p-1},X_p).$$

Parametric terms can also be added to this model specification.

For more detailed descriptions of the GAM approach, see Hastie and Tibshirani (1990) and Wood (2003). The latter paper details the use of parametric terms and nonparametric interaction terms in GAM models.

Appendix B

DISTRIBUTION OF FAILURE DATA

This appendix elaborates on the overdispersion that was characteristic of the data in this study.

To get a better understanding of the underlying failure distribution, we examined failures at the battalion level. Over the course of one year, the tanks in a battalion represent a very homogeneous set. They are typically about the same age, go through about the same training, have similar usage, operate on similar terrain, and are supported by

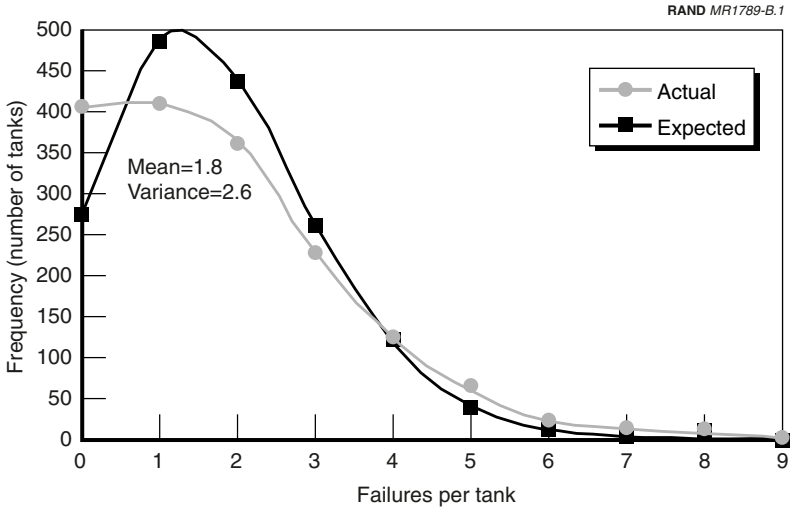


Figure B.1—Illustration of Failure Data Overdispersion

the same maintenance system and command structure. Looking at the battalion level therefore allowed us to see the failure distribution after accounting for the nonrandom factors that produce failure rate variation. (It is precisely the effects of these nonrandom factors that we were trying to isolate in this study.) Thus, examining failures by tank in one battalion over the course of a year enabled us to examine the underlying distribution of the tank failure process.

As Figures B.2 through B.7 indicate, every Armor battalion in the Army but two (i.e., 28 out of 30 battalions) closely matched a Poisson distribution.[1] The figures display each battalion's actual failure distribution (the "actual" curve), the type of M1 in the battalion, the mean failures per tank in the battalion, and a Poisson distribution applied using the actual mean number of failures per tank (the "expected" curve). The horizontal axis in each figure is the number of failures per tank, and the vertical axis is the number of tanks in the battalion experiencing those failures in one year. For example, a graph would be read as "6 tanks in battalion X had 3 failures over one year." When we excluded partial-year tanks, which were only in battalions for a few months, statistical tests showed no significant difference ($p = .05$) between the actual and the expected (Poisson distribution–based) failure curves of 28 battalions.[2]

It is academically interesting to note that the oft-cited Poisson distribution assumption held very well when the failure data were extracted from nonrandom failure-affecting factors. We only saw overdispersion when we combined all battalion data and included partial-year tanks. The primary source of that overdispersion was the partial-year tanks. Many tanks were replaced due to new equipment fielding, for depot overhaul, for other severe maintenance problems, or, in some cases, because units were reorganizing from 58- to 44-tank battalions. Many of these tanks had zero or low failures because of a shorter amount of time present in the dataset.

---

[1] A χ^2 test showed that the failure distributions of two battalions in Europe (Figure B.7) were significantly different from the Poisson distribution.

[2] Plots in the second row of Figure B.6 correspond to the same two battalions as those in the first row; plots in the second row, however, excluded partial-year tanks (i.e., tanks with only a few months of data). With partial-year tanks removed, the two battalions' failure distributions were indistinguishable from Poisson distributions.

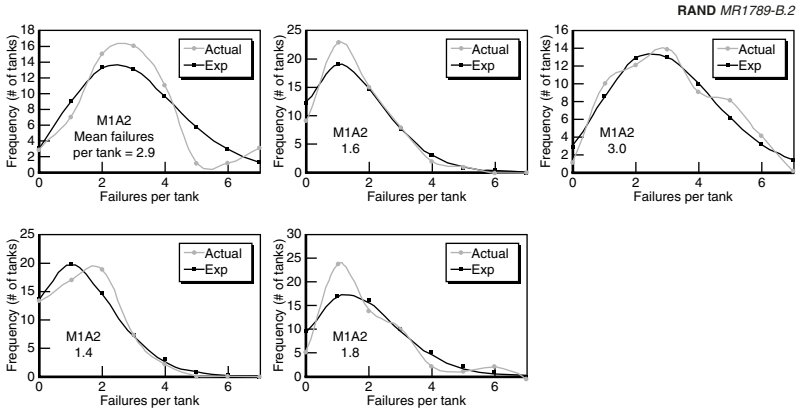
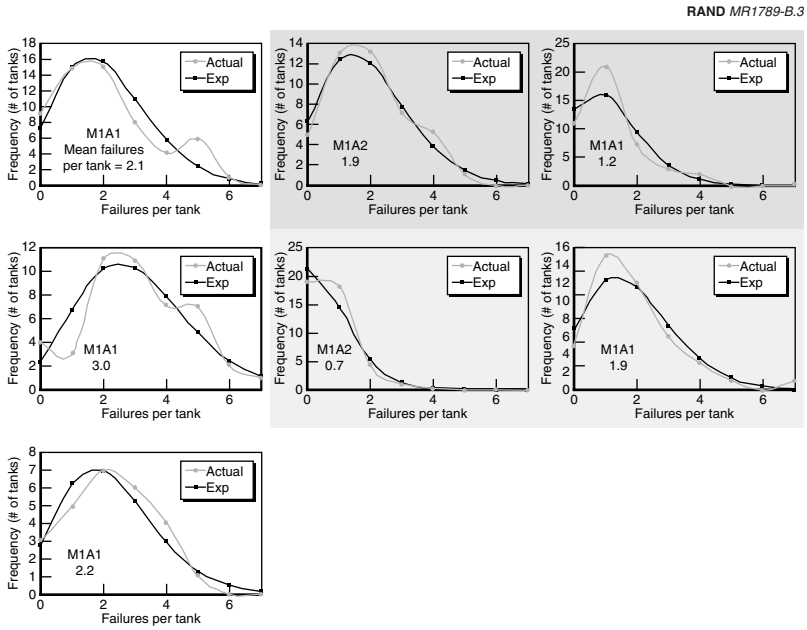


Figure B.2—Comparison of Battalion Failure Distributions and Poisson Distribution in 1st Cavalry Division



NOTE: Both dark gray plots correspond to a single battalion. One displays partial-year data for M1A1s, and the other displays partial-year data for M1A2s, which replaced the M1A1s. Similarly, both light gray plots correspond to a single battalion, and each plot displays partial-year data.

Figure B.3—Comparison of Battalion Failure Distributions and Poisson Distribution in 4th Infantry Division

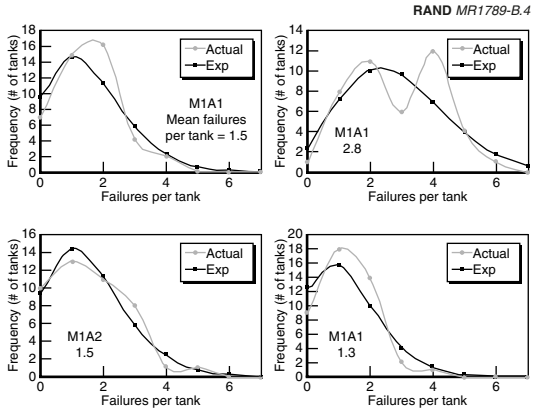


Figure B.4—Comparison of Battalion Failure Distributions and Poisson Distribution in 1st Infantry and 1st Armor Divisions: Fort Riley

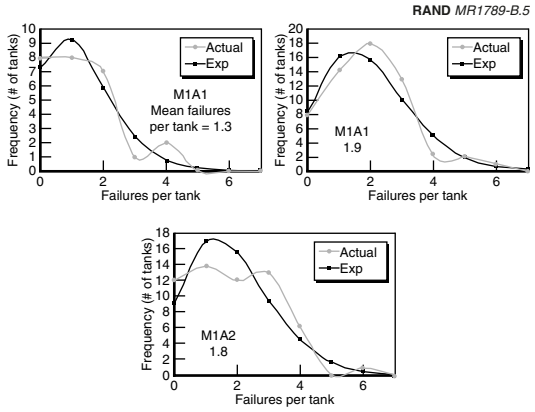


Figure B.5—Comparison of Battalion Failure Distributions and Poisson Distribution in 2nd Infantry Division

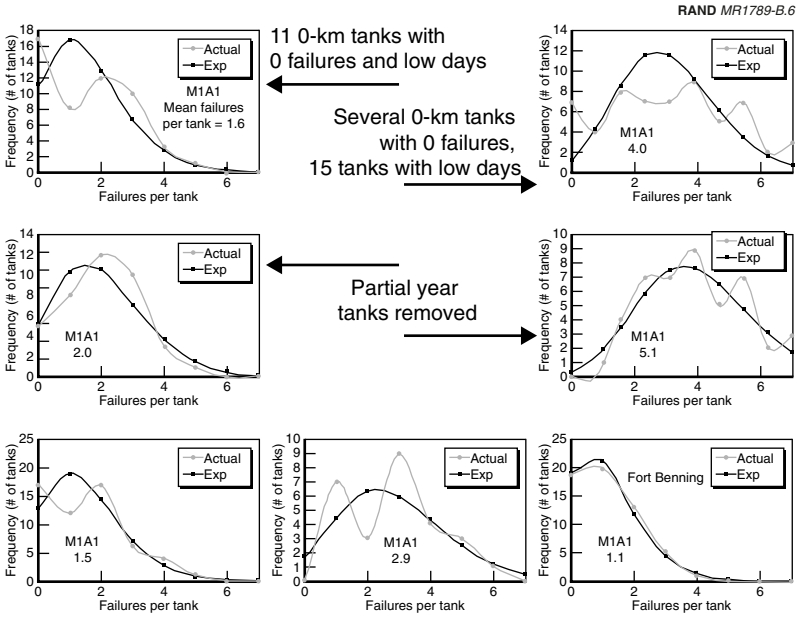


Figure B.6—Comparison of Battalion Failure Distributions and Poisson Distribution in 3rd Infantry Division

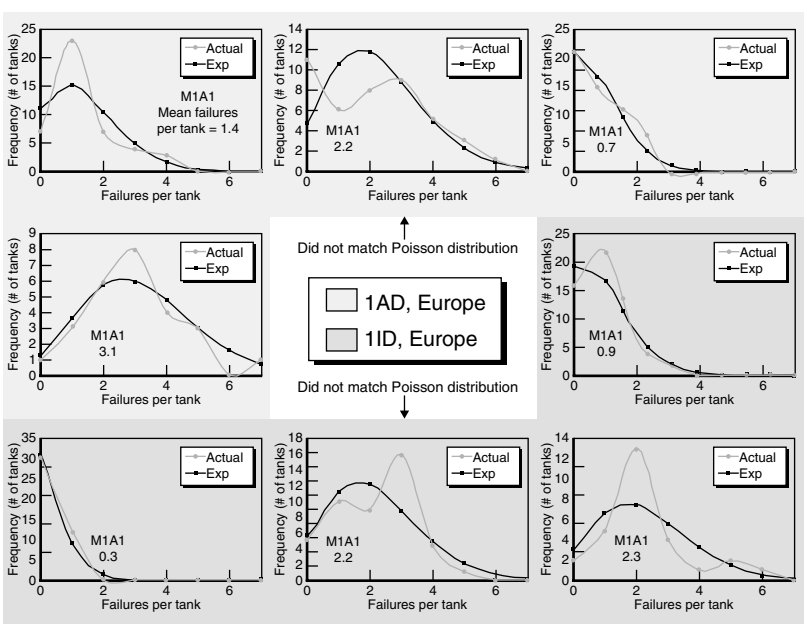


Figure B.7—Comparison of Battalion Failure Distributions and Poisson Distribution in 1st Infantry and 1st Armor Divisions: Europe

Appendix C

CROSS-VALIDATION OF TANK STUDY MODEL

We used a *leave-one-out* cross-validation approach (Burt and Barber, 1996; Geisser, 1975; Stone, 1974) to assess the predictive accuracy of the Tank Study model. Considered "[one] of the most important" cross-validation techniques (Racine, 1997:169), the leave-one-out approach entails removing one observation at a time from a dataset. The removed observation becomes a single-element "test set" and the remaining observations become the "training set."[1] A model is fit to the training set and the result is used to predict the dependent variable value associated with the test set observation. The removed observation is then returned to the dataset, and another observation then becomes the test set. This process is repeated until all observations have been removed and returned to the dataset. At that point, the predicted and observed dependent variable values allow calculation of the model's prediction error.

A standard measure of model prediction error is the predicted residual sum of squares (PRESS) statistic (Neter, Wasserman, and Kutner, 1989:450; Weisberg, 1985:217), defined as

$$\text{PRESS} = \sum_{i=1}^{n} \hat{e}_i^2 \ ,$$

[1] As Neter, Wasserman, and Kutner (1989:466) point out, "By far the preferred method to validate a regression model is through the collection of new data. Often, however, this is neither practical nor feasible. A reasonable alternative when the dataset is large enough is to split the data into two sets." The training set is used to estimate the model, and the test set is used to evaluate the predictive ability of the model.

where $\hat{e}_i$ is the difference between (a) the observed outcome for observation i and (b) the outcome predicted from a regression model calibrated on the set of observations that excluded i. The total number of observations in the full dataset is n. In comparisons of models, the preferred model is the one with the smallest PRESS statistic.

Table C.1 lists the models that our cross-validation study compared to the final model and the PRESS statistic associated with each. Note that our final model, with its PRESS of 4,019.33 for 1,567 observations, performed better than models containing either fewer or more terms. Thus, the cross-validation study supports selection of our model based on its predictive accuracy.

Table C.1

PRESS Statistics for Models in Cross-Validation Study

Model	Equation	PRESS
Final Tank Study Model	ln (mean tank failures during study period) = $\beta_0 + \ln(\text{updays}) + \beta_1(\text{location 2}) + \beta_2(\text{location 3}) + \beta_3(\text{location 4}) + \beta_4(\text{location 5}) + \beta_5(\text{location 6}) + \beta_6(\text{usage}) + \beta_7(\text{usage}^2) + \beta_8(\text{age})$	4,019.33
Full Cubic Model	ln (mean tank failures during study period) = $\beta_0 + \ln(\text{updays}) + \beta_1(\text{location 2}) + \beta_2(\text{location 3}) + \beta_3(\text{location 4}) + \beta_4(\text{location 5}) + \beta_5(\text{location 6}) + \beta_6(\text{usage}) + \beta_7(\text{usage}^2) + \beta_8(\text{usage}^3) + \beta_9(\text{age}) + \beta_{10}(\text{age}^2) + \beta_{11}(\text{age}^3) + \beta_{12}(\text{usage} \times \text{age}) + \beta_{13}(\text{usage} \times \text{age}^2) + \beta_{14}(\text{usage}^2 \times \text{age})$	4,050.45
Cubic Model, without Interactions	ln (mean tank failures during study period) = $\beta_0 + \ln(\text{updays}) + \beta_1(\text{location 2}) + \beta_2(\text{location 3}) + \beta_3(\text{location 4}) + \beta_4(\text{location 5}) + \beta_5(\text{location 6}) + \beta_6(\text{usage}) + \beta_7(\text{usage}^2) + \beta_8(\text{usage}^3) + \beta_9(\text{age}) + \beta_{10}(\text{age}^2) + \beta_{11}(\text{age}^3)$	4,030.28
Full Quadratic Model	ln (mean tank failures during study period) = $\beta_0 + \ln(\text{updays}) + \beta_1(\text{location 2}) + \beta_2(\text{location 3}) + \beta_3(\text{location 4}) + \beta_4(\text{location 5}) + \beta_5(\text{location 6}) + \beta_6(\text{usage}) + \beta_7(\text{usage}^2) + \beta_8(\text{age}) + \beta_9(\text{age}^2) + \beta_{10}(\text{usage} \times \text{age})$	4,034.47
Quadratic Model, without Interactions	ln (mean tank failures during study period) = $\beta_0 + \ln(\text{updays}) + \beta_1(\text{location 2}) + \beta_2(\text{location 3}) + \beta_3(\text{location 4}) + \beta_4(\text{location 5}) + \beta_5(\text{location 6}) + \beta_6(\text{usage}) + \beta_7(\text{usage}^2) + \beta_8(\text{age}) + \beta_9(\text{age}^2)$	4,023.00
Linear Model	ln (mean tank failures during study period) = $\beta_0 + \ln(\text{updays}) + \beta_1(\text{location 2}) + \beta_2(\text{location 3}) + \beta_3(\text{location 4}) + \beta_4(\text{location 5}) + \beta_5(\text{location 6}) + \beta_6(\text{usage}) + \beta_7(\text{age})$	4,136.32
Linear Model in Usage only	ln (mean tank failures during study period) = $\beta_0 + \ln(\text{updays}) + \beta_1(\text{location 2}) + \beta_2(\text{location 3}) + \beta_3(\text{location 4}) + \beta_4(\text{location 5}) + \beta_5(\text{location 6}) + \beta_6(\text{usage})$	4,173.33
Location Model	ln (mean tank failures during study period) = $\beta_0 + \ln(\text{updays}) + \beta_1(\text{location 2}) + \beta_2(\text{location 3}) + \beta_3(\text{location 4}) + \beta_4(\text{location 5}) + \beta_5(\text{location 6})$	4,239.45
Intercept Model	ln (mean tank failures during study period) = $\beta_0 + \ln(\text{updays})$	4,248.56

Appendix D

PLOTS OF SUBSYSTEMS' PREDICTED MEAN FAILURES BY AGE AND USAGE

The following plots display predicted mean failures versus tank age and usage for the first-tier subsystems (hull and turret) and second-tier subsystems of the M1 Abrams. Within the hull, second-tier subsystems are the chassis and power train. Within the turret, second-tier subsystems include the gun and fire control. The remaining second-tier subsystems (electrical, hardware, and hydraulic) can be classified as either hull or turret components. For the hull, chassis, power train, and fire control subsystems, the age-failure curves are those resulting from fitted GAM models, rather than log-quadratic models. Sensitivity analyses suggested that the GAM curves were more plausible for those subsystems.

HULL FIRST- AND SECOND-TIER SUBSYSTEM PLOTS

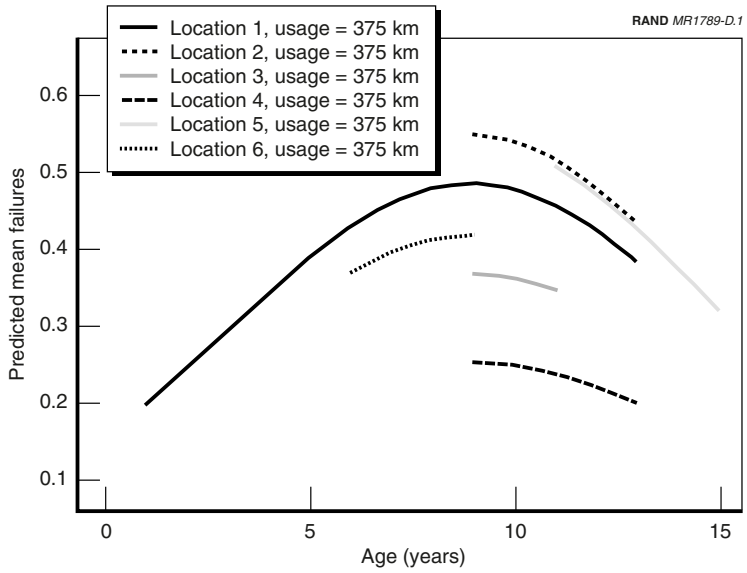


Figure D.1—Predicted Mean Hull Failures by Tank Age

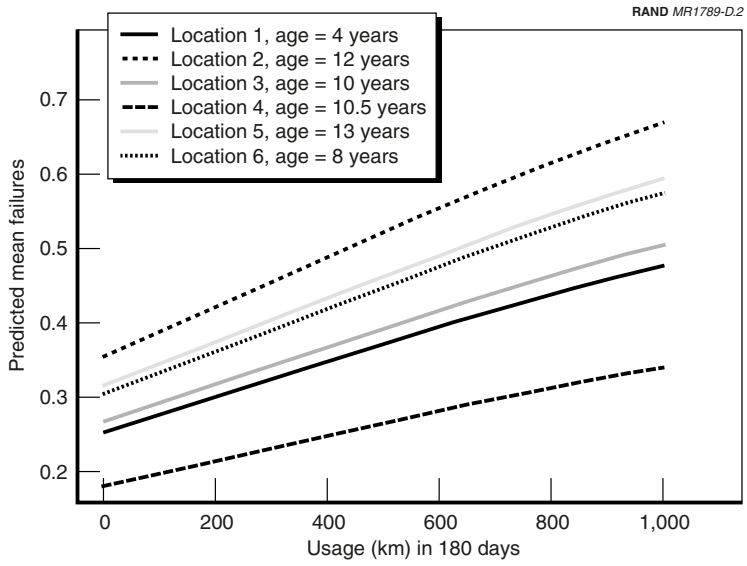


Figure D.2—Predicted Mean Hull Failures by Tank Usage

Figure D.3—Predicted Mean Chassis Failures by Tank Age

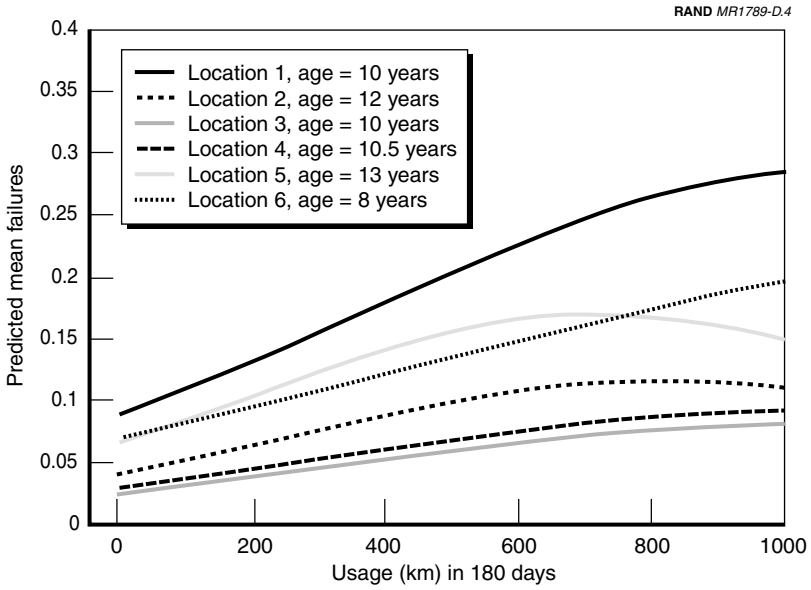


Figure D.4—Predicted Mean Chassis Failures by Tank Usage

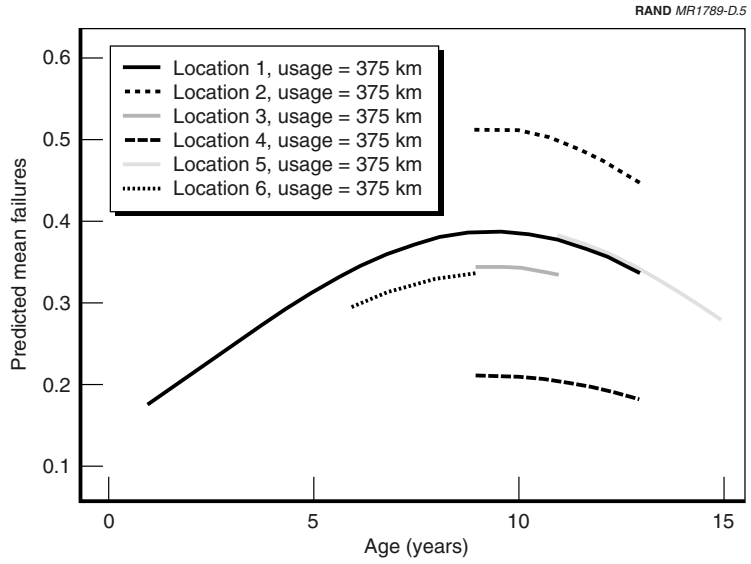


Figure D.5—Predicted Mean Power Train Failures by Tank Age

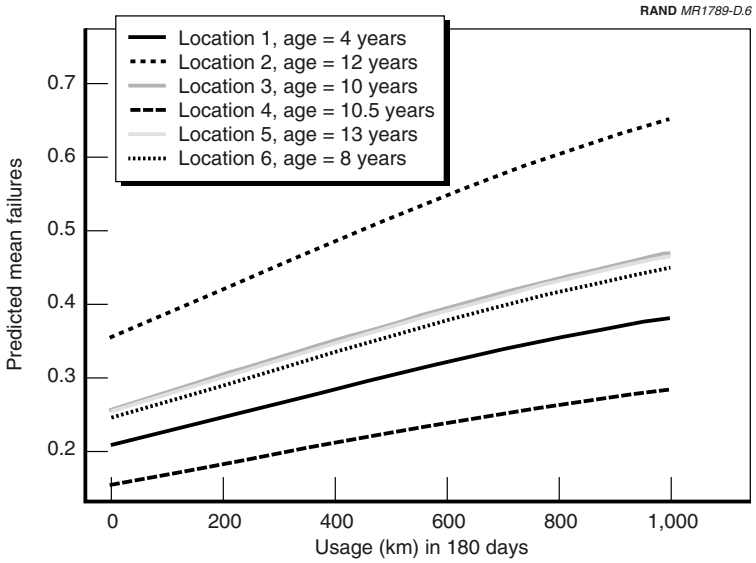


Figure D.6—Predicted Mean Power Train Failures by Tank Usage

TURRET FIRST- AND SECOND-TIER SUBSYSTEM PLOTS

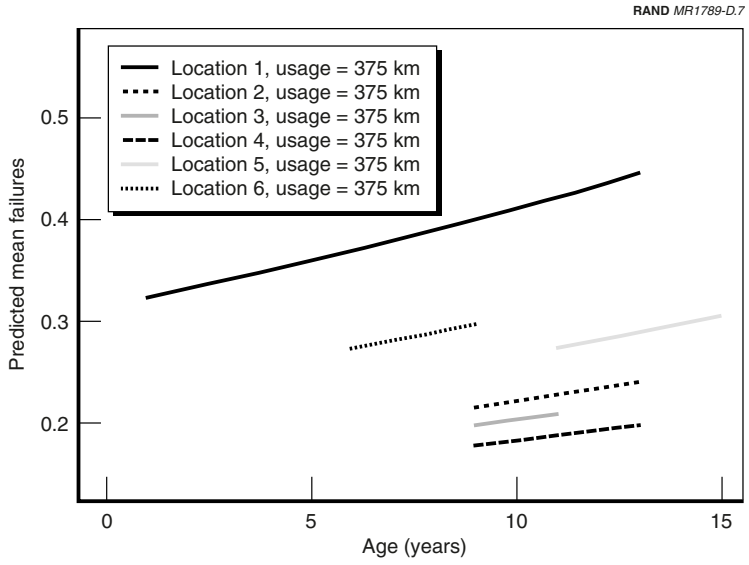


Figure D.7—Predicted Mean Turret Failures by Tank Age

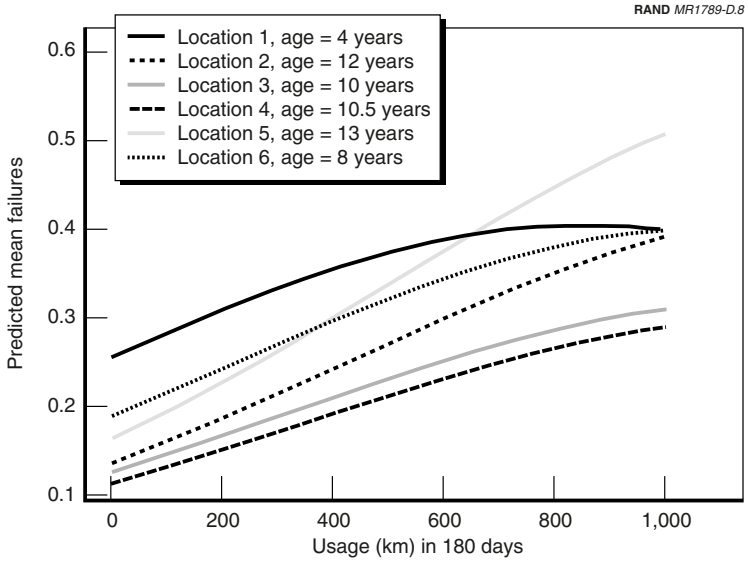


Figure D.8—Predicted Mean Turret Failures by Tank Usage

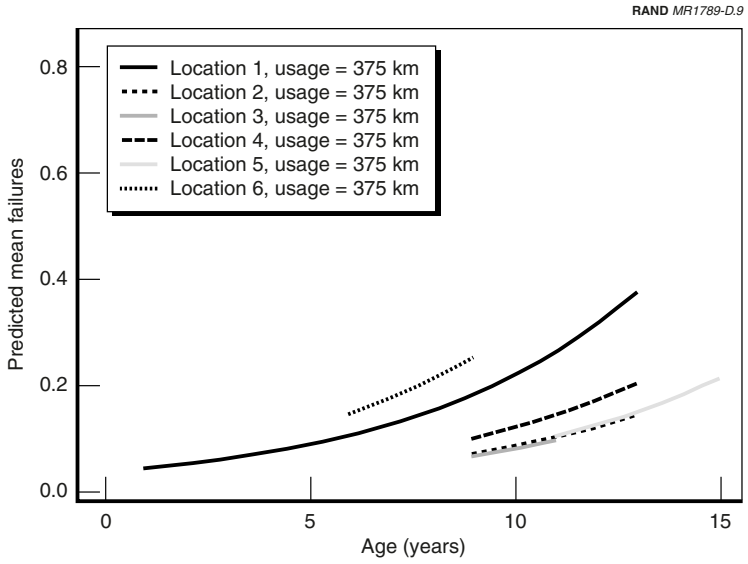


Figure D.9—Predicted Mean Gun Failures by Tank Age

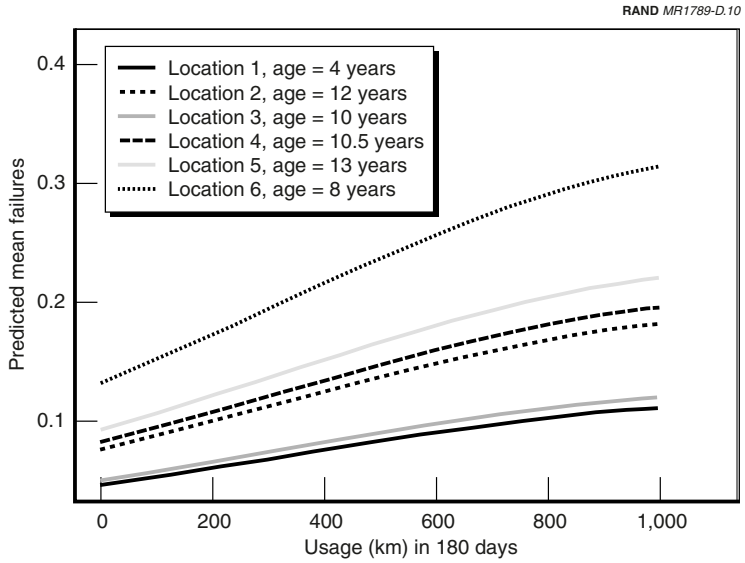


Figure D.10—Predicted Mean Gun Failures by Tank Usage

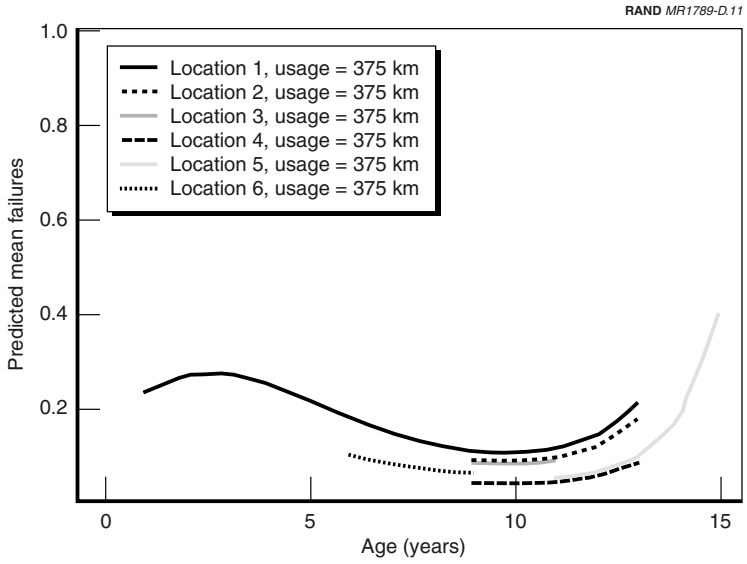


Figure D.11—Predicted Mean Fire Control Failures by Tank Age

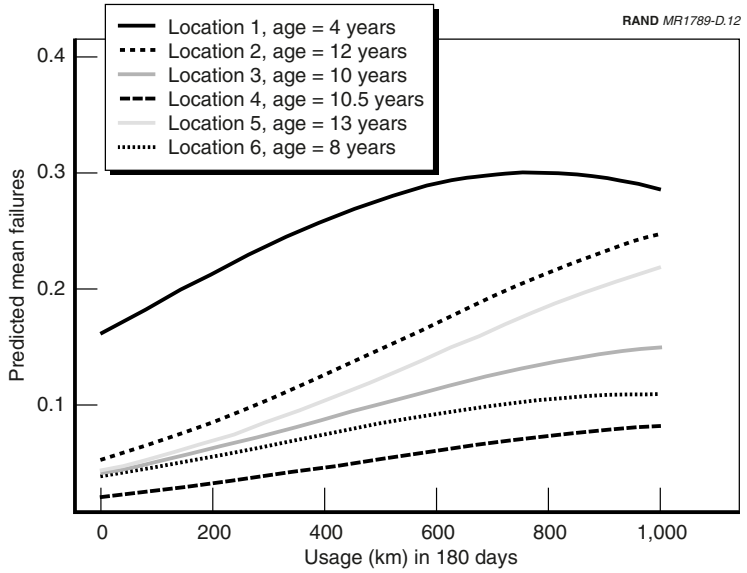


Figure D.12—Predicted Mean Fire Control Failures by Tank Usage

PLOTS FOR OTHER SECOND-TIER SUBSYSTEMS WITHIN EITHER HULL OR TURRET

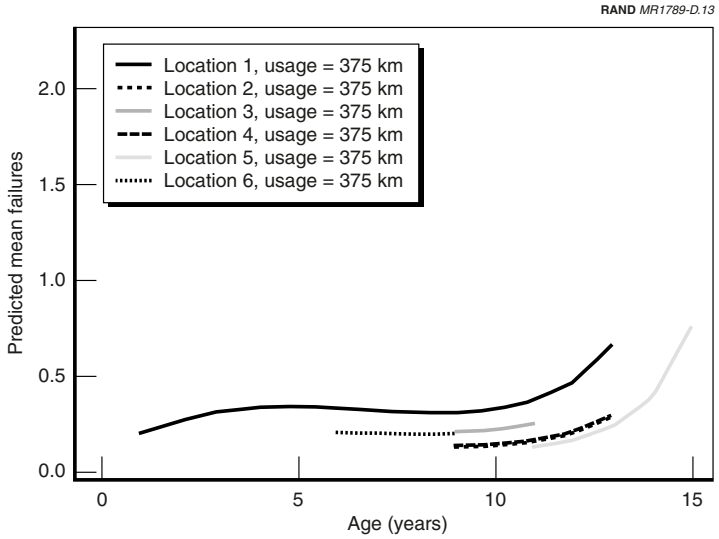


Figure D.13—Predicted Mean Electrical Failures by Tank Age

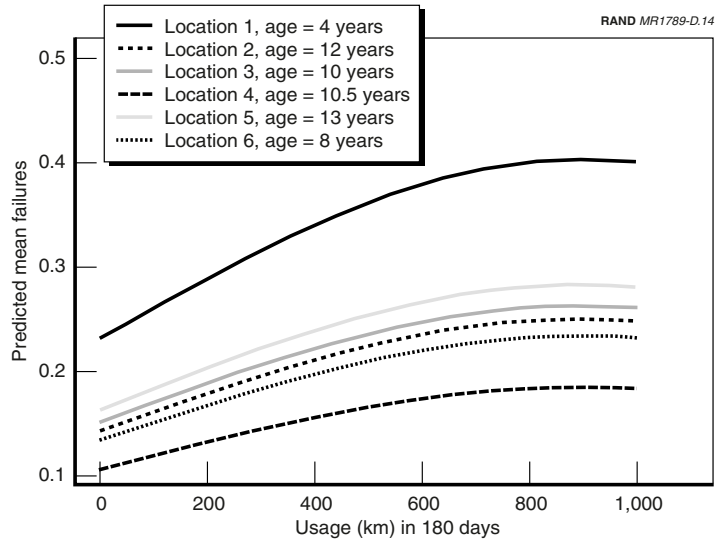


Figure D.14—Predicted Mean Electrical Failures by Tank Usage

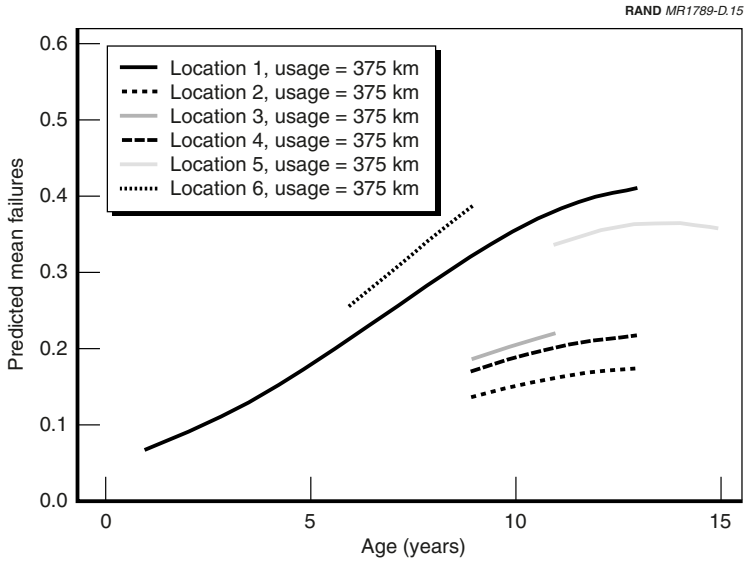


Figure D.15—Predicted Mean Hardware Failures by Tank Age

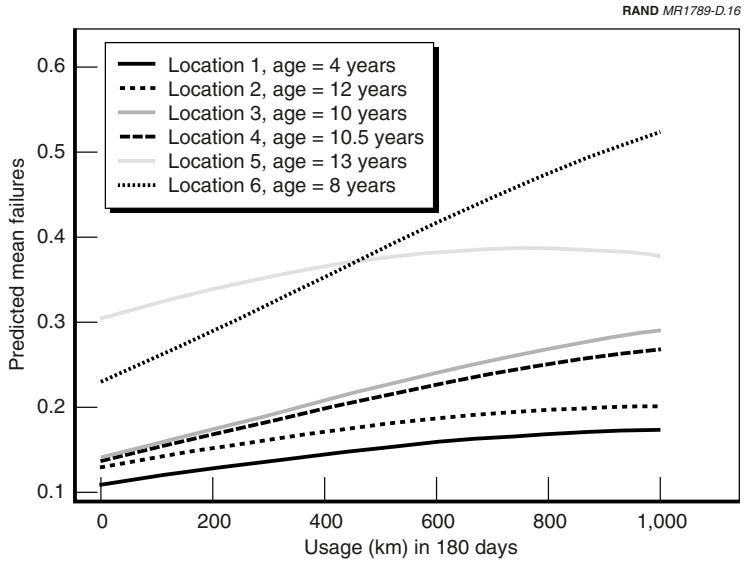


Figure D.16—Predicted Mean Hardware Failures by Tank Usage

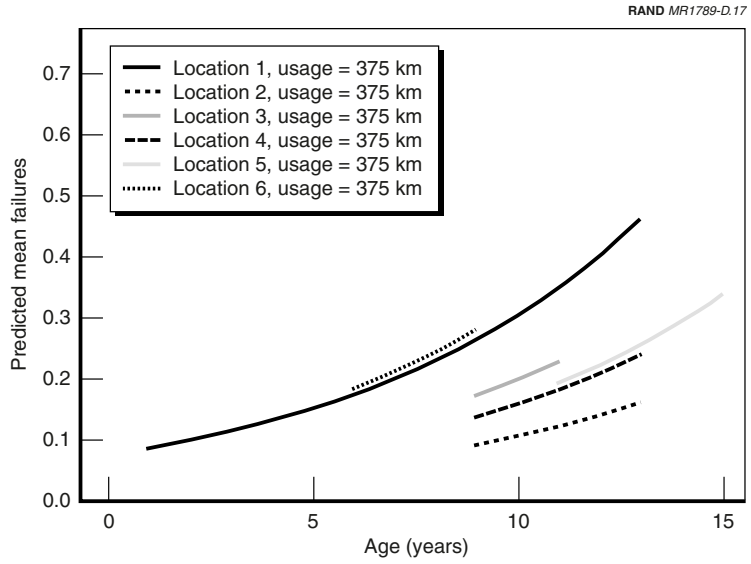


Figure D.17—Predicted Mean Hydraulic Failures by Tank Age

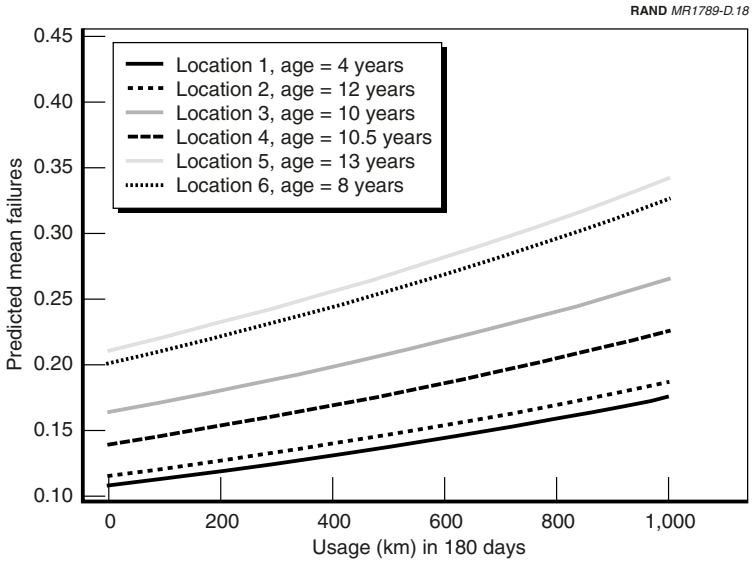


Figure D.18—Predicted Mean Hydraulic Failures by Tank Usage

BIBLIOGRAPHY

Advisory Group on Reliability of Electronic Equipment. 1957. *Reliability of Military Electronic Equipment,* Washington, D.C.: Office of the Assistant Secretary of Defense.

Aiken, L.S., and West, S.G. 1991. *Multiple Regression: Testing and Interpreting Interactions.* Newbury Park, CA: Sage.

Alexander, A.J. 1988. *The Cost and Benefits of Reliability in Military Equipment*. Santa Monica, CA: RAND, P-7515.

Army Recapitalization Management. 2003. *Army Recapitalization Management Performance Plans and Agreements*. https://webportal.saalt.army.mil/armytoc/statuspage02.htm.

Barringer, H.P. 1998. "How to Use Reliability Engineering Principles for Business Issues," *Proceedings, YPF Reliability Symposium,* La Platta, Argentina.

Becker, R.A., Chambers, J.M., and Wilks, A.R. 1988. *The New S Language.* New York: Chapman & Hall.

Blischke, W.R., and Murthy, D.N.P. 2000. *Reliability Modeling, Prediction, and Optimization,* New York: Wiley.

Brick, J.M., and Kalton, G. 1996. "Handling Missing Data in Survey Research," *Statistical Methods in Medical Research,* 5:215–238.

Brownlee, L., and Keane, J.M. 2002. *Joint Statement by The Honorable Les Brownlee, Under Secretary of the Army, and General John M. Keane, Vice Chief of Staff—United States Army, on Army Modern-*

ization and Transformation and Review of the Defense Authorization Request for Fiscal Year 2003.

Burrill, C.W., and Ledolter, J. 1999. *Achieving Quality Through Continual Improvement.* New York: Wiley.

Burt, J.E., and Barber, G.M. 1996. *Elementary Statistics for Geographers,* 2nd Edition. New York: Guilford Press.

Congressional Budget Office. 2001. "The Effects of Aging on the Costs of Operating and Maintaining Military Equipment," *CBO Paper,* August 2001:1–32.

Gansler, Hon. J.S. 1999. "Acquisition Reform Week Remarks," June 7, 1999.

Geisser, S. 1975. "A Predictive Sample Reuse Method with Application," *Journal of the American Statistical Association*, 70:320–328.

Gosling, W. 1962. *The Design of Engineering Systems,* New York: Wiley.

Gourley, S.R. 2001. "Recapitalization," *Army Magazine*, 51(6), available at http://www.ausa.org/www/armymag.nsf.

Gu, C. 2002. *Smoothing Spline ANOVA Models.* New York: Springer.

Harrington, R. 2000. "How to Install a Simple, Effective Reliability-Centered Maintenance (RCM) Program," *SMRP Solutions*. On-line newsletter of the Society for Maintenance and Reliability Professionals. Chicago, IL.

Hastie, T.J., and Tibshirani, R.J. 1990. *Generalized Additive Models.* New York: Chapman and Hall.

Hillier, F.S., and Lieberman, G.J. 1986. *Introduction to Operations Research,* 4th edition. Oakland, CA: Holden-Day, Inc.

Hornik, K. 2003. *The R FAQ.* http://www.ci.tuwien.ac.at/~hornik/R/. ISBN 3-901167-51-X.

Ihaka, R., and Gentleman, R. 1996. "R: A Language for Data Analysis and Graphics," *Journal of Computational and Graphical Statistics*, 5(3):299–314.

Kales, P. 1997. *Reliability: For Technology, Engineering, and Management.* New Jersey: Prentice-Hall.

Kapur, K.C., and Lamberson, L.R. 1977. *Reliability in Engineering Design.* New York: Wiley.

Kern, LTG P.J. 2001. "Recapitalization and Unit Set Fielding," *Army AL&T,* January–February:5–6.

Kitfield, J. 1997. "Biting the Bullet," *Government Executive,* 29(1):18–22.

Konwinski, L.M., and Wilson, LTC P.M. 2000. "Modifying the Abrams Fleet," *Army Logistician*, 32(2):10–13.

Krohn, C.A. 1969. "Hazard Versus Renewal Rate of Electronic Items," *IEEE Transactions on Reliability,* R-18, 64–73.

Li, K.H., Raghunathan, T.E., and Rubin, D.B. 1991. "Large Sample Significance Levels from Multiply Imputed Data Using Moment-Based Statistics and an F Reference Distribution." *Journal of the American Statistical Association,* 86:1065–1073.

Li, K.H., Meng, X.L., Raghunathan, T.E., and Rubin, D.B. 1991. "Significance Levels from Repeated P-Values with Multiply-Imputed Data." *Statistica Sinica,* 1:65–92.

Machol, R.E., Tanner, W.P., and Alexander, S.N. (eds.). 1965. *System Engineering Handbook,* New York: McGraw-Hill.

Maddala, G.S. 1988. *Introduction to Econometrics,* New York: Macmillan Publishing Company.

McCullagh, P., and Nelder, J.A. 1989. *Generalized Linear Models,* 2nd edition. New York: Chapman and Hall.

Meng, X.L., and Rubin, D.B. 1992. "Performing Likelihood Ratio Tests with Multiply Imputed Data Sets." *Biometrika*, 79:103–111.

Morris, S., Dudley, B., Caroli, J., MacDiarmid, P., Nicholls, D., Coppola, A., Criscimagna, N., and Farrell, J. 1995. *Reliability Toolkit: Commercial Practices Edition,* Rome, NY: Reliability Analysis Center.

Moubray, J. 1997. *Reliability-centered Maintenance,* New York: Industrial Press.

Mudholkar, G.S. 1995. "The Exponentiated Weibull Family: A Reanalysis of the Bus-motor-failure Data," *Technometrics,* 37:436–445.

Neter, J., Wasserman, W., and Kutner, M.H. 1989. *Applied Linear Regression Models,* 2nd Edition. Boston, MA: Irwin.

Nowlan, F.S., and Heap, H.F. 1978. *Reliability-centered Maintenance,* San Francisco: Dolby Access Press.

O'Connor, P. 1998. "Standards in Reliability and Safety Engineering," *Reliability Engineering and System Safety,* 60(2):173–177.

Omdahl, T.P. (ed.). 1988. *Reliability, Availability, and Maintainability (RAM) Dictionary,* West Terre Haute, IN: Quality Council of Indiana.

Orsini, E.A., and Harrold, COL G.J. 2001. "Recapitalization: A Key Element of the Army Transformation," *Army AL&T,* January–February:2–4.

Peltz, E., Boren, P., Robbins, M., and Wolff, M. 2002. *Diagnosing the Army's Equipment Readiness: The Equipment Downtime Analyzer,* Santa Monica, CA: RAND, MR-1481-A.

Pyles, R. 2003. *Aging Aircraft: USAF Workload and Material Consumption Life Cycle Patterns.* Santa Monica, CA: MR-1641-AF.

Racine, J.S. 1997. "Feasible Cross-Validatory Model Selection for General Stationary Processes," *Journal of Applied Econometrics*, 12(2):169–179.

Robinson, R.M., Anderson, K.J., and Meiers, S. 2003. *Risk and Reliability—An Introductory Text,* 4th edition, available at http://www.r2a.com.au/publications/4th_Edition/4th_edition.html.

Rubin, D.B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

SAS Institute Inc. 1999. *SAS/STAT® User's Guide,* Version 8. Cary, NC: SAS Institute Inc.

Schafer, J.L. 1997. *Analysis of Imcomplete Multivariate Data,* London: Chapman and Hall.

Stevenson, W.J. 1993. *Production/Operations Management*, Boston, MA: Irwin.

Stone, M. 1974. "Cross-Validatory Choice and Assesment of Statistical Predictions (with Discussion)," *Journal of the Royal Statistical Society*, 36:111–147.

Streilein, J.J. 1984. *Tactical Wheeled Vehicle Replacement Management Program Methodology,* Aberdeen Proving Ground, MD: AMSAA Technical Report No. 379.

U.S. General Accounting Office. 2001. *Report to the Secretary of Defense: Tactical Aircraft Modernization Plans Will Not Reduce Average Age of Aircraft,* Washington, D.C.: GAO-01-163.

Weibull, W. 1939. "A Statistical Theory of the Strength of Materials," *Ingeniors Vetenskaps Akademien Handlingar* [Proceedings of the Royal Swedish Institute for Engineering Research], 151:1–45.

Weisberg, S. 1985. *Applied Linear Regression*, 2nd Edition. New York: Wiley.

Wood, S. 2003. "Thin Plate Regression Splines," *Journal of the Royal Statistical Society, Series B*, 65(1):95–114.

Xiang, D. 2001. *Fitting Generalized Additive Models with the GAM Procedure,* Cary, NC: SAS Institute, Inc., paper P256-26.