SYM-AM-21-040

Proceedings
of the
Eighteenth Annual Acquisition Research Symposium

Tuesday, May 11, 2021 Sessions
Volume I

Acquisition Research:
Creating Synergy for Informed Change

May 11–13, 2021

Published: May 5, 2021

Prepared for the Naval Postgraduate School, Monterey, CA 93943.

Acquisition Research Program
Graduate School of Defense Management
Naval Postgraduate School

The research presented in this report was supported by the Acquisition Research Program of the Graduate School of Defense Management at the Naval Postgraduate School.

To request defense acquisition research, to become a research sponsor, or to print additional copies of reports, please contact any of the staff listed on the Acquisition Research Program website (www.acquisitionresearch.net).

Table of Contents

SYM-AM-21-040

PROCEEDINGS
OF THE
EIGHTEENTH ANNUAL
ACQUISITION RESEARCH SYMPOSIUM

TUESDAY, MAY 11, 2021 SESSIONS
VOLUME I

Acquisition Research:
Creating Synergy for Informed Change

May 11-13, 2021

Published: May 5, 2021

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF DEFENSE MANAGEMENT
NAVAL POSTGRADUATE SCHOOL

THIS PAGE INTENTIONALLY LEFT BLANK

WELCOME: DR. ROBERT (BOB) MORTLOCK, PRINCIPAL INVESTIGATOR, ACQUISITION RESEARCH PROGRAM

Dr. Robert Mortlock, PhD, CMBA, PMP, PE, COL USA (Ret), — Dr. Mortlock is the Principal Investigator, Acquisition Research Program, Naval Postgraduate School, managed defense systems development and acquisition efforts for the last 15 of his 27 years in the U.S. Army, culminating in his assignment as the project manager for Soldier Protection and Individual Equipment in the Program Executive Office for Soldier. He retired in September 2015 and now teaches defense acquisition and program management in the Graduate School of Business and Public Policy at the Naval Postgraduate School in Monterey, California. He holds a Ph.D. in chemical engineering from the University of California, Berkeley, an MBA from Webster University, an M.S. in national resource strategy from the Industrial College of the Armed Forces and a B.S. in chemical engineering from Lehigh University. He is also a recent graduate from the Post-Doctoral Bridge Program of the University of Florida's Hough Graduate School of Business, with a management specialization. He holds DAWIA Level III certifications in program management (PM), test & evaluation (T&E), and systems planning, research, development & engineering (SPRDE).

WELCOME: DAVID H. LEWIS, VICE ADMIRAL, U.S. NAVY (RET), ACQUISITION CHAIR, ACQUISITION RESEARCH PROGRAM

Vice Admiral David H. Lewis, USN (Ret.) took the helm as the Naval Postgraduate School Chair of Acquisition. As chair, he will lead the Acquisition Research Program (ARP) in the Graduate School of Defense Management and connect NPS with leaders and policymakers in the acquisition community.

Lewis graduated from NPS in 1988 with a Master of Science in Computer Science, and we're pleased to welcome him back to campus in this leadership role. Lewis is replacing the founding Chair of Acquisition, Rear Admiral, USN (Ret.) Jim Greene, who retired this June.

Most recently, Lewis served as Director of the Defense Contract Management Agency, managing over $7 trillion in defense contracts. In this role, he oversaw the agency's efforts to ensure that supplies and services contracted for by the Department of Defense are delivered on time and in line with contract performance requirements.

During his career at sea, Lewis served as a communications officer, fire control and missile battery officer, and combat systems officer aboard destroyers and guided-missile cruisers.

Upon selection to flag rank in 2009, Lewis served as Vice Commander, Naval Sea Systems Command and then served four years as Program Executive Officer, Ships, where he directed the delivery of 18 ships and procurement of another 51 ships. From 2014-2017 he served as Commander, Space and Naval Warfare Systems Command where he led a global workforce of 10,300 civilian and military personnel who design, develop and deploy advanced communications and information capabilities.

Lewis's extensive experience in shipbuilding has given him a unique understanding of the full acquisition lifecycle. He has delivered ships as a program manager and program executive officer, then later sustained and modernized them as a fleet engineer and systems commander. He will bring valuable perspective to NPS students and faculty, as well as the broader acquisition innovation community working to get superior capabilities into the hands of our warfighters.

Lewis's expertise in product delivery will amplify ARP's ability to execute its mission of delivering the real-time information and analytical capabilities needed by today's acquisition professionals and policymakers. Adding VADM Lewis to the team also demonstrates NPS's continued commitment to providing world-class defense-focused education and research...

WELCOME: ANN E. RONDEAU, Ed.D, VICE ADMIRAL, U.S. NAVY (RET.), PRESIDENT, NAVAL POSTGRADUATE SCHOOL

Ann E. Rondeau, Ed.D, Vice Admiral, U.S. Navy (Ret.), was appointed as President, Naval Postgraduate School on January 29, 2019. She brings to the assignment an unparalleled record of leadership and achievement within the military and academia in the areas of education, training, research, executive development, change management, and strategic planning. Prior to her appointment, Adm. Rondeau served as the sixth president of the College of DuPage. Her most recent military position was as the President of the National Defense University, a consortium of five colleges and nine research centers in Washington, DC.

Rondeau has extensive leadership experience in significant military and educational roles. In 1985, she was selected and served as a White House Fellow in the Reagan Administration and went on to serve as the Deputy Commander of the U.S. Transportation Command in Illinois, Pentagon Director/Chief of Staff for the U.S. Navy Staff, Commander of the Navy Personnel Development Command in Virginia, Commander of the Naval Service Training Command at Great Lakes, Ill., Pacific Fleet Staff Chief of Staff in Hawaii, Commanding Officer of Naval Support Activity in Tennessee and other staff and commanding responsibilities with policy, planning, Fleet support, joint logistics, training and education. Rondeau retired from the U.S. Navy as a three-star admiral in 2012 and was the second woman to have achieved that rank in the Navy. She then served as a partner and later an independent consultant with the IBM Watson group.

President Rondeau's leadership has served many, both past and present, to include: Board of Directors, United States Institute of Peace; Board of Directors, German Marshall Fund; Board of Directors, The Atlantic Council; Board of Directors, National Museum of the American Sailors; Board of Directors, Council of Higher Education Accreditation; Board of Directors, Chicago Regional Growth Corporation; Board of Directors, Choose DuPage (regional development organization for Chicago northwest suburbs); Tennessee/Mid-South Economic Development Board; DoD liaison to the Center for the Study of the Presidency; Military Advisory Board (studying energy and environment impacts on national security); Flag Officer Advisory Council for Arizona State University, the National Naval Officers Association Senior Advisory Panel, the Eisenhower Memorial Commission and the National Cold War Veterans Memorial Design Steering Committee among others.

Rondeau holds a B.A. from Eisenhower College (NY), an M.A. from Georgetown University (DC) and an Ed.D. from the College of Education at Northern Illinois University in DeKalb. She also holds an honorary Doctorate in Public Service from Carthage College (Kenosha, WI) and an honorary Doctorate in Humane Letters from Rosalind Franklin University of Medicine and Science (Chicago, IL)....

KEYNOTE SPEAKER: STACY CUMMINGS – ACTING, UNDERSECRETARY OF DEFENSE FOR ACQUISITION AND SUSTAINMENT

Ms. Stacy Cummings is a career member of the Senior Executive Service, and is currently performing the duties of the Under Secretary of Defense for Acquisition and Sustainment (USD(A&S)). In this position, she is responsible to the Secretary of Defense for all matters pertaining to acquisition; contract administration; logistics and materiel readiness; installations and environment; operational energy; chemical, biological, and nuclear weapons; the acquisition workforce; and the defense industrial base.

Prior to this temporary role, she served as the Principal Deputy Assistant Secretary of Defense for Acquisition (PDASD(A)). In this position, she advised the Assistant Secretary of Defense for Acquisition (ASD(A)) on matters relating to the Department of Defense Acquisition System while advancing innovative, data-driven approaches across the acquisition enterprise.

Previously serving as the Program Executive Officer, Defense Healthcare Management Systems (PEO DHMS), Ms. Cummings managed the delivery of healthcare and advance data sharing through a modernized electronic health record for service members, veterans, and their families.

Ms. Cummings previously held senior executive positions at the Department of Transportation, where she established strategic direction, provided executive leadership, and managed daily operations as the Executive Director for the Federal Railroad Administration and the interim Executive Director for the Pipeline and Hazardous Material Safety Administration.

Beginning her career with the Department of the Navy, she held senior positions at the Naval Air Technical Data and Engineering Services Command; Commander, Fleet Readiness Centers; Program Executive Office for Command, Control, Communications, Computers and Intelligence; and the Space and Naval Warfare Systems Command.

Ms. Cummings holds a Master of Science in National Resource Strategy from the Industrial College of the Armed Forces and a Master of Science in Management/Information Systems from the Florida Institute of Technology. She received her Bachelor of Science in Business Logistics from the Pennsylvania State University.

Certified in both Program Management and Acquisition Logistics, Ms. Cummings is a graduate of the Naval Air Systems Command's Senior Executive Management Development Program and the Defense Senior Leader Development Program. Ms. Cummings received Meritorious and Superior Civilian Service Awards from the Department of the United States Navy, Meritorious Public Service Award from the United States Coast Guard, and the Office of the Secretary of Defense Medal for Exceptional Civilian Service. Stacy Cummings is a career member of the Senior Executive Service, and currently serves as the Principal Deputy Assistant Secretary of Defense for Acquisition (PDASD(A)). In this position, she advises the Assistant Secretary of Defense for Acquisition (ASD(A)) on matters relating to the Department of Defense Acquisition System while advancing innovative, data-driven approaches across the acquisition enterprise.

Previously serving as the Program Executive Officer, Defense Healthcare Management Systems (PEO DHMS), Ms. Cummings managed the delivery of healthcare and advance data sharing through a modernized electronic health record for service members, veterans, and their families.

PANEL 1. IMPLICATIONS OF THE NEXT ADMINISTRATION FOR THE DEFENSE ACQUISITION SYSTEM

Tuesday, May 11, 2021	
7:45 a.m. – 9:00 am.	**Chair: Todd Harrison,** Senior Fellow, International Security Program and Director, Defense Budget analysis and Aerospace Security Project, Center for Strategic and International Studies **Panelists:** **Hon. David Berteau,** Former Assistant Secretary of Defense for Logistics and Materiel Readiness and Current President & CEO, Professional Services Council **Hon. Elaine McCusker,** Former Acting Under Secretary of Defense (Comptroller) and Current Resident Fellow, American Enterprise Institute **Hon. Peter Levine,** Former Deputy Chief Management Officer of the Department of Defense and current Senior Research Fellow, Institute for Defense Analysis

Todd Harrison—is the director of Defense Budget Analysis and director of the Aerospace Security Project at CSIS. As a senior fellow in the International Security Program, he leads the Center's efforts to provide in-depth, nonpartisan research and analysis of defense funding, space security, and air power issues. He has authored publications on trends in the defense budget, military space systems, threats to space systems, civil space exploration, defense acquisitions, military compensation and readiness, and military force structure, among other topics. He teaches classes on military space systems and the defense budget at the Johns Hopkins School of Advanced International Studies.

Hon. David Berteau—Mr. Berteau is PSC President and CEO, with 400 member companies of all sizes providing federal contract services. Mr. Berteau was ASD for Logistics and Materiel Readiness and served 14 years in the Defense Department, under six defense secretaries. Earlier, Mr. Berteau was at the Center for Strategic and International Studies (CSIS), Syracuse University's National Security Studies Program, and SAIC. He is a Fellow of the National Academy of Public Administration and taught graduate courses for 14 years at the Maxwell School, Georgetown, and the LBJ School.

Hon. Elaine McCusker—is a resident fellow at the American Enterprise Institute (AEI), where she focuses on defense strategy, budget, and innovation; the US military; and national security. She has a background in defense planning and budgeting, military campaign assessments, defense data analytics, contingency operations and science and technology, and substantial experience in resolving complex strategic and tactical-level challenges, including those with international dimensions.

Hon. Peter Levine— Peter Levine is Senior Research Fellow with the Institute for Defense Analyses. Prior to joining IDA, Peter most recently performed the duties of the Under Secretary of Defense for Personnel and Readiness. In this capacity, he served as principal assistant and advisor to the Secretary and Deputy Secretary of Defense on readiness; National Guard and Reserve component affairs; health affairs; training; and personnel requirements and management, including equal opportunity, morale, welfare, recreation, and quality of life.

PANEL 2. SHIP MAINTENANCE AND ACQUISITION

Tuesday, May 11, 2021	
9:15 a.m. – 10:00 a.m.	**Chair: Glen Sturtevant,** Director for Science & Technology, Program Executive Office for Ship ***A Governance Model and Safety Management System Framework for Industrial Fire Safety during Naval Ship Maintenance Availabilities*** Cheryl Marek, University of Maryland John McGowan, NAVSEA Qingbin Cui, University of Maryland Joseph Bradley, Main Sail LLC ***Functional Hazard Analysis (FHA) and Subsystem Hazard Analysis (SSHA) of Artificial Intelligence/Machine Learning (AI/ML) Functions within a Sandbox Program*** Bruce Nagy, NAVAIR, China Lake Guna Sivapragasam, Naval Sea Weapons Division, Dahlgren Loren Edwards, NAVAIR, Weapons Division, China Lake ***Changing Course to a 21st Century Acquisition Strategy: Navy-Industry Collaborative Design*** Tim Nichols, Ubiquitech Llc Peter Jaquith, Consultant Robert Keane, Ship Design USA Barry Tibbitts, CAPT USN (ret.), Consultant

Glen Sturtevant— is the Director for Science and Technology assigned to the United States Navy Department's Program Executive Office for Ships in Washington, D.C.

He graduated from College du Leman in Geneva, Switzerland, earned a Bachelor of Science in Civil Engineering from the University of Delaware and a Master's degree in Management from Indiana University. He has completed Program Management and Engineering programs of study at National Defense University, Webb Institute, Virginia Tech and MIT.

Mr. Sturtevant began his career in 1978 as a Project Engineer at Philadelphia Naval Shipyard. In 1983 he was assigned to the Surface Ships Directorate at Naval Sea Systems Command Headquarters in Washington D.C. where he was responsible for the conversion of USS BELKNAP (CG 26) to U.S. 6th Fleet Flagship in Gaeta, Italy, the overhaul of USS STERETT (CG 31) in Subic Bay, Republic of the Philippines and the modernization of USS BIDDLE (CG 34), the first ship of the Navy's New Threat Upgrade Program.

In 1987 he was assigned to the Aegis Shipbuilding Program (PMS 400) where he held positions as Plans and Programs Manager, Cruiser Conversion Program Manager, and from 1998 to 2004 was Program Manager for the Navy's Smartship Program where he led the introduction of commercial technologies to the Fleet. Mr. Sturtevant was the Navy's Response Team leader activated as a result of the al Qaeda bombing of USS COLE (DDG 67) in Aden, Yemen in October 2000.

As Chief Technology Officer, his duties include senior advisor to the PEO and NAVSEA's Deputy Commander for Surface Warfare for National Shipbuilding Research Program, Flexible Ships, Arctic Operations, Energy Security, Acquisition Research at the Naval Postgraduate School, Small Business Innovation Research Program, Allied Navy collaboration, the Ship Design & Analysis Tools Strategic Sealift R&D and Unmanned Systems-Ship Integration.

He is a member of the American Society of Naval Engineers Programs Committee, World Scientific Engineering Academy and Society, Surface Navy Association, American Management Association, Navy League of the United States, National Defense Industrial Association, and has served on the Association of Scientists and Engineers Professional Development Committee and as Chair of the Science and Education Committee. He is a Project Management Institute certified Project Management Professional and is a certified Defense Acquisition Corps Professional in Program Management and Systems Engineering. Mr. Sturtevant has received the Association of Scientists and Engineers Professional Achievement Award, Office of the Secretary of Defense's Aegis Cruiser Reduced Total Ownership Cost Award, the individual Aegis Excellence Award and is a member of the Pi Alpha Alpha National Honor Society.

Mr. Sturtevant has provided professional consultation to multiple allied Navies, has written numerous technical papers and has been published in several professional society journals. He is married to Karen Birkofer Sturtevant and they have 3 grown children.

A Governance Model and Safety Management System Framework for Industrial Fire Safety During Naval Ship Maintenance Availabilities

Cheryl Marek—is a licensed Professional Fire Protection Engineer and PhD Candidate in Civil Engineering at the University of Maryland, College Park. Her research interests lie at the intersection of fire protection engineering, systems engineering, and the social sciences. She has several years of experience in the shipbuilding industry, is an active member of the Society of Naval Architects and Marine Engineers (SNAME), and is studying acquisition system influences on shipboard industrial fire safety. [cmarek@umd.edu]

Joseph Bradley—is Chief Scientist at Main Sail, LLC, and an Adjunct Associate Professor at Old Dominion University (ODU). Prior to joining ODU, he was Deputy Director for Force Maintenance at Commander, Submarine Force, Atlantic Fleet. He has served in various consulting roles, including Program Manager's Representative for the SSGN conversion of the USS *Ohio* and USS *Michigan*. His research interests include complex system governance and decision-making using modeling and simulation. He received a Degree of Mechanical Engineering and a Master of Science in mechanical engineering from the Naval Postgraduate School, and a BE from The Cooper Union. [josephbradley@leading-change.org]

John (Jack) McGowan—is a Certified Safety Professional who works to help NAVSEA safety professionals reduce injuries and mishaps at NAVSEA Field Activities, which include Warfare Centers, Regional Maintenance Centers, and Naval Shipyards. Before starting at NAVSEA Headquarters, he worked as a Safety Manager at Puget Sound Naval Shipyard, San Diego Detachment, and as a Project Safety Manager at Puget Sound Naval Shipyard, Bremerton. McGowan is now the lead Safety Professional for the Information and Analysis Branch in NAVSEA 04, Industrial Operation Directorate. [john.mcgowan@navy.mil]

Qingbin Cui—is a Professor at the University of Maryland, College Park. He has over 20 years of industry and research experience and specializes in the areas of project management, public–private partnerships (P3), and sustainability. He is a technical consultant to the U.S. Department of Transportation for the P3 Capacity Building Program and a member of the expert panel for the U.S. Department of Energy on energy performance contracting, and built a university–industry partnership for data-driven construction research and education. [cui@umd.edu]

Abstract

Managing shipboard industrial fire safety during a depot or intermediate level maintenance availability on a commissioned naval vessel can be viewed as a complex system bounded by the defense acquisition system. Sociotechnical factors define the hazard and associated risk and risk management practices, and this complex system governs the resulting level of fire safety. Poor industrial fire safety practices during naval ship maintenance availabilities can directly impact project cost and schedule. If a fire is severe enough, these effects can trickle throughout the ship class's maintenance program, adversely impacting fleet readiness and operations. Traditional viewpoints on fire safety prescriptively regulate the fire hazard. Rote compliance with this type of requirement does not provide clear mechanisms for measuring safety performance, resulting in uneven risk management. This paper presents a safety management system (SMS) framework for shipboard industrial fire safety based on the Complex Systems Governance (CSG) reference model developed by Keating and Bradley (2015). The value of a clearly defined governance model and SMS framework in conjunction with industry standardization and information sharing is the emergence of trends. This supports feedback loops between requirements and outcomes, allowing more effective management of fire safety across the broad stakeholder groups involved.

Introduction

Shipboard fire incidents during industrial construction and maintenance activities pose a significant threat to fielding the required naval force strength. Naval vessels transition in and out of construction and maintenance periods throughout their life cycle, and protecting them from fires during industrial work is a complex issue. Fire risk involves not only risk associated with the fire itself, but also the risk associated with loss of an operational asset. The impact of small fires during ship construction or overhaul could be minor property damage or injury to a worker or sailor. Larger fires can result in schedule delays or loss of life, and a major fire can result in the total loss of a ship.

When the USS *Miami* (SSN 755) was decommissioned after suffering a major fire while in dry dock at Portsmouth Naval Shipyard in 2012, five deployments over her remaining 10-year operational life were lost (McDermott, 2013). Investigation of the July 12, 2020, fire aboard the USS *Bonhomme Richard* (LHD 6) while pier side at Naval Base San Diego, CA, towards the end of a GD NASSCO overhaul availability is ongoing, with the ship declared a total loss in early 2021 (Ziezulewicz, 2021). Shortly after this fire, Assistant Secretary of the Navy for Research, Development, and Acquisition James Geurts (2020) issued a memorandum to the entire shipbuilding and ship maintenance enterprise stating, “Preventing shipboard fires is a team sport, no matter where the ship is in its life cycle, and no matter who is working on the ship,” and “There is no place in our Navy for complacency – the lives of our teammates and the accomplishment of our mission depends upon it” (p. 1).

The USS *Miami* (SSN 775) Fire Panel Recommendations (Gortney, 2012, p. 1) also cite complacency, stating “the MIAMI investigation paints a picture of multiple processes within several organizations going through the motions, with no particular failure, but lacking focused attention and oversight, and missing the mark in the aggregate”, and “it is clear that the Navy has unintentionally accepted a reduced margin to fire safety when a ship enters an industrial environment -- where the risk of fire is at its highest.” In the weeks following the major fire on USS *Bonhomme Richard*, minor fire incidents related to hot work occurred on the USS *John F Kennedy* (CVN 79) at HII-Newport News Shipbuilding and USS *Kearsarge* (LHD 3) at GD NASSCO in Norfolk, VA (Eckstein, 2020), serving as a wake-up call to the industry.

A few months prior to the USS *Bonhomme Richard* fire, Naval Sea Systems Command (NAVSEA) presented the results of an effort initiated in June 2019 in response to 2018 fires aboard USS *Oscar Austin* (DDG 79) and USS *Fitzgerald* (DDG 62) to assess industrial fire safety, fire prevention, and control programs based on self-assessment responses and reported data on fires that had occurred over the previous 30 months at public and private maritime maintenance facilities (McGowan & Smith, 2020). Their effort is summarized herein and built upon to propose the framework for a data-driven direction to addressing the Navy’s shipboard industrial fire problem.

In June 2020, the Navy issued a new Safety and Occupational Health Manual that established the Navy Safety Management System (SMS) to “facilitate a transition from reactively managing safety to proactively managing safety and risk, and ultimately, to become predictive” (U.S. Department of the Navy [DoN], 2020, p. A1-2). This evolving direction provides an opportunity for deliberate implementation of a novel approach to managing fire safety during naval ship maintenance availabilities.

Our approach aims to assess common systemic threads in major safety mishaps—contributions of the system that frames management of risk to outcomes and the relationship between complacency and responsibility. The dual incidents of the recent major fire aboard USS *Bonhomme Richard* (LHD 6) and total loss of the USS *Miami* parallel the

National Aeronautics and Space Administration's (NASA's) dual losses of space shuttles Challenger in 1986 and Columbia in 2003, albeit separated by a much shorter period of time. The Columbia Accident Investigation Board (CAIB) identified organizational system failures and flawed organizational practices, including complacency and cultural beliefs. Specifically, "history again at work: how past definitions of risk combined with systemic problems in the NASA organization caused both accidents" (Columbia Accident Investigation Board [CAIB], 2003, p. 195).

We propose an SMS framework for shipboard industrial fire safety based on the Complex Systems Governance (CSG) reference model developed by Keating and Bradley (2015) to inform the transition from regulating fire hazards to systematic management of fire safety during maintenance availabilities on commissioned naval vessels. The focus is on developing feedback loops between requirements and outcomes through a data-driven approach to support proactive management of shipboard industrial fire safety, a reduction in lost operational days due to fire incidents, and identification of the gaps that must be addressed to implement this framework.

Table 1. Research Goals and Objectives

Goals	Objectives
Dissect the shipboard industrial fire safety problem, including acquisition system influences	Evaluate the problem within the sociotechnical safety perspective on risk
	Evaluate the problem with respect to defense-in-depth
Develop an SMS framework for Shipboard Industrial Fire Safety based on the CSG reference model	Evaluate the problem within the Cynefin framework
	Evaluate how the Navy SMS framework integrates into the CSG reference model
	Identify necessary feedback loops between requirements and outcomes
	Identify gaps between framework architecture and the current state

NAVSEA Shipboard Industrial Fire Incident Data Analysis Summary

In June 2019, "COMNAVSEA directed Naval activities and requested private maintenance facilities to report all fires over the last 30 months. NAVSEA received responses from public and private maritime maintenance facilities. These responses were analyzed by 04RS for completeness, self-reflection, and innovation/solutions" (McGowan & Smith, 2020, slide 2). Data from 339 fire incidents were reviewed, and NAVSEA 04RS (Industrial Operations, Safety) performed an analysis of causal factors using the Department of Defense's Human Factors Analysis and Classification System (HFACS) to identify trends and lessons learned. The top three sources of shipboard fires during industrial work were hot work, electrical, and temporary sources. Shipboard industrial fire incidents were also analyzed with regards to the Cognizant Activity and controlling document for fire safety requirements. Results are summarized in Figure 1.

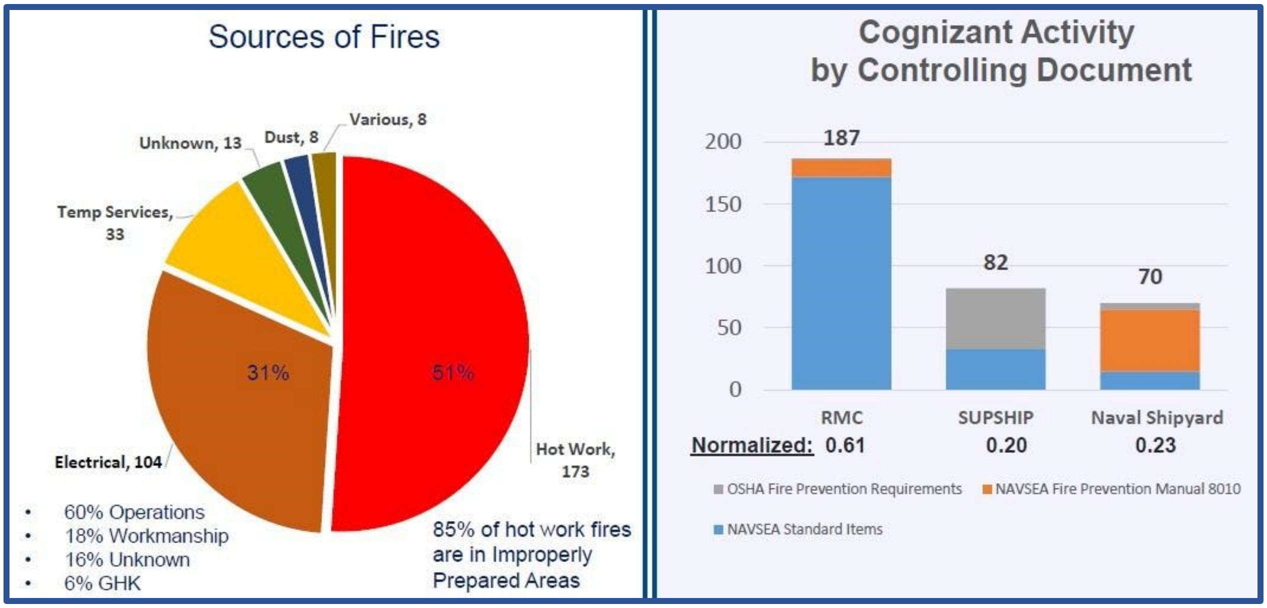


Figure 1. Fire Reporting Data
(McGowan & Smith, 2020, slide 4)

Regardless of the source of the requirements, non-compliance related to fundamental fire prevention practices was identified as the cause of 85% of hot work fires, which occurred in improperly prepared areas. The distribution of fires among Cognizant Activities is related to the volume of work overseen by each, suggesting that the problem is universal across the naval maritime repair industry.

Significant findings outlined next "were remarkably similar to other fires experienced by the Navy," including "1942 SS Normandie, Total loss of the ship, 285 Injuries" and "1944 USS Saturn, 15 fatalities, 20 injuries." "Both events were the result of improper hot-work controls" (McGowan & Smith, 2020, slide 8). These findings include the following:

- Failure to follow the "35-foot rule" is the most frequent cause of hot work fires.
- Port loading variations and high influx of temporary employees with low experience increase the likelihood of fire safety non-compliance.
- Communication between fire watches and hot workers, and inspections by hot work supervisors, are less than adequate.
- Failure to comply with the invoked standard (OSHA, NAVSEA Requirements) was common. (McGowan & Smith, 2020, slide 7)

In addition to the analytic side of this effort, actions taken by NAVSEA included establishing new hot work requirements in NAVSEA Standard Item (NSI) Fiscal Year 2020, Change 2, and "VADM Moore wrote a letter to all parties stating that a ship repair contractor's non-compliance with fire prevention standards and regulations is a contractual non-compliance" (McGowan & Smith, 2020, slide 10). A variety of recommendations to improve compliance were made for Naval maritime facilities (Naval shipyards, Regional Maintenance Centers (RMCs) and the Supervisor of Shipbuilding (SUPSHIP/SOS). Future initiatives identified by NAVSEA include sharing best practices across the industry and establishing a multi-functional committee to address issues uncovered through information sharing (McGowan & Smith, 2020). Uniformly implementing good and best practices in private shipyards requires a combination of approaches, largely due to contractual influences.

Methodology

Defining the Problem

- ***Sociotechnical Safety Perspective on Risk***

The concept of a sociotechnical system is derived from the context of work systems and was defined by Ropohl (1999) as

> established to stress the reciprocal interrelationship between humans and machines and to foster the program of shaping both the technical and the social conditions of work, in such a way that efficiency and humanity would not contradict each other any longer. (p. 59)

Sociotechnical systems in this sense are made up of a hierarchy of action systems, or subsystems, with three primary levels that perform the work, oversee the work, and set goals for the work. These subsystems can further be divided into sub-subsystems, and significant feedback loops and couplings exist between them. In the case of overhaul work on a commissioned naval vessel, these lines can sometimes be blurred, and a single organization can have functions within multiple subsystems. Sociotechnical is further defined by Aven and Ylönen (2018)

> to include the following dimensions: 1) two or more persons, interaction with some form of 2) technology, 3) and internal work environment (both physical and cultural), 4) external environment (can include political, regulatory, technological, economic, educational and cultural sub-environments), 5) an organisational design and management subsystems. (p. 14)

Continuing the parallels to the dual NASA accidents, the CAIB devoted chapters of its report to discussing organizational causes, history as a cause, and decision-making within the organization. The report identified a broken safety culture and pointed to Naval Reactors and the Navy's SUBSAFE programs as strong examples of a good culture, with key differences being requirements ownership (technical authority) and emphasis on lessons learned (CAIB, 2003).

- ***Safety vs. Risk***

A goal of the Navy SMS is to proactively manage both safety and risk (DoN, 2020). To do this, it is important to distinguish between these two interrelated objectives. The Society for Risk Analysis Glossary defines safe as "without unacceptable risk," and safety as "the antonym of risk" (Society for Risk Analysis [SRA], 2018, p. 7). Risk is given several qualitative definitions, but for the context of this paper we will use "the occurrences of some specified consequences of the activity and associated uncertainties" (SRA, 2018, p. 4). Möller et al. (2006) argue that safety goes beyond the antonym of risk due to epistemic uncertainty and proposed the intersubjective concept of safety: "(1) it is based on the comparative judgments of severity of harm that the majority of humans would agree on, and (2) it makes use of the best available expert judgments on the probabilities and uncertainties involved" (p. 427). Aven (2009) argues that safety is the antonym of risk for certain perspectives (definitions) of risk, specifically, when risk is the two-dimensional combination of uncertainty and consequences, uncertainty is integral to the definition, safety is the antonym of risk, and safe can be defined as "acceptable risk and acceptable safety" (p. 929). This is the perspective adopted in this paper.

- ***Defense-in-Depth***

Sorensen et al. (1999) identify that defense-in-depth as a nuclear industry safety strategy began development in the 1950s. Their review of the history of the term indicated that there was as of yet no official or preferred definition, but that when the term is used and if a definition is needed, “one is created consistent with the intended use of the term. Such definitions are often made by example” (p. 1). By 1999, the term had come to have two different meanings, roughly corresponding to the perspective of the particular model. Those perspectives were cast as either denoting “the philosophy of high level lines of defense, such as prevent accident initiators from occurring, terminate accident sequences quickly, and mitigate accidents that are not successfully terminated” (p. 3). The other portrays “the multiple physical barrier approach, most often exemplified by the fuel cladding, primary system, and containment” (p. 2). These two model perspectives are cast as either structuralist or rationalist:

> The structuralist model asserts that defense in depth is embodied in the structure of the regulations and in the design of the facilities built to comply with those regulations. The requirements for defense in depth are derived by repeated application of the question, “What if this barrier or safety feature fails?”. (pp. 3–4)

Sorensen et al. (1999) portray that “the rationalist model asserts that defense in depth is the aggregate of provisions made to compensate for uncertainty and incompleteness in our knowledge of accident initiation and progression” (p. 4). They also assert that “the structuralist and rationalist models are not generally in conflict. Both can be construed as a means of dealing with uncertainty,” and further, they note that “neither incorporates any reliable means of determining when the degree of defense in depth achieved is sufficient” (p. 5).

As more nuclear power plants were built and more service experience acquired, new rules were progressively added, yielding a very complex set of requirements for the last part of the existing fleet of reactors to be built. A variation of technical debt was building up; even with “the accumulation of experience with various incidents and accidents, a growing list of unresolved safety issues emerged” (Fleming & Silady, 2002, p. 206). Fleming and Silady (2002) highlight that even as requirements were increased,

> many additional incidents occurred, including literally hundreds of common cause failures in redundant safety systems. This experience casts doubt on the wisdom of excluding common cause failures from the design basis envelope, thereby exposing a serious limitation of the single failure criterion as a tool to help define what is credible. (p. 206)

A footnote in Fleming and Silady (2002) notes that

> In the peer review of an earlier draft of this paper it was pointed out that the regulations governing nuclear power include one definition of defense-in-depth in 10 CFR Part 50 Appendix R which sets rules for fire protection in older plants. This definition sets forth the following objectives for the defense-in-depth of fire protection. Prevent fires from starting, detect rapidly, control and extinguish the fires that do occur, and to protect SSCs needed to safely shutdown the plant from the effects of the fire and firefighting activities. (p. 207)

Saleh et al. (2014), discussing the Texas City refinery fire, noted “a fundamental failure mechanism in this accident, namely the absence of observability or ability to diagnose hazardous states in the operation of the refinery, in particular within the raffinate splitter tower and the blowdown drum of the isomerization unit" (p. 1). They go on to "propose a general safety–diagnosability principle for supporting accident prevention, which requires that all safety-degrading events or states that defense-in-depth is meant to protect against be diagnosable, and that breaches of safety barriers be unambiguously monitored and reported." Further "violation of the safety–diagnosability principle translates into a shrinking of the time window available for operators to understand an unfolding hazardous situation and intervene to abate it." They go onto conclude that "defense-in-depth be augmented with this principle, without which it can degenerate into an ineffective defense-blind safety strategy" (Saleh et al., 2014, p. 1).

Cynefin Framework

Cynefin framework is a sense-making framework first developed by a group of researchers at IBM "conducting a program of disruptive action research using the methods of narrative and complexity theory to address critical business issues" (Kurtz & Snowden, 2003, p. 462). This work was partially funded through the Defense Advanced Research Project Agency (DARPA) with an interest in "new approaches to support policy-making" (Kurtz & Snowden, 2003, p. 462). The group challenged three basic assumptions of decision-making and policymaking—order, rational choice, and intentional capability—believing that while commonly available tools and techniques assume they are true, they are not universally true. With regard to order, they discuss situations where lack of order isn't a bad thing and the concept of emergent order (un-order) that is self-organizing rather than controlled and emerges from the interaction of many entities. Ordered-systems thinking has limitations because it assumes "we can derive or discover general rules or hypotheses that can be empirically verified and that create a body of reliable knowledge, which can then be developed and expanded" (Kurtz & Snowden, 2003, p. 466), and this does not hold true for all domains.

Unlike the more traditional categorization framework with a two-by-two matrix where the most desirable condition exists in the upper right-hand quadrant, the Cynefin sense-making framework "is used primarily to consider the dynamics of situations, decisions, perspectives, conflicts, and changes in order to come to a consensus for decision-making under uncertainty" (Kurtz & Snowden, 2003, p. 468) and does not favor any of the domains as more desirable or imply value axes. The five domains, currently referred to as clear, complicated, complex, chaotic, and confused, are depicted in Figure 2.

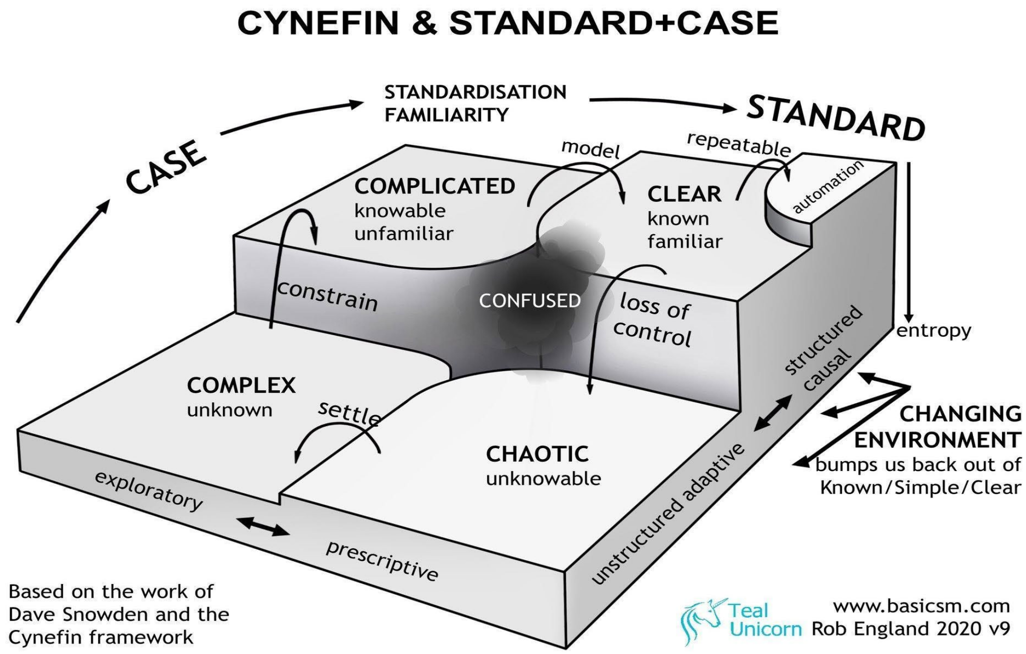


Figure 2. Cynefin Framework and Relationships

Descriptions of Domains

Clear (simple) is the domain of the known, with clear cause and effect relationships that are undisputed and generally empirical in nature. It is the domain of structured techniques and process engineering, with focus on consistency and efficiency (Kurtz & Snowden, 2003). In this ordered and obvious domain, optimal solutions can be identified. In other words, it is the domain of best practices (Fierro et al., 2018).

Complicated is the domain of the knowable, or “known unknowns” (Fierro et al., 2018, p. 6), with stable cause and effect relationships that are either not fully known or only known to a small group of people (Kurtz & Snowden, 2003). With enough time and resources, anything in this domain could become known and move to the clear domain. Kurtz and Snowden (2003) described this as “the domain of methodology, which seeks to identify cause-effect relationships through the study of properties which appear to be associated with qualities” (p. 468). In this domain of reductionism, the goal is to analyze, versus categorize, and decompose a system or problem into its constituent parts with approaches governed by things like standard rules, procedures, and protocols manuals. Because there may be multiple right answers and multiple options must be considered, good practices are preferred to best practices (Fierro et al., 2018).

Complex is the domain of the “unknown knowns” (Fierro et al., 2018, p. 8), where “there are cause and effect relationships between the agents, but both the number of agents and the number of relationships defy categorization or analytic techniques” (Kurtz & Snowden, 2003, p. 469). This is the domain of complexity theory, and “emergent patterns can be perceived but not predicted.” Fierro et al. (2018) refer to this domain as “unordered—obvious in hindsight” (p. 8), where the facts can be understood through reconstruction and rationalized after the fact. In this realm, there a range of potential failures, and emergent behaviors between highly interconnected subsystems can result in the emergence of different failure modes as actions are applied. The best approach in this domain is to "probe, sense, and then respond" (p. 8), and detailed planning is of minimal value due to the dynamic nature of sub-system interactions. An evolutionary strategy is recommended where solutions are developed in builds, and unlike an incremental strategy, it is acknowledged "that the user need is not fully understood and not all requirements can be defined up front" (p. 9).

Chaotic is the domain of "unknown unknowns" (Fierro et al., 2018), where there are no perceivable relationships between cause and effect and there is insufficient response time to evaluate change because the system is turbulent. Here, best practices can contribute to the chaos because there is nothing to analyze and "waiting for patterns to emerge is a waste of time" (Kurtz & Snowden, 2003). This is an unordered domain where "we don't even know which are the relevant aspects related to the problem, and no information is available even to be able to define the problem" (Fierro et al., 2018). Response in this domain is to act quickly to reduce turbulence, sense where stability is or is not present based on this action, and then "respond by working to transform the situation from chaos to complexity, where the identification of emerging patterns can both help prevent future crises that discern new opportunities" (p. 12). This is also known as "the domain of novel practice" (p. 12) and can sometimes be desirable on the path to innovation.

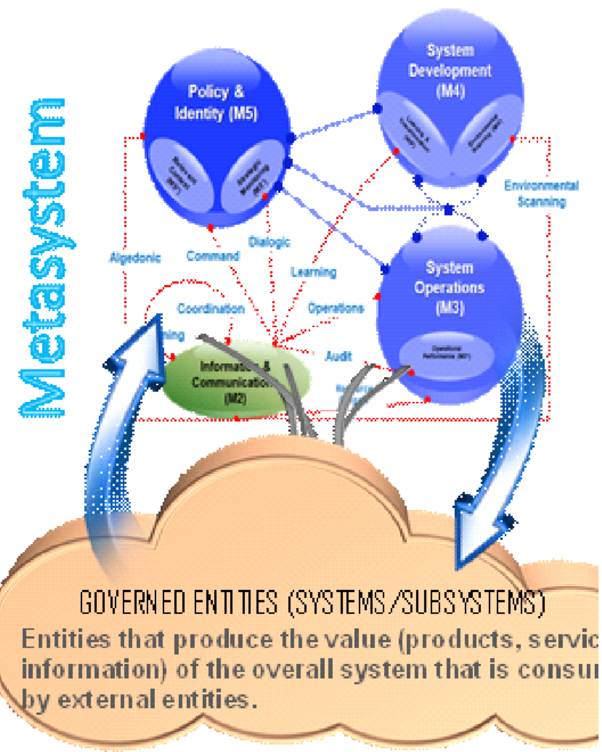
Confused or Aporetic, formerly Disorder, is at the interaction of the other domains of the Cynefin framework, reflective in the conflict of decision-makers approaching a problem from different points of view (Kurtz & Snowden, 2003). In this domain, each decision-maker relies on their own comfort zone to pull the issue into the domain that plays to their strengths or desires. The goal in this domain is to adapt leadership styles based on context and shift the problem into the appropriate domain given the nature of the problem and decision-making context (Fierro et al., 2018).

Complex Systems Governance

Complex System Governance (CSG) is an emerging field formally defined by Keating et al. (2014) as the "design, execution, and evolution of the metasystem functions necessary to provide control, communication, coordination, and integration of a complex system" (p. 274). Theory and concepts are derived from the fields of systems theory, management cybernetics, and system governance. The domain of complex systems includes the characteristics of complexity, contextual dominance, ambiguity, and holistic nature. The governance perspective of CSG is rooted in the cybernetic perspective of "design for 'regulatory capacity' to provide appropriate controls capable of maintaining system balance" (Keating & Katina, 2019, p. 690) and differs from management by emphasizing outcomes versus outputs with a higher degree of separation between action and response. The CSG model provides a framework for improving system performance (Keating et al., 2019). Discovery, classification, and engagement are neither mutually exclusive nor interdependent aspects of the framework, facilitating emergence of unabsorbed pathologies and the ability to improve resilience of the system. Included meta-functions (pathologies), definitions, and their relationships are depicted in Table 2.

Table 2. CSG Reference Model and Metasystem Definitions

Policy and Identity (M5) – overall steering and trajectory for the system includes context and monitoring of the governance system.	
System Development (M4) – models current and future system; includes learning and environmental scanning.	
System Operations (M3) – day to day execution of the metasystem includes performance monitoring of the governed systems.	
Information and Communications (M2) – flow and interpretation of information necessary to execute metasystem functions; the "spine" of the system.	
Governed Systems (S1s) – the organizational components of the larger system that are governed by the meta-systems. They are identified here as Regional Maintenance Commands and the Contractors executing work.	



CSG provides a suitable framework for modeling the complex relationships between the ship/ship's force, contractor/activity performing the work, and various Navy activities involved in the management of shipboard industrial fire risk. A model-centric approach to evaluating these relationships and the delegation of ownership for each aspect of the sociotechnical problem reveals emergence of the system and reduces ambiguity, providing answers to questions such as *who owns the risk* and *are all aspects of the risk accounted for*?

Navy Safety Management System Framework

The Navy Safety program updated its policy in June 2020 with the release of a revised *Navy Safety and Occupational Health Manual* (DoN, 2020). This manual "establishes the Navy Safety Management System (SMS), a comprehensive framework that will ensure operational readiness through continuous improvement and risk-based decision making processes and procedures" (DoN, 2020, p. A1-2). It defines an SMS as "a system of processes that proactively manages day-to-day safety and risk management in an organization across all operations and business lines. It is not a single written policy or database" (DoN, 2020, p. A1-4). The Submarine Safety (SUBSAFE) Program is given as a Navy community-level example, with policy supporting the operational safety functional area of an SMS.

The Navy SMS is a high-level framework intended to be both transparent and scalable. The policy applies directly to Navy civilian and military personnel and operations worldwide but is not applicable where Navy contractors are responsible directly to state or the Department of Labor (DOL) Occupational Safety and Health Administration (OSHA) for the occupational safety and health of its employees. However, the requirement for Navy SMS framework is an enterprise-level policy that encompasses areas beyond occupational safety, including industrial ship safety, industrial ship fire protection and prevention, and safety mishap reporting and investigation (DoN, 2020). Shipboard industrial fire safety is unique in that requirements and outcomes can dually influence both occupational safety and ship safety. Maintaining ship conditions goes above and beyond what is required to protect workers, similar to how building and life safety codes distinguish between life safety and property protection requirements.

The manual is a requirements document that outlines the minimum requirements for an SMS framework, consisting of "an iterative continuous improvement cycle, four pillars, and one or more minimum fundamental elements that underpin those pillars" (DoN, 2020, p. A1-6). The four pillars and their fundamental elements are depicted in Figure 3.

Figure 3. Four Pillars of a Navy SMS and Fundamental Elements

Note that some of the fundamental elements cross, or exist, in multiple pillars. Personnel awareness, education, and training are fundamental to both "Policy and Organizational Commitment" and "Promotion." Similarly, risk monitoring and change management cross "Risk Management" and "Assurance." These particular elements are key to establishing feedback loops between requirements and outcomes, or the continuous improvement cycle.

With regards to appointment of SMS personnel, "SMS-related responsibilities and authorities must be defined, documented, and communicated throughout the organization" and "Safety management system personnel must be appointed with the authority to execute SMS processes and programs." Although aspects of shipboard industrial fire safety are distributed among various safety functional areas, NAVSEA 04RS[1] (Industrial Operations, Safety) the technical warrant holder (TWH) for safety policy, is emerging in a leadership role in addressing fires throughout the naval enterprise. Previously NAVSEA 04RS limited its scope to collecting data and sharing that data within the NAVSEA community. During the past 2 years, NAVSEA 04RS has been engaged with the private shipbuilding community and the maritime industry at large to bring attention to the problem of fires onboard naval vessels during construction and maintenance availabilities. The focus of this paper is on identifying ownership, accountability, and communication channels as they relate to roles and responsibilities within the framework.

Research Results

The Sociotechnical Problem

Complexity emerges from decomposing contributing factors to the Navy's shipboard industrial fire safety problem. Figure 4 outlines the sociotechnical factors that contribute to industrial fire safety during ship maintenance availabilities in each of the five contributing dimensions and corresponding sub-environments.

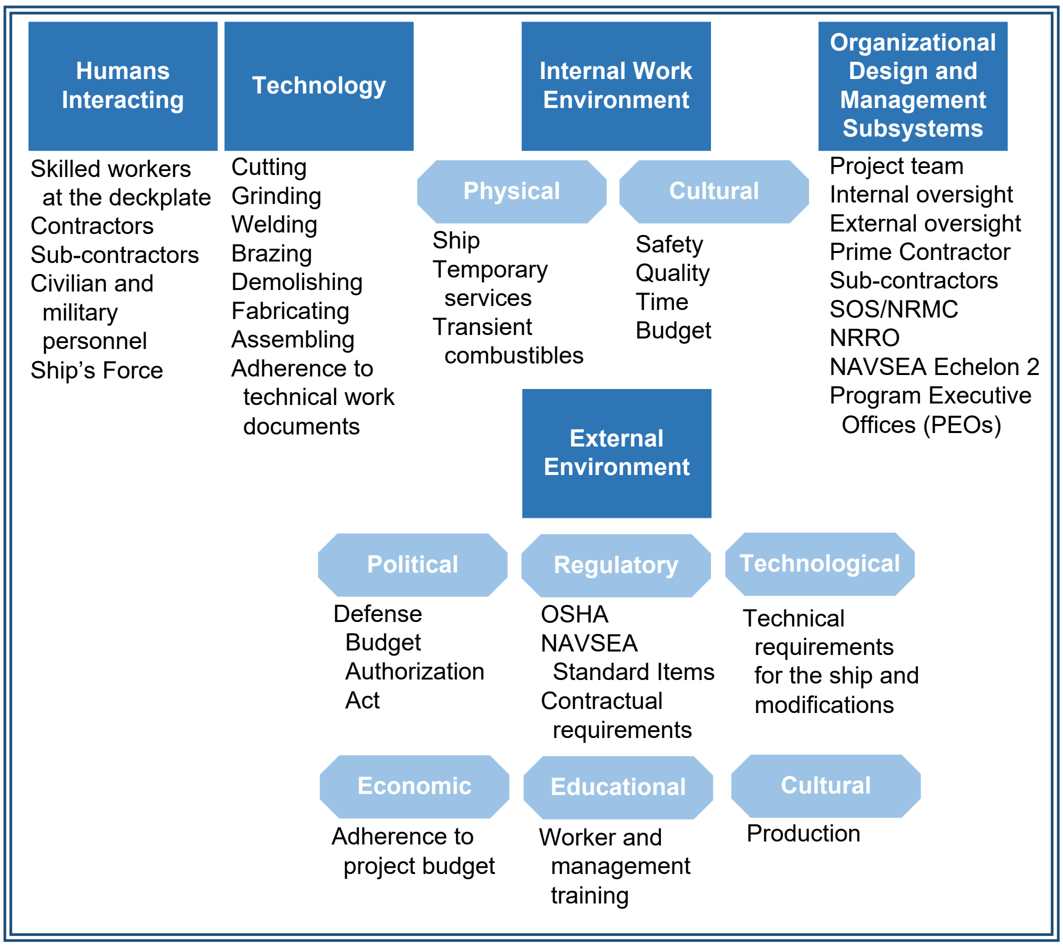


Figure 4. Sociotechnical Decomposition of Shipboard Industrial Fire Safety

In addition to the underlying technical factors, multiple parties are involved with managing the risk. Shipboard industrial fire risk management occurs at the intersection of the contracting authority overseeing the work (NAVSEA), the contractor performing the work, and the ship that is being worked on or constructed. For a commissioned ship, U.S. Navy Regulations dictate that when work is performed in a private shipyard under a contract being administered by the Supervisor of Shipbuilding Conversion and Repair, responsibility for safe execution of the work is assigned via contract to the contractor. The commanding officer retains responsibility for safety of the ship and requesting services necessary to maintain safety of the ship (DoN, 1990). This relationship is depicted in Figure 5.

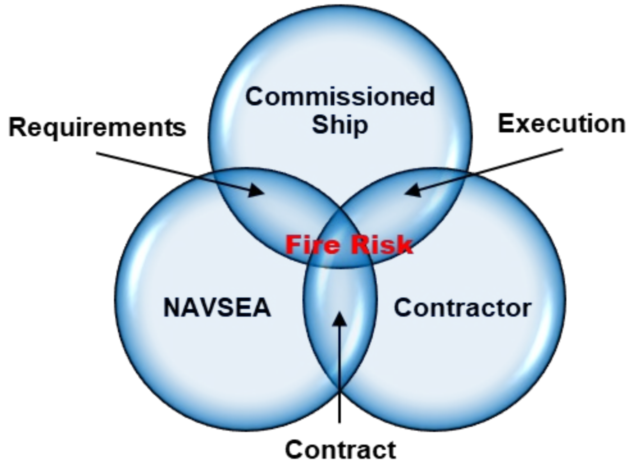


Figure 5. Relationship Between Risk Management Parties

While the responsibility for ship safety is shared, the contractor is directly responsible for occupational safety (DoN, 2020). Unlike in the commercial shipbuilding and repair industry, Navy ship maintenance contracts do not require the contractor to retain a significant marine builder's risk insurance policy, thus limiting the options for fire risk transfer between parties. This also limits the solution set for determining how to manage fire risk to what is in the contract, necessitating that the Government be specific in identifying fire protection and prevention measures above and beyond what is required by OSHA Standard 1915 Subpart P, *Fire Protection in Shipyard Employment.* The SUPSHIP Operations Manual states,

> Under vessel fixed-price contracts, the Government customarily assumes the same risk of loss or damage as would have been assumed by private insurance underwriters had the contractor obtained and maintained marine builders risk insurance. This risk is subject to a deductible as identified in the contract. (Naval Sea Systems Command, 2021, p. 3-101)

Transferring more of the liability for determining how to provide adequate fire protection and prevention for commissioned naval vessels to a contractor would have the unintended consequence of increasing the amount of insurance the contractor would need to carry in order to assume this liability.

Defense-in-Depth

Ships are most vulnerable to fire during construction and overhaul periods because the normal layers of protection are not present. Permanent fire protection systems may be out of service, fire resistant boundaries may be compromised, less of the ship's force is present, and many industrial workers are present. As opposed to strict controls on material aboard, temporary services are run throughout the ship, and a significant amount of combustible materials are brought on board. When a ship is at sea, sources of fire ignition are planned for in the design of protective measures. During industrial work, sources of fire ignition include industrial work evolutions, permanent or temporary services installed in the ship, and human causes such as discarded smoking materials or criminal acts. There is a high reliance on human intervention rather than a layered approach to protection, which can be a critical single point of failure in preventing the escalation of an incident.

Cynefin Domains

Within the "Policy and Organizational Commitment" pillar of the Navy SMS framework, it is important to recognize that different management approaches are required for different aspects of the problem. Approaching decision-making modalities from the correct domain is as important as making decisions at the right level. This could mean best practices (clear domain), good practices (complicated domain), evolutionary or novel approaches, or decomposing the problem in order to shift it into a domain where it is more easily managed. Table 3 provides examples of shipboard industrial safety problems that currently exist within each Cynefin domain. Note that from a programmatic standpoint, this topic currently exists in the confused domain at the intersection of other domains primarily due to identified gaps that will be discussed in more detail in a later section of this paper. The goal of implementing an SMS framework based on the CSG model and using the Cynefin framework to analyze it is to decompose the problem in a manner that makes it manageable.

Table 3. Industrial Fire Safety Examples in Each Cynefin Domain

Complicated	**Clear**
The work done by the welding engineer to determine the welding requirements and produce the technical work documents for a particular work evolution.	Skilled work (tasks) performed by a welder.
Complex	**Chaotic**
Interaction between welding (hot work) and the surrounding environment. Factors such as type of welding, proximity to combustibles, fire resistant and non fire-resistant separations, and adequacy of the fire watch all contribute to the safe execution of this work evolution.	Introducing a transient and unequally trained workforce with a few workers that may randomly decide to follow no rules into ship repair work evolutions. While emergence between sub-systems in the shipboard fire safety problem should be discernable, this is only possible when all agents are playing by an identifiable and uniform set of rules.
Confused/Aporetic	
From a programmatic standpoint, this is the current domain of shipboard industrial fire safety.	

System Architecture

The system of interest (SOI) in this paper, depicted in Figure 6, is a maintenance availability on a commissioned naval vessel (otherwise known as a project). Boundary conditions (other systems) that interact with this system include the acquisition system that contracts the work, the ship maintenance activity that undertakes the work, and the ship itself. The context includes the shipboard industrial fire safety system and associated requirements that are transmitted into the SOI by contract items and include reference to Manuals and Instructions.

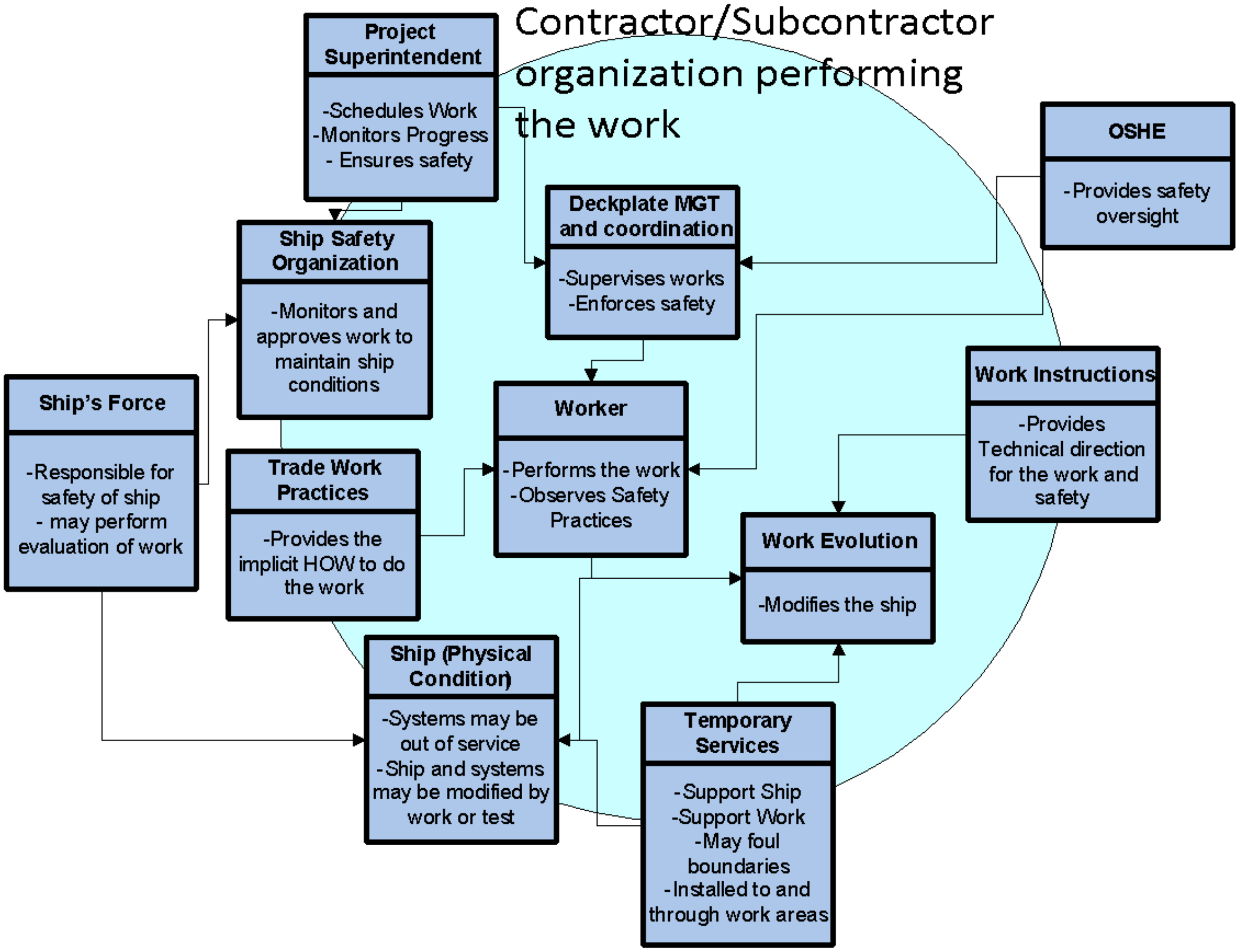


Figure 6. Architecture of System of Interest

SMS Framework and Complex System Governance Model

The governance model meta-function descriptions in Keating and Bradley (2015) provide the reference model assignments of requisite responsibilities and products. In Figure 7, we identify a preliminary assignment of the responsibilities with the corresponding meta functions. We also identify explicitly the governed systems (referred to as S1s) to place them in the appropriate location in the schema.

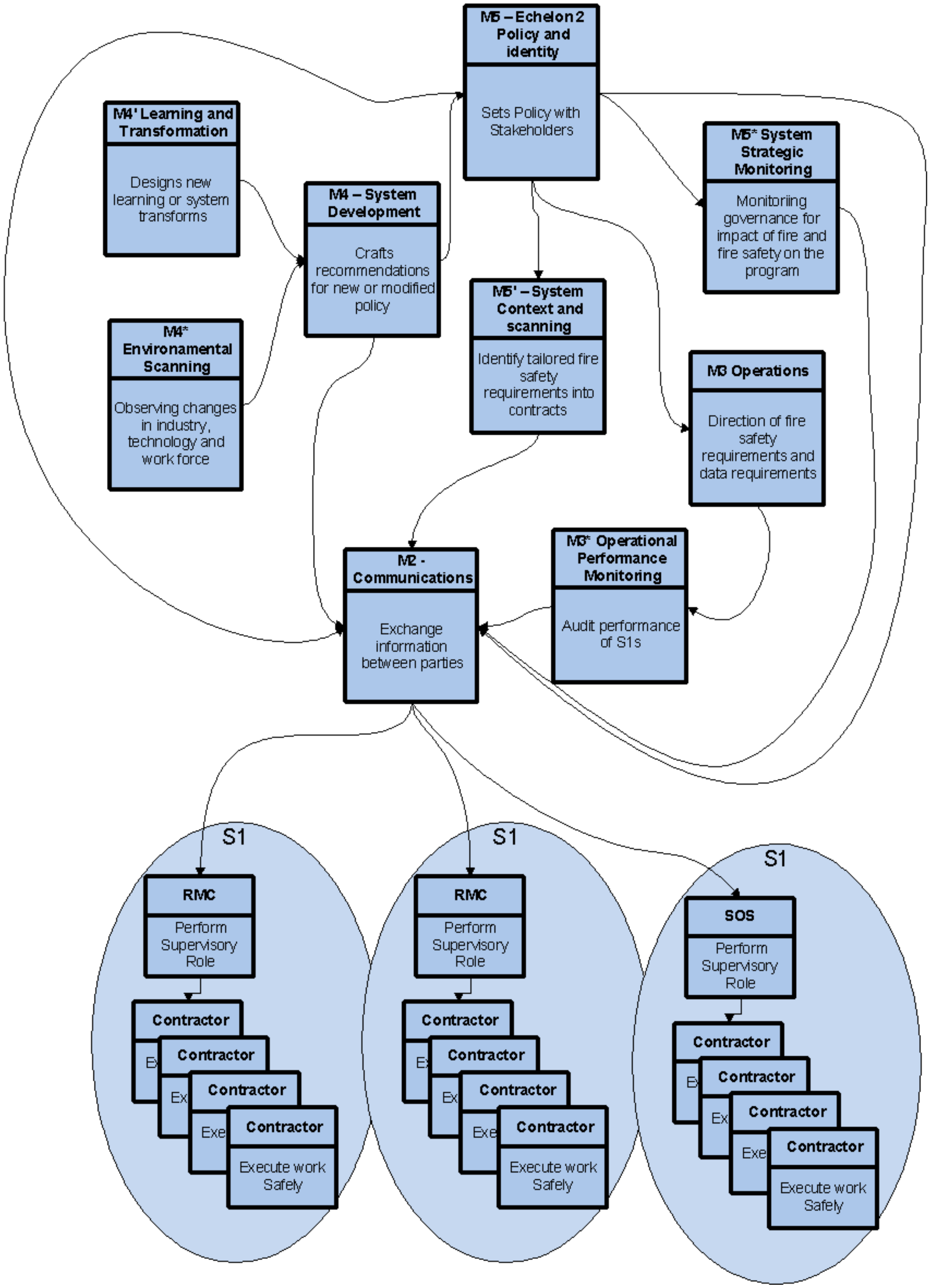


Figure 7. Metasystem Functions Identified for Shipboard Industrial Fire Safety Perspective

Setting policy at the NAVSEA Echelon 2 level that incorporates all the functions of the CSG model, which is inherently a risk communication and iterative continuous improvement cycle, can result in an SMS that is compliant with the new requirements for the Navy SMS framework. Developing procedures that adapt decision-making modalities and risk management practices based on system decomposition using the Cynefin framework support utilizing limited resources in a manner that is more likely to bring about meaningful change in shipboard industrial fire safety than seeking "one size fits all" solutions. Given that significant fires in the past decade on the USS *Gunston Hall* (LSD 44) and USS *Oscar Austin* (DDG 79) had similar root causes to historical fires on the SS *Normandie* and USS *Saturn* (AK 49), (McGowan & Smith, 2020), the current state mirrors NASA's "'failures of foresight' in which history played a prominent role" (CAIB, 2003, p. 195).

Identified Gaps

- ***Technical Authority***

There is not a single technical authority or TWH with responsibility for shipboard industrial fire safety. This role is split between NAVSEA 04RS responsibilities as the TWH for safety policy, NAVSEA 04X6 responsibility for maintaining the NAVSEA *Industrial Ship Safety Manual for Fire Prevention and Response (8010 Manual)*, and NAVSEA 05P responsibility for damage and survivability, which includes the ship's fixed fire protection systems. Other TWHs have input within their respective areas of responsibility, but decision-making and responsibilities do not roll up to a single entity. The lack of a TWH with overall responsibility can lead to issues such as contract negotiations over industrial fire protection and prevention requirements without the involvement of all relevant stakeholders.

- ***Technical Cognizance***

In conjunction with the lack of an overarching TWH, there is not a clear flow down of technical requirements or a single entity responsible for adjudicating technical issues that arise in the domain of shipboard industrial fire safety. Nor do the activities performing the work have specific requirements to house this technical expertise internally. Where requirements are not clearly defined and rooted with a technical basis, decisions are made at the deckplate by individuals who may or may not have the expertise to make them.

- ***Weaknesses in Defense-in-Depth***

Our initial work has indicated that the unwanted occurrence of significant fire events may be an indicator of an ineffective system to provide defense-in-depth. This potential gap will need to be explored further in later phases of this work.

- ***Lags in Incorporating Lessons Learned Into Contracts***

Technical requirements related to shipboard industrial fire safety are typically found in NAVSEA Standard Items (standard specifications for ship repair and alteration), either directly or by reference to other documents, which may or may not be incorporated into every contract. Lessons learned from the USS MIAMI fire are still not fully incorporated into NSIs 9 years later, and even when new requirements are invoked, the multi-year nature of ship maintenance availabilities means that contractual requirements typically lag current recommendations.

- ***Contract Requirements Are Not Driven by Data***

Trends are collected and analyzed by data; collection is not standardized and does not directly influence what is required in future contracts or contract modifications. Use of data is critical to the risk communication, risk monitoring, and change management fundamental elements of the Navy SMS framework. Within the CSG model, this reflects feedback loops between metasystems.

Conclusions

There is not currently a uniform level of industrial fire safety during ship maintenance availabilities, primarily because there is not currently a cohesive governance model or framework that is driving specificity of requirements to manage risk. Rather than the current state of rote compliance (or noncompliance) to general requirements, contractual requirements should be data driven and vary based on risk, and there should be clear technical authority over setting these requirements and technical cognizance in ensuring they are met. Feedback loops between requirements and outcomes in conjunction with faster routes (such as contract modifications) to incorporate lessons learned into ship maintenance contracts support a higher level of safety through better management of the risks involved. Note that the underlying goal is not necessarily to avoid all fire due to the nature of the work but have defense-in-depth and "right-sized" work controls to prevent major fires and reduce the impact of minor ones while not unduly impacting production schedule or project cost.

Planned Future Work

This paper is the first step in a concerted effort towards implementing data driven decision-making for industrial fire safety during ship maintenance availabilities by defining the governance model and SMS framework. NAVSEA 04RS has already done a significant amount of work analyzing human factors in historical fire incident data, and the intent is to continue to build upon their efforts. The next step is to further analyze available historical data to identify causal factors in why small fires become large, forming the basis of determining what standardized data needs to be collected to analyze future trends and inform contract requirements. We will also evaluate where and how more robust defense-in-depth principles can be incorporated. Then, we intend to create a standard data architecture and viewpoints for data-driven decision-making that could be implemented through creation of a new data repository held by a neutral third party.

The long-term vision is for decision support systems that are a model-based engineering cross between tools like the National Fire Incident Reporting System (NFIRS) database used for land-based fire reporting, the Aviation Safety Reporting System (ASRS) maintained by NASA for the Federal Aviation Administration (FAA), and more traditional project and program management dashboards. Having a tool such as this in the toolbox would allow the Navy, contractors, and the broader maritime repair industry to learn and evolve based on data from shipboard industrial fire incidents and near misses.

References

Aven, T. (2009). Safety is the antonym of risk for some perspectives of risk. *Safety Science, 47*(7), 925–930. https://doi.org/10.1016/j.ssci.2008.10.001

Aven, T., & Ylönen, M. (2018). A risk interpretation of sociotechnical safety perspectives. *Reliability Engineering & System Safety*, *175,* 13–18. https://doi.org/10.1016/j.ress.2018.03.004

Columbia Accident Investigation Board. (2003). *Report of Columbia Accident Investigation Board, volume I.* National Aeronautics and Space Administration. https://www.nasa.gov/columbia/home/CAIB_Vol1.html

Eckstein, M. (2020, July 22). *Stop-work order lifted at NASSCO-Norfolk after fire safety walk-through, training*. USNI News. https://news.usni.org/2020/07/22/stop-work-order-lifted-at-nassco-norfolk-after-fire-safety-walk-through-training

Fleming, K. N., & Silady, F. A. (2002). A risk informed defense-in-depth framework for existing and advanced reactors. *Reliability Engineering & System Safety*, *78*(3), 205–225. https://doi.org/10.1016/s0951-8320(02)00153-9

Geurts, J. (2020). *Memorandum for the shipbuilding and ship maintenance enterprise. Subject: Fire safety.* https://assets.documentcloud.org/documents/7007810/Fire-Safety-20200724-1.pdf

Gortney, W. (2012). *USS* Miami *(SSN 755) fire panel recommendations* (USFF ltr Ser N00/200). https://archive.org/details/ussmiamissn755firepanelrecommendations/mode/2up

Keating, C. B., Katina, P. F., & Bradley, J. M. (2014). Complex system governance: Concept, challenges, and emerging research. *International Journal of System of Systems Engineering*, *5*(3), 263–288.

Keating, C. B., & Bradley, J. M. (2015). Complex system governance reference model. *International Journal of System of Systems Engineering*, *6*(1/2), 33–52.

Keating, C. B., & Katina, P. F. (2019). Complex system governance: Concept, utility, and challenges. *Systems Research and Behavioral Science*, *36*(5), 687–705.

Keating, C. B., Katina, P. F., Jaradat, R., Bradley, J. M., & Hodge, R. (2019). Framework for improving complex systems performance. In *29th Annual INCOSE International Symposium Conference Proceedings*.

Kurtz, C. F., & Snowden, D. J. (2003). The new dynamics of strategy: Sense-making in a complex and complicated world. *IBM Systems Journal*, *42*(3), 462–483.

McDermott, J. (2013, November 13). USS *Miami* crew not exactly idle as ill-fated sub gets new skipper. *The Day*. https://www.theday.com/article/20131115/NWS09/131119792

McGowan, J., & Smith, S. (2020, September 29–October 2). *Industrial fire safety, fire prevention and control January 2017 - June 2019 self-assessment and recommendations* [PowerPoint presentation]. SNAME Maritime Convention, Virtual.

Möller, N., Hansson, S. O., & Peterson, M. (2006). Safety is more than the antonym of risk. *Journal of Applied Philosophy, 23*(4)*,* 419–432.

Naval Sea Systems Command. (2021). *Supervisor of shipbuilding, conversion, and repair (SUPSHIP) operations manual (SOM)* (S0300-B2-MAN-010; Revision 2, change 32). https://www.navsea.navy.mil/Portals/103/Documents/SUPSHIP/SOM/SOM2008-14Jan2021.pdf

Ropohl, G. (1999). Philosophy of socio-technical systems. *Society for Philosophy and Technology Quarterly Electronic Journal*, *4*(3), 186–194.

Saleh, J. H., Haga, R. A., Favarò, F. M., & Bakolas, E. (2014). Texas City refinery accident: Case study in breakdown of defense-in-depth and violation of the safety–diagnosability principle in design. *Engineering Failure Analysis*, *36,* 121–133. https://doi.org/10.1016/j.engfailanal.2013.09.014

Society for Risk Analysis. (2018). *Society for Risk Analysis glossary.* https://www.sra.org/wp-content/uploads/2020/04/SRA-Glossary-FINAL.pdf

Sorensen, J., Apostolakis, G., Kress, T., & Powers, D. (1999). On the role of defense in depth in risk-informed regulation. *Proceedings of PSA*, *99,* 22–26.

U.S. Department of the Navy. (1990). Chapter 8: The commanding officer. In *U.S. Navy regulations.* https://www.secnav.navy.mil/doni/US%20Navy%20Regulations/Chapter%208%20-%20The%20Commanding%20Officer.pdf

U.S. Department of the Navy. (2020). *Navy safety and occupational health manual* (OPNAV M-5100.23)*.* https://www.secnav.navy.mil/doni/SECNAV%20Manuals1/5100.23.pdf

Ziezulewicz, G. (2021, January 3). New in 2021: *Bonhomme Richard* will be turned into scrap. *Navy Times.* https://www.navytimes.com/news/your-navy/2021/01/03/new-in-2021-bonhomme-richard-will-be-turned-into-scrap/

Functional Hazard Analysis and Subsystem Hazard Analysis of Artificial Intelligence/Machine Learning Functions Within a Sandbox Program

Bruce Nagy—is a Research Engineer at the Naval Air Warfare Center, Weapons Division, at China Lake. His research focuses on advanced game theory techniques, artificial intelligence, and machine learning applications for tactical decision aids. Nagy has earned four degrees: one in mathematics, two in electrical engineering, and one in biology from The Citadel and the Naval Postgraduate School. He led the development of advanced algorithms and metrics that resolved national defense issues in satellite communications for the DoD. At UCLA during postgraduate work, he investigated modeling brain stem communication with muscle groups at the cellular level, in cooperation with the NIH. [bruce.nagy@navy.mil]

Gunendran (Guna) Sivapragasam—earned his BS in electrical engineering from the University of Technology, Malaysia, and then received his MBA from the College of William & Mary in Virginia. Sivapragasam joined Naval Surface Warfare Center Dahlgren Division (NSWCDD) in August 2009 and is currently the Technology System Safety Lead for the System Safety Division (R40). In this role, he has been participating in and leading several efforts focused on developing system safety capability to do hazard analysis of new technologies being introduced in U.S. Navy weapon programs, including systems that employ AI. [gunendran.sivapragas@navy.mil]

Loren Edwards—is an Aerospace Engineer working for System Safety at Naval Air Warfare Center, Weapons Division, at China Lake. He graduated from Cal Poly San Luis Obispo in 2017 with a BS in aerospace engineering. He performed Computational Fluid Dynamics analysis work before moving into System Safety. He has attended classes on and performed Functional Hazard Analyses and supports multiple programs in different stages of development. He is working towards solving key safety questions involved with the use of Machine Learning and Artificial Intelligence to better understand and guarantee the safety of these evolving technologies. [loren.edwards@navy.mil]

Abstract

Development of advanced Artificial Intelligence (AI)/Machine Learning (ML) system-enabled weapons and combat systems for deployment in the U.S. Navy has become a reality. This is also true for the other armed forces, as well as in homeland security and even the Coast Guard. From the Navy standpoint, the Naval Ordnance Safety and Security Activity (NOSSA) is attempting to get ahead of the acquisition cycle by focusing on the development of policies, guidelines, tools, and techniques to assess mishap risk in Safety Significant Functions (SSF) that are identified. NOSSA's efforts have the potential of influencing the acquisition community, including in requirements, development, and test and evaluation engineering. This paper makes recommendations for the Functional Hazard Analysis (FHA) and Subsystem Hazard Analysis (SSHA) analysis templates and focuses on ways to decrease autonomy within system operations and increase its correlated Software Control Category (SCC). The questions and discussions devised from this research aim to form guidance and offer best practices to address AI/ML system safety issues.

Introduction

The Department of Defense (DoD) is rapidly approaching the point where system safety practitioners will need to conduct mishap risk assessments on AI functions within upgraded systems being deployed in the Fleet. These systems will be crucial to ensure the DoD retains its dominance in military power (Brose, 2020). The safety community will soon be required to conduct system safety analysis on systems, including weapon systems that contain Artificial Intelligence (AI)/Machine Learning (ML) functionality (National Defense Authorization Act [NDAA], 2021; National Security Commission on Artificial Intelligence [NSCAI], 2021). AI functions present unique challenges to system safety practitioners to identify hazards, assess risk, and identify risk mitigation measures. This includes how to properly employ a system with AI capability in an operational or tactical environment while reducing the probability of a mishap. Currently, no guidance exists on how to conduct system safety analysis on AI/ML functions, and this will prevent the certification of these systems for deployment (Naval Sea Systems Command [NAVSEA], 2008; National Institute of Standards and Technology [NIST], 2019).

The Problem

Assuring safety in AI/ML systems is a considerable challenge to current safety processes for traditional software. Traditional software can be assessed for safety through code review, and traditional software outcomes can be analyzed through automated code analysis techniques such as Modified Condition/Decision Coverage (Joint Software Systems Safety Engineering Workgroup, 2017). Together, these and other methods can provide a rigorous understanding of how software will function in a given situation, assuring some desired level of safety. However, a developed and trained AI/ML system cannot be analyzed with current analysis methods, and though it is theoretically possible for some ML designs (and completely impossible for others) to exhaustively test all inputs and outputs of an AI/ML software function, the calculation time required makes even small systems almost impossible to analyze. These issues, combined with the unique challenge of unpredictable real-world corner cases, result in AI/ML functions having an inherent lack of safety due to unknown, unanalyzable, and untestable factors (Sodhani, 2018).

Within the DoD, MIL-STD-882E guides the software safety process. This standard provides a method for categorizing safety significant software based on its level of autonomy, called the Software Control Category (SCC). SCC 1, Autonomous, defines the highest level of autonomy, while SCC 4, Influential, defines the least autonomous category of safety significant software. These SCCs are combined with the severity of related hazards to define a Software Criticality Index (SwCI). Each SwCI level requires a requisite Level of Rigor (LOR), or a specific set of tasks to be completed before that safety significant software is considered “safe,” or representing a certain level of acceptable risk for the system. SwCI 1 requires the most effort to achieve LOR, while SwCI 4 requires the least amount of effort to achieve LOR.

For software where functional failure could lead to catastrophic hazards and that either has control over safety significant hardware or provides safety-critical information, the safest SCC possible is SCC 3, Redundant Fault Tolerant, which results in SwCI 2 (MIL-STD-882E; Defense Standardization Program Office, 2012). If this function were instead SCC 1, Autonomous, or SCC 2, Semi-Autonomous, the resulting SwCI would be 1. In addition to the SwCI 2 LOR tasks, SwCI 1 LOR tasking additionally requires code level analysis, such as including MC/DC or equivalent testing (JS-SSA, 2017). This means that if the Safety Significant Function (SSF) that could lead to a catastrophic hazard is an AI/ML function, it would likely be impossible to perform full LOR tasking on that function, creating a considerable gap in software safety.

The Need

Unless new analysis techniques are developed that can address the specific issues described previously, the most effective way to increase confidence of safe operations in AI/ML systems is to decrease the safety significance of AI software. Per MIL-STD-882E, this can be accomplished by lowering the potential mishap Severity or the SCC of the function. The SCC is used to define the level of control that software has over SSFs. The higher the number (from 1 to 5), the less safety impact the software has. For catastrophic hazards where the software either has control over safety significant hardware or provides safety-critical information, SCC 3 is the safest possible category and should be the goal for all traditional and AI software functions. There is increased difficulty in reaching this SCC for AI/ML, however, due to the fact that in many applications of AI/ML, the AI system is independent (autonomous) and may be the only system that reviews data and makes decisions based on that data (Sodhani, 2018).

In addition to this, the many uses of AI/ML throughout government and industry do not follow defined procedures for guaranteeing safe operations. The current processes used to determine how safe AI/ML software is, and the processes used to decrease the risk of hazards due to or involving AI/ML, vary widely and are not consistent between companies and government agencies (NSCAI, 2019). These varied approaches not only result in inconsistency and lack of safety rigor in deployed systems, but they also decrease trust in AI/ML technology. To address both of these issues, consistent approaches to AI/ML system safety analysis must be developed.

In many modern implementations of AI/ML, there are neither components nor systems in place that actively decrease the autonomy level of these specially developed software functions.

Figure 1 describes the process for performing a Functional Hazard Analysis (FHA), which is the primary analysis used to determine SCC and SwCI determinations for safety significant software. In reviewing Figure 1, several questions are posed with regard to AI/ML:

- Are unique tools needed because of the presence of AI/ML to complete this analysis?
- How would we complete this determination for an AI/ML deployed system?
- Are current SwCI definitions appropriate for AI/ML?

These questions, alongside proposed answers and solutions to them, are presented in this paper.

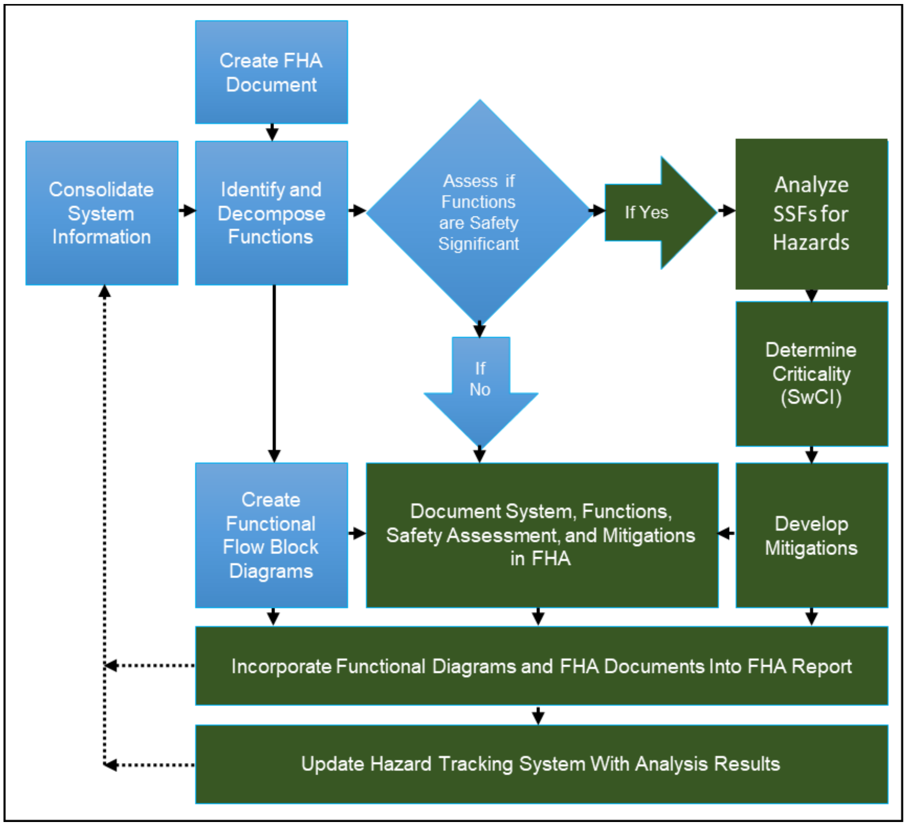


Figure 1. FHA Workflow

The Goal

This paper focuses on how to identify an AI safety critical function, gives recommendations to reduce the function's autonomy level, provides a format on how LOR for an AI/ML function can be identified, and includes some initial examples of AI/ML unique tasks for LOR. The goal of the research is to provide processes, questions, discussion points, and insights regarding the organization of the safety analysis in the form of tables and specifically labeled column headings. These structures and format recommendations are in support of system safety practitioners, providing guidance on how to conduct rigorous safety analysis on AI/ML software functions being deployed in weapon (MIL-STD-882E; Defense Standardization Program Office, 2012) or aircraft (DO-178C; Radio Technical Commission for Aeronautics, 2012) systems. The paper provides, in table format and related recommendations, examples of two system safety analyses, the FHA and Subsystem Hazard Analysis (SSHA).

A complete list of SSHA LOR task descriptions that arose from our analysis will be available through other Naval Ordnance Safety and Security Activity (NOSSA) documented sources. This research, funded by NOSSA, will provide recommendations to be considered by NOSSA. These recommendations are to help better understand the robustness of the model being developed, especially if that model resides within an SSF. This paper makes recommendations for FHA and SSHA templates and related considerations to facilitate system safety analyses on AI/ML functions with a focus on ways to decrease the autonomy within system operations and increase their correlated SCCs. The questions and discussions devised from this research aim to form guidance and offer best practices to address AI/ML system safety issues.

Use Case to Investigate

When considering an operational use case to implement within our sandbox development environment, our first step was to create a stakeholder's analysis table, as shown in Table 1.

Table 1. Stakeholder's Analysis Table

#	Name/Org	Type	Want/Need	Concern/Loss	Notes
1	Safety Engineer/NAWCWD D511000	Analyst	Suite of defined LOR tasks and OQE	Guilt/Liability from loss of life	Knows that AI system is Safety Significant but no LOR tool set available
2	Safety Engineer/Contractor (Weapon System Supplier)	Analyst	Suite of defined LOR tasks and OQE	Guilt/Liability from loss of life	Knows that AI system is Safety Significant but no LOR tool set available
3	Warfighter	User	Assurance of weapon system safety	Guilt/Liability from loss of life	Assumes that AI system is safe; unaware of lack of safety rigor
4	WSESRB Member	Analyst	Suite of defined LOR tasks and OQE	Guilt/Liability from loss of life	Knows that AI system is Safety Significant but no LOR tool set available
5	Program Manager	Sponsor	Assurance of weapon system safety	Guilt/Liability from loss of life	Pressured to meet military requirement; accepts safety risk
6	Civilian or Military Victim of Mishap	Neutral Observer	Safety in Battle Space as Non-Target	Personal Death or Injury	Unaware of Latent Safety Hazard
7	American Public	Neutral Observer	Assurance that weapon systems will not kill or injur friendlies or non-combatants	Anger/Disapproval	"How could this tragedy happen?" "Who is responsible?" "Why was a dangerous weapon system deployed by the US
8	NOSSA, PM	Sponsor, Developers	What processes and policy associated with the various phases of the acquisition cycle will be needed to support system safety for AI/ML software?	NOSSA: Unsafe deployed system, PM: Added cost to retrofit safer	
9	NOSSA	Sponsor	What tools, guidance and documentation would need to be created to support the processes and policy per each group's needs? Groups: Developers need from system safety, System safety practitioners from system safety and Oversight folks from system safety.	NOSSA: Unsafe deployed system	
10	NOSSA	Sponsor	Along with the processes, what analytics need investigation for each user group?	NOSSA: Unsafe deployed system	
11	NOSSA	Sponsor	How would various AI/ML software designs affect the analytical approach?	NOSSA: Unsafe deployed system	
12	NOSSA	Sponsor	What kind of OQE is required per a given AI/ML technique and implementation structure to support a program moving forward?	NOSSA: Unsafe deployed system	
13	NOSSA	Sponsor	Will data and analytics be considered as separate pieces to inspect?	NOSSA: Unsafe deployed system	
14	NOSSA	Sponsor	During a WSESRB or Technical Review Panel review that involves AI/ML, how would systems, data and numbers be presented to allow for proper investigation and analysis to ensure contextual accuracy based on group technical background?	NOSSA: Unsafe deployed system	
15	NOSSA	Sponsor	What are the factors and limitations associated with confidence of numbers presented regarding AI/ML performance?	NOSSA: Unsafe deployed system	
16	NOSSA	Sponsor	AI/ML performance is always associated within the context of the training data?	NOSSA: Unsafe deployed system	
17	NOSSA	Sponsor	What does it mean to perform architecture, design, or code analysis (see MIL-STD-882E Table V) with an AI/ML system, especially when, for example, even the developer has limited understanding on how the neural network works?	NOSSA: Unsafe deployed system	
18	NOSSA	Sponsor	How will confidence be assured for each user group in terms of how the software will perform as specified to AI performance requirements (see MIL-STD-882E paragraph 4.4.1.b)?	NOSSA: Unsafe deployed system	
19	NOSSA	Sponsor	What would be the type of contractual language associated with AI/ML integration/deployment?	NOSSA: Unsafe deployed system	
20	NOSSA	Sponsor	Should it include the complete system because of potential reduction in overall system maturity?	NOSSA: Unsafe deployed system	
21	NOSSA	Sponsor	Will AI/ML algorithms exponentially increase the complexity of the system under review affecting hardware issues involved with processing, bandwidth and storage? If not considered, will performance degrade, causing system safety concerns? How will this be analyzed? What are the limitations associated with confidence of numbers presented regarding AI/ML performance? Note: AI/ML performance is always associated within the context of the training data.	NOSSA: Unsafe deployed system	
22	NOSSA	Sponsor	What format will allow technical and non- AI/ML technical stakeholders to support discussion, understanding and eventual application for their particular AI/ML situation? This sets the requirement for how processes and policy should be technically written and displayed while still supporting the necessary detail. It is anticipated that each group will have a different set of requirements for communicating and displaying technical detail related to guidance. What will be the training requirements for each group?different set of requirements for communicating and displaying technical detail related to guidance. What will be the training requirements for each group?	NOSSA: Unsafe deployed system	
23	NOSSA	Sponsor	1. how do we build confidence in the AI black box?	NOSSA: Unsafe deployed system	
24	NOSSA	Sponsor	2. How do we build rigor into, or is it necessary to build rigor into, the training code for the AI?	NOSSA: Unsafe deployed system	
25	NOSSA	Sponsor	Is this the appropriate AI technique to use and is there an non-AI technique that could be used?	NOSSA: Unsafe deployed system	

From a review of Table 1, it became obvious that current system safety analysis methodologies (MIL-STD-882E; Defense Standardization Program Office, 2012) were inadequate to address the unique system safety needs of AI/ML functions. New methodologies would be required to ensure comprehensive system safety analysis of AI/ML functions. In order to conduct research to develop new methodologies, the effort focused on developing a fictional system that implemented various AI/ML functions, allowing the system safety team to investigate gaps in current methodologies when analyzing these specially developed functions. The fictional system would have to replicate parts of the acquisition cycle in detail. To support realism in our fabricated system, various ongoing development efforts within various project and programs in early to late research phases were modified and then combined into an ML hybrid mission planner and a multi-ML algorithm robot technology. Again, it should be emphasized that although the program is fictitious, technology architecture, design, code, and test were based on existing research in the field of AI/ML currently being performed by various naval commands.

An Operational Use Case was identified that had the potential to provide answers described in Table 1's stakeholder analysis. Not only did it need to support answers to the questions posed in Table 1, but the operational used case also needed to be constructed from realistic aspects of AI/ML technology. An operational view of the Use Case is shown in Figure 2.

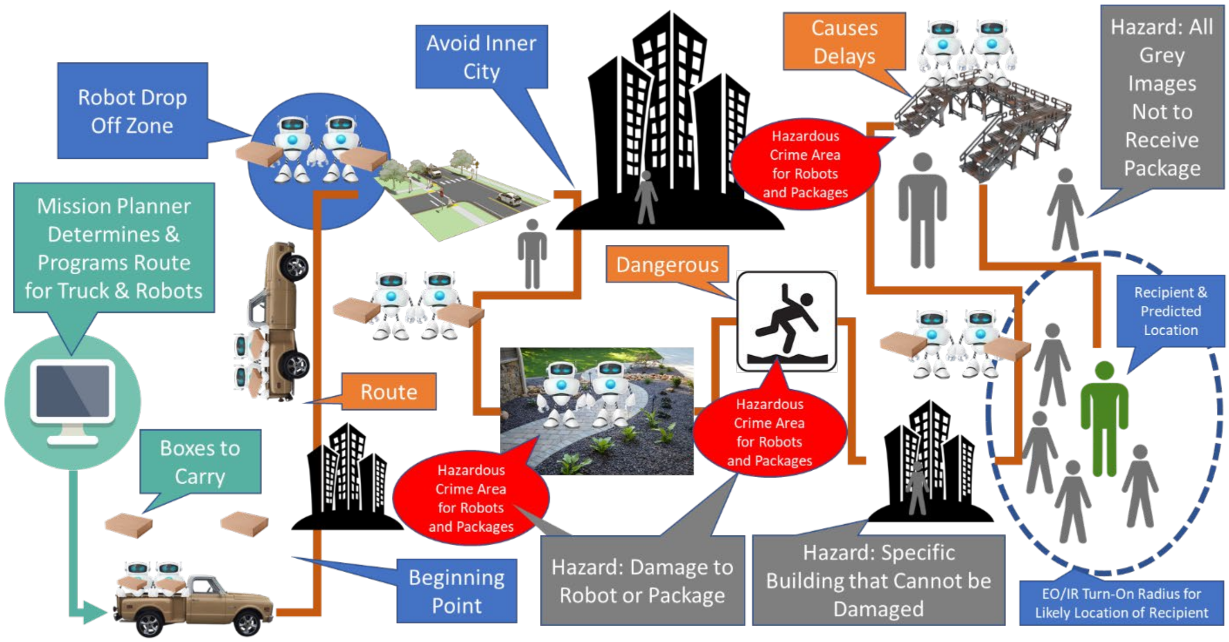


Figure 2. Operational Use Case of Two Robots Delivering Packages

Figure 2 is based on the following considerations involving the Operational Scenario, the Operational/Deployed Environment, and Key SSFs:

- ***The Operational Scenario***
- The design consists of the following subsystems: (1) Two Robots, which are identical in performance, (2) Two Pickup Trucks, which are identical in performance, and (3) a Mission Planner.
- The two robots are carried a partial way to their destination on the two pickup trucks. After the pickup trucks arrive at their destination, each robot will be unloaded from their respective pickup trucks. From the unload point, the robots will walk synchronously to arrive and deliver their packages to the single intended recipient at the same time. The two robots are able to walk long distances using GPS navigation, and as the robots gets closer to the subject receiving the package, a Convolutional Neural Network (CNN) image recognition algorithm using color spectrum and infrared images from an EO/IR sensor on the robot will take over navigation. The special GPS navigation is pre-loaded with waypoints produced by an AI/ML trained mission-planning tool.
- ***The Operational/Deployed Environment***
- It could be a rainy day when the robots are deployed. Weather conditions, houses, and buildings all result in background clutter and obstacles for the navigation system. There are cars and pickup trucks on the road, including other robots and people walking, complex highway systems, and city-like sidewalks and walkways that need to be

navigated by the two robots. Many other people, some looking very similar in side profile to the recipient, in this scenario are part of the environment. It is important that people should not receive this package by mistake, as the packages are hazardous and very valuable. Delivering a package to the wrong recipient will be a Catastrophic mishap. Thus, it is important to system safety that sufficient mitigations are incorporated into the system to minimize the risk of an incorrect person receiving the package by mistake from either robot.

- ***Key Safety Significant Functions***

- Navigation. The navigation system of the robots uses a special GPS function following waypoints and then switches to an AI-based seeker function at a certain range from the recipient. The AI-based seeker function uses polar coordinates instead of waypoints to navigate. The AI/ML navigation system must avoid obstacles, as mentioned previously, during navigation. Some obstacles might include other people attempting to steal the package carried by a robot. The AI-based seeker does not take over navigation until a separate switching function determines when the robot is at a certain distance from the recipient. Once this switch is activated, robot navigation is turned over to the CNN function for final navigation to the recipient. The non-AI navigation is responsible for avoiding obstacles in route until the CNN takes over navigation. Again, once within a certain range of the recipient, the seeker function is switched on to take over the complete navigation of the system. The seeker function consists of a CNN, designed to recognize side profiles of the recipient, and avoid obstacles. The CNN is trained using synthetic images of side profiles within a synthetic clutter environment. The CNN is trained to navigate in such a way as to avoid people attempting to steal its package. While in a traditional system, package theft would be considered a security issue and not a safety issue, the team decided to identify this as a system safety concern to allow investigation of the CNN function.
- Sensor Data. Each robot is receiving non-curated data from a 3-D sensor. The sensor streams a color scaled set of images that contain complex backgrounds at a certain sampling rate for database storage and CNN processing. Images are stored in a separate database for each robot with no data sharing between robots.
- Image Recognition**.** There is a large amount of synthetic data available for training and some actual images of recipients' side profiles. Unfortunately, the added clutter to the image is also synthetic (i.e., building and house backgrounds, day and night lighting, rain, etc.; see previous Operational/Deployed Environment section.) The developer is also considering a transfer learning approach to add another classification layer to the CNN to increase the probability of successful recognition.
- Timing Synchronization. Timing synchronization is implemented using reinforced learning, with real time updates to the Mechanics Reinforced Learning (RL) Dynamics Manager neural network that affects the physics of the robots in terms of direction and speed. Both robots must deliver their packages at the same time, but they can take different routes to avoid environmental conditions. Once a robot delivers a package, the recipient will not wait for the second package. Therefore, it is important that both packages be delivered at the same time to avoid the recipient leaving and the possibility of the second robot delivering its package to someone who should not receive the package. Again, this is a significant system safety concern because delivery of a package to an incorrect recipient is considered a catastrophic mishap.

Sandbox Development Environment Approach

Within the sandbox development environment, a variety of AI/ML algorithms supporting a mission planner and autonomous vehicle selection and navigation were developed—again inspired by existing AI/ML projects and programs. Formal Department of Defense Architecture Framework (DoDAF) (Department of Defense Chief Information Officer [DoD CIO], n.d.; Dam, 2006) and Unified Modeling Language (UML; Booch, 2017) diagrams were created to support a System of Systems (SoS) design, including interface messages, SQL commands and Application Programming Interfaces (API). Using this sandbox approach allowed an initial top-down safety analysis, starting with system decomposition and traceability from an Operational View (OV), through a Systems View (SV), and then finally down to the algorithm's code level supporting the specific AI/ML being implemented. This process was used to provide a broad scope representation of a potential DoD program implementing AI/ML and to "realistically" investigate conduct of the FHA and SSHA methodologies on a variety of AI/ML functions.

Figure 3 represents the subsystems associated with the mission planner and robot.

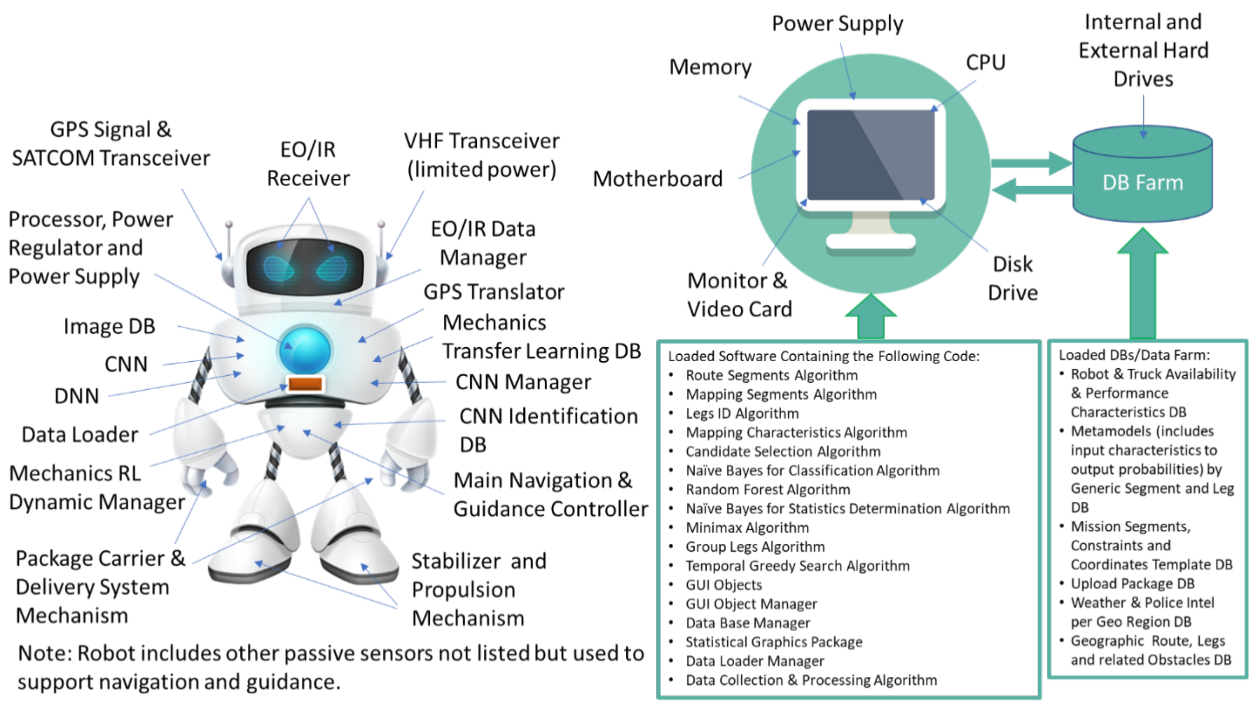


Figure 3. Robot and Mission Planner Subsystems

The goal of the sandbox development environment was to implement a variety of AI/ML technologies that worked together as an SoS, offering a variety of ML approaches to investigate (Hastie et al., 2017). The mission planner provides analysis of the following AI/ML technologies:

- The Naïve Bayes and Logistic Regression algorithms receive sensor and human input data in order to select the appropriate type of meta-model from a repository.
- The Random Forest algorithm, by creating a "Similarity Table" between branches of trees (and its counterpart "Distance Table" for other ML algorithms using distance analysis), allows for data estimation for missing data within the meta-model tables that was not originally identified in the Design of Experiment (DOE) simulation

requirements. Therefore, the Random Forest can account for the challenges when the Modeling and Simulation (M&S) is missing inputs needed for a meta-model. Predicting the deployed operational future is difficult for a variety of reasons. Random Forest clustering allows estimation in situations with both limited inputs and an unknown statistical output. This allowed for greater flexibility in the mission planner's ability to adapt in non-ideal operational environments.

- The minimax, per meta-model selected, looks at the factors needing to be addressed in completing the mission, including tactical capability, external challenges and delivery issues. Based on sensor data, it selects "worst" case scenario and then finds the "best" case autonomous performance combined with tactical sequence needed to successfully complete the mission. "Worst" case and "best" case are represented by statistical structures within the route leg's meta-model counterpart.
- The meta-model tables are processed through the Random Forest approach and a minimax algorithm to determine the statistics for each leg in the route. Then a non-linear optimization program determines the optimal selection of robots and routes.

The robots' autonomous platform provides analysis of the following AI/ML technologies:

- Deep Neural Network (DNN) - supports the mechanical motion of the robot given various states. The states are determined based on real time input conditions of the robot movement. Input conditions include traveling surface and conditions. Based on the attributes received, the DNN would use a control feedback mechanism to adjust its walking mechanics. For example, if the robot, through the CNN, recognizes that it is about to approach the package recipient, it would slow down, unsecure the package, and extend its arms to show the package. While traveling, the package would be secure. Upon understanding its current travel state, the robot would use DNN to determine which mechanical state to implement. It should be noted that the sandbox will also be investigating how Deep RL might apply.
- Convolutional Neural Network (CNN) - supports the recognition of the recipient. The input would be based on facial recognition. The result of this analysis would be input to the DNN in terms of its mechanical functions, as discussed previously.

It should be noted that the sandbox is still in development. Design analysis of all AI/ML functions described previously has been completed. Implementation is ongoing. This paper's findings are based on sandbox development environment investigations, as well as from previous game theory, DNN and CNN research with other projects.

AI Type Definition

There are many opinions surrounding the definition of AI. For this research, we defined the term "AI Type." We defined an AI Type to be identified by objective measurements. Therefore, if a function is determined to be an AI Type based on its score from the following objective measurements, it requires special FHA and SSHA investigation. The definition and scoring are as follows:

AI Type (Working Definition): For system safety concerns, an AI Type of function means that an algorithm will be developed:

(1) using data approximations to build its algorithm (e.g., from simulations and synthetic data vs. an equation that accurately represents real world physics) and/or

(2) when data samples used to build its algorithm are a subset of the actual population size (e.g., training data samples from population to support machine learning, training data samples requiring clutter backgrounds).

Scoring: For all functions that are candidates for being implemented using an AI/ML algorithm (examples in table), then each function must be graded using criteria (1) or (2), with corresponding points awarded. A final score of 1 or greater indicates that the function is an AI Type.

Table 2 represents examples of scoring based on various AI Types. It is not intended to be a complete list. The goal here is not to provide the system safety practitioner with a complete list, but to aid the practitioner based on a practical scoring approach.

Table 2. Example of Scoring Based on AI Type Definition

AI Type Examples of Specific Algorithms	Algorithm built based on using data approximations	Algorithm built based on using data samples from larger population	Final Score
CNN	x (if synthetic data used for CNN)	x (training data samples)	0 to 2
DNN + SL	x (if synthetic data used for DNN)	x (training data samples)	0 to 2
DNN + RL	x (if synthetic data used for DNN)	x (training data samples)	0 to 2
RNN (LSTM)	x (if synthetic data used for RNN)	x (training data samples)	0 to 2
RNN (Simple)	x (if synthetic data used for RNN)	x (training data samples)	0 to 2
Naïve Bayes	x (if modeling and sim data used to produce statistics for Naïve Bayes)	x (if RL used during opponent interaction to train algorithm)	0 to 2

The AI Type definition allows for a productive, focused discussion between the system safety practitioner and the function developer. Questions that might initiate the conversation would include:

- What parts of the system need special rigor consideration (as compared to traditional algorithm code development)?
- Does the SSF identified qualify as an AI/ML function?

It is especially important to investigate if the function is an AI Type when an algorithm is identified as a safety-critical function using an FHA approach. Again, the hypothesis is that it is an AI Type function if

- (Consideration 1) it uses data approximations to build/train its algorithm (e.g., data approximations can come from simulations and synthetic data vs. an equation that accurately represents real world physics), and/or
- (Consideration 2) data samples are used to build/design its algorithm and these data sample are a subset of the actual population size (e.g., training data samples from population to support machine learning, training data samples requiring clutter backgrounds).

One way to think about consideration (1) is to ask, "Could another developer create a different set of statistics under the same conditions?" If no, then maybe this algorithm is not an AI Type. If yes, then it meets the condition. As an example, if a statistical model of the function was developed, how accurate were the approximations used in creating the function. In other words, how close do these approximations fit the real world physics regarding operational deployment? If the function is based on simulation results, then the concern is the "garbage in, garbage out" issue—poor real world representative synthetic data will result in an inferior model. The goal is to have good quality and comprehensive training data that would result in a robust model.

One way to think about consideration (2) is to ask, “What is the actual population size of the training set?” If the training set is equal to the actual population size, then it is not an AI Type. Consider the most basic ML algorithm, a regression line. If all the points that will ever occur for this function are on the scatter plot used to approximate the curve, why use a regression line? If all the ML algorithm inputs and outputs are known, why use ML and not traditional code? Again, if traditional code can address the needs of the function, then that would be the goal.

Notice that both considerations are related. It is like looking at two sides of the same coin: how ML algorithms are developed and why they need to be avoided in critical functions.

Figure 4 describes the need to separate out AI Type designated functions from traditionally developed software coded functions.

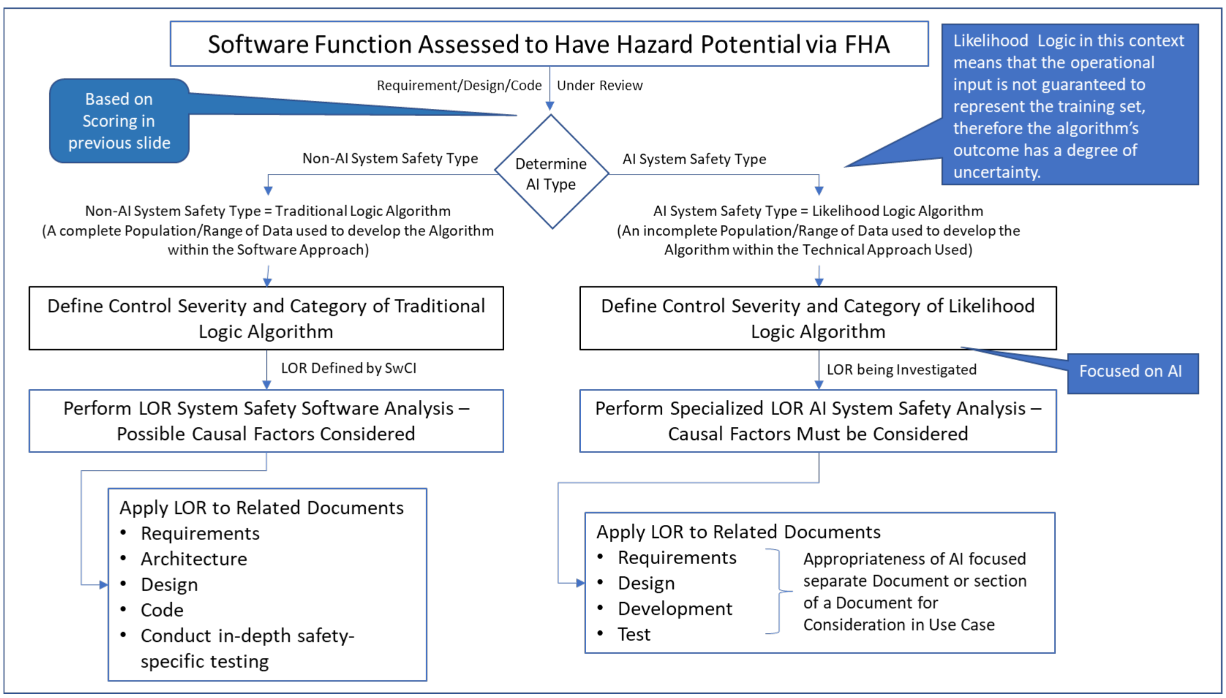


Figure 4. Flow to Assess AI Type Using Special FHA and SSHA Rigor

Six Recommendations to Assess AI Type Functions

This paper describes six recommendations. The first three are associated with the FHA assessment, and the last three involve the SSHA analysis. The FHA recommendations are provided in the form of a complete list of columns to use during the development of the FHA. It is noted that the current FHA approach works well in identifying the safety significance of a function. As shown in Table 3, three columns have been added in support of the three recommendations focused on reducing autonomy. The last three recommendations are in support of doing an SSHA. One recommendation is in structuring the table, with an added column regarding the focus of the analysis. The final two recommendations offer a list of questions listed as line items within the table focused on (1) API/MSG/SQL interface corruption to the ML algorithm and (2) modality of how well the data training the algorithm represents deployed conditions. The fifth recommendation addresses the interface to the algorithm and asks questions involved with how data corruption might need to be addressed within the training data composition of the ML algorithm. The sixth recommendation is in regards to the modality of how the training data is gathered, organized and managed, known as curation. If the training data isn't curated properly, the ML algorithm may not function properly. In either case, interface corruption or poor curation, a safety critical ML function could have a greater likelihood of becoming a hazard.

- **Functional Hazard Analysis**

Prior to addressing the first three recommendations, the initial steps of the functional hazard analysis (FHA) need to have been completed to identify the safety significance of the functions (JS-SSA Rev. A, 2017). It should be noted that the system safety analysis found no gaps in the FHA process in identifying the safety significance of a function when dealing with an AI/ML function. This is because the identification of the Severity, SCC, and resulting SwCI does not change when dealing with an AI/ML function. At this stage, it does not matter whether the function has AI/ML capability in it or not.

An example of an FHA regarding our Use Case example for mission planning is described in Table 3.

Table 3. FHA Example for Mission Planner From Sandbox

Each Column represents a recommendation

Recommendations 1, 2 and 3

Haz ID #	Phase	Activity	State/ Mode	Function	Functional Failure	Hazard Title	Hazard Description	AI Type Scoring	AI Type Justification	AI Type Autonomous Justification	System Item(s)	Causal Factor Description
	Test & Deployment	Metamodel Selection	Setup	Mission Planner - Route/Robot Selection (Naïve Bayes Metamodel Selection)	Function Unavailable	No hazard since no metamodel will be selected		2				
					Function Malfunctions (degraded, partial, or unexpected results of the function, signal too small/strong/intermittent)	Wrong Metamodel selected	If a wrong metamodel is selected, then the wrong robots could be selected resulting in packages not being delivered or being too early or late and so delivered to wrong receipient, or the package could be lost. (Assumption: Package not delivered means package will be lost - Catastrophic hazard)		Input to function has infinite number of combinations - no traditional approximation is sufficient	Semi-Autonomous -- Man in Loop for approval of route	Mission Planner	Training Data is incorrect or incomplete, corner case occurence
					Function out of sequence/In another combination (Same thread/separate thread)	No hazard - if metamodel is not selected at the right sequence then route selection will not progress.						
					Function at incorrect time (too early, too late, outside defined window, function never ends)	No hazard since function is only relevant at a certain sequence, not time based.						

Scoring Grade of 2

Continued

Mishap	Effects	Existing Mitigations	Software Control Category (SCC)	Criticality Index	Software Criticality Index (SwCI)	Recommended Mitigations	Component(s)	Follow-On Actions	Comments
Delivering the package to the wrong receipient could be catastrophic since the material is hazardous and/or very valuable.	Personnel Injury / Equipment Loss	1. Operator review of this function output is required before system operations can progress 2. Test Data is used to ensure ML is comprehensive	2- Semi-Autonomous	1	SwCI 1	1. Implement threshold on propotionality used for selection e.g. if propotionality of selected item is not significantly different from next choice (10x), then declare fault or select default metamodel. 2. Implement redundant, independent non-ML functions that review route/robot selection for compatibility	Business Rule Manager Data Manager Database Farm Graphics User Interface (GUI)	Implement recommended mitigations if possible	User review of this function's output is required before the system can progress operation. The user must approve of the route/robot selection, including compatability between robot and route. With a trained user, detailed procedures, and appropriate time windows for review and approval, the user increases the SCC from 1 (Autonomous) to 2 (Semi-Autonomous), though the lack of other redundant interlocks does not further increase the SCC. Implementing of recommended mitigations would increase the SCC.

Recommendation 1: Once an FHA identifies the SSFs, each function needs to be assessed using the AI Type definition following these three steps:

(1) For each identified SSF that potentially has AI/ML within, identify the algorithm that will be used to support that function, and then document the grade each function receives in terms of AI Type definition (see previous section on AI Type scoring and definitions).

(2) If the function scores either 1 or 2 (i.e., it is an AI Type), describe what specific AI/ML algorithm is going to be trained/implemented.

(3) If the specific algorithm is using DNN structures (i.e., three or more layers), identify if there is enough training data to support this approach and, if not, were older ML algorithms considered, like Logistic Regression, kNNs, etc.

If function qualifies as an AI Type, follow these two recommendations:

Recommendation 2: Verify that an AI/ML function is needed by asking the following questions:

a) For AI Type definition 1: Can the algorithm be traditionally built using data approximations? Why or why not?

b) For AI Type definition 2: Can the algorithm be broken into subpopulations to allow development of traditional code? Why or why not?

Recommendation 3: Justify that an AI/ML function needs to be Autonomous by documenting the following:

a) Document how the design can or cannot include a human in the loop or traditional hardware/software technology to provide checks and balances.

b) If it cannot provide checks and balances, provide documentation as to operational limitations by:

1. Describing weaknesses of each AI/ML technique (e.g., expected success rate of the function). For example, if AI/ML is built on data approximations (using AI Type definition), how much bias will the data approximations add to the functional outcome? Or if AI/ML is built on data samples (using AI Type definition), how representative are the samples to the population?
2. Determining how the training data is being generated (e.g., truth, synthetic, combination). Are these sources valid? Why?
3. Where is the training data coming from? Is it enough? (Remember the more sophisticated the AI/ML software, the more likely that it needs larger amounts of training data.)
4. Will an outside independent source review the training, validation and test data created? Why or why not?
5. Will an outside independent source validate the success rate of the AI/ML function as compared to other AI/ML functions used in industry? Why or why not?

SSHA

Table 4 and Table 5 show analysis of one of the AI/ML functions in the sandbox, a 17-attribute, five-class Naïve Bayes algorithm for meta-model selection that implements statistically independent instances representing missing and sparse data operational issues. In our sandbox, we used 15,000 samples of training data to support classification training of Logistic Regression and Random Forest algorithms. Our analysis is independent of the categorization algorithm selected but focuses on how to assess it in terms of identifying hazards at the subsystem level. Naïve Bayes is used in the example to remove focus on the algorithm complexity and place it on the recommendations being offered.

Recommendation 4: Table 4 identifies the hazard description and, again, provides a similar table approach to non-AI functions under investigation.

Table 4. SSHA for Meta-Model Selection Algorithm Within Mission Planner From Sandbox

Haz ID#	Phase	State/ Mode	System	Subsystem	Component	Element	Hazard Title	Hazard Description	Causal Factor Description	Mishap	Effects	Existing Mitigations
Identifier used to reference specific hazard.	*The life-cycle phase for which the risk and risk assessment apply. Multiple phases can be specified if the risk & mitigations are equivalent.*	*The State and/or Mode of the system for the hazard of concern.*	*The composite at any level of complexity of interworking parts (personnel, procedures, equipment, hardware, software, et al) used together to perform a task or accomplish a mission.*	*A functional or physical portion of a system designed, used or integrated to accomplish one aspect of the system task or mission.*	*A functional or physical portion of a subsystem designed, used or integrated to accomplish one aspect of the subsystem task or objective.*	*A functional or physical portion of a component designed, used or integrated to accomplish one aspect of the component.*	*Short title of the hazard*	*The detailed description of the conditions under which hazardous energy may be released in an uncontrolled or inadvertent way.*	*The detailed description of the failures, conditions, or events that contribute either directly or indirectly to the existence of the hazard.*	*The event or series of events where hazardous energy release could negatively effect equipment, personnel or environment; accident*	*The results of the mishap to include injury or death, damage to equipment and property, or damage to the environment.*	*Controls that are already planned existing to mitigate the risk.*
SSHA-001	Test & Deployment		Mission Planner				AI Function for metamodel selection (Naïve Bayes) failure	If a wrong metamodel is selected, then the wrong robots could be selected resulting in packages not being delivered or being too early or late and so delivered to wrong receipient, or the package could be lost. (Assumption: Package not delivered means package will be lost - Catastrophic hazard)	Inadequate quality or quantitiy of training data.	Delivering the package to the wrong receipient could be catastrophic since the material is hazardous and/or very valuable.	Personnel Injury / Equipment Loss	
									Incorrect algorithm selection			
									Improper curation of data			Multiple sources (primary, secondary and tertiary sources) accommodate sources that fail/missing.
									Too much or too little data (Underfitting and Overfitting of model)			

Recommendation 5: For system safety practitioners, Table 5 might also look familiar. The recommendation is to add a single column labeled "Focus" that categorizes the LOR list of descriptions that might be unique to ML algorithm development. Notice that the list of LOR Descriptions is based on the "Focus" described. It is a simple recommendation, but from our research within the sandbox, it helps organize the range of issues that might occur. For every AI Type identified in the FHA, it is recommended that an interface analysis be considered, as described in Table 5, using the LOR Description column. Each row in this LOR Description column provides API/MSG/SQL interface questions that might affect the algorithm's performance during deployment, as will be explained next.

Table 5. SSHA LOR Table Example Based on Data Flow Analysis of Meta-Model Selection Within the Mission Planner

Level or Rigor (LOR) Activity	Phase	Focus	LOR Description	Primary Responsibility	Support Responsibility	Baseline	Software Criticality Index (SwCI)				Representative Artifacts Produced
							4	3	2	1	
ALG6: Data Flow Analysis for the Mission Planner	Algorithm Design, Algorithm Code and Test and Evaluation	API/MSG/SQL Interface	Would the corruption of API/MSG/SQL/Other affect data variations requiring additional training of the Target Algorithm?	AI/ML Algorithm Developer	Data Analytics Engineer						Data Analytics Report
			If so, will quality (composition/complexity/structure) of Training Data significantly increase? Explain specific to the API/MSG/SQL/Other.	AI/ML Algorithm Developer	Data Analytics Engineer						Data Analytics Report
			Will these variations be part of the analysis for selecting the "best" algorithm? Explain.	AI/ML Algorithm Developer	Data Analytics Engineer						Data Analytics Report
			Because of this issue, will quantity (more instances) of Training Data significantly increase? Explain specific to the API/MSG/SQL/Other.	AI/ML Algorithm Developer	Data Analytics Engineer			R	R	R	Data Analytics Report
			Will creating/finding enough training data replicating the corruption be an issue? Explain.	AI/ML Algorithm Developer	Data Analytics Engineer						Data Analytics Report
			Are you confident that any additional data created/found will adequately represent the effects associated with replicating the corruption? Explain.	AI/ML Algorithm Developer	Data Analytics Engineer						Data Analytics Report
		ML Modality: During Deployment & Curation Congruency	Based on Modality Table: Describe Data Source Precedent for Improving Success Rate (ranking of primary, secondary tertiary… n attributes) -- by addressing related question in the table.	AI/ML Algorithm Developer	Data Analytics Engineer						Training Data Curation Report
			Based on Modality Table: Describe how missing and sparse data issues are modeled -- by addressing related question in the table.	AI/ML Algorithm Developer	Data Analytics Engineer			R	R	R	Training Data Curation Report
			Based on Modality Table: Describe how the quality of Training Data Characterized -- by addressing related question in the table.	AI/ML Algorithm Developer	Data Analytics Engineer						Training Data Curation Report
			Based on Modality Table: Describe how the quantity of Training Data Characterized -- by addressing related question in the table.	AI/ML Algorithm Developer	Data Analytics Engineer						Training Data Curation Report

- PR: Prerequisite Requirement – Required regardless of LOR or required in order to assess and determine LOR
- AD: As directed by Customer/Contract
- R: Required for assigned LOR
- IV&V: Independent Verification and Validation
- N/A: Not Applicable for this program or LOR

To understand whether the ML Algorithm was trained properly to handle issues based on interface corruption, the following six questions (in sequential rows) are recommended based on our sandbox analysis:

1. Would the corruption of the API/MSG/SQL/Other affect data variations requiring additional training of the AI/ML Algorithm? This is a yes or no answer.
2. If yes, will quality (composition/complexity/structure) of Training Data significantly increase? Will it affect the ML Training Modality? Explain this specific to the API/MSG/SQL/Other. Corruption might result in a need to add secondary or tertiary sources. It might also affect how data is collected from various sources, potentially changing the ML Training Modality.
3. Will these variations be part of the analysis for selecting the "best" algorithm? Explain. ROC sweet spot analysis might be used with hyper parameter changes based on the type of variation.
4. Because of this issue, will quantity (more instances) of Training Data significantly increase? Explain this specific to the API/MSG/SQL/Other. This could result in a need to have more of a certain type of instance to train on based on mixes of primary, secondary, or tertiary attribute requirements.
5. Will creating/finding enough training data replicating API/MSG/SQL/Other corruption be an issue? Explain. If it is synthetic, is may not be an issue, depending on the model. If it comes from “live” data, then would there be more training data associated with the effects of the corruption?
6. Is there confidence that any additional data created/found will adequately represent the effects associated with replicating the corruption? Explain. This is an important statement related to the quality (composition/complexity/structure) of the Training Data.

Recommendation 6: Another series of rows has been added to Table 5 based on modality associated with the training data. Table 6 provides additional questions for the algorithm developer and data analytics engineer to address based on modality regarding how the ML is trained.

Table 6. Investigation Questions Based on Modality

Investigation Topic	(Modality 1) multiple data sources, where each source contains one or more attributes	(Modality 2) single data source containing multiple data attributes, e.g., CNN	(Modality 3) combination of multiple data streams, where each stream contains one or more attributes and from a single data stream containing multiple aggregated data attributes, e.g., Naïve Bayes aggregated with CNN
Describe Data Source Precedent for Improving Success Rate (ranking of primary, secondary tertiary… n attributes)	Which sensor, communication link or human input content elements take precedent over others for improving success rate when training the ML algorithm under normal to stressed operational conditions?	Which attributes within the single data source take precedent over others for improving success rate when training the ML algorithm under normal to stressed operational conditions?	What data source content is more significant with regard to normal to stressed operational conditions? When dealing with separate streams, which sensor, communication link or human input content elements take precedent for improving success rate when training the ML algorithm under normal to stressed operational conditions? When dealing with combined streams, which attributes within the single data source are identified as primary, secondary and tertiary regarding importance for ML algorithm to improve success rate under normal to stressed operational conditions?
Describe how missing and sparse data issues are modeled	How is sensor malfunction, message corruption and human input errors on the higher precedent attributes forcing lower level attribute mixes of training data to ensure algorithm can deal with “real” operational issues?	Corruption in parts of image, especially containing higher precedent attributes forcing secondary and tertiary attribute mixes of training data to ensure algorithm can deal with “real” operational issues.	Combinations on modalities 1 and 2 regarding training of algorithm to deal with “real” operational issues.
Describe how the quality of Training Data Characterized	What is the precedent list (from highest to lowest) of attributes being used for training.	Same as Modality 1 for this row.	Same as Modality 1 for this row.
Describe how the quantity of Training Data Characterized	How much more emphasis Is placed on quantify of training data variations that have higher precedent than lower?	Same as Modality 1 for this row.	Same as Modality 1 for this row.

A brief summary of modality types that support training data composition and size are described next:

- ML Training Data Modality 1: This modality supports training data sets that are based on an operational environment from multiple data sources, where each source contains one or more attributes. The various sources of separate data attributes are either found from live events or synthetic simulations created to match the deployed operational scenario. Therefore, the input for the ML algorithm for training needs to replicate the input that will be received during deployment.
- ML Training Data Modality 2: Training data sets that are based on an operational environment from a single data source, where the single data source contains multiple data attributes. The one stream set of aggregated attributes is either found from live events or synthetic simulations created to match the deployed operational scenario. Therefore, the input for the ML algorithm for training needs to replicate the input during deployment.
- ML Training Data Modality 3: Training data sets that are based on an operational environment from a combination of multiple data sources where each source contains one or more attributes form various sources and from a single source containing multiple aggregated data attributes. It is a combination of Modality 1 and 2 that the algorithm uses for categorization or regression.

Conclusions and Final Best Practice Recommendations

Our findings indicate that the FHA and SSHA for AI/ML SSFs need to be addressed differently from traditional functional analysis methods. To address these differences, the AI Type definition and scoring approach was introduced, along with six recommendations regarding the FHA and SSHA. The research describes questions/issues needing to be addressed when conducting safety analysis on AI/ML function types. It includes discussion on how the current FHA process is still valid for AI/ML functions and only requires three additional columns to support added justification that an AI/ML function is required to meet operational goals. The SSHA discussion provides a simple table modification and two examples of LOR Descriptions that need to be addressed when dealing with AI/ML critical functions: (1) interfaces to the algorithm to understand the impact of potential data corruption, and (2) modality issues to ensure robust curation of the data to ensure the algorithm is trained to meet deployment challenges.

Along with the AI Type scoring and six recommendations associated with the analysis of critical functions, there were complementary "Best Practice" questions that arose from our sandbox development environment when developing AI/ML algorithms within a deployed weapons system.

"Best Practice" areas to consider specific to AI/ML critical functions include:

General AI/ML Questions:

- When a critical function is identified, does it meet the AI Type definition criteria: (1) Is the algorithm built based on using data approximations, and/or (2) is the algorithm built based on using data samples from larger populations?
- When doing M&S to create the training data, does the simulation adequately represent Classes for the ML process? If not, how are Classes represented?
- Does each Class have a sufficient number of attributes that can be learned by the algorithm for that Class? Are overfitting and underfitting considered for that Class

with regards to the quantity of attributes simulated and does that reflect real world operations?

- How do we know that the M&S creating the training data is aligned with the mission parameters? Was a traceability study performed to ensure adequate coverage? Have statistics been developed to show how many configurations exist and how many were trained using primary, secondary, and tertiary data sources? How are we avoiding overfitting and underfitting based on primary, secondary, and tertiary training data mixes and sets? Is the training data organized in terms of primary, secondary, and tertiary attributes to be able to represent missing and sparse data priorities from related sources?
- How are we ensuring that the algorithm being deployed provides the correct answer when data input issues occur? Is the algorithm success rate determined by primary, secondary, and tertiary attributes?
- Can other control entities (such as a human operator) be inserted into the loop to reduce the SCC?

Operational “Realism” Questions:

- Is the M&S able to create training data that represents reality when sparse and missing data issues occur?
- Does the architecture, design, and code support sparse and missing data management; specifically, does it filter or select less significant attributes to do the calculations?
- Does the data management support filtering to ensure the ML algorithm is provided accurate data input, avoiding “garbage in, garbage out?” Has what constitutes “garbage in, garbage out” been defined?
- How well does the particular ML algorithm support increased complexity, and how does that affect sparse and missing data issues?

Selected AI/ML Algorithm:

Note: Individual types of AI/ML require specific questions that address their method and application. Naïve Bayes is used as a simple example, but Logistic Regression, Random Forest, DNNs, or other categorization algorithms could have used the training data produced within our sandbox environment.

Some questions to guide the examination of Naïve Bayes are:

- How do you trust the behavior in the real world for this Naïve Bayes function? Success Rate? Quality of Training Data? How did it compare with other categorization algorithms?
- Was Naïve Bayes the correct selection for this function vs. other algorithms? The choice should be based on what gives you the best operational performance and understanding of operational limits (Potential OQE: k-fold cross validation comparisons, etc.). How reliable will the answers be in the real world?
- How do you assess the operational limits of this Naïve Bayes (or alternate algorithm) categorization function?
- Did the training set model enough noise/clutter (in this case, less significant attributes determined by SMEs for a particular meta-model class) for each class that allows for the function to work properly when deployed? Are there sparse data and/or missing data issues? How is the bias of the training set and variance of the test determined?

- How would you ensure simulation configurations (i.e., the training data) are adequately covering the real world experiences? Consider optimizing bias (how well it fits the training set) and variance (how well it predicts using the test set)—overfitting/under-fitting.
- How many types of simulations and how much training data is really enough?
- Are the attributes used for the assessment really independent?
- Is the size of the alpha correct? Is this hyper parameter optimally used?
- Is MAP or Maximum Likelihood better for this calculation?

Developing defined lists of questions/issues, as described in this paper, allows system safety professionals to identify how to increase the inherent safe operation of safety significant AI/ML functions. By following the guidance provided, the system safety practitioner can drive important discussions on the development of the AI/ML function and thereby potentially influence design of the overall system to decrease the mishap risk associated with these specially developed functions.

References

Booch, G. (2017). *Unified modeling language user guide* (2nd ed.). Addison-Wesley Professional.

Brose, C. (2020). *The kill chain: Defending America in the future of hi-tech warfare*. Hachette Books.

Dam, S. (2006). *DoD architecture framework: A guide to applying system engineering to develop integrated, executable architecture*. SPEC.

Defense Standardization Program Office. (2012). *System safety* (MIL-STD 882E). Pentagon.

DoD Chief Information Officer. (n.d.). *The DoDAF architecture framework version 2.02.* https://dodcio.defense.gov/library/dod-architecture-framework/

Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The elements of statistical learning data mining, inference and prediction* (2nd ed.). Springer.

Joint Services – Software Safety Authorities. (2017). *Software system safety implementation process and tasks supporting MIL-STD-882E* (JS-SSA-IG Rev. A).

Joint Software Systems Safety Engineering Workgroup. (2017). *Software system safety implementing process and tasks supporting MIL-STD-882E* (JS-SSA-IF Rev. A). Pentagon.

National Defense Authorization Act (NDAA) for Fiscal Year 2021, Pub. L. No. 116-283 (2021).

National Institute of Standards and Technology. (2019). *U.S. leadership in AI: A plan for federal engagement in developing technical standards and related tools*. https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf

National Security Commission on Artificial Intelligence. (2019). *Final report: Establishing justified confidence in AI systems* (Ch. 7).

National Security Commission on Artificial Intelligence. (2021). Autonomous weapon systems and risks associated with AI-enabled warfare. In *Draft final report* (pp. 45–53). https://www.nscai.gov/wp-content/uploads/2021/01/NSCAI-Draft-Final-Report-1.19.21.pdf

Naval Sea Systems Command. (2008). *Department of the Navy Weapon Systems Explosives Safety Review Board* (NAVSEAINST 8020.6E).

Radio Technical Commission for Aeronautics. (2012). *Software considerations in airborne systems and equipment certification* (DO-178C). Federal Aviation Administration.

Sodhani, S. (2018). *A summary of concrete problems in AI safety*. https://futureoflife.org/2018/06/26/a-summary-of-concrete-problems-in-ai-safety/

Changing Course to a 21st Century Acquisition Strategy: Navy-Industry Collaborative Design

Tim Nichols—has extensive experience in warship design and construction both in international and domestic shipyards. After completing 9.5 years on active duty in the U.S. Navy, he joined GE Marine & Industrial Engine Division where he spent the next 15 years developing the global marine propulsion markets for GE marine gas turbines. Subsequently, he held international leadership positions at GE supporting the growth of their aerospace products. Then he transited to the aerospace information technology industry eventually retiring from Siemens in 2017. He is now president of Ubiquitech LLC, an independent consulting company. [tim.nichols@ubiquitechllc.com]

Robert G. Keane, Jr.—has more than 55 years of ship design and acquisition experience. During 21 years as a member of the Federal Senior Executive Service (SES), he served as the U.S. Navy's Chief Naval Architect, Chief of Ship Design, Director of Survivability and Director of Naval Architecture. He was responsible for the design of surface ships, aircraft carriers, and submarines from pre-acquisition concept studies, through design, construction, and in-service. His Navy-led design teams successfully completed 35 major contract designs leading to the acquisition of 200 ships. He now heads his own consulting firm, Ship Design USA, Inc. He is well published and has received numerous professional society and Navy honorary awards. [Keanerg@comcast.net]

Captain Barry Tibbitts, USN (Ret.)—has many decades experience in R&D, design, construction, maintenance, and operations in both submarines and surface ships. He served 5 years at sea and 8 years in shipyards, was commander of the David Taylor Research Center, Director of the NAVSEA Ship Design Group for 6 years, and professor of Naval Construction and Engineering at MIT. He has published more than 30 technical papers and authored chapters in four books. He was an adjunct professor at both Virginia Tech and Stevens Institute of Technology, and for 25 years lectured on ship design integration each summer at MIT. [bftibbitts@yahoo.com]

Peter E. Jaquith—has shipyard executive experience in ship design, production engineering, planning, and production. He is a recognized leader in improving North American shipbuilding processes including the application of advanced outfitting, modular engine room techniques, the Toyota Production System, and world-class Lean Design and Production Engineering practices to naval and commercial ship design and construction. Peter received a B.S. in Naval Architecture and Marine Engineering from Webb Institute of Naval Architecture with advanced studies at the University of Maine and Harvard Business School. [petejaquith@earthlink.net]

Abstract

The U.S. Navy (USN) and U.S. naval shipbuilding industry are facing an historic inflection point to realize the growth in the number of warships in the fleet over the next 2 to 3 decades. And a demanding shipbuilding program demands a new 21st-Century Acquisition Strategy: Navy–Industry Collaborative Design. This new strategy will enable and promote open, substantive collaboration between the U.S. Navy and naval shipbuilding Industry and will ensure the design, construction, and sustainment of a more affordable, adaptable, and durable fleet. The team of four authors of this paper with collectively more than 200 total years in naval ship acquisition, design, and construction management believes strongly the time is long overdue for such a bold strategy. No longer will the recent failed acquisition approaches enable the USN and U.S. shipbuilding industry meet and/or surpass the existential and growing challenges of its naval adversaries. Based on the team's significant experience and insight into naval ship design and shipbuilding as well as a decade of American Society of Naval Engineers (ASNE) Global Shipbuilding Executives Summits (GSES), the authors have compiled in this paper a set of recommendations for a bold new acquisition strategy for the USN.

Executive Summary

The U.S. Navy and U.S. naval shipbuilding industry are facing an historic inflection point to realize the growth in the number of warships in the fleet over the next 2 to 3 decades. And a demanding shipbuilding program demands a new 21st-Century Acquisition Strategy: Navy–Industry Collaborative Design. This new strategy will enable and promote open, substantive collaboration between the U.S. Navy and naval shipbuilding Industry and will ensure the design, construction, and sustainment of a more affordable, adaptable, and durable fleet. The team of four authors of this paper with collectively more than 200 total years in naval ship acquisition, design, and construction management believes strongly the time is long overdue for such a bold strategy. No longer will the recent failed acquisition approaches enable the USN and U.S. shipbuilding industry meet and/or surpass the existential and growing challenges of its naval adversaries. Based on the team's significant experience and insight into naval ship design and shipbuilding as well as a decade of American Society of Naval Engineers (ASNE) Global Shipbuilding Executives Summits (GSES), the authors have compiled in this paper a set of recommendations for a bold new acquisition strategy for the USN that is well-grounded on successful acquisitions for the Cold War fleet, proven recently in allied navies' acquisitions and, moreover, will help avoid the mistakes of the past 2 decades.

Additionally, best practices and lessons learned on naval ship design, engineering, construction, and sustainment are reviewed based on the innovations and breakthroughs global naval-shipbuilding leaders have implemented over the past 2 to 3 decades. For example, lean process re-engineering, digital shipyard process simulation and optimization, and enterprise-wide digital transformational have produced double-digital improvements in construction productivity, cycle reduction, and capacity throughput. Increase in shipbuilding capacity could be crucial in the U.S. shipbuilding industry to satisfying the dramatic increase in naval shipbuilding rates that are projected over the next 2 decades.

The team finally lists a set of acquisition-related recommendations to build a long-term commitment to naval shipbuilding continuous improvement and to create a pipeline of seasoned naval shipbuilding professionals to guide the future of the U.S. Navy through-out the 21st century and beyond.

USN SITREP

Figure 1 is our assessment of the current situation. What has been called the post-Cold War era is over. China is now a peer Navy, and Russia is a near-peer Navy. Both are growing in size and capability. U.S. sea control is being challenged for the first time in many years.

SITREP

- Post-Cold War era is over.
- Peer navies growing in size and strength.
- COCOMs' requirements overstress ships and personnel
  - Number of battle force ships stuck at 300 for over a decade
  - Many ships offline for maintenance/modernization for long periods
  - New ships "deliver" years before deployable and over budget
- Future force **level** uncertain: 350 or 400 or 500
- Future force **mix** also uncertain: DDG(X), CVLs, Light Amphibious Warship, and broad array of Unmanned Systems

Figure 1. USN at Historic Inflection Point

Our ships and personnel are overstressed, and even so, not all of the combatant commanders' requirements can be met. USS *Nimitz's* most recent deployment stretched to 340 days, yet gaps in aircraft carrier coverage—once unheard of—now occur. The number of battle force ships dropped below 300 15 years ago and has hovered near 300 since then. The high tempo of ops creates a backlog of maintenance. Many ships are offline for maintenance and modernization, and most complete late—often quite late. Some new ships are delivered and commissioned but are years away from being deployable.

The Navy is therefore at an historic point in time, but when it comes to the future fleet, there is uncertainty about both force level and force fix. Is the goal still 355, or is it much higher as the last administration proposed? As for the mix of ships, there are knowns like DDG(X), and unknowns like CVLs, light amphibious ships, and a host of unmanned vehicles.

Over 30 years, DDG51 capability has substantially increased in a series of flights, but further growth is no longer feasible. Future weapons will require substantially more electric power and the ship space to accommodate larger missiles. This requires a new hull that incorporates the time-tested service life allowances, which enabled DDG51 Flight I to evolve into Flight III. DDG(X) will provide this future growth capability.

CVNs are costly, and less expensive solutions are periodically examined and, up until recently, rejected. The F-35B, far more capable than the venerable AV-8B, operating off big deck amphibs, is seen by some as a game changer.

The Marine Corp's recent shift in emphasis from inland wars in the Middle East to Indo-Pacific littorals, has translated into the need for light amphibious warships.

Finally, the broad array of unmanned vehicles presents both challenges for design and acquisition and opportunities for new concepts of operations.

Build on Success

There are lessons to be learned from successful acquisition programs for the Cold War fleet, including recent acquisitions of our allies (see Figure 2).

BUILD ON SUCCESS

- US submarine community
  - Design-build strategy
  - Shipbuilders' collaboration early in Navy - led design
- Asian Aegis Shipbuilders
  - Proven design-build strategies
  - No arbitrary displacement/size constraints thus less dense ships
- NATO Shipbuilders
  - Long-term commitment (>30 years) to more adaptable surface combatants
  - Shipbuilders' early collaboration (including other nations)
  - Modular combat systems
- 1980s build up to 600-ship Navy
  - Tailored approaches with Navy-led designs: FFG 7, CG 47/52, DDG 51 FLTs I/II/IIA, CVN 76, LPD17
  - SEA 05 controlled ship design resources/capabilities

Figure 2. Leverage Best Practices and Proven Principles of Good Design

In the *Virginia* class, the U.S. submarine community built on lessons learned from the *Seawolf* (SSN 21) program. In SSN21, there had been a winner-take-all approach and as was also common in surface ship acquisitions, production started well before detail design was complete. A "design-build" strategy was adopted in *Virginia* to control costs, and both building yards acted as collaborators vice competitors. The *Virginia* class is built through an industrial arrangement designed to maintain both GD Electric Boat and HII Newport News, the only two U.S. shipyards capable of building nuclear-powered submarines. Under the present arrangement, the Newport News facility builds the stern, habitability, machinery spaces, torpedo room, sail, and bow, while Electric Boat builds the engine room and control room. The facilities alternate work on the reactor plant as well as the final assembly, test, outfit, and delivery.

In comparing U.S. and foreign shipbuilders, different ship types built to different requirements and facing different threats may present difficulties. A major exception appears to be the Asian shipbuilders, Japan and South Korea, which have built their own versions of DDG51 Flights I and IIA. The Aegis Combat System is common to all three navies, but the ships are quite different. Both displacement and size constraints had been imposed on DDG51 in a (vain) attempt to control cost. This was due in part to undue emphasis placed on weight (vice work content) when estimating cost. *Kongo*, the Japanese version of Flight I, is substantially larger than *Arleigh Burke* but being less dense was much easier to build.

Our NATO allies will be covered later in the paper, and there is much to learn from them. They have pioneered in developing adaptable ships. In part to meet the needs of their own services, but also to appeal to a variety of foreign customers. Shipbuilders frequently collaborate, and this can include multinational programs.

During the buildup to the 600-ship Cold War Navy, a number of tailored collaborative acquisition strategies were adopted. When it came to ongoing programs (and even some new starts) this included business practices such as multiyear procurements and the emphasis on fixed price contracts. Clean sheet of paper designs such as FFG7 and DDG 51 were in-house designs with co-located NAVSEA-led design teams, assisted by the Navy labs (current Warfare Centers), industry (shipbuilders), and local design agents. The time from earliest concept studies until delivery of a lead ship takes a decade or more (source selection alone may take up to a year), the combat system may still be evolving, and yet the first of class must be “fully successful.” Indeed, in the case of DDG51, seven follow ships were on order or under construction at two shipyards **before** *Arleigh Burke* delivered, and three more before IOC was achieved. This is inherently risky, but the time scale for ships prevents the use of prototypes and “fly-offs” common to other DoD acquisitions. To minimize risk the Navy became **“self-insured,”** meaning NAVSEA (and predecessor organizations) developed a comprehensive shipbuilding specification which the potential shipbuilders bid on and followed. Combat systems were developed by the Navy, tested at sea and ashore, and delivered to the shipbuilder(s) as GFM. Land-based testing of new propulsion plants was also common, and their procurement was specified. Time has shown that this approach ensured that the new ship met its operational performance requirements and minimized the risk of incorporating new technologies.

AVOID PAST MISTAKES

- Acquisition Reform
  - Gutted NAVSEA 05- Shifted early design responsibly to industry
  - Reassigned ship design funds from SEA 05 to PMs
  - Wasted scarce resources on designs never built
- DDG 1000
  - Many high-risk developmental systems
  - Took too long- basic mission became OBE
- LCS
  - Bypassed checks and balances determining requirements & costs
  - Rushed ships into production a decade before mission systems ready

Figure 3. Focus on Building a World Class Team and Reducing Risks

Avoid Past Mistakes

Over the last 75 years, the Navy has employed many different warship acquisition strategies—often directed by higher authorities. In particular, the relative roles and responsibilities of the government and the shipbuilders have differed. Total Package Procurement (TPP) shifted major design responsibilities to industry in the 1970s, then during the buildup to the 600-ship Navy in the 1980s the Navy (NAVSEA) regained design responsibility, and then in the early 2000s under the banner of Acquisition Reform, design responsibilities were again shifted to industry (TPP reborn).

As shown in Figure 3, as a result of Acquisition Reform, NAVSEA 05 was reduced from 1,200 naval engineers to 300, and lost control of ship design funding including RDT&E design funds for future ships in the Navy’s Shipbuilding Program, as well as funds to sustain the Navy’s ship design capabilities. Time-consuming industry design competitions were conducted, and as with any competition, this resulted in some designs that were never built—wasting scarce national resources. (One DDG1000 design and one LCS design were never built).

To ensure a fair competition, and to avoid protests, the Navy's ability to influence the designs was limited. Unfortunately, the winning designs were seriously flawed. Axe (2021) states, "The U.S. Navy spent a decade in the early 2000s building warships that either don't work, cost too much to build in large numbers, or whose designs are fundamentally flawed on a conceptual level. Or all three."

DDG 1000 must be deemed a failure. The acquisition strategy was to hold a design competition between two industry teams. The Navy requirements (speed, number of missile cells, rounds of gun ammunition, manning, etc.) were expressed as thresholds and goals. The goals could not be met without exceeding the cost constraints; the Navy did not indicate their preferences, and so each team decided which requirements to emphasize. Thus, industry assumed an inherently governmental responsibility.

In addition, the government requirement ultimately necessitated development of far too many "critical technologies." Radical new systems onboard this first-of-class warship that had to be fully integrated include: Integrated Propulsion System (IPS), Integrated Fight Through Power (IFTP), Advanced Perimeter Vertical Launch System (PVLS), Advanced Gun System (AGS), Advanced Signature Control across all spectrums, Large Composite Enclosed Deckhouse with Embedded Sensors, Total Ship Computing Environment (TSCE), Advanced Survivability and Recoverability, and Wave Piercing Tumble Home (WPTH) Hull Form.

Construction started with the majority of these systems still immature; costs grew, and schedules slipped. The rise of peer navies, discussed earlier, and shore-based anti-ship missiles, rendered the basic mission of close in land attack in a "benign" environment moot. The program was truncated, and construction of DDG51s restarted.

LCS. The problems with these two classes are well known and need not be dwelled on here. Reports from GAO and CRS are depressing to read. The LCS program has been controversial over the years due to cost growth, design and construction issues with the first LCSs, concerns over the survivability of LCSs (i.e., their ability to withstand battle damage), concerns over whether LCSs are sufficiently armed and would be able to perform their stated missions effectively, and concerns over the development and testing of the modular mission packages for LCSs.

The program was flawed from the start. The time-tested cost versus capability studies were compressed, and NAVSEA was basically ignored. Cost was grossly underestimated. The potential risks and large ship impact of requiring 40 knots were also ignored.

The ships (basic sea frames) were delivered on an accelerated schedule, but with many deficiencies. The government has experienced many delays in developing the three major mission modules, and the requirement to rapidly change modules had to be abandoned. Crew manning was far too low. Many years later a valid CONOPS **still** does not exist.

The latest CRS report (RL33741 dated December 17, 2019) states:

> They could argue that the LCS program validated, for defense acquisition, the guideline from the world of business management that if an effort aims at obtaining something fast, cheap, and good, it will succeed in getting no more than two of these things, or, more simply, that the LCS program validated the general saying that haste makes waste.

Bold New Acquisition Strategy

The Navy is at a crossroads. The fleet is too small to counter the increasing geopolitical threat(s). There is uncertainty about what types of ships to buy, how many of each, and how quickly. Between new construction, maintenance, and modernization the industrial base is stretched thin. Between DDG1000, both LCSs, and, yes, CVN 78, the Navy's reputation has suffered. To meet all these challenges, we urge that the Navy adopt a bold new acquisition strategy.

The key is **increased collaboration** between the Navy and the shipbuilders, capitalizing on the strengths of each. For programs like DDG(X) where there will be two shipbuilders, they must also collaborate with each other for the good of the nation. Both the Navy and the shipbuilders must agree, but there is precedent. The Navy, HII, and General Dynamics have been collaborating on the successful *Virginia* submarine program for years.

Figure 4 illustrates key aspects.

NEED BOLD NEW ACQUISITION STRATEGY

- Involve industry early
  - AOA cost/capability/risk studies which establish requirements
  - Incorporate production planning into Navy - led ship design teams
  - Ensure design decisions facilitate manufacturing and construction
    - In both shipyards when construction will be split
  - Review/comment on shipbuilding specifications/contract drawings
  - Expand contract design to include aspects of functional design
  - Assist in developing 3D product model
- Leverage digital twin/digital thread to minimize Total Ownership Costs

Figure 4. Navy–Industry Collaborate for Innovation Good for the Nation

Industry would be involved in the cost, capability, and risk studies, which establish balanced and realistic requirements. These studies are conducted before there is even a "program of record," but they lock in up to 80% of the cost and performance. The shipyards will bring a unique perspective.

The Navy's record for involving industry in the design process is mixed. With rare exceptions, FFG7 and SWATH T-AGOS19 for example, this has been too late to materially influence the design. When the acquisition strategy is one where multiple shipyards will compete against each other, they are often reluctant to share proprietary data. This reluctance must be overcome in order to develop a tailored vice a generic build strategy.

Ultimately, it is the shipbuilding specifications and contract/contract guidance drawings (now, 3-D CAD product model) which define what gets built, and the shipbuilders should be active players in their development. In addition to participating in reading sessions, this could include assigning them responsibilities for preparing selective sections and early development of the 3-D CAD product model.

The scope of the traditional contract design should be expanded to include elements of functional design—the initial step in developing the detail design.

Best practices include developing a 3D product model. There are many benefits, but past efforts have failed due to incompatibility with shipyard systems and/or requiring them to "learn" a new system after the construction contract is awarded. The 3D product model should eventually reflect the "as delivered" ship and be maintained through its service life.

Last, but far from least, greater emphasis must be placed on making decisions based on Total Ownership Costs (TOC). This sounds obvious but has proved to be difficult in practice. The elements that comprise TOC span many budget claimants. Acquisition managers focus on delivering ships on time and under cost. Others are responsible for maintaining and modernizing them during their long service lives, yet decisions made years earlier are major factors. Finally, the fleet operators live in the real world, and words like *sustainment* and *availability* are not just buzzwords to them. The GAO (2020) found that shipbuilding programs did not consistently address sustainment risks in acquisition planning documents. For example, for six shipbuilding programs whose costs GAO could assess, the Navy had underestimated sustainment costs by $130 billion. The application of digital twin/digital thread should be leveraged here.

More Affordable, Adaptable and Sustainable Naval Ships

Asian navies are building larger warships that are easier to construct, easier to maintain, and that have greater service life allowances for future combat system upgrade. They are also doing this at significantly lower acquisition and life-cycle costs than U.S. practice. A critical element in their ship design process is early application of production engineering and design optimization analysis that improves ship performance, life-cycle maintenance, and future combat system upgrade while reducing work content, reducing design variation, and ensuring design alignment with their warship manufacturing processes. Best early-stage warship design optimization practices include the following:

- **Superior Performance at Lower Cost.** Early production engineering and lean design optimization analysis supports development of superior and more robust warship designs with reduced work content, reduced variation, design alignment with manufacturing processes and lower total ownership costs (TOC).
- **Integrated Product Team (IPT).** Modern design practice harnesses the combined knowledge and experience of the Navy, shipbuilders, key suppliers, operators, and maintainers to consider alternative design solutions and select superior designs that can be produced at the lowest possible time and cost.
- **Production Engineering Focus.** Stressing production engineering analysis in early-stage design supports lean design (reduced work content) and design for manufacture and assembly (design alignment with manufacturing process); strategy improves design quality and reduces time and cost.
- **Robust Contract Design**. Utilizing a robust design process with functional design level definition that meets current and future requirements, minimum work content and variation, and design alignment with manufacturing processes provides a stable basis for design execution and reduces time and cost.
- **Increased Displacement, KG, and Service Life Allowances.** Modern warship designers use significantly larger displacement, KG, electric power, and cooling system service life allowances and design margins than U.S. practice. This strategy reduces design rework and program execution time and cost.
- **Reduced Outfit Density.** Modern warships have significantly lower outfit density than recent U.S. practice. This design strategy provides significantly improved access resulting

in reduced construction, life-cycle maintenance, and future combat system upgrade time and cost.

- **Design Optimization.** Utilizing modern engineering analysis tools (e.g., FEA, CFD, M&S, Functional Affinity Analysis, etc.) to optimize designs for functional performance, reduced work content, and design alignment with the manufacturing process improves design quality and reduces time and cost.
- **3D Product Model.** Utilizing 3D Product Modeling Systems starting in early-stage design provides a single voice of truth for design development, configuration management, production planning and control, and resource/material requirements planning; strategy improves design quality and reduces time and cost (see Figure 5).

MORE AFFORDABLE, ADAPTABLE AND SUSTAINABLE NAVAL SHIPS

- 21st Century Acquisition Strategy
  - Industry collaboration starting in early -stage design
  - Increased design quality & reduced cost/work content
- Design for Performance & Reduced Cycle Time
  - Design for performance & reduced cycle time
  - Robust contract design definition strategy
  - Increased service life allowances
  - Early consideration of maintenance & upgrade
- Early 3D Product Model Development
  - 3D Product Model initiated in early -stage design
  - M&S of maintenance & CS equipment loadout
  - Build Strategy included in 3D product model

Figure 5. Design for Performance, Construction, Sustainment, and Upgrade

Use of such a collaborative design organization supports direct engineer-to-engineer communication; rapid design decision-making; and development of higher quality more producible, maintainable, and upgradeable designs than traditional methods. Implementation of these best warship design practices requires early industry involvement and a new 21st century warship acquisition strategy. Recent NEJ articles by Keane et al. (2018, 2019) describe such a strategy. Further details on best early-stage warship design practices are addressed in the recent NEJ article “Asian vs. U.S. Warship Design, Production Engineering, and Construction Practice” (Jaquith, 2019).

Optimize Design-Build Process

An excellent case study is Bath Iron Works’ implementation of a comprehensive warship manufacturing plan on its FFG-7 Class program in the late 1970s (Jaquith, 2020). The plan focused on first-time quality, cost, and schedule reduction. Since the 1970s, technology developments including use of 3D product models, “interim product by stage-of-construction” based work instructions, and digital reporting of installation and test status provide further improvements in productivity. Best manufacturing practices for warship construction include:

- **Integrated Hull, Outfit, and Paint Build Strategy.** Develop integrated build strategy starting in early-stage design; target >95% metal outfit and >90% equipment, piping, ventilation, and local cable installation in pre-outfit; strategy provides the basis for requirements/material planning and reduces time and cost.[1]
- **Design and Material Support.** Ensure >98% design and material support to the manufacturing plan as measured at the work package level and >98% work packages completed without design or material change; strategy empowers production work teams and supports strategic reduction in construction time and cost.
- **Facilities and Tooling Support.** Ensure >98% key facilities and tooling support of the manufacturing plan (e.g., major cranes, transporters, assembly and pre-outfit halls, panel lines, burning machines, etc.); strategy empowers production work teams and supports strategic reduction in construction time and cost.
- **Repeating Workstations and Work Teams.** Utilize both real and virtual workflow to plan assembly, pre-outfit, ship erection, onboard outfit, test, and trials activities using repeating workstations and repeating work teams; strategy increases ship-to-ship learning, improves quality, and reduces construction time and cost.
- **Change Management.** Manage both internal and customer change with time fencing rules to avoid production impact; accomplish critical changes in Post Delivery Availability (PDA); strategy empowers production work teams and allows construction to proceed under planned/controlled conditions, reduces time and cost.
- **3D Product Model.** Utilizing 3D Product Modeling Systems starting in early-stage design provides a single voice of truth for design development, configuration management, production planning and control, and resource/material requirements planning; strategy improves design quality and reduces time and cost.
- **Work Instruction Design.** Utilize "interim product by stage-of-construction" based work instructions in lieu of traditional system drawings; strategy provides production work teams with clear direction reducing rework, construction time, and cost.[2]
- **Continuous Improvement, Accuracy Control, and Quality Management.** Based on Deming Quality Management System; focus on improving work sequence, increasing pre-outfit levels, reducing schedule durations, and addressing systemic quality issues; strategy improves quality and reduces time and cost.
- **Supply Chain Integration.** Supplier and sub-contractor fabrication and installation schedules fully aligned to manufacturing plan; strategy empowers production work teams, improves quality, and supports strategic reduction in construction time and cost.
- **Navy and Shipbuilder Collaboration.** Program milestones, funding for long-lead material, GFI, GFE, combat system installation and test schedules, Navy certification schedules, sea trials, and crew training fully aligned to the manufacturing plan; strategy supports strategic reduction in ship construction schedules (see Figure 6).

---

[1] Pre-outfitting includes the installation of outfit equipment and systems during the hull block (unit) assembly process prior to ship erection. Pre-outfitting also includes the use of shop assembled outfit modules (rafts).

[2] "Interim product by stage-of-construction" work instructions include all drawings, material lists, and instructions required for efficient construction. They are prepared automatically using the 3D Product Modeling System.

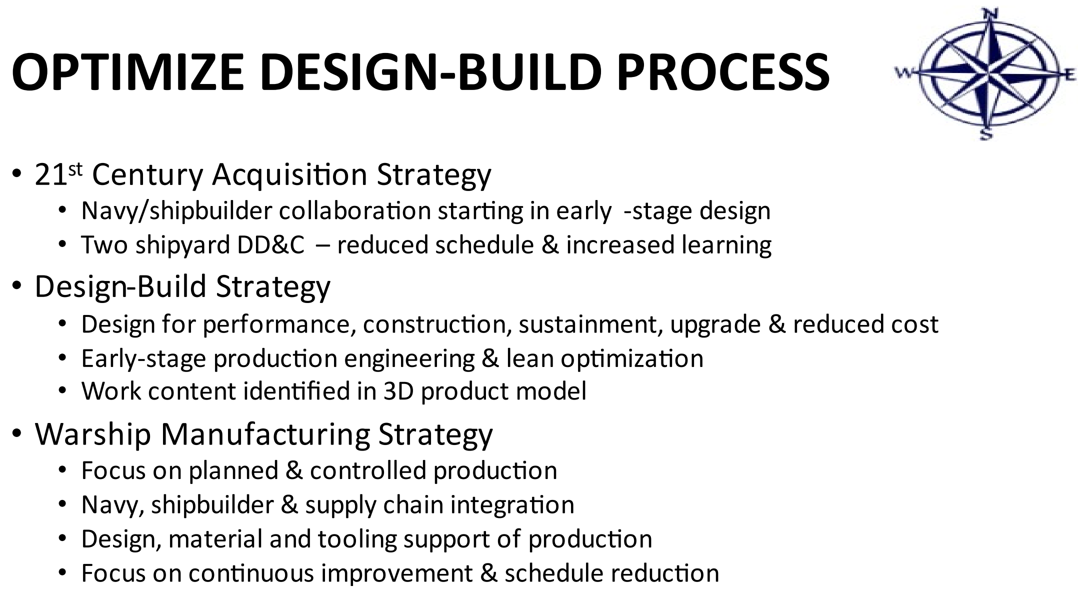


Figure 6. Design-Build to Enable Durability and Longevity > 40 Years

For critical warship programs such as DDG(X), the use of two shipyards for Detail Design and Construction (DD&C) will reduce design and construction durations, increase production throughput, and increase ship-to-ship learning. Further details on best warship manufacturing practices are described in the recent NEJ article “Modern Warship Manufacturing Practice: Impact on Acquisition Cost, Schedule, and Industrial Mobilization” (Jaquith, 2020).

NATO Navy Acquisition Innovations

The international acquisition innovations that have been implemented over decades in Europe and Asia have focused in large part on the mid-range surface combatant, the backbone of most international navies. In several countries, this continuous focus on the naval frigate has spanned over 30 to 40 years. Navy officials and shipbuilders have joined forces to design and build naval ships that satisfy domestic as well as global markets. Designing a common solution to multiple markets has resulted in designs that are more versatile and adaptable. The latter being critical for a naval class to maintain its operational relevance for more than 40 years with two major upgrades. Moreover, the focus has increased shipbuilding design and construction productivity, cycle reduction, and throughput.

In Germany, the Netherlands, and Denmark, there has been an emphasis on lean processes, elimination of non-value-added tasks and the implementation of a comprehensive digital transformation to provide a single, reliable source of knowledge about a ship class and each hull number. Now this digital backbone is being extended to the supplier network and sustainment infrastructure. The latter being an essential step to improving sustaining engineering and upgrade engineering efficiency and cycle reduction.

The results of these innovations that have been dramatic improvements, for example, **design versatility,** 1,500 payload modules for one family of naval warships; **production productivity**: 50% reduction in design and construction cycle time, and **upgradability**: 75% to 90% module replacement cycle reduction and 33% to 67 % reduction in major ship systems upgrade cycle time. Additionally, leading naval shipyards in Europe have developed the processes and management versatility to enable international customers to build their designs efficiently in local shipyards. This has permitted these shipyards to expand their markets and production base to further reduce the price of their ships.

The discipline of a robust, comprehensive, and secure 3D model with all associated technical information has enabled shipbuilders in Europe to form coalitions with other shipyards that can transcend international borders to pursue new programs. For example, Luerssen in Germany and Damen in The Netherlands formed an alliance to pursue and win the Germany Navy, F126 Frigate Program [displacement 9,000 tons; four ships in the class], and Babcock Marine International won the UK RN F31 Frigate Program with an OMT Danish Stanflex hull design [displacement 5,700 tons; five ships in this class]. The total price of a UK RN F31 Frigate appears to have established a new benchmark for a naval frigate of $100,000 USD/ton. The new USN FFG62 Frigate at $135,000 USD/ton is very competitive with other current international frigates.

Successful coalitions that have been formed in Europe are examples of what can be accomplished when leading shipyards join forces and share best practices with a secure digital backbone that synchronizes processes and change management and maintains a fidelity across all operations throughout the entire design and build cycle. Moreover, a 3D Product Life Cycle Management technical definition when maintained continuously can improve sustaining engineering efficiency which in turn can boost class availability and eventually enable operational relevance for more than 40 years (see Figure 7).

NATO NAVY ACQUISTION INNOVATIONS

- **Acquisition Initiatives:**
  - Public-Private Focus on Surface Combatant Development: >30 years
  - Joint Shipbuilding Coalitions: in Germany, UK, Denmark, France, Italy, etc.
  - Versatile Designs: 1,500 payload modules for a family of warships designs
  - Durable Designs: 40+ years of operational relevance/superiority with multiple upgrades
  - Pursuit of both domestic and international naval programs with collaborative design/platform/specs.
- **Key Drivers/Motives for Initiatives:**
  - Meeting concurrently naval operational obligations and defense budgetary constraints
- **Results:**
  - Average price of first 10 FFG62 frigates = $135,000 USD/long ton(FL)
  - Average price of first 5 RN F31 frigates = $100,000 USD/long ton(FL)
  - Successful collaborative programs:
    - German Navy: F125 Frigate: tkMS and Luerssen;
    - German Navy: F126 Frigate: Luerssen and Damen
    - UK RN: F31 Frigate: BMI and OMT
    - Danish Navy: Frigate and Supply Ship common hull

Figure 7. Double Digit Improvements in Productivity, Cycle Reduction, and Production Throughput

Acquisition-Related Recommendations

The progress that has been made in acquisition strategy and design, construction, and sustainment process improvements have had significant impacts on naval shipbuilding programs, but this trend needs to be adopted more widely if the USN is going to satisfy current and future challenges while remaining compliant with budget constraints. Much of the success of European allies can be attributed to a long-term commitment by government and naval shipyard officials to designing and building more capable and cost competitive naval warships for both domestic and international markets (see Figure 8).

ACQUISITION RELATED RECOMMENDATIONS

- Foster Substantive Collaboration Between USN & US Shipbuilding Industry Officials Including WSI & key Naval Suppliers
- Form Group of Industry Experts to Accelerate Digital Transformation Across the USN enterprise based on best practices from related industries, e.g., aerospace, defense, utility
- Form Flag-Level Committee to Develop a Long-Term 50-year Naval Warship Design, Construction and Sustainment Strategy
- Form Indo-Pacific Naval Special Interest Group to Maximize Return on Total Investment and Synergy Among New Naval Shipbuilding Programs in the USA, Canada, Australia, Japan, South Korea, India, etc.
- Build career development program to develop future naval Ship Design and Program Managers focused on naval shipbuilding best practices, innovations and lessons learned

Figure 8. Long-Term Commitment to the Next Generation of Naval Warships

After reviewing the progress and results over the last decade during annual naval shipbuilding summits, it was concluded, as a **First Recommendation**, that USN shipbuilding programs could benefit from forming a panel of shipbuilding experts from government, shipbuilding industry, and academia to develop a long-term shipbuilding strategic plan to improve the affordability, adaptability, sustainability, and durability of future USN warships. This panel could provide balanced and continuous advice to USN leadership to ensure that there is a focused campaign to implement the best practices and lessons learned from the leading shipbuilders in the world as well as breakthroughs from adjacent, relevant industries and academia.

As a **Second Recommendation** from the shipbuilding summits is the implementation of appropriate and proven Information Technology for USN program management offices, U.S. shipyards, and principal suppliers in early-stage design and throughout the program life to synchronize knowledge management across the extended shipbuilding enterprise. It is evident that when advanced and proven Information Technology is implemented, it can have a significant impact on design and construction productivity and quality control throughout the entire build cycle, and moreover it can boost sustaining engineering and major upgrade efficiency and cycle reduction. Since the implementation of a comprehensive IT transformation can be very complex and time consuming, it is recommended that a small team of qualified experts be formed to review best practices and lessons learned from the implementation of IT enterprise-wide solutions at naval shipyards and related industries like aerospace and automotive to ensure the success of future U.S. shipyard implementations.

The preponderance of the findings and results of the first decade of Global Shipbuilding Executives Summits (GSES) have been from shipyards in Europe and the United States. Over the next decade, **Third Recommendation**, more focus should be made on Asian naval shipyards and in particular shipyards in Japan and Korea where it appears that relatively modest increases in design dimensions may have a much greater impact on construction efficiency and reduce total construction labor. Additionally, the incorporation of automotive production and quality best practices in these shipyards may have also shortened construction cycle time and boosted throughput. Both issues are crucial to the

U.S. naval shipbuilding industry as it prepares to launch a significant increase in naval shipbuilding.

The increase in U.S. naval shipbuilding for both traditional ship classes as well as some new and somewhat non-conventional designs will place enormous stress on the Navy–Industry collaborative design teams that will be tasked with managing these design and acquisition programs. It is imperative that a program, **Fourth Recommendation,** be implemented to accelerate the training and career development of the professionals who will be responsible for managing these programs over the next 30 to 40 years and avoiding costly mistakes of the recent past (Keane & Jaquith, 2021).

Finally, a dedicated team needs to be tasked to efficiently upgrade the existing, strategic USN fleet assets, like the family of DDG51 destroyers, if the USN is going to realize a **net gain** of 50 to 150 ships over the next 2 decades.

References

Axe, D. (2021, January). The U.S. Navy wasted a whole decade building bad ships. *Forbes Aerospace & Defense*. https://www.forbes.com/sites/davidaxe/2021/01/05/the-us-navy-wasted-a-whole-decade-building-bad-ships/

GAO. (2020, March). *Navy shipbuilding: Increasing focus on sustainment early in the acquisition process could save billions.* (GAO-20-2).

Jaquith, P. (2019, December). Asian vs. U.S. warship design, production engineering, and construction practice. *Naval Engineers Journal.*

Jaquith, P. (2020, December). Modern warship manufacturing practice: Impact on acquisition cost, schedule, and industrial mobilization. *Naval Engineers Journal.*

Keane, R. & Jaquith, P. (2021, March). Ship design and manufacturing professional development program: Teaching experience and judgment. *Naval Engineers Journal*.

Keane, R., Jaquith, P., Tibbitts, B., Mierzwicki, T., Hough, J., & Arcano, J. (2018, December). Rapid warship acquisition: A case for fundamental change in design and acquisition policy. *Naval Engineers Journal.*

Keane, R., Tibbitts, B. & Jaquith, P. (2019, September). The Navy's ship design factory: NAVSEA—The golden goose. *Naval Engineers Journal.*

PANEL 3. STRATEGIES FOR AWARDING CONTRACTS BETTER, FASTER, AND CHEAPER

Tuesday, May 11, 2021	
9:15 a.m. – 10:30 a.m.	**Chair: Brig Gen Alice Trevino, USAF,** Commander, Air Force Installation Contracting Center ***How Long Does it Take to Award a Government Contract? Understanding PALT timeframes with Big Data Analytics*** David Gill, Internal Revenue Service Tim Hawkins, University of North Texas ***Why Marketing Matters: Strengthening the Defense Supplier Base Through Better Communication with Industry*** Amanda Bresler, PW Communications Alex Bresler, PW Communications ***Non-Competitive Contracting: Lessons from Contracting Personnel*** Latika Hartmann, Naval Postgraduate School Rene Rendon, Naval Postgraduate School Joshua D. Cissell, CAPT, USAF, F-22 Modernization/F-119 Sustainment

Brig. Gen. Alice W. Trevino, USAF—serves as Commander, Air Force Installation Contracting Center, Air Force Installation and Mission Support Center, Air Force Materiel Command, Wright-Patterson Air Force Base, Ohio. She leads over 750 operational acquisition professionals responsible for a $55 billion contract portfolio. In this capacity, she directs enterprise-wide installation strategic sourcing efforts for the Air Force and oversees $9.1 billion in annual obligations in mission and installation requirements. Her contracting authority extends worldwide across AFICC in support of nine major commands, 82 units and the United States Air Force Academy. Additionally, she is designated as the Commander of a Joint Theater Support Contracting Command upon activation. She also directs the contract execution in support of the Defense Technical Information Center, Air Force Medical Readiness Agency and Air Force Civil Engineer Center.

Brig. Gen. Trevino received her commission from the U.S. Air Force Academy in 1993. She is a joint qualified officer and has deployed extensively in support of combat, humanitarian assistance and peace-keeping enforcement operations to Croatia, Turkey, Oman, Kuwait and Afghanistan. Prior to her current position, Brig. Gen. Trevino was the Principal Military Assistant to the Deputy Secretary of Defense. In addition, she has served as a Warranted Procuring Contracting Officer for major defense acquisition programs and was the Senior Contracting Official for a 365-day tour in Afghanistan. She has commanded contracting squadrons at the base and major command specialized levels; and joint units at both the group and wing levels.

How Long Does It Take to Award a Government Contract? Understanding PALT Time Frames with Big Data Analytics

David I. Gill—Internal Revenue Service [david.gill@irs.gov]

Timothy G. Hawkins, PhD—Data and Analytic Solutions, Inc. [Hawkins@DASconsultants.com]

Abstract

Awarding federal contracts is perceived as an excessively lengthy process. The purpose of this research is threefold: (1) to understand the drivers of procurement administrative lead time (PALT), (2) to identify opportunities to reduce PALT, and (3) to predict when specific requirements are likely to be awarded. These analyses will be performed using newly available, government-wide data for over 5 million federal contracts.

Keywords: Contracting, Procurement Administrative Lead Time, Big Data Analytics, PALT, Predictive Modeling, Machine Learning, Data Visualization, Time to Contract Award

Introduction

Half a trillion dollars is spent annually on government contracts that are mission critical for performing all functions of government. The vision of the acquisition system is "to deliver on a timely basis the best value product or service to the customer" (FAR 1.1020). Even in the government, time is money. The average transaction cost of a formal source selection has been estimated at $245,000 (Hawkins et al., 2016). Given the enormous quantity of contract actions across the federal government, the cost in man-hours is enormous. Of course, more time taken to award a contract usually translates to delays to internal requiring activities that rely on the work products of contractors to help meet mission needs. "Complaints of excessive PALT continue to plague the acquisition system and present challenges to both government and industry" (Berteau, 2018). The federal government is addressing procurement administrative lead time (PALT) via the President's Management Agenda. Therein, a cross-agency priority (CAP) goal called "frictionless acquisition" seeks to, among other things, deliver commercial items at the same speed as the commercial marketplace (U.S., 2021).

Nevertheless, "understanding procurement cycle time is sometimes difficult because organizational buyer behavior processes are often dynamic and complicated" (Hult, 1997, p. 403). Understanding PALT is necessary in order to muster and assign the necessary amounts and types of resources to complete required tasks. Once PALT is better understood, managing PALT is needed to reengineer processes that consume PALT and to prevent instances in which PALT exceeds reasonable bounds.

In the context of a supply chain of physical goods, the importance of procurement cycle time cannot be overstated. The government operates numerous and varied instances of such supply chains. For example, the Bureau of Engraving and Printing runs a manufacturing operation to produce currency. The military departments each operate multiple depots wherein weapon systems are overhauled and repaired. The Defense Logistics Agency serves as an inventory control point for military systems. The Department of Veterans Affairs and the military departments operate multiple hospitals and clinics that rely upon the availability of medical supplies and operate pharmacies stocked with inventory. These supply chains rely upon proper inventory management to ensure needed supplies are on hand yet minimize inventory carrying costs. Forecast accuracy partly depends on the planning time horizon. Longer planning horizons caused by longer procurement cycles can increase forecast error resulting in either excess inventory (i.e., inventory carrying costs) or stockouts (i.e., service failures), and therefore also increase safety stock levels (LeSueur & Dale, 1997). Longer lead times also result in larger cycle stock.

Despite early calls by scholars and business leaders in procurement (now referred to as supply management) in forums such as the (then) Center for Advanced Purchasing Studies (CAPS) to reduce procurement transaction costs and purchasing cycle time (Carter & Narasimhan, 1996), little progress has been made in the federal sector.

Some research has explored the antecedents of PALT. Significant factors have emerged such as dollar value of the contract action, type of goods and services, number of offers, number of evaluation criteria, contract type, and source selection method. However, research has been constrained by the unavailability of data at the transaction level (i.e., contract action) rendering models based on limited variables. Several potential predictors have not been explored such as: (1) the time remaining until the end of the fiscal year (i.e., funds availability time), (2) type of set-aside program, (3) orders against existing contracts (i.e., solicitation procedures), (4) interagency contracting, (5) buyer workload, (6) requirements returned to requiring activities (due to omissions, errors, or unresolved issues), (7) type of appropriation (i.e., one-year versus multiple-year funds), (8) buying agency, (9) buying activity, (10) option periods or quantities, (11) number of contract line items, (12) government furnished property, (13) narrative description, (14) contract consolidation or bundling, (15) the formality and rigor of trade-offs applied to task order awards, (16) combined synopsis/solicitations, and mandatory sources of supply (e.g., Ability One and FPI), to name a few. Several models have also been based on small sample sizes with low statistical power. Furthermore, models have been developed in limited contexts such as a few buying activities of only one federal agency. Research also rarely reports a comparison of adjusted R^2 to predicted R^2 and fails to report prediction intervals; thus, we don't know how accurate the estimated models are.

When contracts will be awarded is of significant interest to contracting officers, program offices, and vendors alike. The date a requisition turns into a signed contract is the culmination of the pre-award acquisition process. FAR 7.105 emphasizes the importance of identifying schedule "constraints," "risks," and identifying key "milestones" in the pre-award acquisition process on the way to contract award. Typically, acquisition plans include milestone schedules developed manually by the contracting officer. Award dates are projected without statistical rigor, and the accuracy of award date projections is rarely assessed.

Meanwhile, in federal procurement, data is increasingly collected and made available publicly. Yet this vast and numerous contract award data has not been analyzed in order to build machine learning models that can be trained and result in improved predictive accuracy. Therefore, the purpose of this research is to explore new features (i.e., predictors) of PALT and to utilize them in machine learning models to more accurately predict PALT. These predictors can then be used to generate milestone schedule estimates informing customers when their contract is likely to be awarded. The research questions are as follows:

RQ1: What are the significant unexplored features (predictors) of PALT?

RQ2: Can machine learning models be applied to reliably and accurately predict when a contract action will be awarded?

The remainder of this research is organized as follows. It begins with a review of the relevant literature surrounding PALT, both in the for-profit and not-for-profit sectors. Next, the study presents the methodologies of quantitative data collection and analysis to explore the research questions. Lastly, discussion, limitations, implications, future research directions, and conclusions are offered.

Literature Review

Factors affecting PALT in a government context have been studied, but not extensively. Several early attempts to explore PALT were conducted by graduate students at the Air Force Institute of Technology and the Naval Postgraduate School in the late 1980s and early-to-mid 1990s. However, these early studies predated the explosion of information technology and major changes in federal contracting processes such as the Federal Acquisition Streamlining Act of 1994 and the Federal Acquisition Reform Act of 1995 that instituted PALT-reducing measures such as multiple-award contracts and commercial item procedures.

MacKinnon (1992) found relationships between PALT and contract type, dollar value, and type of purchase (supply, service, or research and development) using a regression model of 559 contract awards by the Naval Air Warfare Center Weapons Division at China Lake, CA. Cost reimbursement contracts are associated with lower PALT. Additionally, contracts for research and development are awarded faster than other types of requirements. Contracts for supplies consumed more time than others (e.g., for services). Contracts for larger dollar values are associated with longer PALT. However, the extent of competition did not impact PALT. MacKinnon (1992) concluded that, due to complexity, it is difficult to accurately predict PALT.

Ng et al. (1997), in their literature review of cycle time, identified several factors associated with procurement cycle time such as electronic commerce, automated reordering, and several practices associated with supplier alliances. Most of the practices pertained to partnering with suppliers; thus, they could only apply to orders once a supplier is selected. Examples included: increased frequency of buyer review of manufacturing schedule and internal requirements, supplier TQM involvement, sharing information, just-in-time ordering, and early supplier involvement in design. In a government context, Ng et al. (1997) said practices could be implemented once a supplier is on-contract. Frequent competition and supplier switching would render these practices impractical.

Lamoureux et al. (2015) examined contract awards ($n = 33$) from two U.S. Air Force installations in Colorado using data manually extracted from contract files (due to the limitations of FPDS-NG data). Using Multiple Analysis of Covariance, they explored whether characteristics of the source selection were associated with PALT and with contractor performance ratings (i.e., contractor performance assessment reports—CPARs). Their definition of PALT encompassed the time from receipt of the requisition to the time of contract award. They found the number of evaluation factors and the number of offers received have a significant effect on increased PALT, accounting for 62.7% of the variance in PALT. The source selection method (i.e., low-price, technically acceptable versus full trade-off) and the number of internal reviews did not affect PALT.

Landale et al. (2017) also explored predictors of PALT and contractor performance. Notably, PALT, in this study, encompassed the time from receipt of a requirement in contracting to contract award. Using a sample of 124 U.S. Air Force and U.S. Navy contracts and controlling for the effect of dollar value, PALT was found to be positively related to the number of offers and to the number of evaluation criteria. The full trade-off source selection method was found to be a marginally significant predictor of PALT ($p < .10$) showing a moderate effect size increasing PALT. The "average [PALT] was approximately 36 percent longer for the [trade-off] supplier selection method than for source selections using an LPTA approach" (Landale et al., 2017, p. 60). Also, the research revealed that an increase of one evaluation factor increased PALT by 28%. Furthermore, the study found that a 10% increase in the number of offers resulted in a 1.9% increase in PALT.

Chung et al. (2018) explored the effects of several antecedents on PALT, but only for U.S. Air Force sole source major systems acquisitions exceeding $500 million (n = 26). Factors found to increase PALT included undefinitized contract actions (UCA) (i.e., the time to definitize a UCA such as a letter contract), the number of major subcontractors and corporate transfers, foreign military sales, and the type of weapon system acquired (bombers and fighters). Other factors decrease PALT such as award to a non-profit (for research and development efforts) and the type of goods (i.e., buying armaments). Notably, several factors had no effect on PALT including price, proposal quality (operationalized as the time of initial proposal minus the time of adequate proposal = 0), aggressiveness of the government's negotiation position (contractor proposal—government objective)/contractor proposal x 100), the number of internal approvals for price being too high, and whether cost or pricing data was available on a previous acquisition.

Some benchmark studies on procurement metrics of for-profit-sector firms provide insights as to the realm of possible PALT. Zycus's (2014) Purchase to Pay Benchmarking study (n = 450+) showed that the average time from requisition to order was 4.6 days for "simple" requirements, 14.3 days for "complex" requirements, and 13 days for "services." However, these categories were not defined; hence, it is unknown what renders a requisition simple or complex. The Center for Advance Procurement Strategy (CAPS Research, 2011) published benchmark metrics in 2011 showing average cycle times across 10 industries. They measured the time from requisition approval to purchase order for both direct goods and indirect goods. The averages for direct goods ranged from 1.52 days to 50.75 days (average 11.75 days). The averages for indirect goods ranged from 2.04 days to 12 days (average 6.36 days). These cycle times are drastically shorter than those prescribed by the various federal agencies.

While efficiency is important, having sufficient PALT is also necessary. The perceived sufficiency of planned PALT (defined as the extent to which the buyer believed he or she had enough time to conduct a proper source selection process) has been shown to improve the sufficiency of the requirement definition, which, in turn, yielded higher service quality ultimately delivered by the contractor (Hawkins et al., 2015). PALT is also important in ensuring compliance with the myriad of laws, regulations, and policies in a federal contracting context. Hawkins et al. (2014) found a positive relationship between the perceived sufficiency of PALT and the perceived level of compliance of contracts. Having sufficient PALT has also been associated with bid protests. A study by Hawkins et al. (2016) showed that sufficient planned PALT reduced the fear of a bid protest. Fear of protest, in turn, increases added PALT (Hawkins et al., 2016).

Methodology

Multiple data sets were analyzed. Contract award data for all federal agencies from the USASpending.gov was used to better understand actual PALT. Newly available government-wide data on PALT time frames provides a large dataset that can be explored for answers to the research questions. PALT data collection in FPDS-NG began in Fiscal Year (FY) 2018. A total of more than 5 million contract actions were compiled for analysis covering the time period from FY2018 to FY2020. The OFPP recently declared its formal definition of PALT as a response to a requirement of Section 878 of the National Defense Authorization Act for 2019, Public Law 115–232. The OFPP's definition of PALT measures a subset of the overall acquisition life cycle, including only "the time between the date on which an initial solicitation or a contract or order is issued by a Federal department or agency and the date of the award of the contract or order" (Wooten, 2020, p. 3429).

Separately, data on shopping carts (i.e., requisitions) from the IRS's Procurement for the Public Sector (PPS) system were also used to construct a prediction model of acquisition lead time, given the minimal characteristics of the requirement known prior to transmission to a contracting activity (dollar amount, workload of assigned contract specialist, and days remaining in the fiscal year).

The remainder of the methodology is organized as follows. First, we describe the method used for the explanatory model. Next, we describe how NLP of the contract award description data was used to enhance the explanatory model. Finally, we describe the shopping cart prediction model.

Explaining PALT with USA Spending Data

We look to explain the number of PALT days using variables available in the USASpending.gov data. We identified 11 data fields present among the data that are relevant to PALT: days remaining in the FY, month of solicitation, number of offers received, NAICS code, dollar value (base plus options), civilian agency, parent award type, small business or other status, solicitation procedure, type of contract, and assisted acquisition.

The scope of the analysis omits contract modifications. The data included contract actions awarded in FY2020. In total, a random sample of 50,000 contract action awards were included. We applied a Random Forest regression model (provided through the randomForest R package) due to its ease of use and performance over other models. Random Forest models are trained using three as the 'mtry,' the number of predictors randomly sampled at each split when creating tree models. All other hyperparameters use default values.

Natural Language Processing of USA Spending Contract Award Description Field

The award description field is one of the few fields in USASpending.gov data that provides contextual information about what service or product is being procured. These descriptions can be key in defining the scope of work, nature of product, or complexity of service that is being procured. To limit the scope, only data on service contracts was analyzed. This analysis explored whether descriptions which are similar in context or share similar words will likely define similar scopes of work and therefore have similar PALT times.

First, re-interpreting the text data was necessary. The chosen method was through text pre-processing, vectorization, Latent Dirichlet Allocation Analysis, and token analysis. We re-interpreted the data in a machine-passable way. This is the role of text preprocessing. In this stage, we simplified the text data. The text preprocessing techniques involved trimming out non-alpha characters, ensuring all documents were entirely lowercase, removing stop-words (i.e., removing: infrequent words fewer than 250 occurrences, highly frequent words with higher than a 25% prevalence, and non-indicative words), and stemming (i.e., reducing words down to their stem). Using the text2vec library, there are two stages for pre-processing. The first is simply called "preprocessing" and it applies a string manipulation function to each entry. Then, the following "tokenization" step applies a string manipulation function to each word in each entry. The resulting tokens are the most common word-stems.

The encodings that we used to indicate whether an award contains a particular token is similar to a one-hot-encoding. Each of the tokens found in the dataset, of which there are 564, becomes a column. Then for each award row, the value in that column indicates how many times the token appears in the award description.

Then, we constructed a generalized linear model (GLM) to assess the “importance” of each token for predicting PALT. The GLM provided coefficients (weights) for the various token columns, and we can assess these coefficients to see whether a token indicates an increase or a decrease in PALT times based on whether it is positive or negative. Table 1 displays a list of the 10 highest importance tokens from a GLM. It includes their coefficients from the linear model to allow investigation of which tokens have positive or negative weightings. The Variance Importance column helps to indicate how “important” a variable is for predicting PALT. In order to get a better idea of how different tokens affect the PALT of a contract action, we use the mean PALT days for all awards that contain that token.

Table 1. Token Importance

Token	Count	Coefficient	Variance Importance	Mean PALT Days
idiq	3452	25.8	12.6	137.4
tuition	736	-90.2	11.2	9.4
macc	443	147.0	28.1	286.2
ae	1963	18.4	7.1	119.3
protect	895	45.2	12.1	143.9
guarante	660	-58.6	7.6	196.8
uss	1563	-35.2	9.7	38.2
repair	9240	-12.4	9.8	70.5
report	1545	27.7	8.4	178.4
express	740	132.2	23.8	266.0

To conduct topic analysis, we used an algorithm called Latent Dirichlet Allocation (LDA). It essentially looks at each document and creates a set of N topics based on the co-occurrences of words. For example, if the words *mechanic* and *vehicle* often end up in the same award description, they are likely to be grouped together once the LDA has been run. In this case, we are defining N = 10. This value can be manipulated based on the analysts’ initial belief of the number of underlying topics. In our case, 10 topics yielded the best results.

In this visualization (Figure 1), on the right we can see the “salience” of each term; salience is the extent a term is about this specific topic. These are the terms that appear the most frequently in concurrence with other words. As shown, the words *base* and *task* are near the top where we expect them to be. The hope is that we can create a distribution for each award, where we can see the likelihood that it belongs to each cluster. That likelihood, we hypothesize, will help in grouping similar awards and hopefully with predicting PALT.

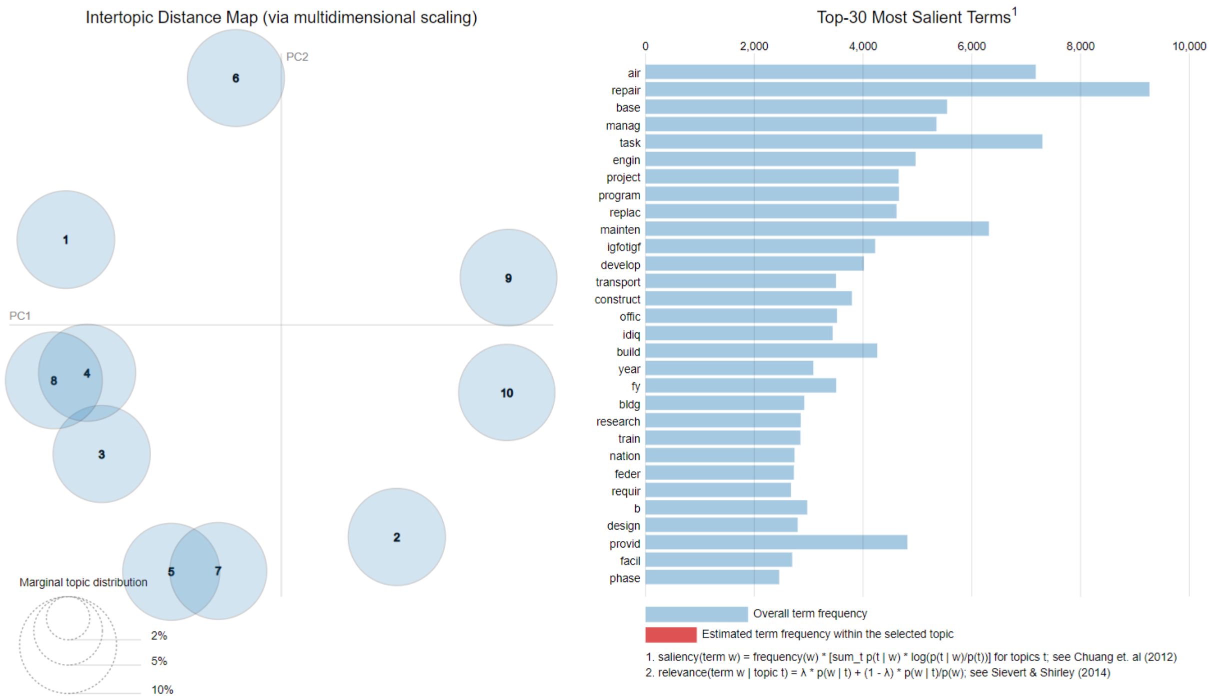


Figure 1. Topic Salience

Table 2 represents the topic distributions. Each row represents an award, and the columns indicate the likelihood for each topic. For example, the V1 column contains likelihoods of an award belonging to topic 1. Now, we can pass these into a GLM and assess the value of using these topic distributions to predict PALT.

Table 2. Topic Distributions.

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
1	0.000000000	0.00000000	0.00000000	0.000000000	0.00000000	0.000000000	1.00000000	0.00000000	0.000000000	0.000000000
2	0.000000000	0.00000000	0.02500000	0.000000000	0.00000000	0.500000000	0.12500000	0.00000000	0.225000000	0.125000000
3	0.000000000	0.00000000	0.00000000	0.000000000	0.00000000	0.000000000	0.00000000	0.00000000	1.000000000	0.000000000
4	0.100000000	0.00000000	0.60000000	0.000000000	0.00000000	0.000000000	0.00000000	0.00000000	0.000000000	0.300000000
5	0.000000000	0.00000000	0.00000000	0.000000000	0.50000000	0.000000000	0.00000000	0.00000000	0.000000000	0.500000000
6	0.000000000	0.33333333	0.03333333	0.000000000	0.00000000	0.300000000	0.00000000	0.33333333	0.000000000	0.000000000
7	0.000000000	0.00000000	0.00000000	0.000000000	0.00000000	0.000000000	0.00000000	0.05000000	0.950000000	0.000000000

The results of the GLM analysis are shown in Table 3. In this table, the Topic column is our interpretation of what a topic might be, based on some of the most salient words provided by the LDA.

Table 3. GLM Analysis of Topics

Topic	Sample Salient Words	Coefficient	Variance Importance
1. Planning and Logistics	Nation, require, plan, study, logistic, assess	78.712	39.0
2. Program Development and Management	Program, manage, develop, office, technical	22.614	11.4
3. Base Awards	Base, year, phase, period, idiq	22.423	10.5
4. Facility Operations	Center, operate, engine, train, facility	24.069	12.8
5. Equipment	Provide, install, medic, labor, equip	-4.9642	2.5
6. Construction and Project Design	Project, construct, design, integrate	4.7078	2.2
7. Construction	Repair, replace, build, water, roof	25.467	13.2
8. New Task Orders	Task, purpose, new, report, nurse, bpa	-10.221	5.2
9. Transportation and Maintenance	Air, igfotigf, transport, repair, test	64.491	32.3
10. General Maintenance	Maintenance, federal, software, supplies	-10.756	6.0

Predicting Contract Award Dates with IRS Shopping Cart Model

We look to predict the number of days until award for a new contract action. Rather than PALT, this model considers the time from the approval of a requisition (i.e., a shopping cart) to contract award, since the IRS dataset we are analyzing does not capture the solicitation issue date (RFX date). The scope of the analysis omits contract modifications. We ensure that contracts have a non-zero obligated value and are of the "Base Award" action type. The model is trained on all IRS obligated awards in FY2020 as of September 30, 2020. Also, the model is deliberately limited to use only data elements that are also available on a new (open) requisition. Many desirable data elements (e.g., contract type, solicitation procedures, etc.) may be unavailable or not yet decided upon early in the acquisition process.

We considered various machine learning models such as generalized linear regression and XGBoost, and settled on using a Random Forest regression model (provided through the randomForest R package) due to its ease of use and performance over the other models. Random Forest models are trained using nine as the 'mtry,' the number of predictors randomly sampled at each split when creating tree models. All other hyperparameters use default values.

We chose features based on which data fields are available in both the IRS's open and obligated ALT reports and provide information relevant to the time of shopping cart award date. We identified 12 data fields present among the data that are relevant to days in procurement. We also created four additional features from these data fields to provide the model with more data on the contract time frame and workload of the contract specialist handling the contract award. In total, the 16 features we use are as follows: Contract Specialist (CS) Section, CS.Branch, CS.Office, CS.Division, Agency, Fiscal.Year, Fund.Expiration, total_Shopping Carts (SC)_completed_at_approval, Funding.Business Unit (BU), workload_proportion, Obligated, Fund, days_until_FY_end, Functional.Area, SC.PM.Approval date, and current_CS_workload.

The four features we created evaluate the number of days until the fiscal year ends when the contract is approved (days_until_FY_end), the current number of contracts assigned to the contract specialist (CS) overseeing the contract (current_CS_workload), the total number of contracts the CS has completed in the last 90 days (total_SCs_completed_at_approval), and the proportion of their current workload to the amount they have completed in the last 90 days (workload_proportion).

Results

The data were analyzed using a combination of several methods—various data visualizations and machine learning models. Summarizing PALT by agency (see Figure 2) shows significant differences in typical time frames and the distribution of PALT for specific contract awards. The following chart sorts agencies by FY2020 PALT time frames with faster agencies appearing at the top. The top five agencies for short PALT time frames are the Small Business Administration, Department of Labor, Department of the Treasury, Department of Agriculture, and the Social Security Administration. Also notable is that the DoD reported over 173,000 and the General Services Administration reported over 126,000 PALT time frames in FY2020—a larger volume of awards than all other agencies combined.

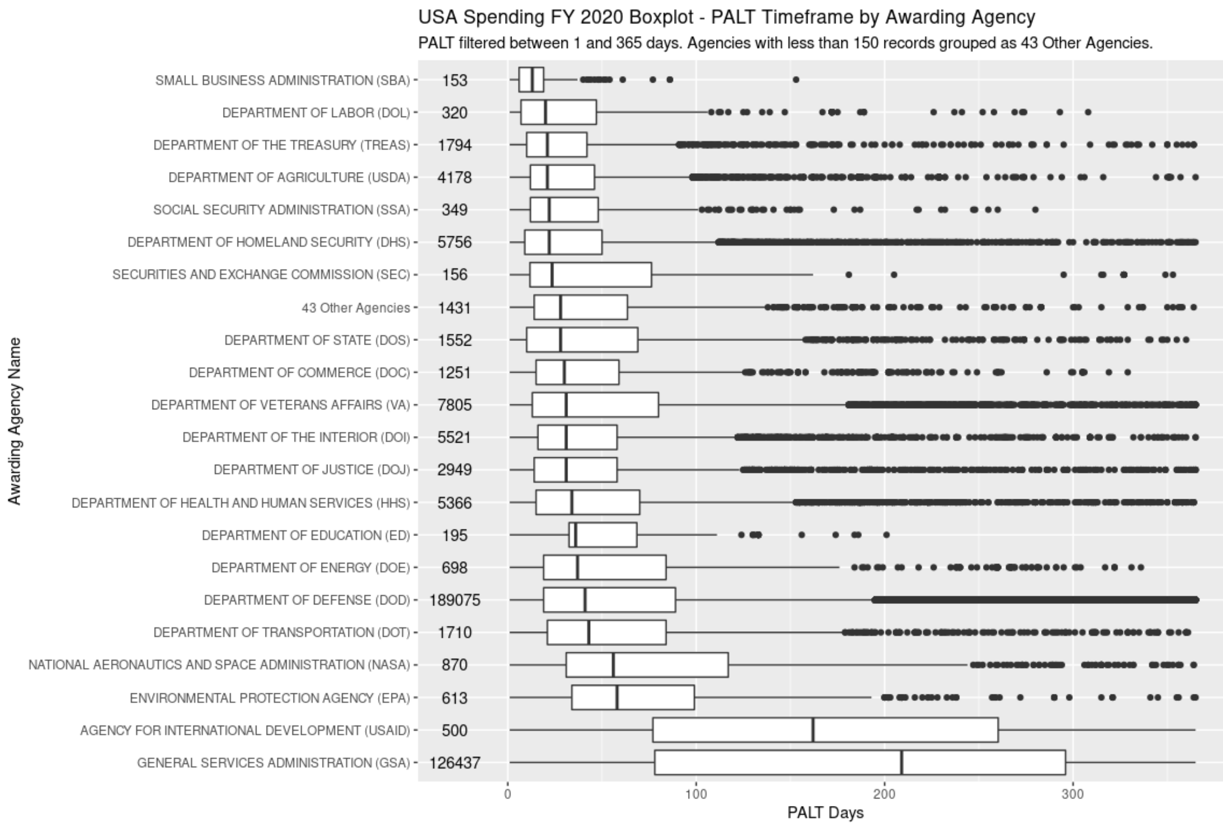


Figure 2. USA Spending FY2020 Boxplot—PALT Time Frame by Awarding Agency

USA Spending Model Explaining PALT Results

A supervised ML model (see Figure 3) was used to train the model on the data. The training data included input data and response values (i.e., PALT days). The algorithm used was a regression Random Forest model, suitable for determining quantities. Ten decision trees were used, and three variables were tried at each split.

Acquisition traits (FPDS data elements) were statistically ranked in descending order of importance as drivers of PALT time. ML models enable understanding PALT time drivers with statistical learning. The model had an explained variance of 92% and a mean of squared error of 3206.701.

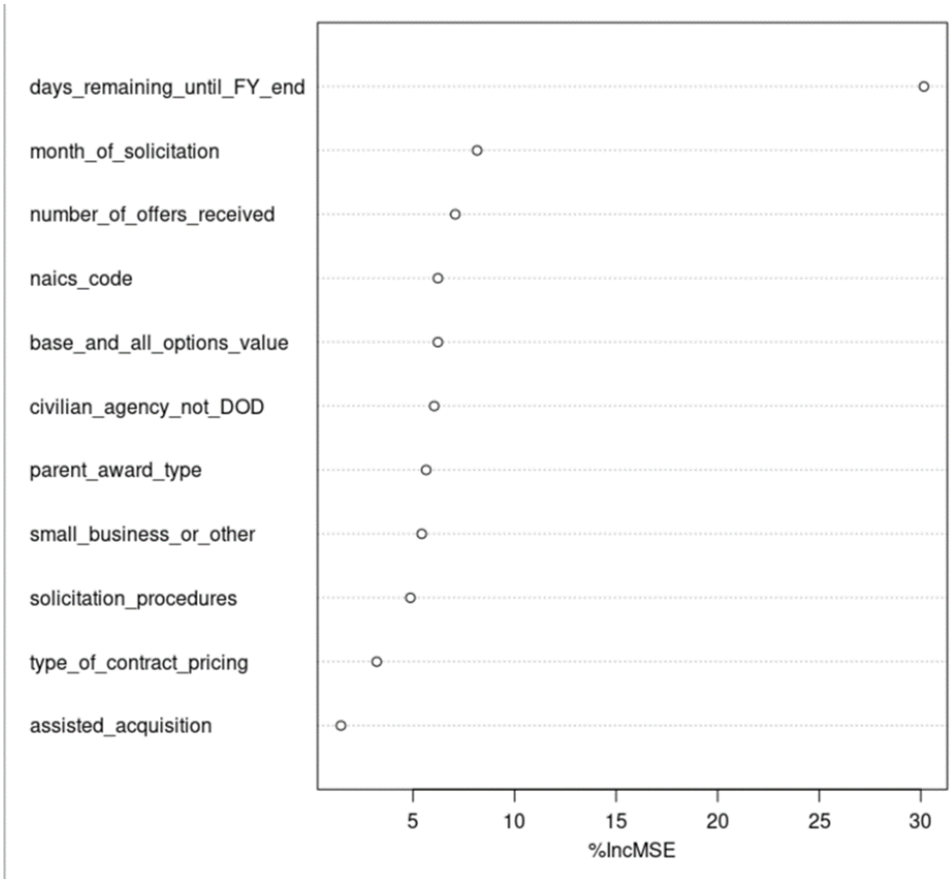


Figure 3. USA Spending PALT Explanation Model Results

- **USA Spending Natural Language Processing of the Contract Award Descriptions Results**

In order to evaluate whether or not our encodings provide us with a more accurate PALT prediction, we appended the encodings features onto our base GLM features. Now, each award is represented by the base GLM features, as well as a series of binary columns that indicate whether or not the award contains any of the most indicative tokens in its description.

This model (see Figure 4) had an explained variance of 72.32% and a mean of squared error of 4250.511. The following table shows us that the most important feature is days_remaining_until_FY_end because it has the highest %IncMSE. This (%IncMSE) can be interpreted as the projected loss in accuracy if the feature is omitted.

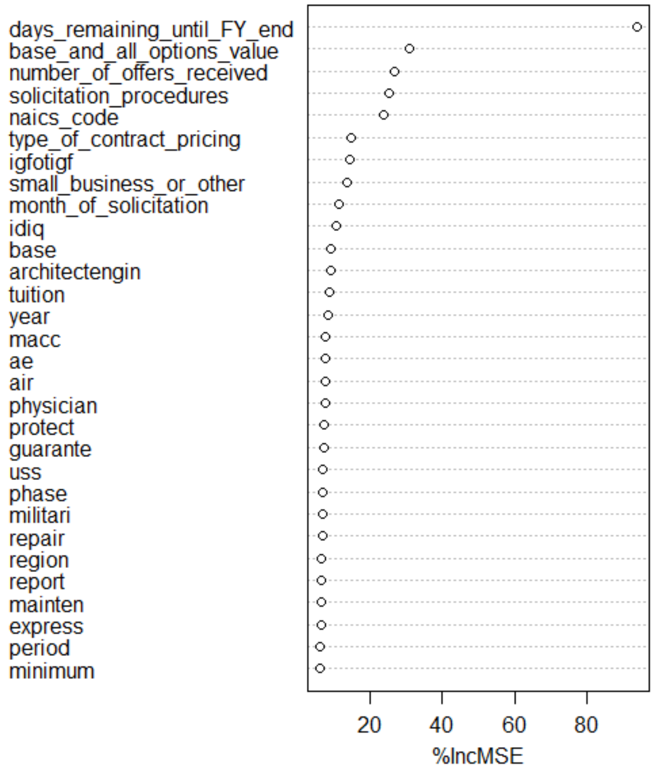


Figure 4. USA Spending Random Forest Model—NLP Encodings

The topic analysis model shown in Figure 5 had an explained variance of 69.67% and a mean squared error of 4654.5.

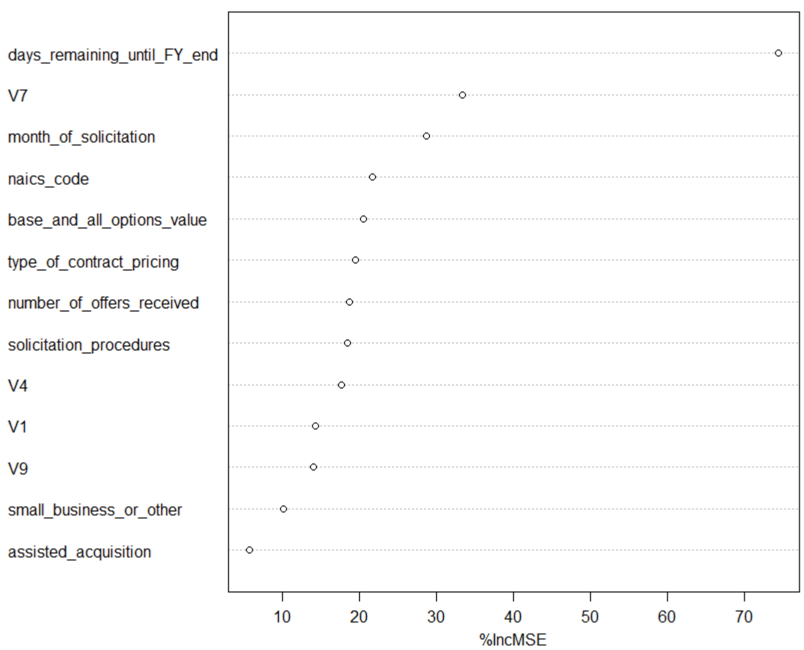


Figure 5. USA Spending Random Forest Model—NLP Topics

Figure 6 shows the token distributions of the two most informative topic features, topics 1 and 7.

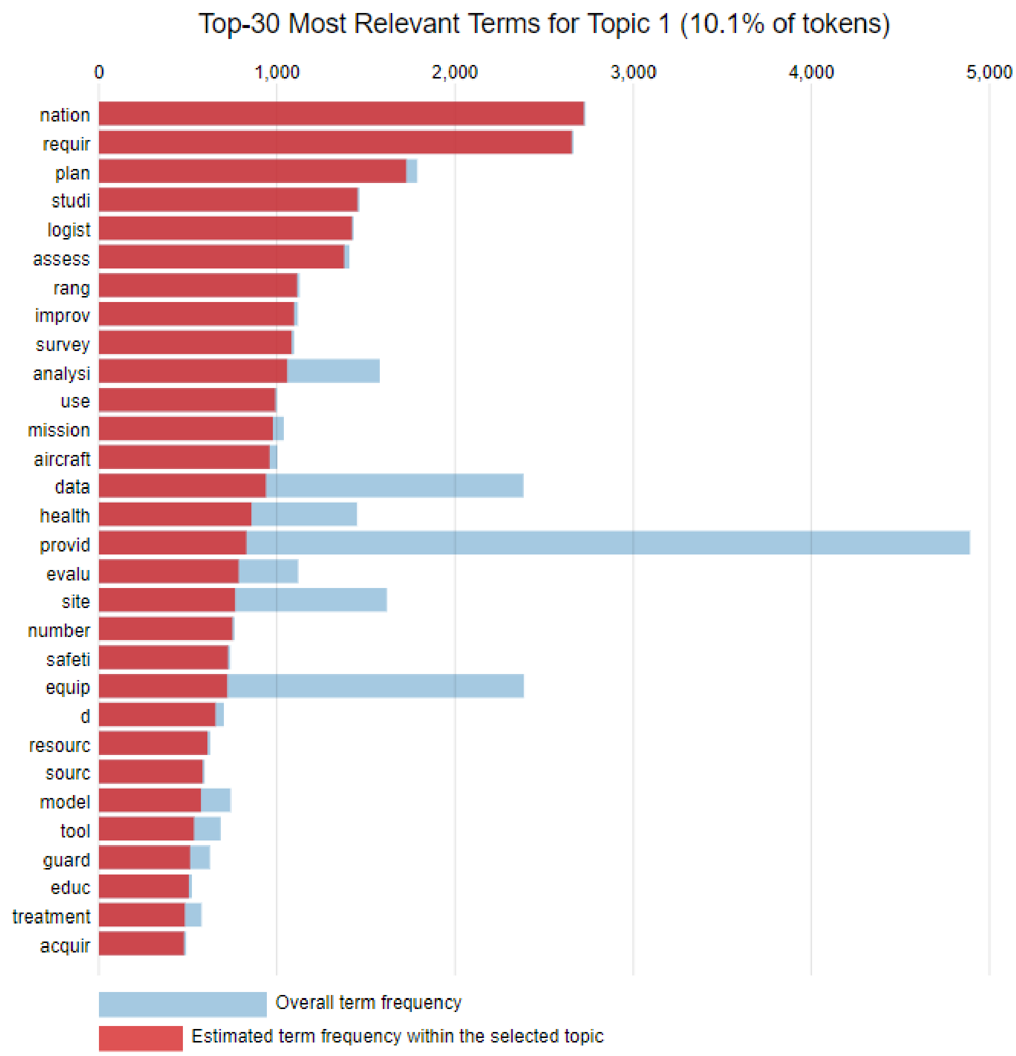


Figure 6. USA Spending—Top-30 Most Relevant NLP Terms for Topic 1

- **IRS Shopping Cart Model Predicting Contract Award Dates Results**

When evaluating our model for feature importance, we find that the number of days until the fiscal year end (days_until_FY_end), the current number of contracts assigned to the CS overseeing the contract (current_CS_workload), and the functional area (Functional.Area) of the contract have the largest impact on model performance. The ordered list of feature importance by the percent increase in MSE when values of a feature are shuffled and the increase in node purity are plotted in Figure 7. We provide hex plots comparing the counts of the pairings of both of the continuous features and a trend line showing how the number of days of award trends with a change in value of these features in Figure 8.

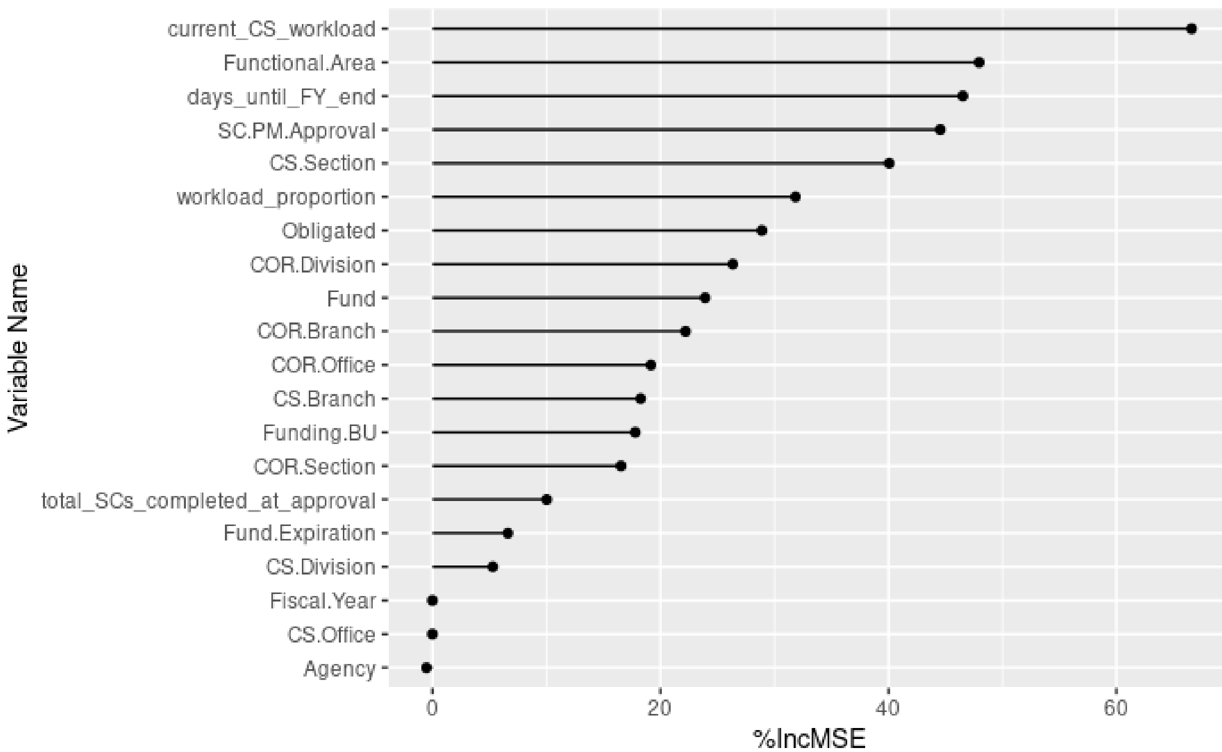
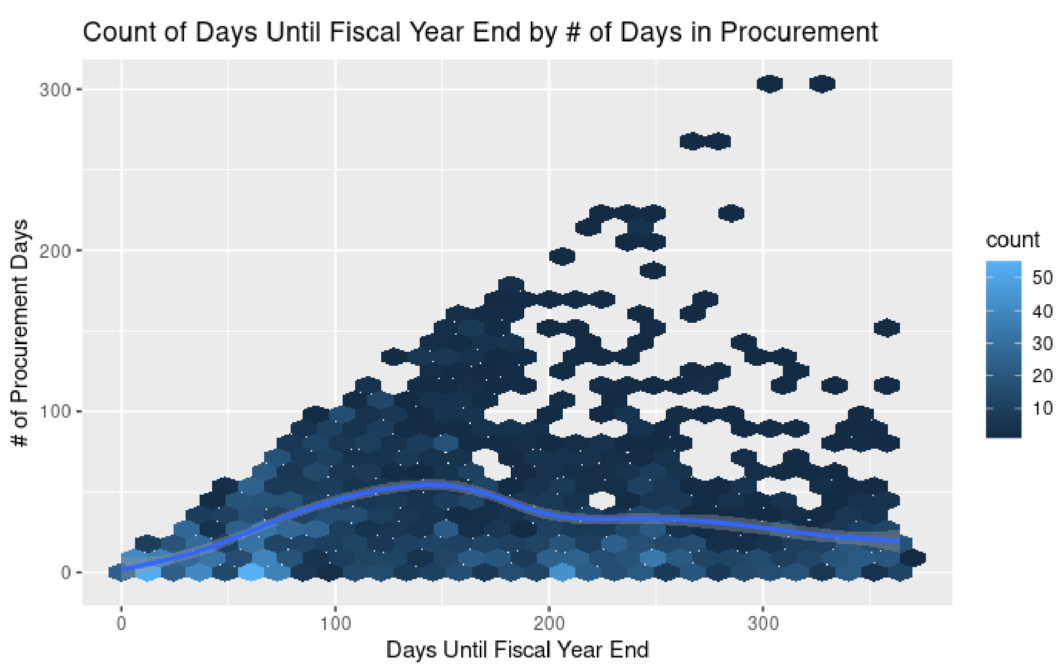


Figure 7. IRS Shopping Cart Random Forest Model



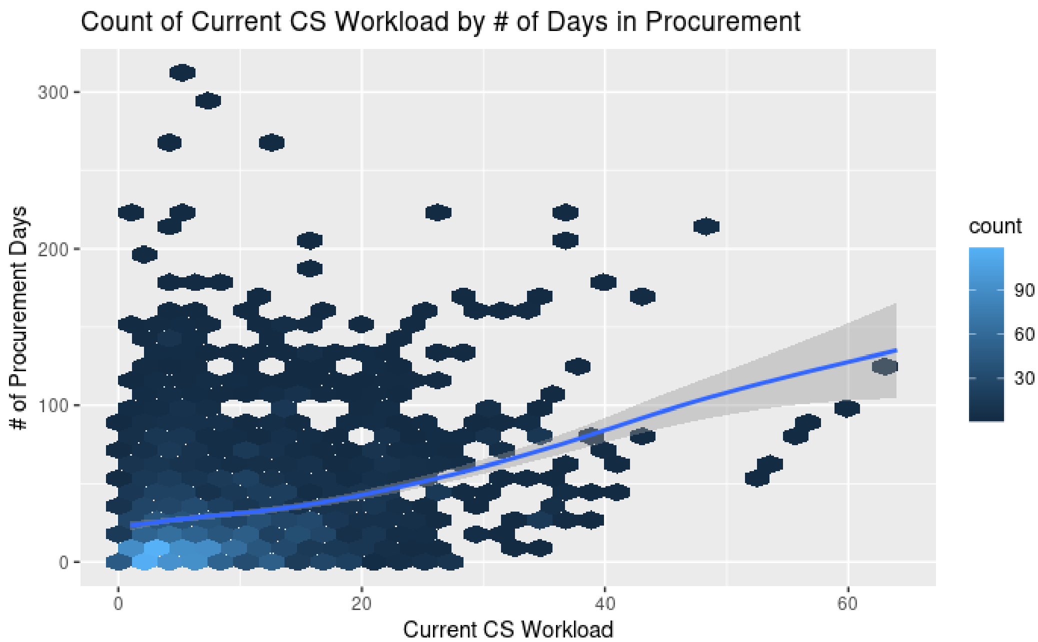


Figure 8. IRS Shopping Cart Hex Plots of Features

- **IRS Shopping Cart Model Evaluation and Tests of Validity**

We evaluated the contract award date Random Forest model using 4-fold cross validation on the IRS's obligation PALT data sheet available in the internal September 30, 2020, PALT report. This sheet contains 2,960 observations of contract awards completed in FY2020. First, we divided these 2,960 observations into four independent subsets of 740 observations (25%) each. Then, we trained a contract award date model on each permutation of three of the four sets and evaluated their performance on the fourth test set. We report the average Mean Squared Error (MSE), Root-Mean-Squared-Error (RMSE), Mean-Absolute-Error (MAE), and R^2 value of these models in Table 4. We also provided contract prediction time frame metrics to better understand the range of time our predictions match the actual values. We provided the percent of contracts within a +/- 30-day range, same working month, a +/- 7-day range, and the same working week in Table 4. When training a model on the entire dataset, we find the out-of-bag (OOB) MSE comes to 541.9 with an R^2 value of .606.

Table 4. IRS Shopping Cart Model Evaluation

Evaluation Metric	Value		Evaluation Metric	Value
Mean Squared Error (MSE) days	571.1		Percent within +/- 30 days	86.75%
Root-Mean-Squared-Error (RMSE) days	23.8		Percent within same month	61.62%
Mean-Absolute-Error (MAE) days	14.5		Percent within +/- 7 days	44.32%
R^2	0. 563		Percent within same week	24. 19%

Discussion

Recently, government and industry leaders have expressed a need to accelerate the procurement process. Source selections consume a significant amount of time and, thus, transaction costs in terms of man-hours. Perhaps more importantly, agencies are delayed in executing their missions. Therefore, the purpose of this research was to explore new features (i.e., predictors) of PALT and to utilize them in machine learning models to more accurately predict PALT.

Several findings emerged from the analyses. First, the number of days until the fiscal year end (days_until_FY_end), the current number of contracts assigned to the Contract Specialist (CS) overseeing the contract (current_CS_workload), and the functional area (Functional.Area) of the contract have the largest impact on the PALT prediction model performance. Next, the Small Business Administration, Department of Labor, Department of the Treasury, Department of Agriculture, and Department of Homeland Security have the lowest PALTs. Several variables seem to affect the amount of PALT including days remaining in the fiscal year, obligation value, number of offers, NAICS code, solicitation procedures, and contract type. Additionally, the award date of requisitions can reasonably be predicted within the same month 61% of the time and within seven days 44% of the time. We also found that tokens that appear in the award description field in FPDS that are useful in explaining PALT included: idiq, base, architectengin, tuition, year, macc, ae, air, physician, protect, guarante, uss, phase, militari, repair, region, report, mainten, express, period, and minimum.

Lastly, useful topics included: Topic clusters V4 "Facility Operations," V1 "Planning and Logistics, and V9 "Transportation and Maintenance," which appear to affect PALT. Refer to Table 3 (GLM Analysis of Topics) for further information about topic clusters. Interestingly, these topic clusters usurped some characteristics of the procurement in importance (e.g., business size/type and assisted acquisition).

Managerial Implications

From the results, several recommendations for addressing PALT are made.

- Make the prediction model available to customers, enabling them to more accurately forecast needs and when those needs will be fulfilled.
- Using the results, consider adjusting the IRS's PALT standards by redefining categories of shopping carts/requirements by solicitation procedure, competition, dollar value, and type of goods/services with commensurate PALT goals.
- Consider expanding the definition of PALT to reflect the time from the identification of the need to contract award. To do so, the date the need was identified would need to be added to FPDS-NG reporting.
- One strategy for reducing PALT is to maximize coverage of requirements by an existing IDIQ contract, basic ordering agreement (BOA; Findenstadt & Hawkins, 2015), or BPA.
- Consider how the IRS forecasts requirements in advance of need (i.e., before they get to contracting, during the customer's budgeting process). If suppliers know the requirements well in advance, they might be able to quote/bid/offer faster. Forecasting requirements could also be useful in consolidating transactions and in ensuring an IDIQ, BOA, or BPA can cover it (i.e., in scope)—or in getting an IDIQ contract, BOA, or BPA in place.
- Evenly distribute workload to CSs so that anyone CS is not overloaded. Evaluate workload models to ensure proper staffing levels, and rebalance across organizations where necessary.

- Benchmark agencies that have lower PALT for best practices.

Study Limitations

As with any research, this study is not without limitations. The lack of solicitation dates limited the contract award actions analyzed. There could be systematic reasons for award actions not including solicitation issue dates, which would introduce bias into the results. Additionally, solicitation issue dates are only available beginning in 2018. Thus, the data is mostly limited to 2019 and 2020. Additionally, the narrow definition of PALT that excludes all of the work after the identification of a need but before the solicitation is issued omits many decisions that affect PALT (e.g., source selection method, extent of market research, contract type, etc.). Additionally, the COVID-19 pandemic in 2020 may have distorted buying patterns, such as goods and services purchased and more rapid procedures (e.g., sole source awards). Finally, data coding errors in the USA Spending data set could distort the results.

Directions for Future Research

This research raises further questions related to PALT. For example, the definition of PALT could be expanded to include all of the pre-award process from need identification to contract award. Then repeat the analyses herein to determine the factors affecting all of pre-award cycle time. Then, explore the end-to-end value chain by further examining the post-award effects of pre-award decisions. For example, what are the effects of shortened PALT on contract compliance? What are the effects of shortened PALT on post-award modifications? Does shortened PALT affect contractor performance? Additionally, for the Shopping Cart predictive model, the current contract award date Random Forest model could be extended to use training data from awarded contract data from both FY2020 and FY2021. With a larger dataset that spans fiscal years, we will look to predict on unawarded contract actions available in the open obligation PALT sheet. Finally, the NLP analysis raises opportunities to further explore the nature of the impactful topics and tokens in order to understand what is it about the appearance of terms such as “tuition” or “phase” in the award description field that either increases or decreases PALT.

Conclusion

Government and industry leaders have recently expressed a need to accelerate the procurement process. PALT was a focus of study in the 1980s and 1990s; however, emergent technology and the availability of big data provide opportunities to apply more robust methods and explore more complicated questions. Using machine learning, newly collected, standardized PALT data was analyzed to better understand factors influencing time to contract award. The goals were to explain the key factors impacting PALT, identify opportunities to increase efficiency and reduce PALT, and use a data-driven approach to generate milestone schedule estimates informing customers when their contract is likely to be awarded.

This study confirms findings from prior PALT research and also provides a more comprehensive, government-wide explanation of factors driving time to contract award. Confirmed factors affecting PALT include dollar value, number of offers received (i.e., extent of competition), the goods or services procured (i.e., NAICS code), and type of contract. A number of new insights into PALT have been quantified using a large dataset. Differences between agencies were found, with some agencies awarding contracts in a particular dollar value range faster than others. The choice of solicitation procedures by the contracting officer impacts time to contract award. Further, the number of days remaining until the fiscal year end is a powerful driver of contract award dates. The contracting personnel's workload also affects PALT, as does the organization which they support. Finally, certain words and word-combinations in the award description field are related to PALT. A better understanding of these factors should help acquisition teams to reduce PALT and help acquisition leaders to set policies and processes to mitigate PALT.

References

Berteau, D. (2018). DoD can reduce time to contract. *Federal Times*. https://www.federaltimes.com/acquisition/2018/03/23/dod-can-reduce-time-to-contract/

CAPS Research. (2011). *Cross-industry report of standard benchmarks*. Institute for Supply Management and Arizona State University.

Carter, J. R., & Narasimhan, R. (1996). Purchasing and supply management: Future directions and trends. *International Journal of Purchasing and Materials Management*, *32*(4), 2–12.

Chung, A. S., Feldman, A. P., & Manuel, S. R. K. (2018). *Factors influencing the PALT of sole-source acquisitions greater than $500 million* (NPS-CM-19-026). Naval Postgraduate School, Acquisition Research Program.

Finkenstadt, D., & Hawkins, T. (2015, September). To contract or not to contract: BOAs vs. IDIQs. *Contract Management*. National Contract Management Association.

Hawkins, T., & Muir, W. (2014). An exploration of knowledge-based factors affecting procurement compliance. *Journal of Public Procurement*, *14*(1), 1–32.

Hawkins, T., Berkowitz, D., Muir, W., & Gravier, M. (2015). Improving services supply management in the defense sector: How the procurement process affects B2B service quality. *Journal of Purchasing and Supply Management*, *21*, 81–94.

Hawkins, T., Gravier, M., & Yoder, E. C. (2016). Federal bid protests: Is the tail wagging the dog? *Journal of Public Procurement*, *16*(2), 152–190.

Hult, T. M. (1997). Measuring cycle time of the global procurement process. *Industrial Marketing Management*, *26*(5), 403–412.

Lamoureux, J., Murrow, M., & Walls, C. (2015). *Relationship of source selection methods to contract outcomes: An analysis of Air Force source selection* (NPS-CM-16-006). Naval Postgraduate School, Acquisition Research Program.

Landale, K., Rendon, R., & Hawkins, T. (2017). Examining the effects of source selection method on procurement outcomes. *Journal of Defense Analytics and Logistics*, *1*(1), 47–68.

MacKinnon, D. J. (1992). *A forecasting model for procurement administrative lead time* [Master's thesis]. Naval Postgraduate School.

Ng, B., Ferrin, B. G., & Pearson, J. N. (1997). The role of purchasing/transportation in cycle time reduction. *International Journal of Operations & Production Management*, *17*(6), 574.

Office of Federal Procurement Policy. (2008). *Interagency acquisitions*.

Wooten, M. E. (2020). Procurement administrative lead time (PALT). *Federal Register*, *85*(13). National Archives and Records Administration.

Zycus. (2014). *P2P benchmark study*. https://sig.org/docs2/Zycus_-_P2P_Benchmark_Study.pdf

Why Marketing Matters: Strengthening the Defense Supplier Base Through Better Communication with Industry

Amanda Bresler—serves as Chief Strategy Officer for PW Communications Inc. She runs the company's Future Capabilities Practice, focused on helping the Department of Defense improve its ability to identify, access, and retain innovative solutions providers. Prior to joining PW Communications, she served as Chief Operating Officer for Maurice Cooper Brands. She serves on the board of directors of PW Communications; St. Dalfour SAS, a French food company; Chatham International Inc.; and AlmaLinks. She graduated cum laude from Georgetown University's McDonough School of Business. [abresler@pwcommunications.com]

Alex Bresler—serves as Chief Data Officer for PW Communications Inc. Bresler is a data-driven technologist, investor, and advisor to world-class high-tech companies. He is an expert data scientist and programmer with experience supporting clients in defense, financial services, law, real estate, and sports. He graduated from the Wharton School at the University of Pennsylvania. [alexbresler@pwcommunications.com]

Abstract

This paper demonstrates that the Department of Defense (DoD)'s primary methods of marketing requirements and soliciting information from prospective suppliers inhibit the military's access to innovative nontraditional companies. To conduct this research, we leveraged qualitative and quantitative research techniques, including assessing the features of more than one million DoD solicitations from https://beta.sam.gov and surveying small businesses on the readability of DoD requirements. Our results concluded that DoD solicitations are not conducive to attracting nontraditional suppliers because they are difficult to discover, lack ample response time frames, are not easy to read or understand, and lack critical information. These and other factors deter innovative, nontraditional companies from participating in the DoD's market research process, in turn limiting the pool of suppliers available to the military. We offer recommendations for how the DoD can improve the way it writes and markets solicitations to attract and engage innovative, nontraditional companies more competitively.

Introduction

Over the last 2 decades, companies outside of the U.S. military's traditional industrial base—rather than entrenched defense contractors—have increasingly driven advancements in areas of critical importance to national defense. This paradigm shift has forced the Department of Defense (DoD) to rethink how it sources and funds new technologies and has prompted continuous investment—to the tune of billions of dollars annually—in innovation initiatives and rapid acquisitions programs whose stated purpose is to accelerate the adoption of commercial technologies. In spite of these efforts, we demonstrated in research we published in 2020 that the vast majority of DoD suppliers, including participants in DoD innovation programs, continue to be legacy contractors (Bresler & Bresler, 2020). We posited that one reason why the DoD does a poor job of attracting innovative new vendors ("nontraditionals") is its failure to adequately market its requirements to communities outside of the traditional defense industrial base (DIB).This research aims to explore that hypothesis in more detail. Specifically, we sought to analyze how the composition and marketing of DoD requirements impacts the military's efforts to attract innovative, nontraditional suppliers.

Research Approach

This paper begins by providing an overview of the DoD's current methods of marketing open requirements ("opportunities" or "requirements") and soliciting information from prospective suppliers ("supplier outreach"), including the Federal Acquisition Regulation (FAR), which dictates these procedures. Next, employing quantitative and qualitative research techniques, we analyzed the extent to which these methods enable the DoD to engage innovative, nontraditional companies outside of the DIB. We primarily focused our analyses on the following criteria:

- **discoverability:** the extent to which nontraditionals can find relevant DoD opportunities
- **response time:** the number of days between when an opportunity is posted and when responses are due
- **content:** the extent to which requirements are written in a clear and readable fashion and the extent to which requirements contain the information needed for nontraditionals to adequately evaluate them
- **redundancy:** the extent to which multiple DoD/government entities are simultaneously seeking similar solutions and how redundancy may affect nontraditionals' ability to prioritize relevant opportunities

In each section, we demonstrate that the DoD's methods of marketing its requirements and conducting supplier outreach substantially inhibit the military's access to companies outside of the DIB. This finding offers important context relative to our 2020 research results insofar as it makes clear a driving factor behind the DoD's failure to introduce a significant number of innovative new suppliers into the defense market over the last decade. Throughout the paper, we offer concrete recommendations for how the DoD can improve the way it communicates with industry to reach and engage a broader and more diverse audience of potential suppliers, thereby ensuring that the warfighter has access to the cutting-edge technologies necessary to fight and win.

How the DoD Markets Requirements: Federal Acquisition Regulation

The primary ways in which the DoD markets requirements and conducts supplier outreach in the procurement process are dictated by Federal Acquisition Regulation (FAR). These regulations create a set of rules that government stakeholders must comply with when procuring products and services. Certain exceptions exist within contract administration that allow contracting personnel to employ non-FAR contract strategies, such as Other Transactions, Procurements for Experiments, and Research and Development (R&D) Agreements (Defense Acquisition University, n.d.). However, the majority of contracts are FAR based, and non-FAR contracts are not always precluded from the marketing-specific requirements most relevant to this research.

For the purposes of this research, it is important to understand FAR Part 5, Part 6, and Part 10. FAR Part 5 requires contracting officers to "disseminate information on proposed contract actions ... expected to exceed $25,000, by synopsizing in the Governmentwide Point of Entry (GPE)" (FAR 5.1, 2021). The website https://beta.sam.gov (hereafter referred to as *beta.sam*), which replaced legacy site FedBizOpps in 2019, serves as the GPE. Thus, to comply with the FAR, all contract actions are made public on beta.sam, and the archived and active data on the site serves as a primary resource for our quantitative analyses. FAR Part 6 requires "with certain limited exceptions, that contracting officers shall promote and provide for full and open competition in soliciting offers and awarding Government contracts" (FAR 6.1, 2021) and

> contracting officers shall provide for full and open competition through use of the competitive procedure(s) contained in [the FAR] subpart that are best suited to the circumstances of the contract action and consistent with the need to fulfill the Government's requirements efficiently. (FAR 6.1, 2021)

In short, FAR Part 6 requires government stakeholders to ensure that opportunities are marketed competitively.

Additionally, and of particular import, FAR Part 10 explicitly addresses the "policies and procedures for conducting market research to arrive at the most suitable approach to acquiring, distributing, and supporting supplies and services" (FAR 10, 2021). These policies dictate that government stakeholders must follow a number of steps during the market research process, including but not limited to the following:

- **Conduct market research appropriate to the circumstances-**
    - Before developing new requirements documents for an acquisition by that agency; …
    - On an ongoing basis, take advantage (to the maximum extent practicable) of commercially available market research methods in order to effectively identify the capabilities of small businesses and new entrants into Federal contracting that are available in the marketplace for meeting the requirements of the agency.
- **Use the results of market research to-**
    - Determine if sources capable of satisfying the agency's requirements exist;
    - Determine if commercial items or, to the extent commercial items suitable to meet the agency's needs are not available, nondevelopmental items are available that-
        - Meet the agency's requirements;
        - Could be modified to meet the agency's requirements; or
        - Could meet the agency's requirements if those requirements were modified to a reasonable extent;
    - Determine the extent to which commercial items or nondevelopmental items could be incorporated at the component level; …
- **When conducting market research, agencies should not request potential sources to submit more than the minimum information necessary. (FAR 10, 2021)**

While the intention of these and other FAR clauses may be to foster competition, we sought to analyze, in practical terms, the extent to which these objectives are met. Furthermore, the importance of broadly marketing requirements and fostering healthy competition go beyond regulatory requirements. As we mentioned previously, now more than ever, the military needs innovative capabilities originating outside of the DIB, yet the DoD has continued to fall short in the critical mission of engaging these types of firms. This trend has persisted in spite of the FAR requirements and in spite of substantial investments into defense-sponsored innovation initiatives.

Discoverability

Beta.Sam Awareness

In our 2020 research, we argued that one reason why legacy contractors continue to receive the vast majority of DoD contracts is because there is a general lack of awareness among companies outside of the DIB on the basics of how to identify and engage with military customers (Bresler & Bresler, 2020). While the majority of this paper is concerned with whether or not specific features of DoD opportunities inhibit the military's ability to engage nontraditionals, it is first important to consider whether or not nontraditionals can discover DoD opportunities at all. Simply put, are nontraditionals aware of beta.sam, and do they know how to leverage it to identify prospective opportunities?

We do not have access to information about website traffic to beta.sam, precluding us from quantitatively assessing the reach and composition of the site's audience. However, we can tell from site embeddings that the government does track critical data, such as overall site traffic, the number of unique visitors, the locations of visitors, and more. We encourage the DoD to make use of this information to assess the effectiveness of its marketing initiatives and to shape the development of future marketing and search engine optimization (SEO) strategies.

In the absence of site traffic data, we nevertheless have reason to believe that many nontraditionals are unfamiliar with beta.sam and/or struggle to navigate it. For instance, in addition to the multibillion dollar lobbying and consulting industry centered around helping firms navigate the defense market, companies such as GovWin, Bloomberg Government, and GovShop charge firms a subscription fee in exchange for repackaged opportunity data from beta.sam. The existence of a secondary market for publicly available government opportunity data suggests that beta.sam fails to serve as a viable resource for this information. The result of this "pay to play" paradigm is that the DoD does not see companies with the most cutting-edge capabilities. Rather, the military's requirements primarily reach only those companies willing to pay for access. While service providers and relationships will always play a role in navigating an organization as large and bureaucratic as the DoD, it is important that basic information about the military's requirements be accessible to a wide and diverse audience.

While the remainder of our analyses make the assumption that nontraditionals can successfully reach beta.sam, there is clearly a need to market the site better overall. Further research is required to determine the appropriate level of investment the DoD should make to broaden awareness of the site, along with how to allocate those resources. For starters, we suggest that they invest in SEO to ensure that beta.sam is returned at the top of all search engine searches for queries related to selling products/services to the government. Additionally, we suggest that the DoD engage a marketing firm to develop a strategy for promoting the site in places heavily trafficked by nontraditionals, like *Bloomberg Businessweek*, Crunchbase, LinkedIn, *The Wall Street Journal*, and more.

Site Design

Assuming companies successfully reach beta.sam to explore potential DoD opportunities, they face yet another obstacle: how to navigate the site. It is clear from the landing page, a snapshot of which is provided in Figure 1, that it is not designed with supplier outreach in mind. Rather, it explicitly states that it is "for people who make, receive, and manage federal awards" (General Services Administration, n.d.). These distinct stakeholder groups have markedly different purposes for visiting the site and have markedly different levels of familiarity with government data and terminology.

Figure 1. Beta.Sam Landing Page

Nontraditionals face an immediate challenge of trying to decipher the wide array of links and drop-down menu options to determine what content is relevant to them. References to topics such as "Wage Determinations'' and "Contract Data Reports'' confuse and intimidate companies unfamiliar with the government market. Furthermore, there is no explicit call to action on the homepage for companies interested in learning more about selling their products/services to the government—only a drop-down menu that allows a user to select "Contract Opportunities" and small text towards the bottom of the page that says "Learn More" followed by "Contract Opportunities (FBO)."

Rather than relying on a single site to serve multiple distinct stakeholder groups, we recommend that the federal government create a separate site specifically for suppliers. The site would speak directly to prospective and current suppliers using simple, clear, and straightforward language. It could be linked to the "New Supplier Portal" we recommended in our 2020 research paper—a resource specifically for companies with no prior experience selling to the government (Bresler & Bresler, 2020). There would be a prominent search feature with an explicit call to action to the effect of "Interested in Selling Your Products/Services to the Government? Search for Open Opportunities Here." Additionally, we suggest that investments made by the DoD to market beta.sam be specifically focused on marketing this offshoot, supplier-specific site.

Search Functionality

If and when users reach the landing page associated with "Contract Opportunities," they can input keywords to conduct Boolean searches for relevant opportunities. Two significant limitations to this search functionality include:

- When inputting a search term, beta.sam only returns matches that reference the exact term searched; it does not stem the search term to generate matches for related terms. For instance, if a company searches "UAV," they will not see matches for "drone" (unless the "drone" opportunity also contains the term UAV). As it stands, the scope of relevant opportunities presented to a company is substantially limited, which in turn limits the pool of prospective suppliers that participate in a given DoD opportunity. We recommend that the federal government at large, including the DoD, incorporate related terms to beta.sam's search function. They can leverage resources such as the Defense Technical Information Center (DTIC) thesaurus to do so in a consistent fashion.
- Beta.sam only searches for the input term in the title and description of that opportunity—it does not search for the term in the attachment data. DoD customers often outline their needs in attachments rather than in the description, particularly in calls for market research. As it stands, companies who rely solely on beta.sam searches miss out on many potentially viable opportunities. We recommend that the DoD either mandate stakeholders to outline their needs areas in the description field or enable queries to search attachment data.

Of note, identifying too many opportunities can also be problematic, so it is important that the opportunities presented are easy to assess and understand. These nuances are addressed in greater detail in the Readability and Redundancy sections below.

Response Time

Acknowledging that a lack of awareness of beta.sam, coupled with challenges posed by the design of the site, greatly inhibit the DoD's ability to reach a broad audience, we now shift our focus to assess the features of DoD opportunities. These analyses make the assumption that companies know beta.sam exists and are using it to search for potential DoD opportunities. The first feature we explored is the length of time a company has to prepare and submit a response from when an opportunity is made public to when submissions are due. Response time is an important metric for competitiveness because companies need adequate time to identify an opportunity, to evaluate whether the opportunity is worth pursuing, and to prepare and submit a compliant response.

To quantitatively analyze the response time frames associated with DoD solicitations, we aggregated the archived solicitation data from beta.sam in each year from 2002 through 2020, starting in 2002 because the data sets become more complete in that year. After joining and cleaning 18 years' worth of data, we filtered the data to isolate solicitations issued by the DoD. To ensure we counted only distinct solicitations, we also filtered the data to include just the most recent solicitation listing associated with a particular solicitation identification (solicitation ID) and title. Additionally, we excluded solicitations that contained no text in the name or the solicitation description and/or listed a response date that occurred prior to the publishing date.

We also excluded

- Solicitations associated with notice types for “Sale of Surplus Property,” “Modification/Amendment/Cancel,” and “Foreign Government Standard”
- “Award Only” notices
- “Justifications”

We excluded these listings because they contain features inconsistent with the majority of the data and are generally unrelated to the market research process.

Our resulting data set of total DoD solicitations for analysis was 1,050,933. Figure 2 shows the total number of DoD solicitations by year.

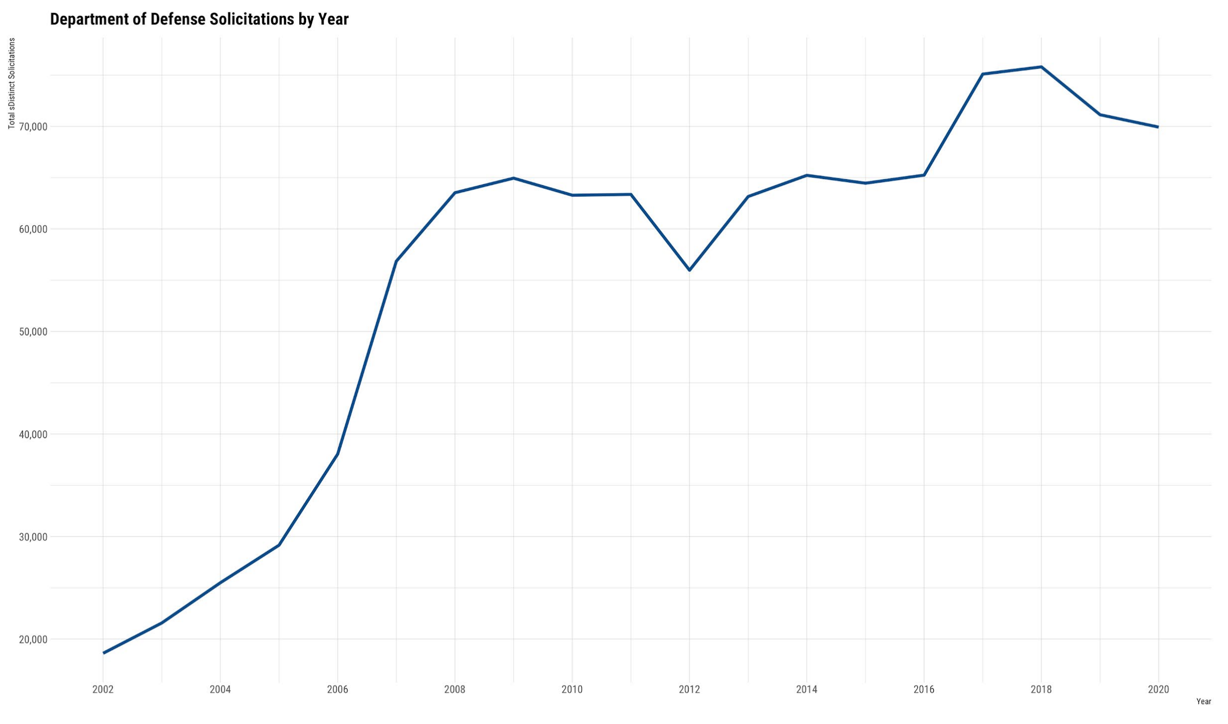


Figure 2. Total DoD Solicitations by Year

We then determined the response time frame for each solicitation by calculating the number of days between the date the solicitation was published and the date by which a response was due, both of which are standard data fields. As shown in Figure 3, every year from 2002 through 2020, 22% to 35% of all DoD solicitations had a response time of 10 days or less, and 45% to 87% of all DoD solicitations had a response time of 21 days or less. In each year over the last decade, 70% or more of all DoD solicitations had a response time of 21 days or less; and with the exception of 2020, at least 30% of all solicitations annually required responses within 10 days.

Department of Defense Solicitation Response Time by Year

Year	Total Distinct Solicitations	Count Response Time <= 10 Days	Count Response Time 11-21 Days	% Response Time <= 10 Days	% Response Time 11-21 Days	% Response Time <= 21 Days
2002	18,612	4,255	4,487	22.86%	24.11%	46.97%
2003	21,569	4,818	4,994	22.34%	23.15%	45.49%
2004	25,496	5,742	6,289	22.52%	24.67%	47.19%
2005	29,157	6,915	9,734	23.72%	33.38%	57.10%
2006	38,042	9,793	15,905	25.74%	41.81%	67.55%
2007	56,854	14,783	26,043	26.00%	45.81%	71.81%
2008	63,532	15,918	28,197	25.06%	44.38%	69.44%
2009	64,953	19,384	26,445	29.84%	40.71%	70.56%
2010	63,289	19,501	26,483	30.81%	41.84%	72.66%
2011	63,368	18,804	28,692	29.67%	45.28%	74.95%
2012	55,975	16,212	26,666	28.96%	47.64%	76.60%
2013	63,162	18,656	31,058	29.54%	49.17%	78.71%
2014	65,229	22,079	31,486	33.85%	48.27%	82.12%
2015	64,462	22,288	28,515	34.58%	44.24%	78.81%
2016	65,248	23,126	29,963	35.44%	45.92%	81.36%
2017	75,106	25,152	35,897	33.49%	47.80%	81.28%
2018	75,808	26,460	33,793	34.90%	44.58%	79.48%
2019	71,139	23,218	30,287	32.64%	42.57%	75.21%
2020	69,933	18,533	42,253	26.50%	60.42%	86.92%

Figure 3. DoD Solicitation Response Time by Year

While these turnaround times may not violate the FAR, it is unreasonable to expect that companies with little or no experience in the public sector will have ample time to participate in the market research process for opportunities open 21 days or less, and a time frame of 10 days or less is that much more challenging. Furthermore, as the data show, the problem has become progressively worse over the last 2 decades. This trend is especially concerning since, over that same time frame, the military has become increasingly reliant on technologies being developed outside of the traditional DIB. In other words, as the need to engage nontraditionals has grown, the process for companies to do so has become more anticompetitive.

Response Time by Notice Type

According to the DoD *Guidebook for Publicizing Notices in Contract Opportunities*, government stakeholders are required to publish notices for "proposed contract actions valued at more than $25,000," which include "announcements through official solicitations in the pre-award process, and up through award" (DoD, 2020, p. 3). As such, each opportunity corresponds to a specific notice type, depending on the purpose of the particular contract action. Each of the 1,050,933 opportunities in our data set corresponded to one of the following notice types, as defined by the *Guidebook for Publicizing Notices in Contract Opportunities* (DoD, 2020, p. 5):

- **Special Notice:** To increase competition and broaden industry participation, a special notice may be used to announce small business conferences, business fairs, long-range procurement estimates, pre-bid or preproposal conferences, meetings, and the availability of draft solicitations or draft specifications for review.
- **Sources Sought:** Use the sources sought notice type for Requests for Information (RFI) and other types of market research. An RFI is used when the Government does not presently intend to award a contract, but wants to obtain price, delivery, other market information, or capabilities for planning purposes. Responses are information only and shall not be used as an offer or proposal.

- **Presolicitation:** In appropriate cases, use a presolicitation notice to advise suppliers on the scope and purpose of the acquisition and to invite potential offerors to submit information. This allows the Government to advise the offerors about their potential to be viable competitors. Responses are information only and shall not be used as an offer or proposal. The FAR requires that a presolicitation notice be published in advance of a solicitation notice unless the combined synopsis/solicitation is used.
- **Solicitation:** Requests for proposals (RFPs) are used in negotiated acquisitions to communicate Government requirements to prospective contractors and to solicit proposals.
- **Combined Synopsis:** Use a combined Synopsis/Solicitation when the procurement meets the applicable conditions outlined in the FAR to reduce the time required to solicit and award contracts for the acquisition of commercial items. This notice type combines the synopsis and the issuance of the solicitation into a single document.

The purpose of Special Notices, Sources Sought, and Presolicitations is to allow the DoD to collect information from a broad range of suppliers about what capabilities they possess and how they would approach solving the DoD's stated problem(s). The DoD then uses the feedback gathered to shape and inform future requirements. It is especially important that nontraditionals participate in these types of information exchanges. Otherwise, the military's view of how problems can be solved is shaped exclusively by entrenched suppliers, which is inherently limiting as they do not always possess the most cutting-edge capabilities and may not be incentivized to encourage the DoD to consider new approaches. As such, we were interested in understanding how response times varied across these different notice types, and—in particular—for Special Notices, Sources Sought, and Presolicitations.

As shown in Figures 4 and 5, the vast majority of Special Notices and Sources Sought, and nearly half of all Presolicitations, have a turnaround of 21 days or less. Based on response time alone, suppliers unfamiliar with the DoD's supplier outreach methods are effectively closed off from participating in these critical calls for market research.

Department of Defense Solicitations, by Notice Type and Response Time

Notice Type	Count Response Time <= 10 Days	Count Response Time 11-21 Days	Count Response Over 21 Days
COMBINED SYNOPSIS/SOLICITATION	197,454	332,578	35,559
PRESOLICITATION	72,472	99,069	208,678
SOURCES SOUGHT	32,431	28,144	18,857
SPECIAL NOTICE	10,818	5,444	2,564
SOLICITATION	2,462	1,952	2,452

Figure 4. Response Time by Notice Type

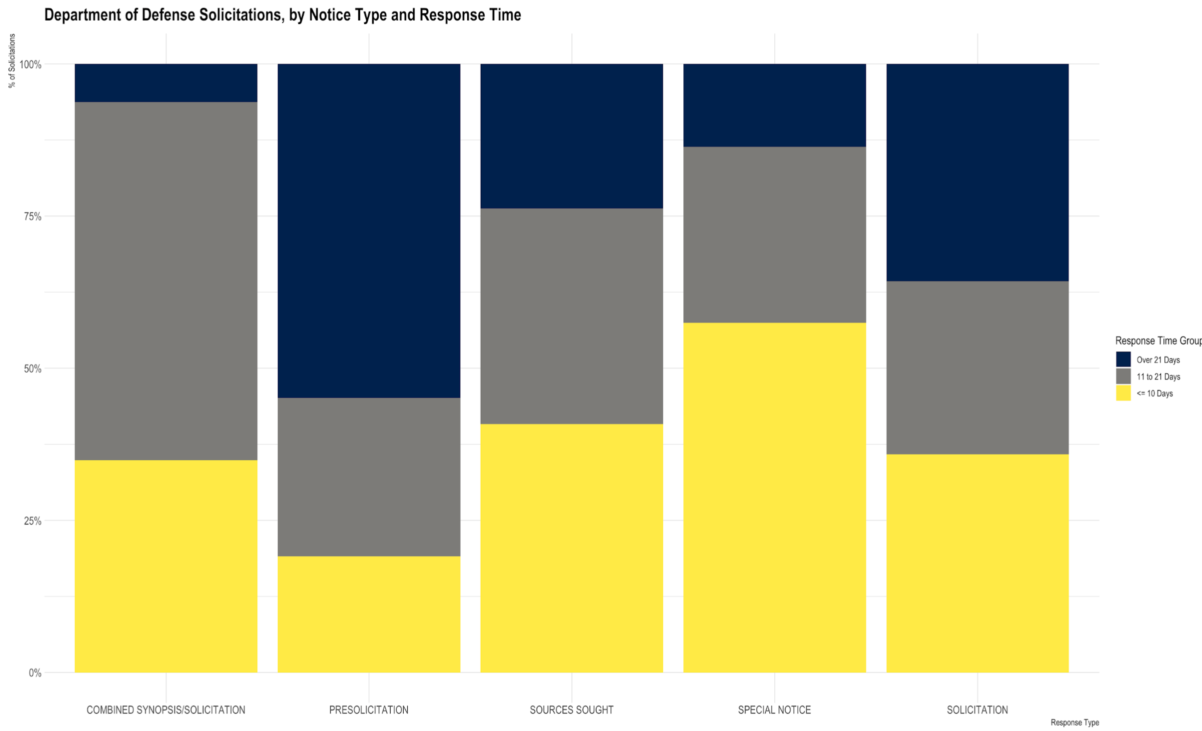


Figure 5. Percentage of DoD Solicitations, by Response Time and Notice Type

While technically speaking, the DoD may comply with FAR Parts 5 and 6 by making these opportunities public, one could argue they fail to meet the objective of FAR Part 10 by virtue of these short turnaround times. Furthermore, DoD opportunities with aggressively short turnaround times are often referred to in industry as “wired” (Walinskas, 2017). A wired opportunity is one where the customer has already identified its vendor, and the formal bid process exists only for compliance purposes. The odds of another supplier winning a future contract are effectively zero.

We recognize the importance of DoD stakeholders being able to engage with suppliers swiftly—in fact, allowing companies to contract quickly is critical for attracting innovators. However, the volume of opportunities with anticompetitive turnaround times indicates a disconnect between the intent of the regulatory standards and how they are employed in practice. To the extent that DoD stakeholders are making opportunities public for 21 days or less as a loophole to award contracts to suppliers they have already identified illustrates that there is a need to allow DoD stakeholders the ability to quickly engage certain suppliers, without doing so at the expense of the military’s overall marketing and outreach strategies. Specifically, we recommend that DoD stakeholders be required to make solicitations active for at least 30 days or be able to formally justify circumventing this requirement to bring a supplier on more quickly, similar to the use of sole-source justifications. If no such justification exists, a suitable response window—coupled with aggressively marketing the DoD requirements in general, as previously discussed—is essential to ensuring that the military has the ability to reach and engage nontraditionals.

Content

Readability: Reading-Ease & Grade Level

In order to ensure fair competition, opportunities not only need a reasonable response time but also must be written clearly so that potential suppliers can understand the requirements. Directly to this point, we sought to evaluate the content of DoD solicitations to determine the extent to which they are readable and easily understood by a wide audience.

To do so, we utilized the Flesch–Kincaid (F–K) readability tests. The two F–K tests, the F–K Reading-Ease test and the F–K Grade Level test, weigh features such as total words, total sentences, and total syllables to indicate how difficult a passage is to understand ("Flesch–Kincaid readability tests," n.d.). For the F–K Reading-Ease test, a low score indicates that a passage is difficult to read, while a high score indicates that a text is easier to read. The F–K Grade Level test scores text based on U.S. grade levels or the number of years of education generally required to understand the text. The scores correspond to one another, insofar as text that is classified as "Difficult to Read" is equivalent to the "College" grade level, "Very Difficult to Read" is equivalent to "College Graduate" grade level, and so forth. Figure 6 lists each F–K Readability Group and its corresponding F–K Grade Level.

To calculate the F–K scores of the 1,050,933 solicitations in our data set, we assessed the text contained in each solicitation description. As shown in Figure 6, which presents the breakdown of the solicitations by F–K Reading-Ease and Grade Level, the majority of solicitation descriptions analyzed were "Difficult" or "Very Difficult" to read. Nearly 59% of all solicitations require some college-level education, and another nearly 20% of solicitations are suited for individuals that graduated from college. By comparison, fewer than 3% of solicitations are written in plain English.

Department of Defense Solicitations, Scored by Reading-Ease & Grade Level

Flesch Readability Group	Flesch-Kincaid Grade Level	Total Solicitations	% of Total Solicitations
DIFFICULT TO READ	COLLEGE	617,516	58.76%
VERY DIFFICULT - READABLE COLLEGE GRADUATES	COLLEGE GRADUATE	209,471	19.93%
FAIRLY DIFFICULT TO READ	10TH TO 12TH GRADE LEVEL	124,176	11.82%
PLAIN ENGLISH - READABLE 13-15	8TH & 9TH GRADE LEVEL	24,111	2.29%
EXTREMELY DIFFICULT - READABLE DOMAIN EXPERTS	DOMAIN EXPERT	23,933	2.28%
EXTREMELY DIFFICULT - READABLE COLLEGE GRADUATES	PROFESSIONAL	18,825	1.79%
FAIRLY EASY TO READ	7TH GRADE LEVEL	17,950	1.71%
EASY TO READ - CONVERSATIONAL ENGLISH	6TH GRADE LEVEL	8,440	0.80%
EASY TO READ - READABLE UNDER 11	5TH GRADE LEVEL	3,834	0.36%
UNSCORED	UNSCORED	2,678	0.25%

Figure 6. DoD Solicitations, Scored by Reading-Ease and Grade Level

Figure 7 provides three examples of solicitation descriptions that were classified as "Difficult to Read," according to the F–K test. They contain esoteric acronyms and range from including excessive information to including almost no information at all.

Description View Changes

3d Dental Battalion/USNDC Okinawa (3D DENBN/USNDC) has a requirement for Dental Implants used in patient treatment . This is a combined synopsis/solicitation prepared in accordance with FAR Subpart 12.6, as supplemented with additional information included in this notice. This announcement constitutes the only solicitation; quotations are being requested. The North American Industry Classification System (NAICS) code is 339114. This acquisition is non-restricted. The incorporated provisions and clauses are those in effect through Federal Acquisition Circular 05-38. The following provisions and clauses apply plus any addenda: FAR 52.212-1 Instructions to Offerors - Commercial Items; 52.212-2 Evaluation - Commercial Items; 52.212-3 Offeror Representations and Certifications --Commercial Items; 52.212-4 Contract Terms and Conditions - Commercial Items; and 52.212-5 Contract terms and Conditions to Implement Statutes or Executive Orders - Commercial Items. The requirements includes the following schedule of supplies: CLIN 0001, BF 12 (12 HOWARD BONE FILE) QTY 8; UNIT PRICE ________________, CLIN 0002, CM18 (#18 MCFARLAND BI BEVEL CHISEL) QTY 2; UNIT PRICE ________________. CLIN 0003, E77 (SERRATED ELEVATOR #77) QTY 40; UNIT PRICE ________________. CLIN 0004, F150K (150 PEDODONTIC FORCEPS) QTY 15; UNIT PRICE ________________. CLIN 0005, F151K (151K PEDODONTIC FORCEPS) QTY 15; UNIT PRICE ________________. CLIN 0006, F17 (#17 HARRIS FORCEPS) QTY 13; UNIT PRICE ________________. CLIN 0007, FAF150 (APICAL FORCEPS 150) QTY 15; UNIT PRICE ________________. CLIN 0008, FAF151 (APICAL FORCEPS 1514) QTY 28; UNIIT PRICE ________________. CLIN 0009, FAFX74N (74N EUROPEAN STYLE APICAL FORCEPS) QTY 12; UNIT PRICE ________________. CLIN 0010, FX13 (#17 EUROPEAN STYLE FORCEP TOOTH EXTRACT) QTY 15; UNIT PRICE ________________. CLIN 0011, R4A QTY 36; (4A CLEVELAND RONGEURS) UNIT PRICE ________________. F.O.B. Destination pricing requested, the contractor shall include the price of shipping into the cost of each CLIN. The contractor shall accept fax or emailed prescriptions from 3D DENBN/USNDC to be filled. Delivery and acceptance point shall be US Naval Hospital Okinawa, PSC 482 BOX 248, FPO AP 96362-1695 seven days after receipt of order. The Government will award a Firm-Fixed Price (FFP) contract resulting from this combined synopsis/solicitation to the responsible contractor whose quote conforming to the solicitation will be most advantageous to the Government, price and other factors considered. The following factors shall be used to evaluate offers: (1) Technical capability of the item quoted to meet the Government requirement (including submission of descriptive literature); (2) Past Performance and (3) Price; (4) Delivery; all to be submitted with quotation. Contractors are responsible for obtaining any and all amendments or additional information. Quotes and all pertinent data (descriptive literature, specifications, etc.) must be received by March 10, 2010 at 0900 (9:00 a.m. Japan Standard Time (JST). Include company name, point of contact, address, phone number, Duns Number, CAGE Code, and Tax ID. As a basis for contract award, contractors must be registered with Central Contractor Registration. (Note: Lack of registration in the Central Contractor Registration will make a Contractor ineligible for award) Registration is available via the Internet at www.ccr.gov. Responses shall be submitted via e-mail in Word, Excel, or PDF format to the POC, Yumi Robb, at yumi.robb@med.navy.mil or faxed to 011-81-989-92-8575. Telephone inquiries or requests will not be accepted.

Description View Changes

The purpose of this announcement is to provide notification of upcoming Industry Days for a potential Cockpit Selectable Output Weapon (SOW) Future Naval Capability (FNC) Broad Agency Announcement (BAA). Industry day discussions and the potential BAA will be limited to technology development in the area of warheads, bombs, kill mechanisms and damage mechanisms. The intent of the potential BAA would not be to improve or develop new weapon to aircraft interfaces, or sensors, or guidance, or control systems. The Naval Air Weapons Center Weapons Division (NAWCWD) will host Industry Days for the potential SOW FNC at the Naval Air Weapons Station Conference Center on 2 March 2010 thru 4 March 2010 in China Lake, CA. The primary purpose of this event is to provide information to industry on the status of NAWCWD's SOW efforts and current planned path forward after review of responses to the Request for Information N6893609R0015. NOTES 1. The Government does not plan to hold Industry Days at any other geographic location at this time. 2. If you have access to JPAS, the SMO Code is 605306 and the POC is Linda Chambless (760-939-2412). In order to receive the necessary badges and attend industry day, each participant's full name, company affiliation, social security number last four digits, date and place of birth must also be sent to Celeste Moore (celeste.moore@navy.mil , phone: 760-939-5979, fax: 760-939-7190) by close of business February 16, 2010. The email, fax or phone call should also indicate if the participant is a U.S. citizen . 3. If the visitor does not have access to JPAS, a copy of the visitor's security clearance should be sent to Linda Chambliss (phone: 760-939-2412, fax: 760-939-0306). In order to receive the necessary badges and attend industry day, each participant's full name, company affiliation, full social security number, date and place of birth must also be sent to Celeste Moore (celeste.moore@navy.mil, phone: 760-939-5979, fax: 760-939-7190) by close of business February 16, 2010. The email, fax or phone call should also indicate if the participant is a U.S. citizen . 4. If the visitor does not have a security clearance, the visitor's full name, company affiliation, full social security number, date and place of birth must be sent to Celeste Moore (celeste.moore@navy.mil , phone: 760-939-5979, fax: 760-939-7190) by close of business February 16, 2010. The email, fax or phone call should also indicate if the participant is a U.S. citizen . 5. Attendance is limited to DOD Personnel and DOD contractors only. 6. Attendance is limited to two (2) representatives per company. 7. Attendance is limited to U.S. citizens only. 8. Contact Celeste Moore (760-939-5979) to schedule a meeting between individual Contractor Teams and the Government. The individual meetings should last less than 1 hour each. Industry Days Program (subject to change): 2 March 20108:00 am - 8:30 am Welcome, Introductions, Meeting Objectives, Meeting Guidelines8:30 am - 9:30 am Update on NAWCWD & ONR SOW efforts9:30 am - 9:40 am Break9:40 am - 11:00 am SOW FNC Status11:00 am - 1:00 pm Lunch (On your own)1:00 pm - 4:30 pm Meetings between Individual Contractor Teams and Government 3 March 20108:00 am - 11:00 am Meetings between Individual Contractor Teams and Government (continued)11:00 am - 1:00 pm Lunch (On your own)1:00 pm - 4:30 Meetings between Individual Contractor Teams and Government (continued) 4 March 20108:00 am - 11:00 am Meetings between Individual Contractor Teams and Government (continued)11:00 am - 1:00 pm Lunch (On your own)1:00 pm - 4:30 Meetings between Individual Contractor Teams and Government (continued) Questions and Answers Q: Does the two-person per company restriction apply to the Contractor/Government sidebars as well?A: Yes. Q: Can teammates be in the sidebar meeting with Government members?A: Yes however, teammates are considered part of the two-person per company restriction. NOTE: All requirements for attendance still apply.

Description

Call for White Paper under Broad Agency Announcement W911NF-17-S-0003, Topic 6; Human Sciences Campaign, subtopic f: CCE-HS-3 Training.

Figure 7. Sample "Difficult to Read" Solicitation Descriptions

To attract a broad audience, requirements must be written in concise, accessible language. Requirements that consist of complex, incomprehensible language limit competition because companies become frustrated by the challenges and effort needed to decipher the text. Furthermore, these poorly written requirements run contrary to the DoD Plain Writing Act of 2010. The act requires federal agencies to write "clear Government communication that the public can understand and use" and stipulates guidelines for compliance that require the DoD to write new documents in "plain language" (Washington Headquarters Services, n.d.).

We recommend that the DoD require all solicitation descriptions to be written in plain English, suitable for an 8th- to 9th-grade reading level. To implement this policy, the government can incorporate a feature on the back end of beta.sam that automatically reads the text of every new solicitation inputted by a DoD stakeholder and calculates its F–K scores. If the scores do not meet the recommended reading level, the system automatically recommends simpler replacement language that the stakeholder can review and approve. Only once the appropriate levels of readability are met can the solicitation be published. We also recommend that text on all public-facing DoD websites, including beta.sam, as well as text in DoD collateral materials meant for public distribution, be written in plain English.

- **Readability: Supplier Feedback**

In addition to analyzing the readability tests on the 1,050,933 solicitations in our data set, we also surveyed 23 small businesses to gather their feedback on government solicitations. The 23 firms are nontraditional dual-use companies that are currently participants in the Air Force's Small Business Innovation Research (SBIR) program. For each company, we utilized publicly available information, including their SBIR award description and related keywords, to establish a basic understanding of their capabilities. We then identified opportunities via beta.sam that appeared to relate to their capabilities, shared the links to the relevant opportunities with a designated company point of contact using Survey Monkey, and asked them to offer feedback on each match. Because they are SBIR participants, they are inherently more familiar with navigating the DoD's solicitation processes than companies with no prior defense business, which would imply a greater comfort level with deciphering additional DoD opportunities presented to them. On the contrary, the companies were frustrated by how challenging it was to decipher the solicitations. Specific feedback included:

- "I cannot tell from the (Areas of Interest) (AOI) what they are asking"
- "I'm struggling mightily to find the AOIs that say what the DoD really wants."
- "Super annoying that I had to comb through attachments to find the AOI's topic."
- "(Broad Agency Announcements) (BAAs) are complex"
- "These BAAs take quite a while to go through and communicate."
- "That was SUPER painful ... because of the opacity with which those SAM postings are written. There are a couple—even AFTER downloading the documents from SAM—that remain mysterious."

To competitively attract and engage nontraditionals, opportunities must be written clearly and provide the detailed information necessary for a company to evaluate whether or not the opportunity is worth pursuing. The aforementioned feedback highlights another problem with the DoD's marketing and outreach methods: critical information is often buried in cumbersome attachments or omitted altogether. Having to sift through complex files to understand the requirements does not inspire a company to respond to a solicitation. Accordingly, in addition to ensuring that the description text of an opportunity is written in plain English, we also recommend that all opportunity descriptions explicitly state the customer's primary areas of interest.

- **Requisite Information**

By assessing the features of the data contained in our solicitation data set and reviewing publicly available opportunities on beta.sam, we found that the DoD often omits critical pieces of information from opportunity listings altogether. Specifically, there are no structured fields requiring DoD stakeholders to indicate on the landing page of beta.sam the value of the opportunity (estimated or actual) or the performance period (estimated or actual).

We recognize that providing specific contract values or performance periods for all notice types is a challenge for DoD stakeholders, because the market research process is intended to help shape the requirements. However, nontraditionals, especially those with robust private sector revenue streams, are unlikely to invest time and resources to explore an engagement with a DoD customer without some sense of the potential upside and/or when the work might begin.

We suggest that the DoD be required to provide an estimated contract value/range for all opportunities. An algorithmic approach can be employed to generate the estimates, including aggregating and weighing factors such as average contract size awarded by the corresponding contracting office over the last 5 years; average contract size for the particular product or service the opportunity corresponds to (for instance, if the opportunity relates to drones, calculating the average size of drone contracts in DoD over the last 5 years); budget estimates for that particular product/service as provided by the Office of Management and Budget (OMB); and other related data points. The solicitation would state that the information provided is an estimate and subject to change. While the process for calculating the estimate requires further research and refinement, we believe it is essential for the DoD to invest in providing this data point as part of its marketing and communication efforts. Companies, particularly nontraditionals attempting to scale with limited resources, also need some sense of performance period to prioritize which opportunities to pursue. DoD stakeholders should, therefore, be required to provide an estimated period of performance as well.

Redundancy

Another challenge that companies face when trying to prioritize DoD opportunities is that many of the same technologies are in high demand by stakeholders across all service branches. The DoD’s 2020 modernization priorities, for example, emphasize the importance of “the development and procurement of high priority systems—such as artificial intelligence, directed energy, small satellites, hypersonics, a 5G network and unmanned aerial systems” (Vergun, 2020) for the whole of military.

For companies with applicable capabilities, a large addressable market may make investing in the defense sector more appealing. However, as discussed throughout this paper, to capitalize on the market, companies must have the ability to navigate it. In cases where multiple DoD stakeholders are seeking similar solutions (“redundancy”), the challenges we have highlighted are compounded by the fact that a company must identify and decipher the relevant opportunities and then decide which ones to pursue. To assess the scope of this redundancy problem, we sought to explore the extent to which multiple DoD stakeholders are simultaneously seeking capabilities related to two of the military’s modernization priorities, unmanned aerial vehicles (UAVs) and artificial intelligence (AI).

Redundancy Analysis

To calculate how many DoD solicitations corresponded to UAVs and/or AI, we employed a more computationally intensive approach that required us to utilize a smaller data set. We focused our analysis on a data set of 69,933 solicitations from the year 2020. Next, we leveraged the DTIC thesaurus to expand the set of terms we used to describe UAVs and AI. The DTIC thesaurus allows for the provision of an input term, such as “unmanned aerial vehicle,” and returns a set of related keywords with varying degrees of proximity to the original term. For the purposes of this research, we limited the results to related terms, which can be understood as synonyms.[3] Next, we algorithmically searched for incidences of these terms in the description, ID, and title for each solicitation. A matched term indicated that an opportunity corresponded to a UAV and/or AI requirement. With this methodology, we identified 42 DoD opportunities in 2020 that corresponded to UAVs and/or AI.

As previously discussed, the DoD often buries critical information, including the areas of interest, in attachments. To more accurately calculate the number of solicitations related to UAV/AI capabilities would, therefore, require searching for the terms in the attachment data. Solicitations can have dozens or even hundreds of pages of attachments across multiple files and file types, and because supporting documents are formatted inconsistently, it was not feasible to incorporate the text and data from attachments for all 69,933 solicitations. Instead, to enhance the search, we decided to incorporate a small subset of attachment data.

Specifically, we first filtered the data to isolate opportunities that corresponded to either a Sources Sought or a Broad Agency Announcement (BAA). While BAAs are not a specific notice type, they—like the DoD's other methods of conducting market research—request "scientific or research proposals from private firms concerning certain areas of interest to the government" (AcqNotes, 2021) and may lead to contract awards. The DoD relies on BAAs to communicate with industry and gather critical market research. For instance, the DoD's SBIR topics are issued as BAAs. To identify BAAs in our 2020 data set, we searched for the terms "Broad Agency Announcement'' and "BAA" in the contract ID, solicitation name, and solicitation description. We then combined the BAAs with the opportunities corresponding to a Sources Sought notice type, excluding any Sources Sought that were already counted as BAAs. In total, we identified 2,519 opportunities in 2020 that were either Sources Sought or BAAs. For these 2,519 opportunities, we incorporated the

---

[3] Keyword Corpus: AI APPLICATIONS, AI COMPUTING, APPLIED COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE, ARTIFICIAL INTELLIGENCE COMPUTING, ARTIFICIAL INTELLIGENCE SOFTWARE

COMPUTATIONAL PROCESSES, COMPUTER VISION, COUNTER-DRONE TECHNOLOGY, COUNTER-UAS

COUNTER-UAV TECHNOLOGY, COUNTER-UNMANNED AERIAL SYSTEMS, DEEP LEARNING, DEEP STRUCTURED LEARNING, DRONE, DRONE CONTROL AIRCRAFT, DRONE SWARMS, DRONES, EXPERT SYSTEMS, HEAVY FUEL ENGINES, HEAVY FUEL UAV ENGINES, HIERARCHICAL LEARNING, INFERENCE ENGINES, INTELLIGENT PERSONAL ASSISTANTS, INTELLIGENT SYSTEMS, LAMP RAY ROV, MACHINE LEARNING, MACHINE PERCEPTION, MICRO AIR VEHICLE, NATURAL LANGUAGE PROCESSING, NATURAL LANGUAGE PROCESSING SOFTWARE, NEURAL NETWORKS, REMOTELY PILOTED AIRCRAFT, SEMI-SUPERVISED LEARNING, SMALL UNMANNED AIRCRAFT SYSTEM, SOFTWARE AGENTS, SUPERVISED LEARNING, SUPERVISED MACHINE LEARNING, SURVEILLANCE DRONES, SWARM INTELLIGENCE, SWARMING DRONES, SWARMING TECHNOLOGIES, SWARMS OF FIXED WING DRONES, TARGET DRONES, UAS, UAV, UGV, UNDERWATER DRONES, UNINHABITED AIRCRAFT VEHICLE, UNMANNED AERIAL, UNMANNED AERIAL SYSTEMS, UNMANNED AERIAL VEHICLE, UNMANNED AEROSPACE VEHICLE, UNMANNED AIR SYSTEMS, UNMANNED AIRCRAFT, UNMANNED AIRCRAFT SYSTEMS, UNMANNED AIRCRAFT VEHICLE, UNMANNED GROUND SYSTEMS, UNMANNED GROUND VEHICLE, UNMANNED GROUND VEHICLE SYSTEMS, UNMANNED SYSTEMS, UNSUPERVISED LEARNING, UNSUPERVISED MACHINE LEARNING text and data contained in their attachments and utilized optical character recognition (OCR) and other methods of text extraction to search this data for UAV/AI terms.

With this methodology, we identified an additional 22 DoD opportunities in 2020 that corresponded to UAVs and/or AI, bringing the total to 64. In other words, utilizing OCR and text-extraction on just 3.6% of the solicitation data increased the number of matched opportunities by more than 50%. Based on these results, one can assume that the total number of DoD stakeholders that posited demand for UAV/AI capabilities in 2020 was substantially more than 64.

Demand Outside of DoD

Furthermore, our analyses did not include solicitations from federal stakeholders outside of the DoD. When assessing the challenges companies face in trying to prioritize DoD customers, it is worth considering the potential effects of demand from non-DoD customers—particularly because, with the beta.sam process, companies discover DoD and non-DoD opportunities simultaneously. We recommend further research to explore the DoD-level findings we have addressed in this paper across the entirety of government, and we recommend that this further research incorporate attachment data to the best extent possible.

In the interim, we conducted a microanalysis to explore the potential impact of non-DoD demand on our research results. To do so, we aggregated all open federal opportunities—DoD and non-DoD, including attachment data—from a single day—October 8, 2020—and searched for UAV terms across this data set. As shown in Figure 8, on that single day, 132 open opportunities corresponded to UAVs.

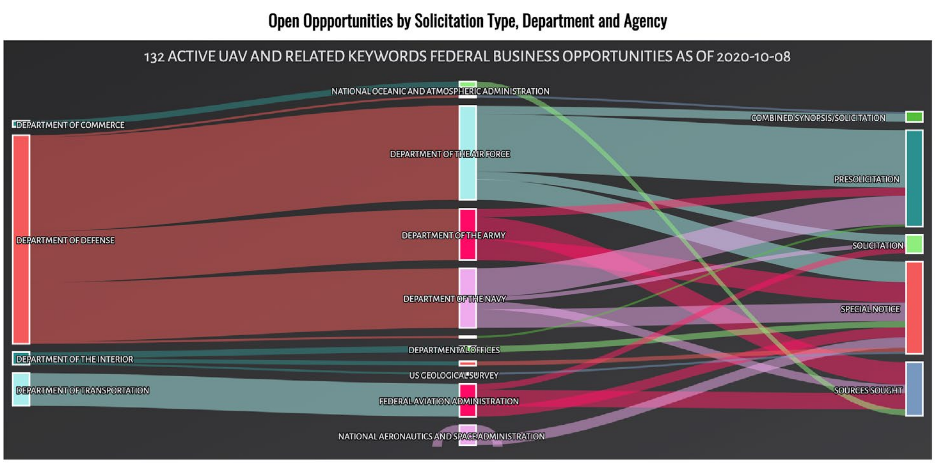


Figure 8. Open Solicitations Related to UAVs on October 8, 2020

Assessing redundancy using any one of the aforementioned methodologies, it is clear that companies with high-priority capabilities can encounter anywhere from dozens to hundreds of prospective DoD and non-DoD opportunities. Therefore, if and when a company identifies and deciphers relevant opportunities, realistically it cannot participate in all of them. The DoD's failure to coordinate its outreach and communication efforts results in negative consequences for both nontraditionals and the warfighter. DoD customers only receive feedback from a small number of firms and are not guaranteed to receive feedback from firms with the most applicable capabilities. As a result, they have a myopic view of how their problems can be solved. Companies interested in serving the needs of government have to decide which customer(s) to engage with the information they have at hand. As a result, they are not necessarily choosing the customers whose use cases align most seamlessly with their capabilities, and they are not necessarily choosing the customers with the most urgent need for their capabilities.

Better intra-government communication would benefit the supplier and the government; thus, it is essential that military stakeholders coordinate their outreach and communication efforts to maximize exposure of their requirements. For priority verticals, we recommend that DoD stakeholders issue joint requirements in the market research/outreach phases. Further research is required to determine the best way to implement this concept, including how to appropriately incentivize DoD stakeholders to take the necessary actions. We suggest that prior to release, the DoD circulate requirements related to priority verticals to designated offices within each service branch. This action will allow DoD stakeholders to incorporate related requirements into the solicitation. In addition to helping the DoD gather information from a wider range of potential suppliers and steer them in different directions more effectively, this approach would allow companies to market their capabilities to multiple prospective customers simultaneously—a major advantage over the current stovepiped system.

Conclusion

In spite of billions in investment for innovation initiatives and unremitting rhetoric from senior leadership about the DoD's commitment to a culture of innovation, our 2020 research proved that the military has failed to attract and engage a significant number of new suppliers over the last decade, which puts the warfighter at risk (Bresler & Bresler, 2020). In this paper, we employed qualitative and quantitative research techniques to illustrate that *how* and *where* the DoD communicates with industry have contributed to this problem. We identified a series of conditions that must be met in order for the DoD's requirements and messaging to reach suppliers outside of the traditional DIB:

- Companies need to know where to go to search for DoD opportunities, and the search process must be user-friendly and intuitive.
- Companies need enough time to identify, assess, and respond to an opportunity.
- Companies need to easily understand what DoD customers are asking for.
- To determine whether or not an opportunity is worth pursuing, companies need certain pieces of critical information, including the potential contract size.
- The DoD needs to coordinate its marketing and outreach efforts, especially for capabilities in high demand across the government.

The absence of any one of these conditions not only fails to meet the objective of the FAR but also creates a bottleneck that limits industry participation in the market research process. The military, in turn, operates with an incomplete picture of how its problems could be solved and what capabilities exist to solve them. The recommendations outlined throughout this paper are intended to help the DoD address each of these bottlenecks as efficiently as possible and to make the process of engaging with the military more seamless for nontraditionals accustomed to operating in the private sector. Ultimately, the military needs access to the best and brightest suppliers to preserve the strength of the warfighter—and to attract best suppliers, the DoD must behave like a better customer.

References

AcqNotes. (2021, March 4). *Contracts & Legal: Broad Agency Announcement.* https://acqnotes.com/acqnote/careerfields/broad-agency-announcements

Bresler, A., & Bresler, A. (2020). The effect of defense-sponsored innovation programs on the military's industrial base. *Proceedings of the 17th Annual Acquisition Research Symposium.* https://calhoun.nps.edu/bitstream/handle/10945/64763/SYM-AM-20-059.pdf?sequence=1&isAllwed=y

Defense Acquisition University. (n.d.). *Contracting cone.* https://aaf.dau.edu/aaf/contracting-cone/

DOD. (2020, August 21). *Guidebook for publicizing notices in contract opportunities: Version 1.0*. https://dodprocurementtoolbox.com/cms/sites/default/files/resources/2020 08/SOP%20Publicizing%20Notices%20-%20FINAL.pdf

FAR 5.1, Dissemination of Information (2021). https://www.acquisition.gov/far/part-5

FAR 6.1, Full and Open Competition (2021). https://www.acquisition.gov/far/part-6

FAR 10, Market Research (n.d.). https://www.acquisition.gov/far/part-10#FAR_Part_10

Flesch–Kincaid readability tests. (n.d.). In *Wikipedia*. Retrieved March 13, 2021, from https://en.wikipedia.org/wiki/Flesch%E2%80%93Kincaid_readability_tests.

General Services Administration. (n.d.). *Welcome.* Retrieved April 5, 2021, from https://beta.sam.gov/

Vergun, D. (2020, September 10). DoD officials describe modernization priorities that will benefit warfighters. *DoD News.* https://www.defense.gov/Explore/News/Article/Article/2343626/dod-officials-describe-modernizaton-priorities-that-will-benefit-warfighters/

Walinskas, K. (2017, August 1). 4 signs the RFP is wired. *Public Spend Forum*. https://www.publicspendforum.net/blogs/karl-walinskas/2017/08/01/signs-rfp-wired-federal-goverment-contracting/

Washington Headquarters Services. (n.d.). *Plain language: Official DoD plain language website.* https://www.esd.whs.mil/DD/plainlanguage/

Non-Competitive Contracting: Lessons from Contracting Personnel

Dr. Latika Hartmann—is an associate professor of economics in the Graduate School of Defense Management at the Naval Postgraduate School. Her research focuses on issues in public finance, regulation, and economic development. She received her PhD in economics from the University of California, Los Angeles in 2006 and a BA in economics from University of California, Berkeley in 2000. She is trained in applied microeconomics and econometrics. She has written technical reports for the Navy (Naval Research Program and Acquisition Research Program) and published peer-reviewed articles on regulation and public finance. In addition, she has advised 25+ student teams at NPS on contracting and manpower issues. [lhartman@nps.edu]

Captain Josh Cissell—is a contracting professional in the USAF. He received a BS from Southern Nazarene University, Bethany, OK, in 2014 and an MBA from the Naval Postgraduate School, Monterey, CA in 2019. His contracting background includes base operational contracting, supply chain management support contracting, weapon system contract administration in DCMA, and weapon system procurement contracting. [joshua.cissell@us.af.mil]

Dr. Rene Rendon—is an associate professor in acquisition management in the Graduate School of Defense Management at the Naval Postgraduate School. He received his BBA from Angelo State University in 1981, MBA from University of North Dakota in 1985, and DBA from Argosy University in 2003. A retired Air Force contracting officer, he has received the NCMA Award for Excellence in Contract Management Research and Writing many times. His research focuses on issues in contract management and strategic sourcing. [rgrendon@nps.edu]

Abstract

In this paper, we survey a small group of Air Force contracting personnel to understand their views on contracting in sole-source environments. Our findings suggest Air Force contracting personnel in this setting know that sellers in non-competitive relationship have more leverage and power than the buyer. Indeed, 90% of respondents feel they operate at a negotiating disadvantage in sole-source contracts. Arming them with certified cost and pricing data does not improve leverage according to majority of them. Rather, they identify two constraints in their qualitative responses. First, the sole-source environment itself contributes to the problem. Second, many respondents feel they operate at an informational disadvantage compared to their private counterparts. This suggests specific training on their contracts would be more valuable than any general training on business acumen.

Introduction

Non-competitive spending in the DoD is large. Per the DoD Competition Scorecard for quarter 3 FY2018, non-competed dollars accounted for 63% of total dollars by the Air Force and 65% by the Navy.[4] A majority of Navy and Air Force dollars are, thus, obligated in non-competitive settings where few large firms dominate. Understanding the views of government contracting personnel on working in such non-competitive markets is critical to DoD acquisition programs. Our paper offers insights by surveying a small number of military

---

[4] Calculated from https://www.acq.osd.mil/dpap/cpic/cp/docs/Publication_of_DoD_Competition_Rerport_-_3rd_Quarter_FY18.pdf

and civilian Air Force contracting personnel at the F-22 program office. The majority of the F-22 program office contract actions are awarded in sole-source environments.

Section 843 of the 2018 National Defense Authorization Act (NDAA) tasked the Under Secretary of Defense for Acquisition and Sustainment to assess whether there are any “gaps in knowledge of industry operations, industry motivation and business acumen in the acquisition workforce” and how such gaps, if any, can be closed with training and development (NDAA, 2018). To that end, Congress asked the Under Secretary to submit a report on the knowledge of “industry operations, industry motivation and business acumen” necessary for each acquisition position and how non-DAU sources (industry and other universities) can be used to improve training. In response, RAND prepared a report drawing on interviews of subject matter experts in the DoD and industry, and published sources (Werber et al., 2019). But the report was unable to speak to the views of acquisition personnel because they did not have the time to survey them.

Our project takes a first step in addressing this gap. We survey Air Force contracting personnel at the F-22 program office on their experience with sole-source contracting, which accounts for a large share of the Air Force budget. We hope our small survey, among others, will help identify gaps in training, if any, that may help acquisition personnel better negotiate contracts and determine “a fair and reasonable” price in these sole-source settings (Federal Acquisition Regulation [FAR], 2018).

By explicitly targeting business acumen, industry operation, and motivation, Congress wants trained acquisition personnel to understand the business of firms. In particular, how do firms price, cut costs, negotiate with buyers and suppliers, and respond to market conditions? A solid foundation in these concepts is likely to help DoD acquisition personnel secure more favorable terms when they buy goods and services. Yet, these concepts are not uniformly defined in the DoD as noted in the RAND report (Weber et al., 2019). Moreover, contracting for common support services such as waste disposal is different than contracting for a major weapons system. A general understanding of firm behavior may be useful to both, but acquisition personnel for a major weapons system need to understand the background and motivations of the big defense firms operating in their field. Identifying gaps in knowledge, thus, requires differentiating between general business learning versus more specific learning tailored to the experience and career path of acquisition personnel.

Our results support this point. In our survey, the majority of Air Force contracting officers know that sellers in non-competitive (i.e., sole-source) relationships have more leverage and power than the buyer. Indeed, 90% of respondents feel they operate at a negotiating disadvantage in sole-source contracts. Giving them certified cost and pricing data does not help, according to the majority of them. Rather, they identify two constraints in their qualitative responses. First, the sole-source environment itself contributes to the problem. Second, many respondents feel they operate at an informational disadvantage compared to their industry counterparts.

Respondents in our study are familiar with the basic constraints of buying in non-competitive markets. They mention an inability to negotiate prices and the associated power asymmetry between them and the sole-source contractor. Yet, many respondents point to specific information gaps. One respondent mentions that defense-focused firms have more nuanced goals such as maximizing revenues that are perhaps not as clearly related to models of profit-maximization. Other respondents say they do not have information on sub-contract terms. And still others say their industry counterparts work on these programs for decades and are familiar with the ins and outs of the contract in ways that government acquisition personnel are not. These responses suggest more specialized training on the contracts themselves and the firm involved is more valuable.

We also find differences between individuals with 7 or fewer years of contracting experience compared to those with more experience. Less experienced individuals are unfamiliar with issues of nonconformance and consideration in sole-source contracts. Consideration in contracts is the idea that both parties benefit from entering into a contract. Given these issues, these differences are perhaps unsurprising because resolving nonconformance involves many steps and can take time.

Our survey focuses on a small population of 57 government contracting personnel at the F-22 program office. That combined with our low response rate of 28% makes us cautious in drawing strong conclusions. That said, our findings are perhaps unsurprising and we offer a few tentative recommendations for training acquisition personnel in non-competitive procurements. Clearly, an individual contracting officer cannot change the market in which they operate. But, we can give them more information on the contract itself—a short readable summary of the contract highlighting any incentives, due dates and changes, a brief history of the negotiations leading to the final contract, and the main players involved. Such details are often included in the "after action report" of seasoned contracting personnel, but we should make them a standard practice. Moreover, we should give more technical and economic information on the particular contract, which would strengthen the commanding officer's negotiating position. On a related note, we may also want to develop in-depth profiles of the principal firms operating in these markets. Finally, we recommend a larger survey of government contracting personnel across multiple commands to better understand and inform the training of acquisition personnel.

Background

We review the RAND report here that motivates our survey. RAND researchers undertook their study in response to Section 843 of the 2018 National Defense Authorization Act (NDAA). On behalf of the DoD, their report assesses the current state of training in "industry operations, industry motivation and business acumen" in the acquisition workforce, documents gaps in that training, and offers recommendations, in particular on the role of non-DAU sources. Their assessment relies on interviews of subject matter experts in the DoD and industry, published competencies for acquisition personnel, and related literature (Werber et al., 2019).

Their findings first highlight that there are no consistent definitions of the terms "industry operations, industry motivation and business acumen" in the competency models associated with different acquisition career fields. Moreover, the competency models themselves are not uniform across career fields. Although the RAND researchers constructed a working definition of these terms, it was difficult to identify which career fields in particular needed knowledge of these terms. Yet, the Report finds that most acquisition career fields need to know these terms and related issues with career fields in contracting requiring more knowledge than science and technology management for example.

Although the acquisition workforce uses both internal sources (e.g., Defense Acquisition University [DAU], Naval Postgraduate School [NPS]) and external sources (e.g., commercial training companies, civilian colleges, and universities) for training and education, the Report was unable to identify precise gaps. This is perhaps due to the study relying on interviews with SMEs as opposed to surveys of acquisition personnel. Finally, the Report finds very little evaluation of the effectiveness of training and development programs. Apart from student responses to courses and some higher-level focus groups, there are no systematic evaluations of whether acquisition personnel make changes in their jobs based on their training and whether those changes translate into improved outcomes.

To address gaps in Section 843 areas of “industry operations, industry motivation and business acumen,” the Report first recommends that the DoD decide what level of knowledge of these areas is required for each career field. Then it recommends the DoD measure the required knowledge of these terms among acquisition personnel. This would enable the DoD to precisely identify gaps, if any, in this knowledge. It also recommends better tracking of training by personnel throughout their career. This would allow for stronger assessments of training and development programs in the acquisition workforce. Our survey addresses an important gap in this study by surveying Air Force contracting personnel on their knowledge of contracting in non-competitive environments. Our findings suggest that most Air Force contracting personnel have a basic understating of industry operations in their field. However, our findings also indicate that the surveyed personnel need more specific knowledge on contracts, sub-contract terms, and contract history,

Survey Methodology

To understand the views of Air Force personnel on sole-source contracting, we surveyed 57 individuals in the F-22 System Program Office at Wright-Patterson Air Force Base, OH, and Hill Air Force Base, UT. Of these 57, five are Air Force officers and 52 are Air Force civilians. After receiving the necessary approval from the Air Force Survey Organization, an administrator emailed a link to the online survey, using the NPS approved Lime Survey web-based survey system. The survey was open from December 6, 2019, to December 31, 2019. It was a voluntary and anonymous survey. We did not ask or collect any personally identifiable information. Moreover, we added the following statement in the beginning: “Your participation in this survey is strictly voluntary. If you choose to participate you can change your mind at any time and withdraw from the survey. You will not be penalized in any way or lose any benefits to which you would otherwise be entitled if you choose not to participate in this survey or to withdraw.” This was to ensure respondents felt comfortable sharing their feedback without fear of retaliation.

Our survey consisted of two parts. Part 1 asked seven demographic and background questions, while Part 2 asked respondents about their experience in non-competitive contracting, whether they feel they are at a disadvantage in such settings, and to identify factors if they do feel they are operating at a disadvantage. We also asked their experience with non-conformance and consideration in non-competitive contracts. The complete survey is available upon request.

Despite multiple email reminders, our response rate was only 28% (16 individuals) dropping to 23% (13 individuals) with complete responses. December is a busy time with year-end deadlines and holidays. This perhaps reduced the number of respondents. Given the small number of respondents, we view the survey results as more qualitative evidence and hope future surveys with larger populations can provide more insight into the issues raised by the F-22 Air Force personnel.

Results

We first describe the background characteristics of the respondents. Unfortunately, 5 of the 16 responses were incomplete. We summarize the responses below for everyone that responded to a question. Of our 16 responses, 3 identified as Air Force officers and 10 as Air Force civilians. Another three chose not to respond. This translates into a higher response rate of 60% among the Air Force officers compared to 19% among AF civilians. Our survey population included 5 Air Force officers, 3 responded, while 10 out of 52 civilians responded. In terms of gender, men accounted for 69% of responses compared to 31% women.

Among the military responses, 2 are of ranks O-1 to O-3, while 1 is of rank O-4 to O-6. Among civilians, a majority of the responses (60%) are at GS-13 and GS-14 grade, compared to GS-12. Figures 1 and 2 summarize the professional background of the respondents. Almost 80% have more than 4 years of contracting experience with the federal government, and 50% have more than 8 years of experience. This suggests our survey responses reflect the views of both entry-level and advanced acquisition personnel. Most respondents (85%) also have a Defense Acquisition Workforce Improvement Act (DAWIA) certification of Level 2 and 3.

Figure 1: Years of Contracting Experience.

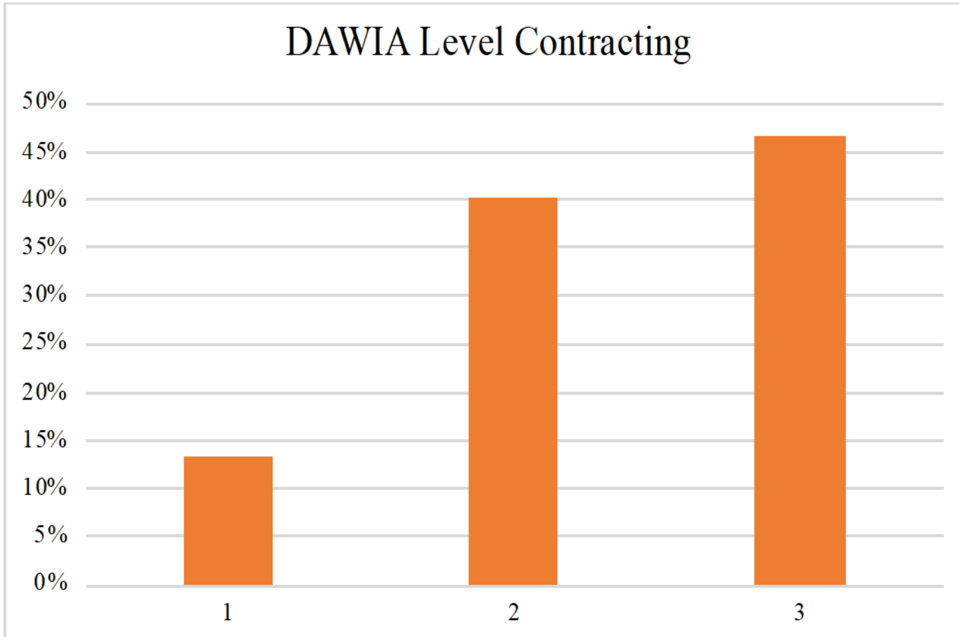


Figure 2: DAWIA Level Contracting

As shown in Figure 3, our respondents have experience in multiple contracting areas with the majority in weapons systems acquisition followed by operational contracting.

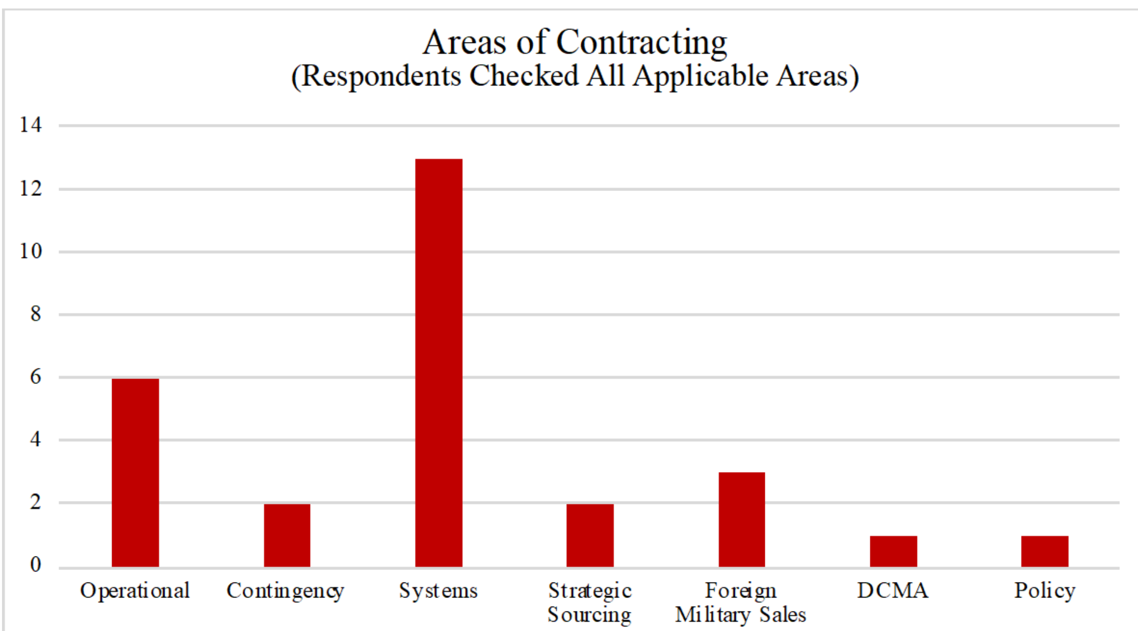


Figure 3: Areas of Contracting

We turn next to our experience questions on non-competitive acquisitions. Given our focus on the F-22 program office, it is unsurprising that 67% of our respondents have been involved (current or past) in non-competitive relationships with incumbent contractors. These would be sole-source contracts. Of these 11 individuals (67%) responding “Yes,” all but one (90%) feel they are at a negotiating disadvantage with the contractor in this setting. Arming them with certified cost and pricing data does not change their views. Of the group that feels they are at a negotiating disadvantage, 90% continue to feel at a disadvantage even with certified cost and pricing information.

Such a lopsided response naturally leads to the question of why contracting personnel feel at a disadvantage in non-competitive settings. Their qualitative responses fit into two categories. First, they feel disadvantaged because private contractors leverage the sole-source environment. Some responses mention the negotiation/power asymmetry in particular of not being "able to walk away and either not purchase the item at all or to purchase it elsewhere." Others mention the inability to negotiate lower prices because the sellers leverage their sole-source position.

This explanation is perhaps unsurprising. We know from economic theories of firm behavior the difference between markets with lots of sellers (i.e., competitive markets) and one seller (i.e., monopoly). Prices are lower in competitive markets because buyers can choose from different sellers. Moreover, buyers can switch from one seller to another if they are unsatisfied with the product or price. This leads to lower prices for goods traded in competitive markets. But the military does not often operate in competitive markets. They buy custom products designed and built to their specific needs. Yet private firms also order custom products in many industries where dual-sourcing is a common response to diversify risk across multiple suppliers and reduce cost. Indeed, Klotz and Chatterjee (1995) show using a theoretical model that dual sourcing can reduce costs in a setting where firms face learning and entry costs, common to most DoD acquisitions. An empirical study confirms the advantages of dual-sourcing. Using a unique though small dataset of 14 tactile missile contracts between 1975 and 1995, Lyon (2006) shows that dual sourcing was undertaken to improve quality, not reduce cost per se, but still led to lower procurement costs.

Other qualitative responses fall under the second category of government contracting personnel operating under an informational disadvantage compared to their industry counterparts. For example, a respondent mentioned that the "Program Office is not trained or knowledgeable [*sic*] on their requirements." Another respondent mentioned they (i.e., government contracting personnel) do not have a "very deep understanding" of "true cost" or the government's "negotiated price." In yet another variant, a respondent described the difficulty in obtaining certified cost or pricing data from sub-contractors. Many respondents mentioned that the Truth in Negotiations Act (TINA) certification does not include data on the terms between contractors and their subcontractors. While sole-source contracts make these information barriers worse, it seems the Air Force should demand more data from their contractors and then disseminate it widely within the DoD acquisition community.

Fifty percent of respondents that used incentive contracts such as Cost-Plus Award Fee (CPAF) or Cost-Plus Incentive Fee (CPIF) felt such contracts better incentivized necessary performance among contractors. But the other 50% of respondents disagreed, describing some problems with incentive-style contracts. One respondent mentioned that contractors are not motivated by fees: rather their business model works around "cash flow" and "quarterly earnings." Another respondent noted that contractors seem to pursue a revenue maximization strategy over profit maximization. Most incentive contracts assume firms seek to maximize profits by cutting costs. Yet these responses suggest these firms have goals that the government does not consider in the negotiation process.

Our final experience questions centered on non-conforming supplies. Fifty percent of respondents had experience with contractor non-conformance. Though 50% nonconformance is high, it is difficult to interpret without more information on non-conformance in other DoD contracts and industry contracts. Of those that experienced non-conformance, 88% said consideration was sought based on their highest dollar value contracts. Figure 4 describes the type of consideration sought with change in schedule being the most common followed by a decrease in price and more additional supplies.

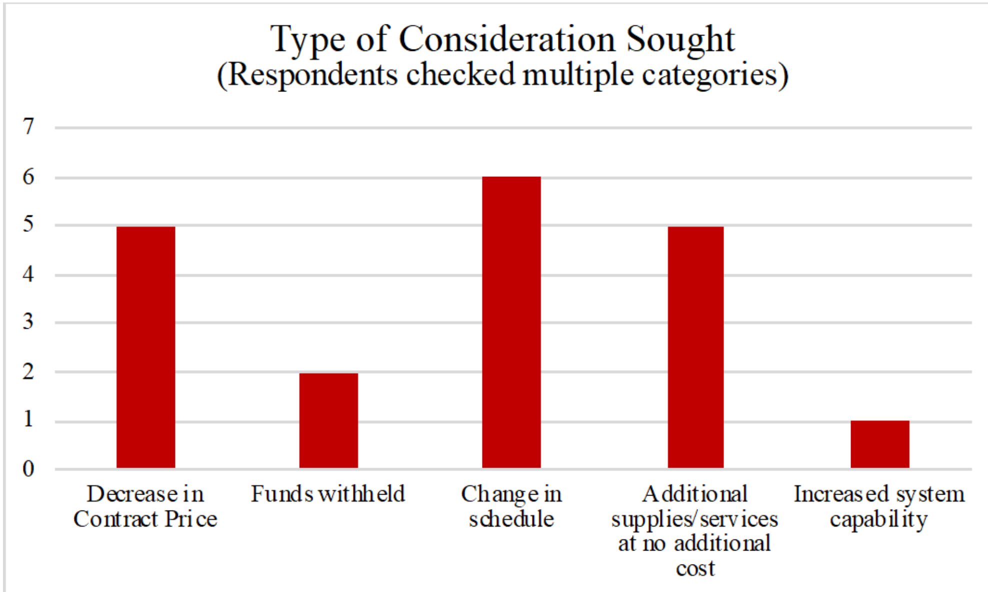


Figure 4: Types of Consideration Sought

When consideration was sought for non-conformance, 57% of respondents said it was recovered. This seems low, but again it is hard to interpret the data without more information on the specific contract and dispute. In cases where consideration was recovered, 75% of respondents felt the consideration was fair. But in cases where consideration was not sought, the respondents again noted power asymmetry problems of sole-source contracts.

Comparing the survey responses across different groups can offer more insight into these overall findings. But we are cautious in analyzing responses by group because of the small number of respondents in some groups. That said, we can use experience to bin individuals into two almost equal–sized groups. This exercise offers qualitative evidence on differences between those with more or less contracting experience.

Among respondents reporting their years of contracting experience, we binned 6 individuals (46%) as those with less experience ranging from 1 to 7 years. We binned the other 7 individuals (54%) as those with more experience ranging from 8 to 25 years. Both groups have similar exposure to non-competitive acquisitions. For example, all but one respondent in each group responded yes to the question on whether they had any experience with sole-source contracting. While both groups had similar backgrounds in sole-source contracting, respondents with more experience felt they were at a negotiating disadvantage in larger numbers compared to those respondents with less experience. Four of those with less experience responded they were at a disadvantage compared to six of seven of those with more experience. Interestingly, respondents with more experience saw less value in certified cost and pricing data compared to those with less experience. Yet both groups mentioned power asymmetry and informational barriers as obstacles to effective negotiating in sole-source environments.

In response to non-conformance, respondents with more experience are more likely to run into non-conformance issues (6 out of 7) compared to those respondents with less experience (2 out of 6). However, more contracting experience also translates into seeing consideration being sought for non-conformance (5 out of 6 individuals) and consideration being recovered (4 out of 6 individuals). None of the respondents with less experience recalled consideration being recovered for their highest dollar value contract. Now these responses may just be a function of time. Non-conformance complaints and considerations can take time to resolve. While some resolve in days, others drag on for multiple years.

Conclusion and Recommendations

We surveyed a small population of Air Force contracting personnel at the F-22 System Program Office at Wright-Patterson Air Force Base, OH, and Hill Air Force Base, UT. Our response rate was 28% or 16 respondents with higher responses among military personnel compared to civilians. Given our small population at one program office and low response rate, we are cautious in extrapolating to the larger government contracting population. That said, our findings likely confirm the perspective of many contracting professionals working in government acquisition. With the caveats in mind, we want to offer tentative recommendations based on our survey.

We recommend the Air Force undertake a large survey in scope and population. In scope, the survey should include (1) basic business and economic questions on non-competitive markets and (2) specific questions on the contracts associated with each program office. Based on our survey responses, most contracting personnel are perhaps familiar with basic business concepts, but do not have the requisite specialty knowledge. If that is the case, a larger survey could help identify those gaps and then give the necessary information. Moreover, the survey should target multiple program offices, or at least those involved in the five largest dollar programs.

Apart from surveys, we also recommend a detailed study on dual-sourcing. This is not a new recommendation. Rather, dual-sourcing has been extensively discussed in the operations academic literature and among DoD decision-makers. But, we suggest research that takes one product such as missiles in the Lyon (2006) article and follows its design and manufacturing history for the DoD to assess the potential for dual-sourcing. Such exercises may exist, and we are perhaps unaware of them. If they do exist, we should consider disseminating them widely so decision-makers are aware if there are precedents for dual-sourcing in their specific domain and how to incorporate them.

References

Adame, D., & Markling, M. R. (2018). *Acquisition strategy in a non-competitive environment: A resource dependency and power-dependent relations perspective.* Naval Postgraduate School.

Blair, R. D., & Harrison, J. L. (2010). *Monopsony in law and economics.* Cambridge University Press.

Day, D. T. (2012). *The limits of monoposony pricing power in the markets for defense goods.* Defense Acquisition University. https://apps.dtic.mil/dtic/tr/fulltext/u2/a580012.pdf

Eisenhardt, K. (1989). Agency theory—An assessment and review. *Academy of Management Review*, *14*(1), 57–74. https://doi.org/10.5465/AMR.1989.4279003

Emerson, R. M. (1962). Power-dependence relations. *American Sociological Review*, *27*(1), 31–41.

Lyon, T. (2006). Does dual sourcing lower procurement costs? *The Journal of Industrial Economics*, *54*(2), 223–252.

Klotz, D. E., & Chatterjee, K. (1995). Dual sourcing in repeated procurement competitions. *Management Science*, *41*(8), 1317–1327.

Mahoney, C. W. (2016). Buyer beware: How market structure affects cntracting and company performance in the private military industry. *Security Studies*, *26*(1), 30–59. https://doi.org/10.1080/09636412.2017.1243912

Merritt, Z. D. (2017). *Defense supply chain: DoD needs complete information on single sources of supply to proactively manage the risks* (GAO-17-786). Government Accountability Office.

National Defense Authorization Act for Fiscal Year 2018, Pub. L. No. 115-91, 131 Stat. 1283 (2017, December 12). https://www.congress.gov/115/plaws/publ91/PLAW-115publ91.pdf

Werber, L., Ausink, J. A., Daugherty, L., Phillips, B., Knutson, F., & Haberman, R. (2019). *An assessment of gaps in business acumen and knowledge of industry within the defense acquisition workforce: A report prepared for the U.S. Department of Defense in compliance with Section 843(c) of the Fiscal Year 2018 National Defense Authorization Act.* RAND.

Williamson, O. E. (1967). The economics of defense contracting: Incentives and performance. In R. N. McKean (ed.), *Issues in defense economics*. National Bureau of Economic Research. https://www.nber.org/chapters/c5165.pdf

PANEL 4. INNOVATING AND NURTURING THE ACQUISITION WORKFORCE

Tuesday, May 11, 2021	
10:45 a.m. – 12:00 p.m.	**Chair: Dyke Weatherington,** Deputy Assistant Secretary of Defense, I&IPM ***Optimal Long-Run Talent Management of the Department of Defense Acquisition Workforce in Response to COVID-19: A Dynamic Programming Approach*** Tom Ahn, Naval Postgraduate School Amilcar Menichini, Naval Postgraduate School ***An Innovative Approach to Assessing DoD Contracting Workforce Competency*** Rene Rendon, Naval Postgraduate School ***Aligning DOD Program Management Competencies with the Project Management Institute Standards*** Robert Mortlock, Naval Postgraduate School Jonathan Karnes, 1st Lt, USAF, NPS Graduate Student ***An Analysis of Turnover Among the Civil Service Components of the Department of Defense Acquisitions and Medical Workforces*** Spencer Brien, Naval Postgraduate School

Dyke Weatherington—is the Deputy Assistant Secretary of Defense, Information & Integration Portfolio Management (I&IPM) in the Office of the Under Secretary of Defense for Acquisition and Sustainment (OUSD(A&S). He manages and is responsible for acquisition shaping, analysis, and oversight of Warfighter capability portfolios across the Department of Defense (DoD) in the nuclear weapons systems; nuclear command, control, and communications; missile defense; and space domains.

He leads assessments of cost, schedule, and performance risks of acquisition programs, and works directly with the Services, the Intelligence Community, and the Office of the Under Secretary of Defense for Research and Engineering to address identified gaps in program strategies. His DoD portfolio includes Ground Based Strategic Deterrence, the Long Range Stand-Off weapon, GPS Enterprise, Space Launch, Space Control, and Missile Defense. His Intelligence Community portfolio includes major system acquisition programs of the National Reconnaissance Office, National Geo-Spatial Intelligence Agency, National Security Agency, and Defense Intelligence Agency. He serves as the I&IPM Senior Acquisition Officer and the primary liaison between Joint Staff, Services, Agencies, and Congress, facilitating actions to achieve cost, schedule, and performance goals and advising the Milestone Decision Authority on program acquisition decisions.

Mr. Weatherington's prior duties included Deputy Director, Intelligence, Surveillance and Reconnaissance in the OUSD(A&S) Space, Strategic, and Intelligence Systems (SSI) directorate. Mr. Weatherington was also the OUSD(Acquisition, Technology, and Logistics) functional lead for the Defense Space Council. Prior to his assignment in SSI, Mr. Weatherington was the Deputy Director, Unmanned Warfare and ISR in the Strategic and Tactical Systems directorate.

Mr. Weatherington holds a Bachelor of Science degree in engineering mechanics from the United States Air Force Academy (1981) and a Master of Arts in National Securities Studies from California State University (1993). He is also a graduate of the Air Force Air Command and Staff College and the Defense Systems Management College. He has been awarded numerous OSD and Air Force decorations including the Airman's Medal, OUSD Exceptional Civilian Service Award, and FY17 Presidential Rank Award.

Mr. Weatherington was born and raised on his family's farm near Burnside, Illinois, and is married with four children. He resides with his family in Northern Virginia.

Optimal Long-Run Talent Management of the Department of Defense Acquisition Workforce in Response to COVID-19: A Dynamic Programming Approach

Tom Ahn—Naval Postgraduate School [sahn1@nps.edu]

Amilcar Menichini—Naval Postgraduate School [aamenich@nps.edu]

Abstract

As the economic impact of the COVID-19 pandemic lingers, with the speed of recovery still uncertain, the state of the civilian labor market will impact the public sector. Specifically, the relatively stable and insulated jobs in the Department of Defense (DoD) are expected to be perceived as more attractive for the near future. This implies changes in DoD worker quit behavior that present both a challenge and an opportunity for the DoD leadership in retaining high-quality, experienced talent. We use a unique panel dataset of DoD civilian acquisition area workers and a dynamic programming approach to simulate the impact of the pandemic on worker retention rates under a variety of recovery scenarios. We find that workers will choose not to exit from the DoD while the civilian sector suffers from the impact of the pandemic. This allows leadership to more easily retain experienced workers. However, once the civilian sector has recovered enough, these same workers will quit at an accelerated rate, making gains in talent only temporary. These results imply that while the DoD can take short-run advantage of negative shocks to the civilian sector to retain and attract high quality workers, long-run retention will be achieved through more fundamental reforms to personnel policy to make DoD jobs more attractive, no matter the state of the civilian labor market.

Introduction

The initial impact of the COVID-19 pandemic on the U.S. civilian labor market was massive, with unemployment spiking to 15% in September 2020. While most world economies contracted in 2020, there is some consensus among economists of a relatively robust recovery in the near future, with average global economic growth projected to be about 5.5% in 2021 (International Monetary Fund, 2021). In the United States, the unemployment rate has already recovered partway since the nadir. However, the trajectory of recovery remains unclear, dependent on a host of public health programs, government stimulus, and the macroeconomic environment.

While the civilian labor market has seen extraordinary swings in employment numbers, the government sector has been somewhat immune to the short-term effects of the pandemic. We examine the potential impacts of the gyrations and continuing uncertainty in the civilian labor market on the labor market decisions of public-sector workers, focusing on the civilian acquisition workforce (AWF) in the Department of Defense (DoD). While senior DoD leadership has historically been concerned with losing qualified senior civilian workers to the private sector, the labor market impact of COVID-19 may present a pressing need to adjust personnel policy as well as an opportunity to leverage the stability of DoD positions to compete against the draw of the private firms.

We solve a dynamic programming model of worker attrition behavior, where long-lasting shocks in the civilian labor market are explicitly modeled. In particular, the model allows for a negative AR(1) shock to the civilian sector, which slowly recovers through time. After calibrating the model parameters to the AWF using a unique panel administrative personnel dataset that tracks the civilian DoD labor force over the span of thirty years, we simulate civilian-side labor market shocks that correspond to economic recoveries of varying speeds and forecast the retention behavior of the workforce.

We find that a persistent negative shock to the civilian sector (plus insulation of the government/DoD labor market from the shock), for our case, the COVID-19 pandemic, leads workers to devalue jobs in the private sector in the short-run and remain in the government sector for a longer period of time. Depending on the severity and persistence of the shock, it may take more than a decade for workers to return to valuing civilian jobs as they did before the pandemic. This relative increase in attractiveness of government jobs is only temporary, however, and workers will accelerate their exit from the government sector into the private sector once the economic recovery is well underway. That is, the attrition rate when the economy recovers turns out to be higher than the rate that would have prevailed had there not been the global pandemic.

The following section describes in more detail the labor market impact of COVID-19 on the private sector and the long-run career trajectories of the typical AWF worker. Then, we explain the dynamic programming model, while the section after that describes the dataset and calibrates the model parameters to the AWF data. Then, we simulate potential COVID-19 scenarios going forward, and project the attrition behavior of the workforce under differing scenarios of economic recovery. Finally, our paper concludes, and the Appendix explores the retention effects of a one-time bonus.

The Impact on Unemployment Arising from COVID-19

The short-run impact of COVID-19 has been extraordinary, with the unemployment rate spiking to almost 15% from near historical lows (3.5%) in 2 months. As Figure 1 shows, even during the Great Recession, the unemployment rate peaked at 10.6%. The Congressional Budget Office (CBO) projects that the U.S. economy will grow 4.6% in 2021, after contracting 3.5% in 2020. These are significantly upwardly revised estimates from its report in July 2020, when the CBO projected a growth rate of 4%. Correspondingly, employment has recovered sharply since September 2020 (Congressional Budget Office, 2021).

However, when the economy can return to business as usual and the vigor with which it can rebound remain unclear. Public health factors such as the efficacy of vaccines and their distribution, the discovery of more infectious variants of COVID-19, and sustained use of masks and social distancing until herd immunity is reached, will all play a role. In addition, the recovery of the rest of the world, additional federal, state, and local fiscal stimuli, as well as permanent changes in the economy such as expanded work-from-home and reconfiguration of global supply chains, may impact the private-sector labor market for years to come.

The impact of such changes to the private sector will inevitably affect the public sector, especially for the civilian workforce within the DoD. The uncertainty in the private sector and the comparably stable government sector is expected to alter their long-term career trajectories. Figure 2 shows the attrition rate of DoD AWF workers, reproduced from Ahn and Menichini (2019). The sample covers September 1987 to December 2018. Approximately 30% of workers leave the DoD after about 8 years of service. By approximately 25 years of experience, roughly three-quarters of employees have left. While these attrition rates are relatively low compared to many civilian industries, DoD leadership still expresses a desire to hold on to highly skilled, senior civilian workers.[5]

Figure 1. Civilian Unemployment Rate. (Bureau of Labor Statistics)

While the shock of COVID-19 has been felt in almost every sector of the labor market, the government sector has notably been shielded from the worst of the impact. As Figure 3 shows, as of November 2020, government workers experienced an unemployment rate around 4%, which is lower than workers in the education and health services fields, which have received much wider media coverage of labor shortages due to the health risks from their proximity to the pandemic.

While job stability has always been a draw for the government sector, the state of the economy as well as the continuing uncertainty about the speed of economic recovery, should make jobs in the DoD relatively much more attractive. Indeed, this argument is parallel to what has been known for a long time in military recruiting: demand for military jobs is countercyclical to the state of the civilian economy. With the backdrop of this large, negative, persistent, and unpredictable shock to the civilian labor market, we model the long-run labor market decisions of civilian DoD employees using a dynamic programming framework.

[5] For example, "Highly educated, skilled, and experienced government acquisition professionals are vital now and, in the future, to provide warfighters the products they need" (DoN, 2018). Or, "All this relies on our most important asset, our people, and the approaches we take to recruit, train, and retain the workforce we need to compete and win in support of our national defense strategy." - The Hon. James F. Geurts, ASN (RD & A)

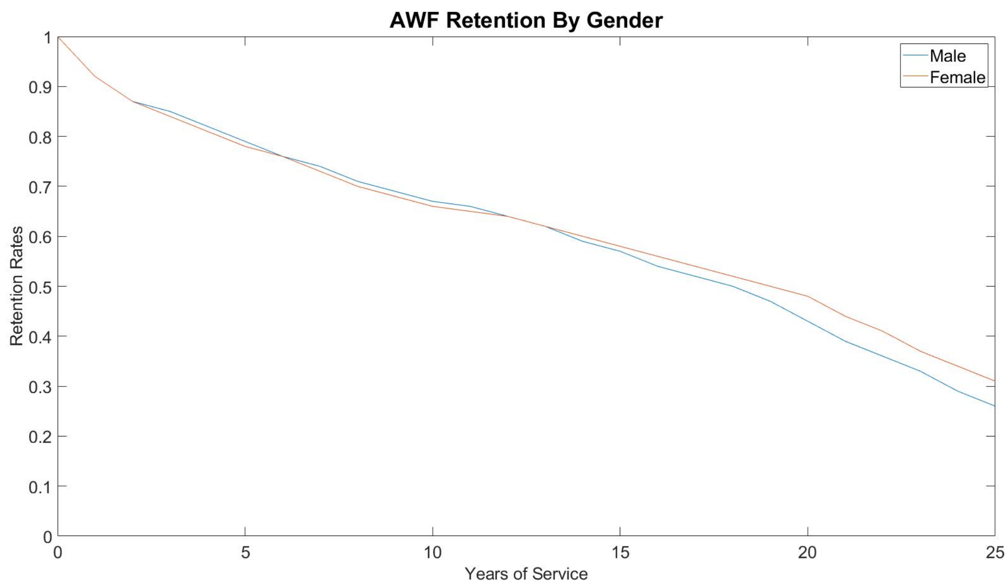


Figure 2. Career Trajectories of DoD AWF Employees. (Ahn & Menichini, 2019)

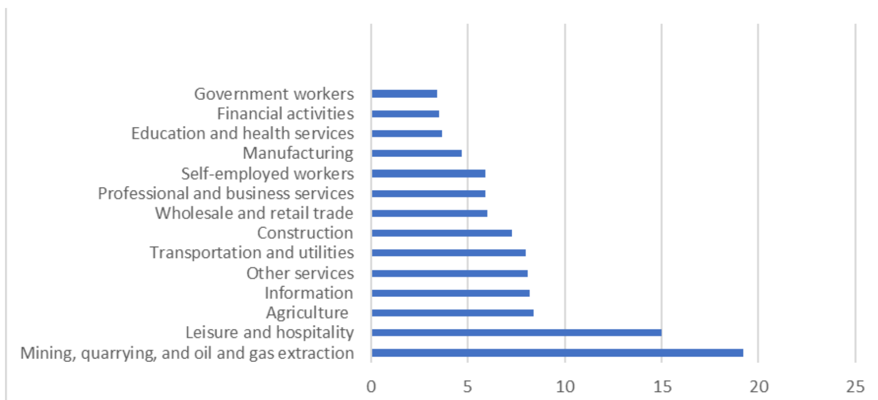


Figure 3. Unemployment Rate by Sector, November 2020. Bureau of Labor Statistics)

Model

In this section, we describe the different parts of the dynamic programming model of employee attrition that will be used to produce policy simulations.

We assume AWF workers are rational decision-makers who make career choices in order to maximize utility over their lifetime. The individual evaluates, at each decision point, all the costs and benefits involved in each possible choice, including pecuniary as well as non-pecuniary elements, which we describe here. At the beginning of each period (i.e., 1 year in this paper) the worker chooses between leaving the AWF to continue their career in the private sector, and staying in the public sector one more period.[6]

We next describe all the costs and benefits (including monetary and non-monetary elements) that the individual trades off at every decision point. We assume that the pecuniary components include

- AWF compensation, including basic pay, health insurance, locality adjustment, bonuses, etc.
- Compensation in the private sector.

We also assume the AWF employee is included in the Civil Service Retirement System (CSRS), and model public retirement accordingly.[7] For employees working in the private sector, we assume they are contributing to a 401(k) plan where the employer matches up to 10% of gross pay.[8]

The non-pecuniary components refer to the individual's taste or preference for a job in the AWF versus a career in the private sector. These components attempt to capture the taste of those agents who prefer the higher predictability and stability of public sector employment, even at the cost of a lower salary compared to the private sector, and vice versa. To capture these relative preferences, we use taste parameters reflecting monetary-equivalent preferences for careers in the private versus the public sectors.

In particular, we use the following notation to construct the dynamic model:

- W_t^m indicates compensation in the AWF (including all pecuniary components) in period *t*
- W_t^c denotes compensation in the private sector in period *t*
- ω^m is the public sector taste parameter, which captures the monetary-equivalent preference for a career in the AWF
- ω^c is the private sector taste parameter, which captures the monetary-equivalent preference for a private sector career
- T denotes the labor time horizon (number of working periods before final retirement)
- $\beta = \frac{1}{1+r}$ is the discount factor, where r represents the subjective discount rate
- $E[.|\varepsilon_{t-1}]$ indicates the expectation operator, given the shock in the previous period
- ε_t^m and ε_t^c are the random shocks affecting government and civilian jobs, respectively, in period *t*

The maximization problem faced by the AWF worker can be described by the following set of equations:

---

[6] We further assume that leaving the AWF is an irreversible decision.

[7] The dataset contains employees from both the extinct CSRS and the current Federal Employee Retirement System (FERS). We model the CSRS because there are more individuals belonging to that system than FERS.

[8] As we note in the data section, the modal AWF employee has a bachelor's degree or above and earns close to $100,000 at their highest paygrade. Workers with these characteristics in the civilian sector most often have employer matching 401(k) options.

$$V_t^L = W_t^c + \omega^c + \beta E[V_{t+1}^L | \varepsilon_t^c] + \varepsilon_t^c \text{ (1)}$$

$$V_t^S = W_t^m + \omega^m + \beta E[V_{t+1} | \varepsilon_t^c, \varepsilon_t^m] + \varepsilon_t^m, \text{ and (2)}$$

$$V_t = Max[V_t^L, V_t^S] \text{ (3)}$$

In these equations, super-index *S* denotes the agent's choice to continue working one more period in the AWF (i.e., *S* = Stay). Alternatively, super-index *L* indicates the individual's choice to quit the AWF job to continue their career in the private sector (i.e., *L* = Leave). Therefore, V_t^S denotes the (present) value of remaining in the public sector one more period, while V_t^L indicates the (present) value of switching to the private sector. Equation 3 implies that the individual will decide to be part of the AWF force in every period in which $V_t^S > V_t^L$, and will leave the force as soon as the opposite is true.

Regarding stochastic variables ε_t^m and ε_t^c, we assume they are independent and mean reverting over time (*t* dimension). The specification of the random shocks is the following:

$$\varepsilon_t^c = \mu_c + \rho_c \varepsilon_{t-1}^c + \tau_t^c, \tau_t^c \sim N(0, \sigma_c^2) \text{ (4)}$$

$$\varepsilon_t^m = \mu_m + \rho_m \varepsilon_{t-1}^m + \tau_t^m, \tau_t^m \sim N(0, \sigma_m^2), \text{ and (5)}$$

$$\tau_t^c \text{ independent of } \tau_t^m \text{ (6)}$$

That is, the shocks evolve independently of each other, oscillating around their own long-run (unconditional) mean over time. In the context of equations 1–3, these innovations could be interpreted as random shocks to salaries in the civilian and private sectors (i.e., W_t^m and W_t^c, respectively) stemming from, for instance, fluctuations in the business cycle. Ashenfelter and Card (1982) find that nominal wages are well represented as AR(1) processes. Accordingly, equations 4 and 5 define AR(1) representations for the error terms. These AR(1) processes play an important role for our main results as they allow shocks to persist over time; that is, to gradually fade as time passes.[9] As we explain in more detail later, we use parameter ρ to define the speed at which the economy (and wages) recovers from a shock (such as from the COVID-19 outbreak). In terms of the optimization problem described in equations 1–3, random shocks ε_t^m and ε_t^c indicate state variables observed by the AWF worker at the time of the decision.

Data Description and Model Calibration

In this section, we describe the AWF sample as well as the selection and calibration of the parameter values necessary to implement the dynamic programming model described previously. In the next section, we show those parameters provide a good approximation of the long-run labor market outcomes for the representative worker in the AWF.

[9] As opposed to white noise processes, where shocks do not persist over time (i.e., they return to the mean immediately), or random walk processes, where shocks do not return to the mean.

DATA: The Acquisition Workforce

The DoD Acquisition workforce is comprised of approximately 150,000 employees, covering the period September 1987–December 2018. Civilians make up about 90% of the workforce, while active duty makes up the remaining 10%. The AWF's mission is the "timely and cost-effective development and delivery of warfighting capabilities to America's combat forces" (DoD, 2015). The AWF is responsible for overseeing equipping and sustaining the military, spending over $1 trillion in FY2021. About 26% of the AWF belongs to the engineering career field, followed by contracting at 19%. Historically, the AWF was sharply reduced in size and capability during the 1990s. The DoD has been working to rebuild the AWF starting in 2008, increasing the AWF by approximately 30,000 employees over 7 years.

For this analysis, we restrict our sample to workers who were ever in the contracting, industrial property management, or purchasing fields.[10] Our sample workers were born after January 1, 1950, but before December 31, 1980. Workers with birthdates outside this range are either too old, in that the environment in which they made their labor decisions may not be reflective of current jobs in the AWF, or too young, in that these workers have not had time to make labor decisions that are pivotal to their careers. Restricting the sample nets us over 2 million worker-month records, with over 13,000 unique workers tracked through their careers. Table 1 presents some summary statistics for our sample.

The workforce is predominantly white and female. Over half the workforce has a bachelor's degree or above. Compared to the civilian sector, careers in the AWF is stable, with the average career length lasting well over a decade. This workforce is also highly paid, with the average worker earning almost $100,000 toward the end of their career. The average worker in this sector begins her/his career at age 33, which indicates that the position in the AWF is not her/his first job. In fact, a large number of these workers have prior military experience.

To rigorously assess the impact of the civilian sector on the attractiveness of the DoD position, every worker in the dataset must be "assigned" a civilian wage that they can expect to earn. To accomplish this, we estimate a hedonic regression using the Outgoing Rotation Group (ORG) of the Current Population Survey (CPS). As this dataset contains a representative sample of workers in the United States, including, most importantly, those who are in the government sector, it is possible to make an apples-to-apples comparison with workers in the private sector.[11]

We run a hedonic regression using the individual sociodemographic characteristics, professional and education experience, and locality indicators from the ORG of the CPS which broadly match the AWF variables summarized in Table 1 to obtain predicted civilian and government sector wages. The difference in the wages across private and public sectors, conditioned on individual characteristics, defines the government sector "wage penalty."

---

[10] These fields correspond to Occupation Codes 1102, 1103, and 1105, respectively.

[11] See Ahn and Menichini (2020) for a detailed description.

Table 1. Summary Statistics

Variables	Mean	Std. Dev.	Min	Max
Female	0.632			
Minority	0.278			
Disability	0.202			
Prior Military Service	0.619			
Has Bachelor's Degree	0.547			
Has Post-graduate Degree	0.332			
Gained Additional Education in AWF	0.441			
Career Length in AWF (in years)	12.0	(8.6)	0.1	25.8
Age at Entry	33.0	(8.2)	15	65
Age at Exit	48.2	(10.55)	20	68
Position Type: Professional	0.657			
(Ever Held) Technical	0.245			
Blue-Collar	0.018			
White-Collar	0.297			
Ever Rated Not Fully Satisfactory	0.575			
Highest Salary	95,143.67	(30,410.74)	27,397	189,600
Observations	13,590			

Calibration Results

Before simulating the model described in equations 1–3, we need to start defining the parameter values, which we show in Table 2 and subsequently describe. We can observe in Table 2 that all parameter values, except compensation, are constant over the career of the AWF employee.

As we described in the section Data: The Acquisition Workforce, estimates from the hedonic regressions suggest that income in the private sector (i.e., W_t^c) is, on average, around 17.61% higher than in the AWF (i.e., W_t^m) for individuals with similar characteristics. For this reason, after initially normalizing W_t^m=1, we let W_t^c=1.1761. We then add the income from the different retirement systems and, thus, compensation changes over time. The data described in that same section show that the longest observed labor time horizon among all individuals is 25 years. For that reason, we let T=25. The subjective discount factor is assumed to be 0.95, implying an interest rate of 5.26%.[12]

Table 2. Parameter Values

Parameter	*Value*
W_t^m	1
W_t^c	1.1761
T	25
β	0.95
ω^m	1.2782
ω^c	1
μ_m	0
μ_c	0
ρ_m	0.90
ρ_c	0.90
σ_m	0.005
σ_c	0.005

Regarding the taste parameters, we calibrated parameter ω^m so that the survival curve predicted by the model approximates the empirical survival curve as closely as possible via grid search (we show the graphical results of this calibration in the next section). In more technical terms, the calibration exercise searches for the value of ω^m that minimizes the summed squared distance between the points of the empirical AWF survival curve and the points of the survival curve predicted by the model. As Table 2 displays, we normalize $\omega^c = 1$ and, from the calibration exercise, we obtain $\omega^m = 1.2782$.[13] These values imply that the representative AWF employee prefers the AWF over the private sector.

[12] This interest rate is similar to the average 30-Year T-Bond Constant Maturity Rate reported by the Federal Reserve Bank of St. Louis for the period covered by the data set.

[13] Ahn and Menichini (2020) estimate a similar dynamic model where economic shocks to the civilian and public sectors are i.i.d. with mean zero. They find the difference between military and private sector taste parameters (i.e., $\omega^m - \omega^c$) to be around 0.2, which, reassuringly, is not far from the result of the calibration exercise.

The remaining parameter values in Table 2 refer to the stochastic process of the error terms ε_t^m and ε_t^c. We follow Ashenfelter and Card (1982) to define the parameter values that govern the AR(1) processes of those terms. Accordingly, we let parameters μ_m and μ_c be equal to zero, we assume values of 0.005 for the standard deviation of the errors, σ_m and σ_c, and let the mean-reversion coefficients ρ_m and ρ_c be equal to 0.9. These values depict the historical behavior of the error terms. In particular, those observed values of the mean-reverting coefficients suggest that wages have a high level of persistence over time and, thus, that the effects of shocks require a long time to disappear.

Model Solution and Policy Simulations

In this section, we describe our policy simulations to forecast the evolution of the behavior of the representative AWF worker under a number of scenarios with differing speed rates of economic recovery from a large, abrupt, and unanticipated negative impact (i.e., COVID-19) to the private sector. This is a major systematic event that adversely affects all sectors of the economy, except for the public or government sector, which we assume keeps its employment constant.[14] The latter is consistent with the assumption of independent errors in equation 6.

Concisely, we “shock” the model with a large negative civilian error draw at a specified point in time. Then, we allow the system to recover and converge back to the steady state. We start analyzing retention behavior assuming the economy recovers according to the empirical historical speed. However, given the observed recovery from the current pandemic seems to be, so far, much faster than normal, we also study the retention implications of different scenarios for the speed of recovery. We “control” the speed of recovery of the economy by setting the autoregressive term, ρ, which controls the velocity at which shocks gradually disappear over time.

While the private sector goes through its gyrations, at every period, the representative AWF agent in our model surveys the current state of the private sector, forecasts the evolution of the state of the economy, and makes the *ex ante* optimal decision to stay or leave the AWF. We describe the simulation procedure in more detail next.

We solve the model described in equations 1–3 via backward induction (see Rust [1987] for an empirical treatment). That is, we start from the final period (i.e., $t=T=25$) and decide whether to stay one more (final) period in the AWF or to leave for the private sector. We then move one period backward (i.e., $t=24$) and select to stay one more period or to leave the AWF, considering the value from the optimal decision in period $T=25$. We continue moving backward, deciding rationally in every period, until we reach the present period (i.e., $t=0$). This solution characterizes the retention behavior of a representative AWF employee in all possible states of the economy.

We then stochastically simulate the model forward (i.e., over the 25 years of work) 100,000 times, which produces the stay/leave decisions of 100,000 employees in all possible different situations over the labor period. These simulations summarize the retention behavior of the representative employee, which we show in Figure 4. The figure

---

[14] While our negative shock is the COVID-19 pandemic, any future unanticipated national shock to the economy and/or public health that is concentrated in the private sector can be expected to operate in a similar manner.

exhibits the calibrated, model-predicted survival curve of the representative individual (blue line) and displays the cumulative probability of the worker staying in the force after a certain period of time. For example, the figure suggests that the likelihood that the employee is still part of the AWF after 10 years is about 65%. The figure also shows the empirical survival curve for the AWF employees (red line) from the data described in the section Data: The Acquisition Workforce, suggesting that the calibrated model predicts actual behavior quite closely.

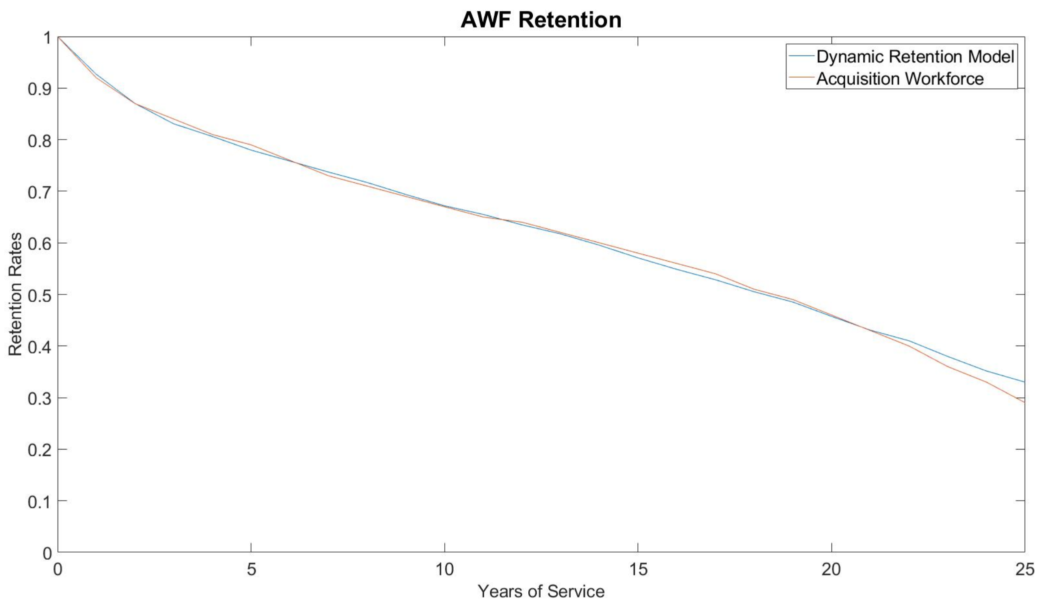


Figure 4. Retention Behavior

Associated with the previous survival curves are the yearly, model-predicted probabilities of leaving the AWF, which we show as the blue line in Figure 5. The attrition rate is relatively low every year, as is shown by the fact that the likelihood of leaving is always below 10% per year, and below 5% in the great majority of years. In addition, the attrition rate is high initially, and diminishes through time before increasing again toward the end of the individual's career. For instance, the probability that the employee departs from the AWF in year 10 is around 2%. As before, we also show the empirical likelihood of leaving (red line) for comparison purposes.

We then proceed to shock the model with a large negative error on the civilian side (i.e., ε_t^c) at year 10. The shock is equivalent to 3 standard deviations below the mean and is intended to capture the large effect of the sudden appearance of COVID-19. In economic terms, given the calibration shown in Table 2, this shock could be interpreted as a roughly 1.5% reduction in the civilian salary, W_t^c, while the public sector salary, W_t^m, remains unchanged. The fact that the error terms (both ε_t^m and ε_t^c) are mean reverting over time implies that the impact of the negative shock on the civilian salary gradually disappears as time passes. As mentioned before, the speed of return to the pre-shock state will depend on the mean-reversion coefficient, ρ.

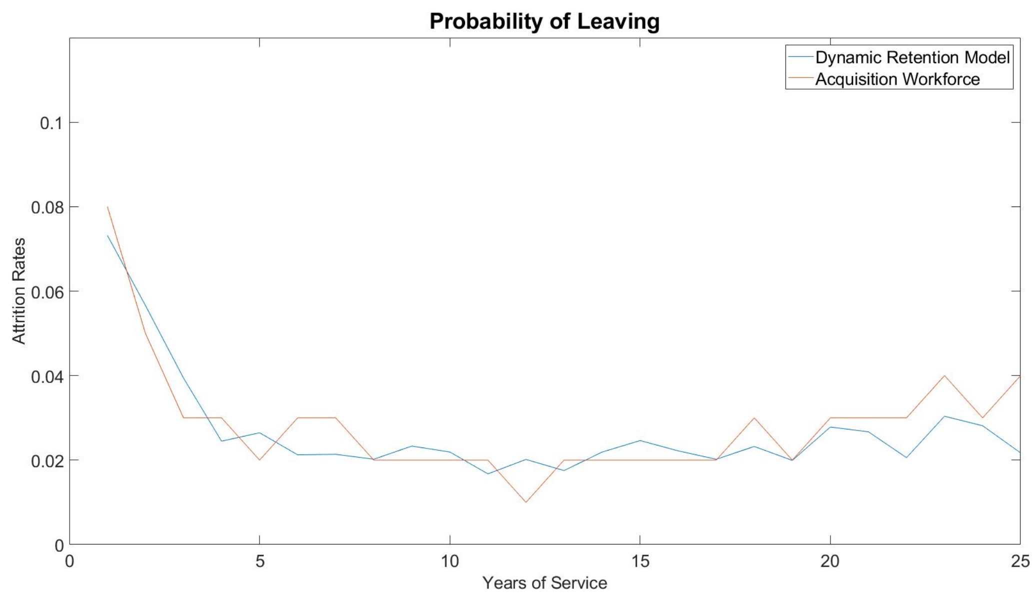


Figure 5. Probability of Leaving

In Figure 6 we show, given the initial negative shock, how the shocks are expected to evolve over time for four different values of the coefficient of mean-reversion. The purple bars depict the historical case, which is based on the observed historical mean-reversion coefficient of ρ=0.9. It is clear that the historical coefficient implies it would easily take a decade or more to return to normality. However, after a year of the appearance of the virus, the economy seems to be recovering much faster than suggested by historical terms. We attempt to capture the faster rebound by reducing the coefficient of mean-reversion (i.e., via a quicker dissipation of the shock). Accordingly, we analyze three different scenarios featuring dissimilar speeds of recovery, all of which are faster than the historical speed. Scenario 1, with the blue bars and ρ=0.3, represents the case of a relatively faster return to the pre-COVID economy. On the other hand, the yellow bars in scenario 3, with ρ=0.7, reflect a slower recovery to normality as compared to scenario 1. In between are the red bars of scenario 2, showing an intermediate speed of recovery with ρ=0.5. Even in the more optimistic recovery scenario 1, it is clear that the effects of the large negative shock remain in place for some years.[15]

[15] The magnitude and persistence of shocks are speculative, although they are informed by very recent (and ongoing) research. Many scholars are currently attempting to forecast the long-run impact of COVID-19 on the economy. See Petrosky-Nadeau and Valetta (2020), for example.

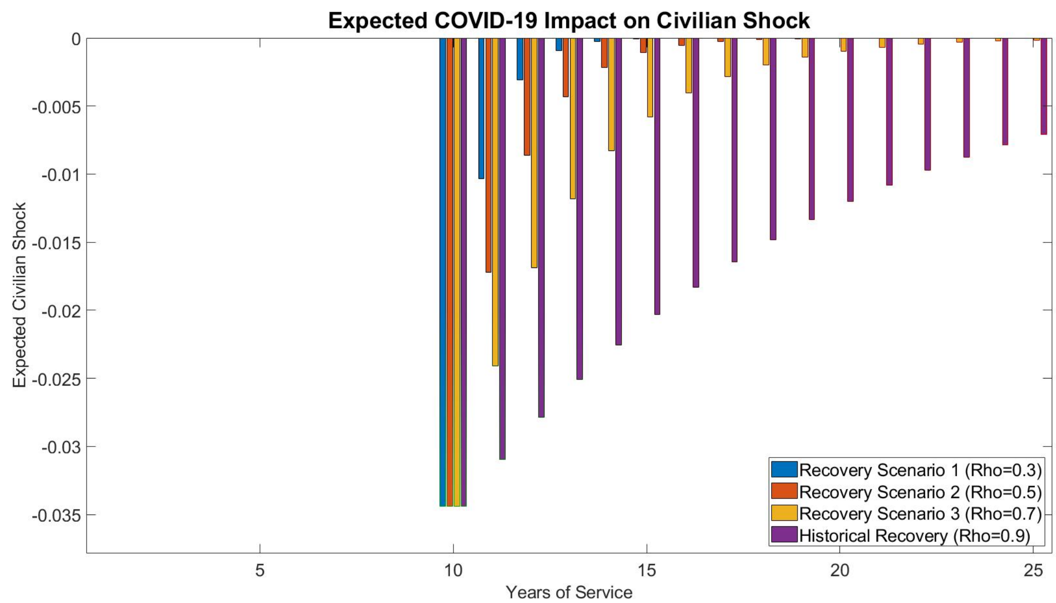


Figure 6. Expected Impact of COVID-19 on Civilian Shock

The effect on retention behavior of the representative AWF worker can be observed in Figure 7. The figure shows that, during the initial 10 years, the retention behavior is equivalent to the blue line in Figure 4. At year 10, the COVID-19 shock happens, and the retention behavior changes considerably. As we mentioned before, we study the attrition behavior in four different contexts. The green line shows the retention impact of the virus under historical terms (i.e., ρ=0.9). The other lines depict the expected retention behavior for three faster rates of economic recovery (i.e., ρ=0.3, ρ=0.5, and ρ=0.7 for recovery scenarios 1, 2, and 3, respectively). In all cases there is a kink and sudden flattening of the curve, suggesting that the individuals stay longer in the AWF, in an attempt to avoid the sharp negative effect of the virus shock on the civilian labor market. Depending on the speed of recovery, it might take a substantial amount of time for the employee to return to the pre-shock retention behavior. For instance, in the historical case it takes around 10 years for the representative employee to return to the pre-virus retention behavior, while in scenarios 1, 2, and 3, the return to normality takes roughly 2, 3, and 5 years, respectively. These long-lasting effects on retention behavior have important implications for the hiring policies of the public sector.

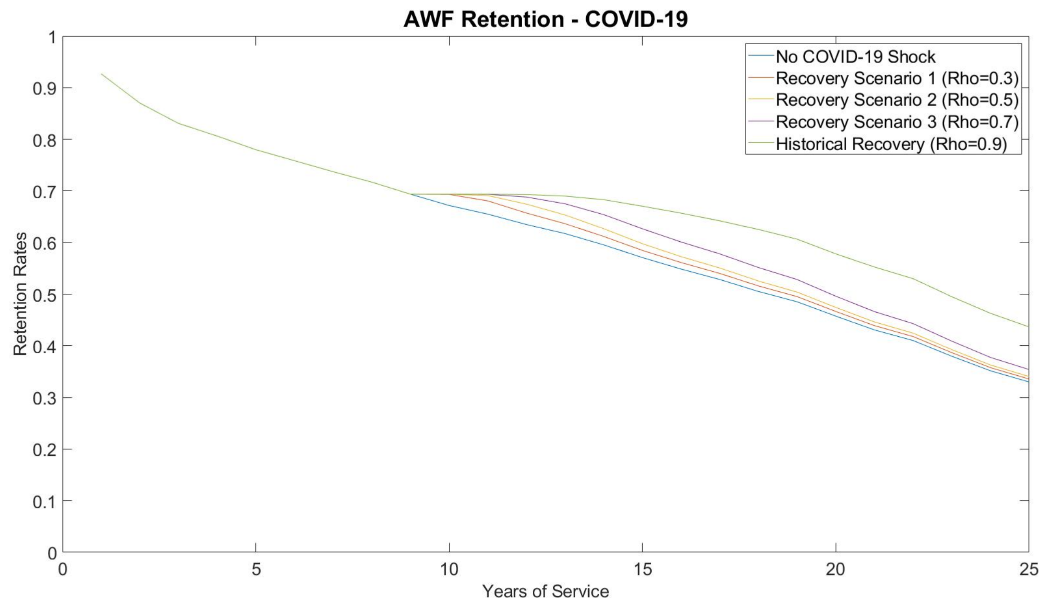


Figure 7. Retention Impact of COVID-19

It is worth noting that the time required to return to the "original" behavior specified above does *not* mean that all workers will choose to delay leaving the AWF by several years due to the impact of COVID-19. Instead, all workers will process the negative shock in the civilian economy as making the AWF job more attractive. Until the shock fully dissipates, the DoD position will be more attractive than had there not been the global pandemic. However, given the substantial wage premium in the civilian sector, the pandemic shock does not need to completely disappear before workers who were planning to move to the civilian sector resume their plans.

To complement the analysis of the return to the pre-COVID context, we present Figure 8. The figure shows the model-predicted yearly probabilities of leaving the AWF for the four different values of parameter ρ. The green line shows the attrition behavior in the historical recovery scenario, confirming that it takes around 10 years to return to the pre-COVID retention behavior (the latter is represented by the no-COVID-19-shock blue line). The red, yellow, and purple lines, reflecting faster speeds of economic rebound, suggest that around 2, 3, and 5 years, respectively, are required to eliminate the effects of the COVID-19 shock on retention. In all four scenarios, the likelihood of leaving the AWF goes roughly to zero in the year of the shock, and then slowly starts to return to the no-shock levels as time passes and the effects of the shock dissipate.

It is also important to note that, after the return to normality, the probability of leaving is higher in the slower recovery scenarios and lower in the faster rebound scenarios. More generally, after the COVID-19 shock dissipates, in all cases with shock, the likelihood of leaving is higher than in the no-shock case, with that probability increasing in parameter ρ. Indeed, the slower the recovery from the pandemic (i.e., higher ρ value), the larger the magnitude of exit probability after the recovery. This outcome suggests that, as more people decide to stay longer in the AWF during the pandemic, when the economy returns to normal, the pent up demand to leave for the private sector is expressed as higher attrition rates in the later years. This implies an opportunity as well as a problem for the AWF leadership. While a slower recovery may induce more employees to stay longer, it cannot be a permanent solution to retain high ability workers. A higher ρ will result in a much sharper exit of workers from the AWF once the civilian economy recovers.

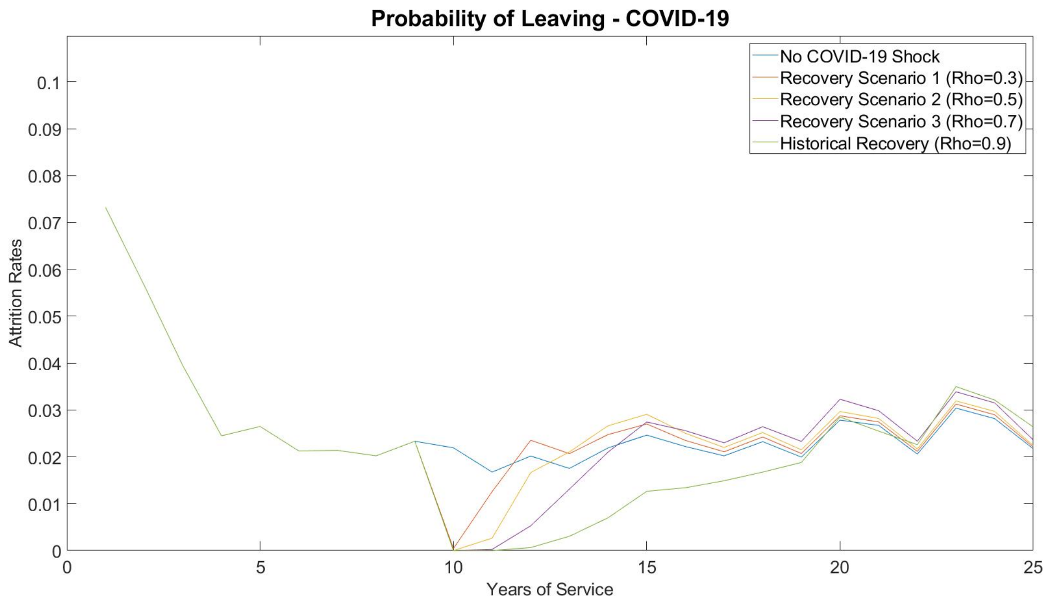


Figure 8. Likelihood of Leaving with COVID-19

In order to retain these workers, fundamental (and traditional) personnel policy reforms will be required. For example, a pay increase or expansion of benefits before the civilian sector fully recovers may permanently induce senior workers to remain in the AWF. Similarly, a one-time retention bonus, set far enough into the future when the civilian economy is back to normal, could prevent that exit.[16]

Conclusions

As of early 2021, the overall unemployment rate in the U.S. stands at 6.7%, an 8 percentage point decrease in just 8 months from the worst unemployment rate in almost 90 years arising from the COVID-19 global pandemic, yet still almost double the unemployment rate from just 1 year ago. While the recovery has been as dramatic as the decline, the future remains very much in doubt. For example, in December 2020, payrolls shrank by 140,000. Outlook has considerably brightened since, but the whiplash in long-run forecast of economic recovery itself adds uncertainty to future labor market prospects in the civilian market.

In this environment, we analyzed the potential impact of the economic recovery on the labor market trajectory of the AWF. The contrast in stability of jobs in the government compared to the private sector should increase the attractiveness of DoD jobs, especially if

[16] While it is beyond the scope of this paper to calculate optimal policy to retain workers as the civilian sector recovers, Figure 9 in the Appendix shows the attrition rates of AWF workers with a) no change in compensation after the COVID-19 shock and b) a one-time bonus of 25% of monthly salary at the 25 year of service mark. The bonus induces experienced workers to remain longer in the AWF.

the recovery proves to be slow or unpredictable. We built and calibrated a dynamic programming model of employee retention behavior, analyzed the impact of a negative persistent shock to the civilian sector, and simulated different recovery paths.

Our results show that government positions become more attractive the larger the magnitude of the negative shock to the civilian economy, and the slower the economic recovery, such that workers may value government positions more highly compared to the pre-pandemic period for several years.

While this environment can reduce attrition of the average worker from the AWF, leadership should understand that, eventually, recovery of the civilian sector will push down the relative desirability of government jobs. This may lead to a speedy exodus of many senior-level workers who were being held back due to economic uncertainty. Personnel planning without considering the temporary reduction in attrition at the beginning of the shock may lead to over-hiring, especially at the junior-levels. On the other hand, short-sighted reductions in hiring due to the initial impacts of the negative shock may lead to a hollowing out of the workforce, once the impact of the shock wanes. In addition, as the economy recovers, there may be fundamental structural changes to the labor market that remain, changing the valuation of both government and private sector jobs in unpredictable ways. Forward-looking leaders should regard these simulation results not as predictions of the future, but as guides to help set personnel policies that are flexible enough to adjust to and even take advantage of the gyrations in the civilian economy.

References

Ahn T., & Menichini, A. (2019). *Retention analysis modeling for the acquisition workforce* (NPS-HR-20-001). Naval Postgraduate School, Acquisition Research Program.

Ahn T., & Menichini, A. (2020). *Retention analysis modeling for the acquisition workforce II* (NPS-HR-21-031). Naval Postgraduate School, Acquisition Research Program.

Ashenfelter, O., & Card, D. (1982). Time series representations of economic variables and alternative models of the labour market. *The Review of Economic Studies*, *49*(5), 761–781.

Congressional Budget Office. (2021). *An overview of the economic outlook: 2021 to 2031*.

Defense Manpower Data Center. (2019). *APF civilian master file*.

DoD. (2015). *DoD acquisition workforce strategic plan: FY 2016–FY 2021*.

DoN. (2018). *DoN: Acquisition workforce FY19–24 strategic plan*.

International Monetary Fund. (2021, January). *World economic outlook update*.

Petrosky-Nadeau, N., & Valetta, R. (2020). *Unemployment paths in a pandemic economy* [Working paper]. Federal Reserve Bank of San Francisco.

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, *55*(5), 999–1033.

Appendix

As we describe in Section 5, the AWF will face a higher-than-normal employee attrition rate when the COVID-19 shock disappears and the economy fully recovers. In this appendix, we study one particular way by which that expected effect could be counteracted. In particular, we analyze the effect of a one-time bonus on the probability of leaving the AWF when the economy returns to normality. We assume the bonus is equivalent to 25% of the individual's monthly salary and is paid at year of service 25 (with the virus shock occurring at year 10). Figure 9 shows the main results of this exercise. The expected bonus has a fairly small effect on employee attrition in the early- and mid-career years, as the attrition rates are almost equivalent with and without the bonus. However, as expected, the effect of the bonus is more visible in the final years of the employee's career, when the economy has fully recovered from the COVID-19 shock. Without the bonus (red line), the attrition rates are substantially higher than with the bonus (yellow line), suggesting that, indeed, a bonus would induce experienced workers to stay longer in the AWF after the recovery. To finish, the bonus is just one of the tools available to the AWF to affect individual retention behavior (for instance, salary raises would be another useful tool).

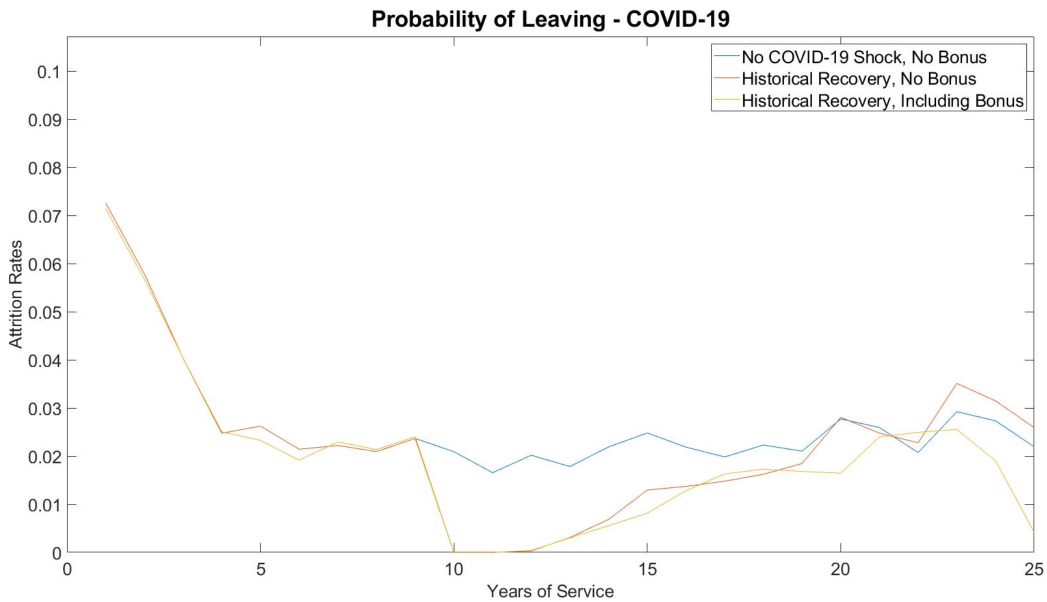


Figure 9. Likelihood of Leaving with COVID-19 and a Bonus at 25 Years of Service

An Innovative Approach to Assessing DoD Contracting Workforce Competency

Rene G. Rendon—is an Associate Professor and Acquisition Management Area Chair in the Graduate School of Defense Management at the Naval Postgraduate School (NPS). A retired Air Force Contracting Officer, he served as a Warranted Contracting Officer for major weapons system programs such as the Peacekeeper ICBM, F-22 Advanced Tactical Fighter, Space-Based Infrared Satellite program, and the Evolved Expendable Launch Vehicle program. He also served as a Contracting Squadron Commander for an Air Force pilot training base. He was presented with the Air Force Outstanding Officer in Contracting Award, and he has received the NPS Hamming Teaching Excellence Award and the Graduate School of Business and Public Policy Research Excellence Award. [rgrendon@nps.edu]

Brett Schwartz—is an active duty Supply Corps Officer and currently serves as a Senior Lecturer in the acquisition management area. He is the Academic Associate for the Master of Science in Contract Management program at the Graduate School of Defense Management at the NPS. He most recently served as a Professor of Contract Management at the Defense Acquisition University and the Director of Strategic Plans in the Office of the Assistant Secretary of the Navy for Research, Development, and Acquisition. Additional acquisition experience includes serving as the Deputy of Contracting for tactical aircraft at the Naval Air Systems Command and as Chief Contracting Officer at a forward operating base in Iraq during Operation Iraqi Freedom. [bmschwar1@nps.edu]

Abstract

The National Defense Authorization Act (NDAA) of 2020 directed the Secretary of Defense to implement a professional certification program for all members of the acquisition workforce that is based on standards developed by a third-party accredited program based on nationally or internationally recognized standards. In response to this NDAA (2019) requirement, the Department of Defense (DoD) senior procurement executives agreed to the establishment of a new contracting competency model and a single level of certification program based on the National Contract Management Association's (NCMA) Contract Management Body of Knowledge (CMBOK; NCMA, 2019a) and American National Standards Institute—accredited Contract Management Standard (CMS; NCMA, 2019b). The purpose of this research is to develop a new competency assessment instrument based on the NCMA CMBOK and CMS to be used in assessing the DoD's contracting workforce competency. This research will answer the following question: How can the CMBOK/CMS competency structure be used as the basis for developing a survey-based instrument for assessing the competencies of the DoD contracting workforce? An additional research question is: Based on the competency assessment results, in which contract management competencies is the workforce less proficient and less knowledgeable? We conduct this research by developing a survey-based assessment instrument for assessing the competencies of the DoD contracting workforce. We then deploy the assessment instrument to DoD contracting organizations and analyze the assessment results to identify contract management competencies that need additional training emphasis.

Introduction

The Government Accountability Office (GAO, 2019) continues to list contract management as a high risk and has done so since 1992. Additionally, the Department of Defense (DoD) inspector general has identified contract management as a top DoD management challenge (Office of Inspector General [OIG], 2019). Both agencies identify the need for increased technical competency in the contracting workforce. Furthermore, recent research on organizational climate assessment on the DoD's contracting workforce indicates that competency management is a critical part of ensuring a trained and experienced contracting workforce (Rendon & Powley, 2017). Thus, how an organization's competency framework is structured may have a significant impact on the competence level of its workforce.

Recent legislative initiatives reflect Congress's concerns about the adequacy of the DoD's acquisition workforce training and competency. For example, the Fiscal Year 2016 National Defense Authorization Act (NDAA, 2015) Section 809 required the Secretary of Defense to establish an independent advisory panel on streamlining acquisition regulations. The Section 809 Panel stated that if the DoD is to achieve its acquisition workforce goals, it will need to prepare and develop its workforce differently (Scott & Thompson, 2019). The Section 809 Panel identified several recommendations for improving the professional development of the acquisition workforce. These recommendations included creating career paths for the contracting functional area that would include those technical competencies and key work experiences as reflected in industry standards. The Section 809 Panel also recommended that the DoD revise its contracting professional development programs to emphasize skills that are transferable across government and industry and focused on a defined set of qualifications connected to contracting positions. Additionally, the panel recommended that the DoD revise its contracting professional development programs to emphasize sufficient domain knowledge, emphasize professional skills, and provide a broad perspective to interact effectively with industry. Finally, the panel recommended that the DoD adopt a common contracting body of knowledge, which would also enhance communication and collaboration between government and industry (Scott & Thompson, 2019).

Even more recently, in the 2020 National Defense Authorization Act (NDAA, 2019), Congress directed the Secretary of Defense to implement a professional certification program for all members of the acquisition workforce that is based on standards developed by a third-party accredited program based on nationally or internationally recognized standards (NDAA, 2019).

Purpose of Research

Recent research has shown that the current DoD contracting competency model may not be sufficient in assessing today's contracting workforce (Rendon & Winn, 2017). Additionally, further research found that the National Contract Management Association's (NCMA) *Contract Management Body of Knowledge* (CMBOK; NCMA, 2019a) and the *Contract Management Standard* (CMS; NCMA, 2019b) may be more suitable and effective in assessing the contracting workforce competency in today's dynamic acquisition environment (Rendon, 2019). The purpose of this research is to develop a new competency assessment instrument based on the NCMA CMBOK and CMS to be used in assessing the DoD's contracting workforce competency. This research answers the following question: How can the CMBOK/CMS competency structure be used as the basis for developing a survey-based instrument for assessing the competencies of the DoD contracting workforce? An additional research question is: Based on the competency assessment results, in which contract management competencies is the workforce less proficient and less knowledgeable? Thus, the objective of the research is focused on adopting the CMBOK/CMS contracting competency model as the basis for assessing the DoD contracting workforce.

Methodology

The methodology for this research consists of two components. The first component is the development of a survey-based assessment instrument for assessing the competencies of the DoD contracting workforce. We draw from the workforce competency literature and survey development literature for this component (Rendon & Schwartz, 2020). The second component of the methodology is the deployment of the assessment instrument to DoD contracting organizations and analysis of the assessment results to identify contract management competencies that need additional training emphasis.

DoD Contract Management Workforce Competency

Recent research compared the DoD contracting competency model with the NCMA CMBOK/CMS (Rendon, 2019; Rendon & Winn, 2017). The CMS has received third-party accreditation by the American National Standards Institute (ANSI) as an ANSI-accredited standard. The CMBOK/CMS is used by both government agencies and industry organizations for managing contracts. The research found that the CMBOK/CMS competency framework may provide an innovative approach for developing and assessing the DoD contracting workforce. The CMBOK/CMS's concise and detailed contract life cycle and greater emphasis and granularity in each of the life-cycle phases and tasks may help develop and fortify the DoD's contracting processes and practices. Providing greater emphasis on each of the contract life-cycle phases and organizing competencies using a hierarchical structure that aligns each competency with processes, tasks, and subtasks would support the development of a professional contracting career path that aligns contracting technical competencies and key work experiences (Rendon, 2019). The recent Section 809 Panel recommended that the DoD create career paths for the contracting functional area that would include such technical competencies and key work experiences as reflected in the CMBOK/CMS.

Additionally, expanding the DoD's contracting workforce knowledge to include industry's side of contracting (e.g., industry operations and processes) as reflected in the CMBOK/CMS will help in developing technical and professional skills that can transfer across government and industry, as well as improve communication and collaboration between government and industry. Including the industry side of contracting would also result in strengthening systems thinking within the contracting workforce (Carlson, 2017). The current DoD contracting competency model may be resulting in linear thinking among the contracting workforce, with contract managers believing that contracting problems have "direct causes and that you can optimize the whole by optimizing each of the parts" (Carlson, 2017). Contract managers using systems thinking will know that contract management "problems can have hidden, indirect causes" and it is the "relationships among the parts that matter the most" (Carlson, 2017). Adopting the CMBOK/CMS competency framework may provide the DoD contracting workforce with a stronger foundational understanding of not only the complete contract life cycle but also an understanding of the different perspectives in contractual relationships (e.g., buyer, seller, subcontractors, suppliers, end users, etc.). Using systems thinking, contract managers will be able to "see the gaps where complications or opportunities can arise" within the acquisition process and understand how their contracting decisions may impact contractors and subcontractors (Carlson, 2017). Including the industry competencies for the DoD contracting workforce may also strengthen "communication, collaboration, problem-solving, and adaptability" skills (Carlson, 2017). The Section 809 Panel recommended that the DoD revise its contracting professional development programs to emphasize skills that are transferable across government and industry and focused on a defined set of qualifications connected to contracting positions (Rendon, 2019).

Furthermore, there may be value in broadening the current DoD contracting competency model to include disciplines such as business management, financial management, project management, risk management, and supply chain management, as reflected in the CMBOK. The inclusion of these disciplines may enhance the DoD contracting workforce's critical thinking, problem solving, and analytical skills—bringing increased efficiency to its contracting processes (Rendon, 2019). The Section 809 Panel recommended that the DoD revise its contracting professional development programs to emphasize sufficient domain knowledge, emphasize professional skills, and provide a broad perspective to interact effectively with industry. A recent RAND study found that, within the defense acquisition workforce, knowledge gaps in business acumen, industry operations, and industry motivation exist. The RAND report also found that the lack of standardized definitions and competency model formats obscures the need for knowledge related to business acumen, industry operations, and industry motivation (Werber et al., 2019).

A greater understanding of these CMBOK/CMS disciplines, as well as understanding both government and industry sides of the contracting relationship, will help develop "T-shaped" acquisition professionals who have both "depth of knowledge in a particular expertise as well as have the ability to work and communicate across disciplines" (Carlson, 2017). T-shaped acquisition professionals will be capable of introducing innovation and process change into the DoD contracting processes. If the DoD would adopt the CMBOK/CMS as its competency framework, it would achieve a desired recommendation from the Section 809 Panel that both the DoD and industry adopt a common contracting body of knowledge, which would also enhance communication and collaboration between government and industry (Rendon, 2019). As previously stated, in the National Defense Authorization Act of 2020 (NDAA, 2019), Congress directed the Secretary of Defense to implement a professional certification program for all members of the acquisition workforce that is based on standards developed by a third-party accredited program based on nationally or internationally recognized standards.

Furthermore, in April 2020, the DoD senior procurement executives decided to establish a new contracting competency model and a single level of certification program. The new competency model is based on the NCMA CMBOK and ANSI-accredited CMS. The new DoD contracting competency model complies with the requirement in Section 861 of the FY2020 NDAA to base a professional certification on standards developed by a third-party accredited program. The CMS uses terms that are relevant and applicable across the DoD, federal agencies, and industry. The model also has an overarching narrative of guiding principles aligned with professional competencies that apply across all phases of the contracting life cycle. The basic top-level structure of the NCMA CMS is reflected in Figure 1 (NCMA, 2019b).

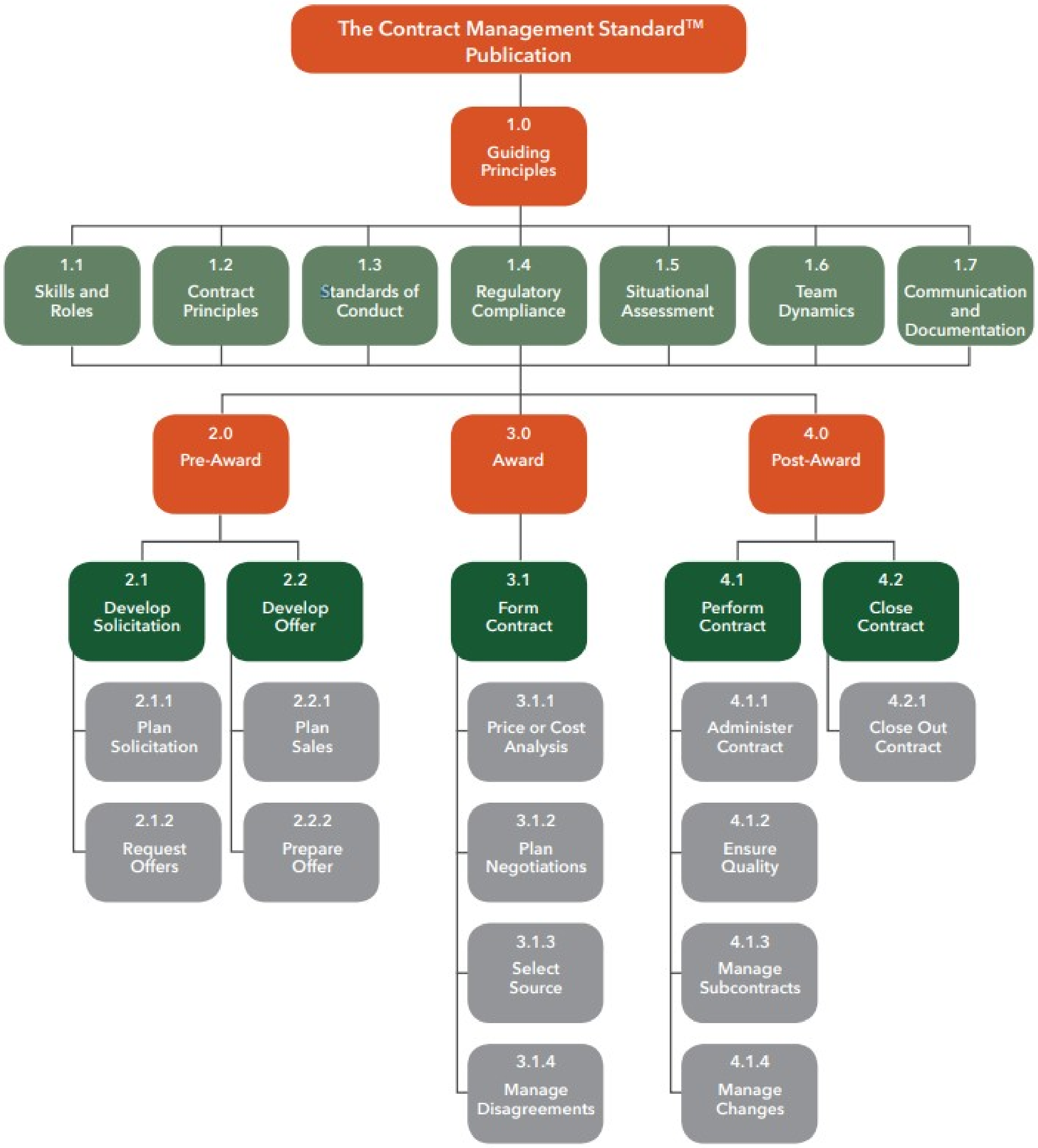


Figure 1. Contract Management Standard.
(NCMA, 2019b)

Development of Competency Assessment Instrument

The development of the contracting competency assessment instrument included structuring contracting competency statements for each of the contract management phases (pre-award, award, post-award), as well as from both contracting perspectives (buyer and seller). More specifically, the contracting competency statements reflect the contracting competencies and the specific job tasks for each contract management phase and for each perspective as reflected in the CMS. The competency statements would be rated by the contracting workforce members using a Likert scale reflecting different levels of proficiency for performing the buyer job tasks and a Likert scale reflecting the different levels of knowledge of the seller job tasks. The proficiency rating scales, for performing buyer job tasks, are identified and defined below:

1. Aware: Applies the competency in the simplest situations and requires close and extensive guidance.
2. Basic: Applies the competency in somewhat difficult situations and requires frequent guidance.
3. Intermediate: Applies the competency in difficult situations and requires little or no guidance.
4. Advanced: Applies the competency in considerably difficult situations and generally requires no guidance.
5. Expert: Applies the competency in exceptionally difficult situations and serves as a key resource and advises others.

N/A: Not applicable/not needed in my job.

The knowledge rating scales, for understanding seller job tasks, are identified and defined below:

1. None: I am not aware of this contractor competency.
2. Aware: I am aware, but have no knowledge of this contractor competency.
3. Basic: I have some basic level knowledge of this contractor competency.
4. Intermediate: I have intermediate level knowledge of this contractor competency.
5. Advanced: I have advanced level knowledge of this contractor competency.

Deployment of Competency Assessment Instrument

Upon development of the survey instrument, the assessment survey was deployed to the Marine Corps Systems Command (MCSC) contracting organization. With the assistance of our graduate students, the assessment survey was deployed using the Naval Postgraduate School (NPS) open-source survey tool LimeSurvey. The web-based LimeSurvey allows participants to respond anonymously to the self-assessment items. The MCSC contracting workforce population consists of 220 government civilian (GS 1102) and military equivalent contracting professionals (Hayashi & Pfannenstiel, 2021).

Findings

Of the MCSC 220 government civilian and military contracting professionals, 43 contracting professionals completed the assessment, equating to approximately 19.5% of the MCSC contracting workforce. The demographic data of the responding population are reflected in Table 1. As can be seen in Table 1, the majority of the respondents were DAWIA Contracting Level 3 certified and had at least 9 years of contracting experience.

Table 1. MCSC Contracting Workforce Competency Assessment Demographics

DAWIA Contracting Certification Level	Number		Years of Contracting Experience	Number
None	1		3 or Less	5
Level 1	3		4 to 8	5
Level 2	5		9 to 13	21
Level 3	41		14 to 18	4
			19 or more	17

Buyer Proficiency

Figure 2 reflects the assessment results of the Buyer Proficiency component of the competency assessment. The figure reflects the categories of buyer tasks, as reflected in the NCMA Contract Management Standard (CMS), along with the average proficiency rating, based on the buyer proficiency rating scales. As can be seen in Figure 2, the average buyer proficiency ratings range between 3.34 (Intermediate) and 4.20 (Advanced). Additionally, the Pre-Award and Award competency rating averages are higher than the Post-Award competency rating averages. Finally, the lowest proficiency rating average was 3.34 for the Manage Disagreement competency.

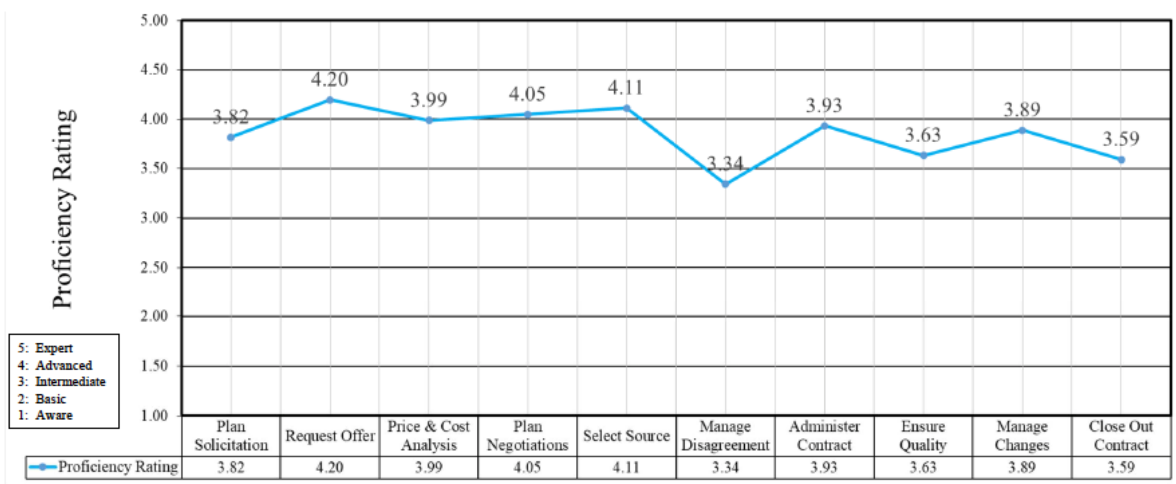


Figure 2. MCSC Contracting Workforce Competency Assessment Buyer Proficiency

Seller Knowledge

Figure 3 reflects the assessment results of the Seller Knowledge component of the competency assessment. The figure reflects the categories of seller tasks, as reflected in the NCMA Contract Management Standard (CMS), along with the average knowledge rating, based on the seller knowledge rating scales. As can be seen in Figure 3, the average seller knowledge ratings range between 2.95 (Aware) and 3.68 (Basic). Additionally, the Pre-Award and Award competency averages are higher than the Post-Award competency averages. Finally, the lowest proficiency average rating was 2.95 for the Manage Disagreement competency.

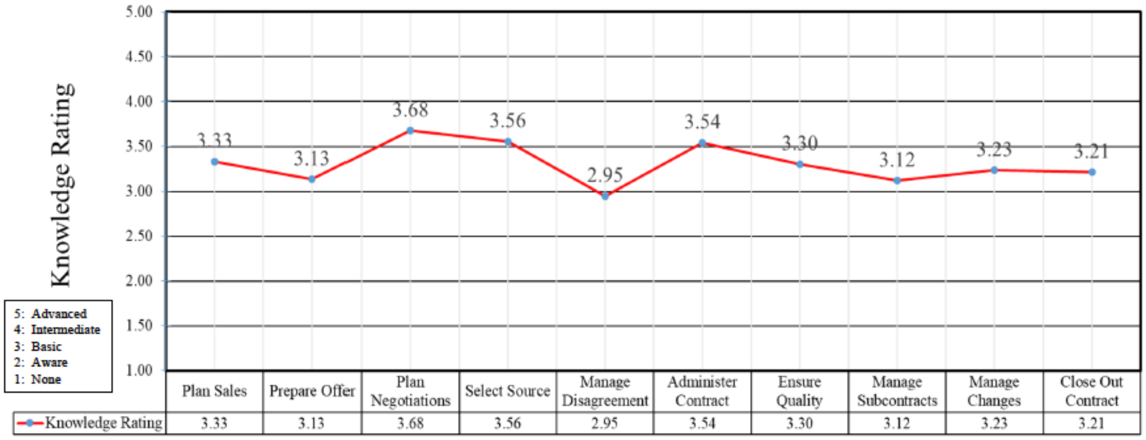


Figure 3. MCSC Contracting Workforce Competency Assessment Seller Knowledge

Discussion

The overall findings from the competency assessment indicate that the organization's competency levels for the buyer proficiency tasks are higher than the organization's knowledge levels of seller's tasks. Specifically, based on the competency assessment, the majority of the buyer proficiency competency ratings are at an Intermediate level, with seven out of 10 competencies rated within this range. The remaining three competencies received ratings of an Advanced level. Additionally, when these competencies are analyzed by contract life-cycle phases, the average pre-award phase competency level is Advanced whereas the average award and post-award phases are both at an Intermediate level. The lowest rated competency was that of Manage Disagreements in the Award phase.

Additionally, based on the competency assessment, seller knowledge competency ratings are at a Basic level with all 10 competencies scoring in this range. When these competencies are analyzed by contract life-cycle phases, the pre-award, award, and post-award phases all rate at a Basic level. The lowest-rated competency was that of Manage Disagreements in the Award phase, which is closely followed by the competency of Prepare Offer in the Pre-Award phase.

The Advanced and Intermediate average competency levels for the Buyer tasks may be related to the background of the surveyed workforce. The majority of respondents are DAWIA Contracting Level 3 certified and have an average of at least 9 years of contracting experience. This level of training and experience may indicate a higher competency level in performing the buyer tasks reflected in the CMS. Additionally, the higher average competency ratings for the pre-award and award competency categories may also be related to past Naval organizational contract management process maturity assessments based on the Contract Management Maturity Model (CMMM). Those CMMM process assessments indicate that Naval contract management process maturity is higher for pre-award and award contracting processes compared to post-award contracting processes (Rendon, 2015).

The higher competency levels for the buyer tasks (Advanced and Intermediate) compared to the lower knowledge levels of the seller tasks (Aware and Basic) may reflect the scope and focus of the contracts training received by the DoD acquisition workforce. The contracts training provided by the Defense Acquisition University (DAU) and based on the current DoD contracting competency framework reflects only the buyer processes and related tasks, specifically dictated by the Federal Acquisition Regulation (FAR). The DAU contracts training courses do not cover the seller (industry) processes and related tasks.

Finally, the consistency in lower proficiency and knowledge levels of the Manage Disagreement competency category for both buyer and seller tasks is indeed an interesting finding. This CMS competency area specifically deals with the seller tasks of submitting protests and appeals and the buyer tasks of responding to protests and appeals. The low proficiency and knowledge levels from the surveyed population in this competency area may reflect a deficiency in the knowledge, skills, and abilities related to these contract management tasks.

Based on these competency assessment findings, our research provides recommendations for the assessed organization for competency development. These recommendations can be used by the organization for developing a training roadmap for targeting competencies and knowledge areas needed for improvement within the contracting workforce.

Recommendations for Competency Development

Based on the findings and discussion on the results of the MSCS competency assessment findings, there are a couple of targeted recommendations for the assessed organization. Additionally, there are a number of recommendations for areas of further research as well as suggestions for uses of the competency assessment as a tool to aid contract management supervisors.

Targeted Recommendations

The first recommendation for the assessed organization is to develop a curriculum for the existing training program focused on the seller processes and job tasks (NCMA, 2019b). Because the survey results indicate only a Basic level of knowledge for all seller tasks, the new curriculum should incorporate seller task information from the CMBOK for all the contract life-cycle competencies (NCMA 2019a). Additional emphasis could also be placed on the Post-Award phase since the results indicate an overall lower knowledge level as compared to the Pre-Award and Award phases.

The second recommendation for the assessed organization is to develop and/or revise the training module covering managing disagreements. This recommendation is based on the survey results indicating that the Manage Disagreements task within the Award phase was the lowest score for both buyer task proficiency and seller task knowledge. Development of this training module could start by incorporating information from Section 5.4 of the CMBOK, Manage Disagreements (NCMA, 2019a). Additional information from the CMBOK could also be incorporated to improve skills such as critical thinking, problem solving, and decision-making related to managing and resolving protests and appeals. Specifically, the CMBOK covers information on these skill sets within the Leadership, Management and Guiding Principles Competencies (NCMA, 2019a).

Areas for Further Research

The primary area for further research is to deploy the currently developed survey instrument to additional contracting activities throughout the DoD. This would increase both the sample size of survey responses and collect data from activities with more diverse contracting mission sets. For example, the organization surveyed for this research, MCSC, has a primary mission focused on Pre-Award and Award tasks for procuring major weapon systems for the Marine Corps. Conducting surveys of organizations whose mission is either procuring base support functions or administrating awarded contracts would likely show data results with different levels of proficiency of buyer tasks and knowledge of seller tasks and produce different targeted recommendations.

Another area for further research would be to revise the existing survey instrument to add questions about the levels of contracting experience in the private sector. The additional information could demonstrate correlation between that level of experience and the data results regarding the level of knowledge of seller job tasks and associated knowledge gaps.

The final area for further research could be conducted once the DoD has established and implemented the new contracting competency model and single level of certification program. The research would compare the buyer and seller tasks of the CMS (NCMA, 2019b) with the objectives of the new certification program to identify any differences or gaps that could be addressed.

Suggestions of Tools for Contract Management Supervisors

One suggestion is to provide the existing survey instrument to contract management supervisors. They could deploy the survey to their subordinates for self-assessment of the buyer tasks proficiency and seller tasks knowledge as well as personally completing the survey for each of their subordinates to assess the supervisor's perception of the same levels of proficiency and knowledge. The self-assessments and supervisor assessments could be compared to improve both the mentoring and personnel evaluation responsibilities.

The other suggestion to aid contract management supervisors would require revising the current survey instrument to add questions on a Likert scale as to the importance of all the buyer tasks. Once revised, the survey could be deployed similarly to the first suggestion and the results could assist the supervisor in developing targeted and personalized individual training plans. The training plans would better align the employees' lower proficiencies of buyer tasks with the organization's mission priorities.

Conclusion

The GAO and the OIG both continue to identify contract management as a high risk and a top management challenge for the DoD. Additionally, research has shown that the current DoD contracting competency model may not be sufficient in assessing today's contracting workforce competencies (Rendon & Winn, 2017). Furthermore, the NDAA 2020 resulted in congressional direction to the Secretary of Defense to implement a professional certification program for all members of the acquisition workforce based on standards developed by a third-party accredited program that is based on nationally or internationally recognized standards (NDAA, 2019). Finally, in April 2020, the DoD senior procurement executives decided to establish a new contracting competency model and a single level of certification program. The new competency model will be based on the NCMA *Contract Management Standard* (CMS), which is accredited by the American National Standards Institute (ANSI).

The purpose of this research was to develop a new contracting competency assessment instrument based on the NCMA CMS to be used in assessing the DoD's contracting workforce. The competency assessment instrument has been developed and is being deployed throughout the DoD. This specific research reflects the application of this competency assessment instrument to the Marine Corps Systems Command (MCSC) contracting workforce. Based on the assessment results, the MCSC can develop a training roadmap for targeting competencies and knowledge areas needed for improvement within the contracting workforce. This research should be expanded by applying the competency assessment tool to other DoD contracting agencies as a way of benchmarking the DoD contracting workforce competencies against the newly adopted NCMA *Contract Management Standard*.

References

Carlson, S. (2017). A new liberal art: How systems thinking prepares students for a complex world. *The Chronicle of Higher Education*, *64*(5).

GAO. (2019, March 6). *High-risk series: Substantial efforts needed to achieve greater progress on high-risk series* (GAO-19-157SP).

Hayashi, S., & Pfannenstiel, A. (2021). *Analysis of Marine Corps Systems Command contracting workforce competency assessment.* Naval Postgraduate School.

National Contract Management Association. (2019a). *Contract management body of knowledge* (6th ed.).

National Contract Management Association. (2019b). *Contract management standard* (2nd ed.).

National Defense Authorization Act for Fiscal Year 2016, Pub. L. No. 114-92 (2015).

National Defense Authorization Act for Fiscal Year 2020, Pub. L. No. 116-92 (2019).

Office of Inspector General. (2019, November 4). *Fiscal Year 2020: Top DoD management challenges.* Department of Defense.

Rendon, R. G. (2015). Benchmarking contract management process maturity: A case study of the U.S. Navy. *Benchmarking: An International Journal*, *22*(7), 1481–1508.

Rendon, R. G. (2019). *Enhancing professional and technical excellence: Analysis of contract management competency models* (SYM-AM-19-036). https://calhoun.nps.edu/handle/10945/63019

Rendon, R. G., & Powley, E. H. (2017). *Navy contracting organizational climate assessment* (NPS-CM18-030). Naval Postgraduate School, Acquisition Research Program.

Rendon, R. G., & Schwartz, B. (2020). An innovative approach to assessing DoD contracting workforce competency. In *Proceedings of the 17th Annual Acquisition Research Symposium* (SYM-AM-20-088). Naval Postgraduate School, Acquisition Research Program.

Rendon, R. G., & Winn, T. (2017). Competency in contract management: A comparison of DoD and CMBOK competency models. *Contract Management, 57*(12), 66–81.

Scott, D., & Thompson, D. (2019). *Toward building the DoD acquisition workforce of the twenty-first century: Recommendations of the Section 809 Panel on professional development*. https://www.ncmahq.org/news/magazine-details/toward-building-the-dodacquisition-workforce-of-the-twenty-first-century-recommendations-from-the-section-809panel-on-professional-development

Werber, L., Ausink, J., Daugherty, L., Phillips, B., Knutson, F., & Haberman, R. (2019). *An assessment of gaps in business acumen and knowledge of industry within the defense acquisition workforce: A report prepared for the U.S. Department of Defense in compliance with Section 843(c) of the Fiscal Year 2018 National Defense Authorization Act* (RR-2825-OSD). RAND.

Aligning DoD Program Management Competencies with the Project Management Institute Standards

Dr. Robert F. Mortlock, COL, USA (Ret.), is a professor of the practice for defense acquisition and program management in the Graduate School of Defense Management at the Naval Postgraduate School in Monterey, CA. He holds a PhD in chemical engineering from the University of California, Berkeley, an MBA from Webster University, an MS in national resource strategy from ICAF, and a BS in chemical engineering from Lehigh University. [rfmortlo@nps.edu]

Jonathan L. Karnes, 1st Lt, USAF, is a contracts specialist for the Air Force Life Cycle Management Center's Medium Altitude Unmanned Aircraft System Division. He holds an MBA in acquisition management from the Naval Postgraduate School, an MBA in project management from the University of Texas at San Antonio, and a BS in communication studies from the University of Texas at Austin. [jonathan.karnes.3@us.af.mil]

Abstract

In 1990, the Government Accountability Office began publishing their high-risk series meant to review federal programs prone to waste, abuse, and mismanagement. Defense acquisitions has appeared in every high-risk list's publication due to the tendency to miss cost, schedule, and performance objectives. In 2019, Congress passed the National Defense Authorization Act, which mandated every acquisition career field to realign their certification requirements to be based on the nationally recognized standards of an accredited third party. This study offers recommendations for improving the DoD program management (PM) training standards by providing traceability between the DoD PM competencies and the Project Management Institute's standards for project, program, and portfolio management. The study elaborates on the extent of alignment, finding that 96% of the DoD PM competency elements align to PMI standards. Areas of misalignment identify opportunities to augment DoD PM training and highlight areas where DoD PM training deviates from industry standards.

Keywords: project management, program management, portfolio management, Defense acquisitions, functional career field competencies, training, industry standards

Introduction

For decades, the Department of Defense (DoD) has been criticized for its inability to manage the various programs funded by the U.S. taxpayers. These repeated failings in the realms of program cost, schedule, and performance have been documented in numerous reports from the Office of the Secretary of Defense and the Government Accountability Office (GAO) and in a myriad of theses and dissertations (Bond et al., 2016; Choi, 2009; H.R. 5211, 1990; GAO, 2019a, 2019b; Kupec, 2013; Pernin et al., 2012; President's Blue Ribbon Commission on Defense Management, 1986; Redshaw, 2011). A debate continues as to whether the acquisition program failings are caused by the DoD's inherently complex acquisition system or the quality of its acquisition personnel. In an article entitled, "Does the Program Manager Matter? New Public Management and Defense Acquisition," the authors claim that until the acquisitions system and processes of the DoD are fixed, the training and education of program managers (PMs) could be considered inconsequential to the success of defense programs (Eckerd & Snider, 2017). However, based on the recommendations in GAO-18-217, which was focused on improving program management, the DoD's program performance would improve if the DoD would "improve practices that do not align extensively with leading practices" (GAO, 2018a, "GAO Highlights" section). This recommendation is further supported by the GAO's annual high-risk list, which lists the DoD career fields that pose a great level of risk to the government if not improved upon or appropriately monitored. DoD weapon systems acquisition has consistently been included on this list since 1990 (GAO, 1995, 2019a). According to the most recent list developed in 2019, DoD program management was considered high risk because of the anticipated $1.66 trillion investments into their acquisition and procurement portfolio (GAO, 2019a). After decades of continuous defense acquisition reform initiatives, there is still no effort guaranteed to resolve the continued shortfalls in meeting cost, schedule, and performance goals for acquisition programs (GAO, 2019a). These three factors are commonly referred to as a program's triple constraint and form the acquisition program baseline for management.

While the DoD has struggled to develop solutions meant to resolve their continued issues with meeting their program's planned cost, schedule, and performance baselines, the DoD has made attempts. One such attempt was implemented under President Reagan's administration. A group of acquisition professionals were assembled under the leadership of David Packard to form President Reagan's Blue Ribbon Commission, also commonly referred to as the Packard Commission. This commission provided a series of recommendations that are still being implemented today. As it pertains to this research study, the Packard Commission's most relevant recommendation was to require business-related education and training for acquisition personnel (President's Blue Ribbon Commission on Defense Management, 1986).

This recommendation led to the passing of the Defense Acquisition Workforce Improvement Act (DAWIA) of 1990, which then led to the establishment of the Defense Acquisition University (DAU). Since its inception in 1991, the DAU has structured its acquisition curriculum in a way that would best prepare PMs to maneuver the complexities of the defense acquisition system, which consists of the interoperation of management processes (the Adaptive Acquisition Framework), requirements processes (like the Joint Capabilities Integration and Development System [JCIDS] for formal programs of record), and a resourcing process (referred to as the planning, programming, budget, and execution [PPBE] system; Office of the Under Secretary of Defense for Acquisition & Sustainment [OUSD(A&S)], 2020a, 2020b). In 2016, the Office of the Assistant Secretary of Defense for Acquisition distributed the functional career field competencies for PMs and broke them down into the following DoD PM categories: Acquisition Management, Business Management, Technical Management, and Executive Leadership (MacStravic, 2016). From the DoD's perspective, these competencies serve as the standards that enable PMs to effectively "deliver mission-critical capabilities in terms of equipment and services" (MacStravic, 2016, p. 2). Furthermore, this list of competencies serves as the basis for the program management (PM) DAWIA certification standards adopted by the services.

The Project Management Institute (PMI) is an independent, private organization that has led the way in establishing the internationally recognized standards for project management, program management, and portfolio management across industries. They offer a variety of certifications to business and management professionals that are recognized globally. Since 1999, the American National Standards Institute (ANSI) has approved PMI's *Guide to the Project Management Body of Knowledge®* (*PMBOK Guide*; PMI, 2017a) as the American national standard for project management (Holtzman, 1999). A contributing factor to the *PMBOK Guide* being ANSI-certified is its wide range of applicability across industries. No matter what industry one is in, the knowledge areas discussed in the PMI's *PMBOK Guide* and the performance domains of *The Standard for Program Management* (*TSPgM*; PMI, 2017c) and *The Standard for Portfolio Management* (*TSPfM*; PMI, 2017b) apply.

In December 2019, Congress passed the National Defense Authorization Act for Fiscal Year 2020 (NDAA). The section of this act that is relevant to this research is Section

861, “Defense Acquisition Workforce Certification, Education, and Career Fields” subsection (c), “Professional Certification.” It states,

> The Secretary of Defense shall implement a certification program to provide for a professional certification requirement for all members of the acquisition workforce. … The certification requirement for any acquisition workforce career field shall be based on standards developed by a third-party accredited program based on nationally or internationally recognized standards. (NDAA, 2019)

This subsection has mandated a refocusing of how the DoD trains its acquisition professionals. Per the NDAA, it is the role of the Office of the Secretary of Defense to produce the realigned certification program based on nationally or internationally recognized standards of an accredited third party (NDAA, 2019). Per the DAWIA (H.R. 5211, 1990), it is the DAU’s role to provide the training that meets the requirements of the acquisition workforce.

The purpose of this research is to understand the extent to which the DoD’s PM functional career field competencies currently align with the internationally recognized standards for project, program, and portfolio management published by the PMI. This research will be used to make recommendations to the DoD on how to best transition from its current PM certification requirements to certification requirements based on the PMI standards. This study answers the following research questions:

- To what extent are the DoD’s program management competency elements at the basic, intermediate, and advanced DAWIA levels aligned with the PMI’s *PMBOK Guide*, *TSPgM*, and *TSPfM*?
- To what extent are the knowledge areas and performance domains in the PMI’s *PMBOK Guide*, *TSPgM*, and *TSPfM* aligned with the DoD’s program management competency elements at the basic, intermediate, and advanced DAWIA levels?

The results of this study provide insight and recommendations for the decision-makers within the Office of the Secretary of Defense (OSD) and the DAU charged with realigning the program management professional certification. This will enable them to make informed decisions on carrying out the modifications to the program management certification requirements as mandated by the NDAA.

This research focuses on the shift in the basis for DoD program management certification requirements. Specifically, this study pertains to the alignment of the DoD’s program management functional career field competencies (MacStravic, 2016) to the PMI’s 10 knowledge areas that comprise the *PMBOK Guide* (PMI, 2017a), the program management performance domains of *TSPgM* (PMI, 2017c), and the portfolio management performance domains of *TSPfM* (PMI, 2017b). This study provides traceability between the DoD program management competencies and the industry standards and elaborates on the extent to which they are aligned. Finally, this study highlights areas of inconsistency and results in recommendations for changes in DoD standards for training and education and potential policy changes.

Literature Review

The study of PM career field competencies can be linked to work in other acquisition workforce career fields. Rendon (2019) states that it is important to make an organization auditable so it is better suited to achieve its mission goals and objectives. The concept of auditability consists of three main components: capable processes, effective internal controls, and competent personnel. The DoD has robust processes within defense acquisition in the form of acquisition management framework, requirements, and resourcing processes. The DoD also has internal controls provided by the GAO, the DoD's Office of Inspector General (DoD IG), congressional oversight, and adherence laws such as annual NDAAs and acquisition acts like the Nunn–McCurdy Act (Schwartz, 2010). This research aids the DoD in improving upon the third component of auditability: competent personnel.

As previously discussed, defense acquisitions have been criticized for failing to meet cost, schedule, and performance program baseline objectives. In response to the deficiencies in these three areas, the DoD has implemented multiple acquisition reform initiatives to improve its processes. The reform initiatives have also modified the acquisition reporting structure and used the power of government watchdogs such as the GAO and the DoD IG to implement effective internal controls. To improve the quality of its acquisition professionals, the DoD has made frequent modifications to the training and education requirements. This literature review covers former acquisition reform initiatives, external findings on DoD acquisition performance, the standards published by the PMI, and scholarly articles that express support and opposition to modifying the alignment of the DoD competencies to the standards of a third party.

In 1985, the Reagan administration appointed former U.S. Secretary of Defense David Packard to lead its Blue Ribbon Commission, which was established to make recommendations on how to improve defense acquisitions. The output of the Packard Commission resulted in nine recommendations; the one addressed in this research study is the recommendation to enhance the quality of acquisition personnel (President's Blue Ribbon Commission on Defense Management, 1986). This recommendation focused on improving the appointment criteria of senior-level personnel to more effectively run programs and portfolios and called for business-related education for civilians and for federal law to allow acquisition personnel to pursue expanded opportunities for education and training (President's Blue Ribbon Commission on Defense Management, 1986). This recommendation was implemented via the passing of the DAWIA in 1990. The DAWIA (H.R. 5211, 1990) resulted in the development of the DAU and the establishment of baseline training requirements for acquisition professionals.

The DAU is the primary source of training for defense acquisition professionals. The DAU provides formal courses as well as continuous learning to promote continuing education and professional growth for thousands of students every year (Woolsey, 2019). To date, these courses are structured to accommodate DAWIA certification requirements and have been broken down into three levels (DoD & DAU, n.d.):

- **Level I:** basic or entry level
- **Level II:** intermediate or journeyman level
- **Level III:** advanced or senior level (DoD & DAU, n.d.)

The content of the training requirements for PMs is based on the DoD PM functional career field competencies, which make up four overarching PM categories that serve as the basis for developing the learning objectives and training materials for PMs (MacStravic, 2016):

- Acquisition Management
- Business Management
- Technical Management
- Executive Leadership (Level III education for unique positions)

In November 2019, the NDAA directed the Secretary of Defense to implement a certification program based on standards developed by a third party (NDAA, 2019). For the DoD's PM training curriculum, this requires adjusting the training standards from being based solely on DoD unique functional career field competencies to instead being founded on the "standards developed by a third-party accredited program based on nationally or internationally recognized standards" (NDAA, 2019, p. 778). This shift from DoD-centric competencies to the widely accepted standards of the private sector is an attempt to improve the quality of defense acquisition personnel by making them more capable to work with industry partners throughout the acquisition process.

Defense acquisition management has been on the GAO's high-risk list since 1990 because of the failure in meeting the five criteria for removal: leadership commitment, capacity, action plan, monitoring, and demonstrated progress (GAO, 2019b). Of those five, defense acquisition management meets the criteria for leadership commitment but only partially meets the other four. This continued pattern of insufficiency makes the DoD vulnerable to budget overruns, schedule slips, and underperformance—observed in major defense acquisition programs like the F-35 Joint Strike Fighter (GAO, 2018b) and the Army Future Combat Systems (Pernin et al., 2012). The poor returns on investment exhibited by these and other programs have led to the acquisition management career field remaining on the high-risk list (GAO, 2019b) and have created a continual demand for acquisition reform (Gansler et al., 2007).

While there is generally consensus among lawmakers and DoD senior leaders that there is room for improvement in how the DoD manages programs, there are different thoughts on how the DoD should work to improve the acquisition career field. Multiple GAO reports have contradicting views on what needs to change to remove defense acquisition from the high-risk list. Some reports recognize that the certification training offered by the DAU is capable of providing adequate training to PMs (GAO, 2010), whereas others state that the issues emanate from those very same training standards not aligning with leading practices (GAO, 2018a). The takeaway is that the DAU has the infrastructure and organizational alignment to provide effective training, but the current training can be more effective if aligned with more widely accepted standards. This issue could be addressed by incorporating the advisement provided by the GAO to the Office of Management and Budget (OMB) by adopting "an existing set of consensus-based standards, such as the widely accepted standards for program and project management from the Project Management Institute" (GAO, 2019a, p. 11).

The PMI is a not-for-profit association that publishes standards for certification programs including the PMP, the PgMP, and the PfMP. Earning these credentials certifies that one is qualified to lead a project, manage a program, and meet strategic objectives in overseeing one or more portfolios, respectively (PMI, 2020). The PMI certifications are recognized globally due to their widely applicable and highly detailed standards that have proven over time to improve the outcomes of projects, programs, and portfolios if applied and resourced appropriately.

In 1999, the ANSI first approved the PMI's *PMBOK Guide* (PMI, 2017a) as the American national standard for project management (Holtzman, 1999). PMI also awards the project management professional (PMP) credential. This credential is ideal for individuals who lead and manage projects, which the PMI defines as "temporary endeavors undertaken to create a unique product, service or result" (PMI, 2017a, p. 4).

The *PMBOK Guide* is broken down into 10 knowledge areas, which are made up of 49 processes. Project management knowledge areas are categorized by their knowledge requirements and are described in terms of their various component processes, practices, inputs, outputs, tools, and techniques (PMI, 2017a). Project management processes are defined as “systematic activities directed toward causing an end result where one or more inputs will be acted upon to create one or more outputs” (PMI, 2017a, p. 18). Figure 1 includes a complete list of the 49 processes that fall under the 10 different knowledge areas in the *PMBOK Guide* (PMI, 2017a).

Knowledge Areas	Project Management Process Groups				
	Initiating Process Group	Planning Process Group	Executing Process Group	Monitoring and Controlling Process Group	Closing Process Group
4. Project Integration Management	4.1 Develop Project Charter	4.2 Develop Project Management Plan	4.3 Direct and Manage Project Work 4.4 Manage Project Knowledge	4.5 Monitor and Control Project Work 4.6 Perform Integrated Change Control	4.7 Close Project or Phase
5. Project Scope Management		5.1 Plan Scope Management 5.2 Collect Requirements 5.3 Define Scope 5.4 Create WBS		5.5 Validate Scope 5.6 Control Scope	
6. Project Schedule Management		6.1 Plan Schedule Management 6.2 Define Activities 6.3 Sequence Activities 6.4 Estimate Activity Durations 6.5 Develop Schedule		6.6 Control Schedule	
7. Project Cost Management		7.1 Plan Cost Management 7.2 Estimate Costs 7.3 Determine Budget		7.4 Control Costs	
8. Project Quality Management		8.1 Plan Quality Management	8.2 Manage Quality	8.3 Control Quality	
9. Project Resource Management		9.1 Plan Resource Management 9.2 Estimate Activity Resources	9.3 Acquire Resources 9.4 Develop Team 9.5 Manage Team	9.6 Control Resources	
10. Project Communications Management		10.1 Plan Communications Management	10.2 Manage Communications	10.3 Monitor Communications	
11. Project Risk Management		11.1 Plan Risk Management 11.2 Identify Risks 11.3 Perform Qualitative Risk Analysis 11.4 Perform Quantitative Risk Analysis 11.5 Plan Risk Responses	11.6 Implement Risk Responses	11.7 Monitor Risks	
12. Project Procurement Management		12.1 Plan Procurement Management	12.2 Conduct Procurements	12.3 Control Procurements	
13. Project Stakeholder Management	13.1 Identify Stakeholders	13.2 Plan Stakeholder Engagement	13.3 Manage Stakeholder Engagement	13.4 Monitor Stakeholder Engagement	

Figure 1. Ten Knowledge Areas of the *PMBOK Guide*
(PMI, 2017a)

The program management professional (PgMP) certification is based on *The Standard for Program Management* (*TSPgM*; PMI, 2017c). The purpose of *TSPgM* is to provide generally recognized guidance to support good program management practices, establish a common understanding of the role of a PM, and offer guidance for PMs’ interactions with portfolio and project managers as well as any other program stakeholders (PMI, 2017c). According to the PMI, a program is made up of “related projects, subsidiary programs, and program activities managed in a coordinated manner” (PMI, 2017c, p. 3).

When programs are run effectively, they can deliver benefits that would not have been attainable had their subsidiary programs and projects been managed independently.

Similar to the *PMBOK Guide* (PMI, 2017a), *TSPgM* discusses five performance domains that are “complementary groupings of related areas of activity or function that uniquely characterize and differentiate the activities found in one performance domain from the others within the full scope of program management work” (PMI, 2017c, p. 23). The purpose of these domains is to provide PMs with a general checklist of tasks and concepts to complete and consider throughout the life of the program (refer to Figure 2).

The portfolio management professional (PfMP) certification is based on *The Standard for Portfolio Management* (*TSPfM*; PMI, 2017b), the purpose of which is to provide portfolio management principles and performance management domains that are considered to be good practices for organizations that manage complex programs and projects. *TSPfM* provides a common understanding of the role of a portfolio manager as well as a unified vocabulary to use across industries (PMI, 2017b). According to the PMI, “a portfolio is a collection of projects, programs and subsidiary portfolios and operations managed as a group to achieve strategic objectives” (PMI, 2017b, p. 3). The purpose of managing a portfolio versus independent programs and projects is to achieve organizational objectives and strategies that could not be met otherwise.

TSPfM is very similar to *TSPgM* in that it consists of seven performance domains and is supported by the *PMBOK Guide*. These seven performance domains, when followed and executed correctly, are what allow for the portfolio management plan to achieve its desired impact on strategy and performance (PMI, 2017b). For a complete list of these domains and what items are associated with them, see Figure 3.

Figure 2. Program Management Professional Performance Domains (PMI, 2017c)

Figure 3. Portfolio Management Professional Performance Domains (PMI, 2017b)

In the early 2000s, the DoD worked with the PMI to develop the *U.S. Department of Defense Extension to: A Guide to the PMBOK* (*PMBOK Guide*; DoD & DAU, 2003). The purpose of the DoD and PMI collaboration was to identify defense applications of the *PMBOK Guide*’s knowledge areas and to meet the objective of the Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]) to build credibility in acquisitions by delivering systems at lower cost and on schedule (DoD & DAU, 2003). However, the *PMBOK Guide*'s extension was never implemented into the DAU certification curriculum (Kupec, 2013).

It has been well established that DoD acquisition programs have struggled to effectively manage program cost, schedule, and performance (GAO 2018a, 2018b, 2019a). The NDAA (2019) addresses this issue by mandating that the DAU modify its existing certification requirements to be based on the standards of an accredited third party with nationally recognized standards. Because of the high visibility of defense acquisitions, there have been many scholarly studies on how the DoD could improve their training standards by mirroring an entity like the PMI (Choi, 2009; Kupec, 2013; Redshaw, 2011). In comparison to the progressive complexity of PMI's certifications for project, program, and portfolio management, the DAWIA certifications for Level I (basic), Level II (intermediate), and Level III (advanced) "correlate to the complexity and responsibilities required for designated positions and different types of assignments in weapon systems, services, business management systems and information technology, and international acquisitions" (Redshaw, 2011, p. 55). Both Choi (2009) and Kupec (2013) concur with this analysis and elaborate further that modeling the new DAU standards after only one of the PMI credentialing standards—PMP for example—would not be sufficient. Individuals who earn the PMP credential have proven themselves to be capable of effectively leading project teams and managing a temporary project. While this credential holds value in the program management industry, the body of knowledge that accompanies it would not be enough to equip an individual to run a complex decade-long program or portfolio. For these reasons, it is essential to base the new DAWIA certification requirements on all three of the PMI credentials.

According to auditability theory, in order for an organization, project team, program office, or portfolio executive officer to meet their specific objectives, it is critical that competent personnel are employed, effective internal controls are maintained, and capable processes are implemented (Rendon & Rendon, 2015). As it relates to defense acquisition reform, there are divergent opinions as to which of the three components of auditability should be focused on to improve program metrics in cost, schedule, and performance. For example, Eckerd and Snider (2017) claim that the defense acquisition processes should be the focal point for reform due to their complexities. They add that the environmental politics that DoD PMs maneuver daily prevent them from being effective, which nullifies any quality training they undergo. Other research comes to a similar conclusion that in order to make significant changes in federal acquisitions, acquisition reform needs to comprehensively consider changes to the management processes (acquisition framework), the resources processes (PPBE system), and the requirements processes (Bond et al., 2016). Mortlock (2020) asserts that providing DoD PMs with professional-level training and adopting internationally recognized industry standards (for example, PMP, PgMP, and PfPM certifications) could help improve the effectiveness of PMs, help gain acceptance for program management as a profession, and help solidify the credibility of the defense acquisition workforce.

Methodology

This research involved a qualitative, lexicographic analysis of the descriptions of the DoD's program management competencies and the descriptions of the PMI's knowledge areas and domains in the *PMBOK Guide* (PMI, 2017a), *TSPgM* (PMI, 2017c), *TSPfM* (PMI, 2017b), the NDAA for Fiscal Year 2020 (NDAA, 2019), and other key sources. This highlighted key words, phrases, and meaning from the description of each knowledge area, domain, and competency and allowed for an informed mapping of the DoD's PM competencies to the PMI's standards.

The OUSD(AT&L) memorandum entitled *Program Management Functional Career Field Competencies* served as the primary DoD source used in analyzing the alignment between the DoD's program management competencies and the PMI's standards (MacStravic, 2016). According to the memorandum, an integrated product team developed the updated competencies while considering the three certification levels: Level I (basic), Level II (intermediate), and Level III (advanced; MacStravic, 2016). The memorandum includes the following information used in this research:

1. **Program Management Competency Units and Competencies:** The PM competencies are organized into the four program management categories and 18 units of competency. Figure 4 demonstrates the distribution of the competencies.

Acquisition Management		Technical Management
Capability Integration Planning	**Program Execution**	**Engineering Management**
Requirements Management (Mgmt)	Risk/Opportunity Mgmt	Technical Planning
Acquisition Program Strategic Planning	Program Planning	Requirements Decomposition
Business Case Development	Teaming	Technical Assessment
Acquisition Law and Policy	Program Oversight	Decision Analysis
Acquisition Policy and Best Practice	Resource Mgmt	Configuration Mgmt
Contractual Laws, Regulations, and Obligations	Technology Mgmt	Technical Data Mgmt
Financial Mgmt Laws, Directives, and Policies	**Services Acquisition**	Interface Mgmt
Program Support Laws, Directives, and Policies	Business Management	**Defense Business Systems**
Technical and Engineering Laws, Directives and Policies	**Contract Management**	DBS Certification
Information Technology Laws, Policy, Best Practices	Market Research	DBS Acquisition Approach Preparation
International Acquisition and Exportability	Pre-Solicitation Planning and Execution	**Test and Evaluation Mgmt**
International Cooperative Programs	Source Selection and Negotiations	Test Planning
Sales and Transfers	Contract Administration	Test Execution
Technology Security and Foreign Disclosure	Contract Closeout	**Manufacturing Mgmt**
Defense Exportability Integration	**Financial Mgmt**	Manufacturing Planning and Transition
	Financial Planning	Manufacturing Shutdown
Stakeholder Mgmt	Programming	**Product Support Mgmt**
Political Savvy	Budget Formulation	Product Support Planning
External Situational Awareness	Budget Execution	Product Support Mgmt
Media Relationships	Cost estimates	Supply Chain Mgmt
Executive Leadership		
Foundational Competencies	**Leading Change**	**Results Driven**
Interpersonal Skills	Creativity & Innovation	Accountability
Integrity / Honesty	Vision	Decisiveness
Communicate Effectively	Flexibility	Entrepreneurship
Continual Learning	Resilience	Customer Service
Public Service Motivation	**Leading People**	Problem Solving
Technical Credibility	Conflict Management	
Building Coalitions	Leveraging Diversity	
Influencing / Negotiating	Developing Others	
Partnering	Team Building	

Figure 4. DoD Program Management Competency Units and Competencies (MacStravic, 2016)

2. **Program Management Functional Career Field Competencies:** Descriptions of the 70 competencies are provided for each of the three DAU certification levels.

The data sources used from the PMI include the *PMBOK Guide*, *TSPgM*, and *TSPfM*. Although the *PMBOK Guide* is the only ANSI-accredited standard, the contents of *TSPgM* and *TSPfM* are recognized internationally and accepted industry practices for program and portfolio managers, respectively. *TSPgM* and *TSPfM* define the standards for the application of their principles and practices, which enhances the likelihood of program and portfolio success (PMI, 2017b, 2017c). The PMI standards were mapped to each of the 190 elements at the basic, intermediate, and advanced level (570 total element descriptions). PMI conference papers served as the primary source for additional information on PMI standards (Alie, 2016; Ross & Shaltry, 2006; Shenhar & Dvir, 2004).

This research required the qualitative analysis of data—the data being the DoD's PM competency descriptions and the contents of the PMI's knowledge areas and performance management domains, and the qualitative analysis being the alignment mapping. Six qualitative analyses of lexicographic comparisons were preformed:

1. The DoD's basic (DAWIA Level I) PM competencies to the PMI's *PMBOK Guide* knowledge areas and processes
2. The DoD's intermediate (DAWIA Level II) PM competencies to the PMI's *PMBOK Guide* knowledge areas and processes
3. The DoD's intermediate (DAWIA Level II) PM competencies to the PMI's *TSPgM* program management domains
4. The DoD's advanced (DAWIA Level III) PM competencies to the PMI's *PMBOK Guide* knowledge areas and processes
5. The DoD's advanced (DAWIA Level III) PM competencies to the PMI's *TSPgM* program management domains
6. The DoD's advanced (DAWIA Level III) PM competencies to the PMI's *TSPfM* portfolio management domains

The analysis resulted in the mapping of 1,085 DoD PM competency elements to PMI knowledge areas and domains. The next step in this research applied a quantitative analysis to the completed competency map (Bernard, 1996). This transition to a matrix format was completed in conjunction with the more qualitative analysis by classifying each element mapping as either aligned (Green/"G"), somewhat aligned (Yellow/"Y"), completely unaligned (Red/"RR"), or not applicable (Black/"N/A").

Data Analysis

This section addresses the extent to which the DoD's 2016 PM functional career field competencies are aligned with the PMI's *PMBOK Guide*, *TSPgM*, and *TSPfM.* The first step taken in the analysis was to count how many DoD competency elements were mapped to the PMI's *PMBOK Guide*, *TSPgM*, and *TSPfM* and were classified as *aligned*, *somewhat aligned*, *completely unaligned*, or *N/A* (refer to Table 1). Categories were created for the *PMBOK Guide*, *TSPgM*, and *TSPfM* by combining the basic, intermediate, and advanced elements that mapped to each standard. A fourth category was included that combined the findings across all three PMI standards to demonstrate the extent of alignment between the DoD PM competencies and the PMI standards for when all PMI standards were applied. For example, if a single element was labeled as aligned under the *PMBOK Guide* but completely unaligned under *TSPgM* and *TSPfM*, it would be classified as aligned under the All PMI category. This method demonstrates the value of applying all three PMI standards in DoD PM training instead of only the *PMBOK Guide*. Finally, a fifth category was applied that shows the number of elements categorized as 100% aligned, somewhat aligned, or completely unaligned with the *PMBOK Guide*, *TSPgM*, and *TSPfM*. This category is significant because it shows that when all three PMI standards are applied, only eight of 190

DoD PM competency elements are completely unaligned with the PMI standards. According to the research, the DoD PM competencies align with the *PMBOK Guide*, *TSPgM*, and *TSPfM* as depicted in Figure 5.

Table 1. Quantity of DoD PM Competency Elements Mapped to the PMI's Standards (Organized by Level of Alignment and DAWIA Level)

	Basic *PMBOK Guide*	Intermediate *PMBOK Guide*	Intermediate *TSPgM*	Advanced *PMBOK Guide*	Advanced *TSPgM*	Advanced *TSPfM*
Aligned	73	65	52	56	47	47
Somewhat Aligned	66	83	98	99	115	116
Completely Unaligned	20	29	27	35	28	27
N/A	31	13	13	0	0	0
	190	190	190	190	190	190

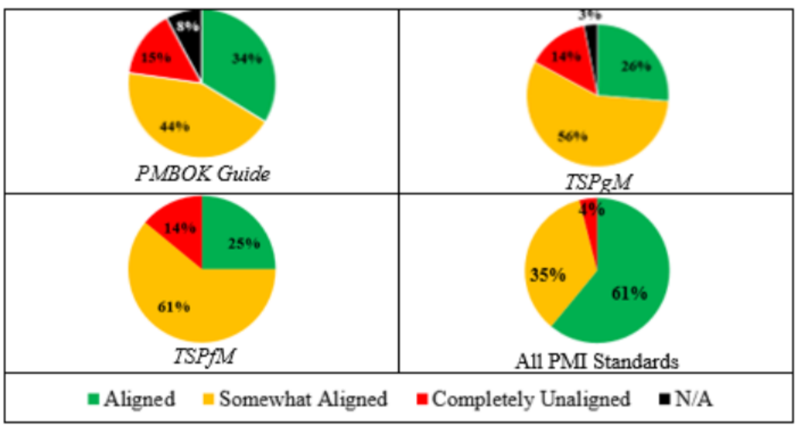


Figure 5. Extent to Which the DoD PM Competency Elements Align to the PMI Standards by Pie Chart

Based on these findings, it is evident that the *PMBOK Guide* is the PMI standard that is most aligned with the DoD PM competency elements. This is expected, as the *PMBOK Guide* serves as the building block for *TSPgM* and *TSPfM* and is the broadest of the three standards. However, by adding *TSPgM* and *TSPfM* standards to the standards of the *PMBOK Guide*, the alignment level of the PMI standards with the DoD PM competencies increases to 96% (61% completely aligned and 35% somewhat aligned). Furthermore, the percentage of elements that are categorized as completely unaligned or not applicable decreased to 4% and 0%, respectively.

Figures 6–9 further elaborate on the impact achieved when applying all three PMI standards to DoD PM competencies in order to provide sufficient detail in determining which DoD PM competency elements need to be improved upon to ensure sufficient alignment with the PMI standards. These figures provide a visualization of the progressive improvement in alignment as all three PMI standards are applied. Figure 6, Figure 7, Figure 8, and Figure 9 demonstrate the different levels of alignment within the Acquisition Management, Business Management, Technical Management, and Executive Leadership DoD PM categories, respectively.

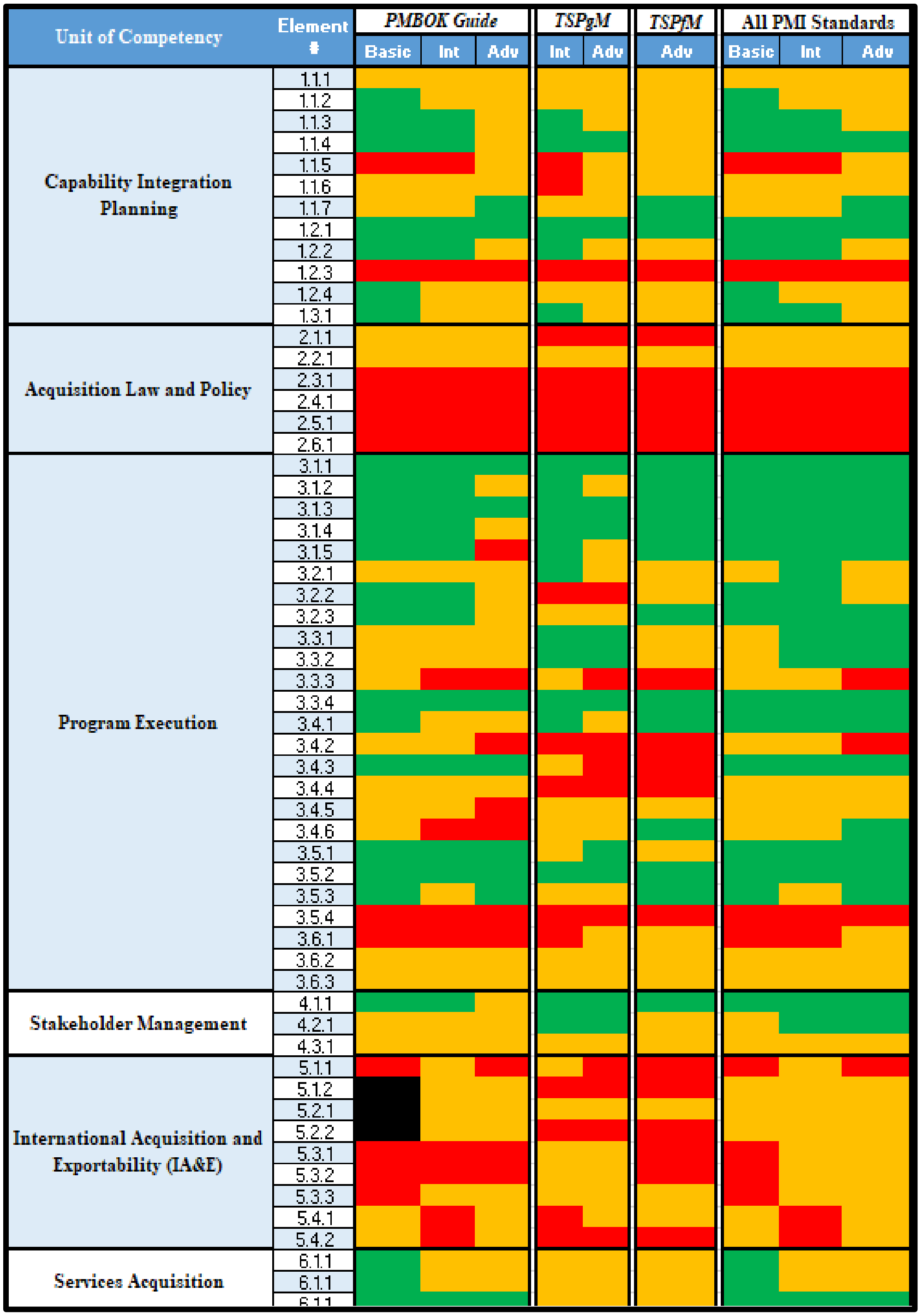


Figure 6. Alignment of Acquisition Management DoD PM Category by PMI Standard

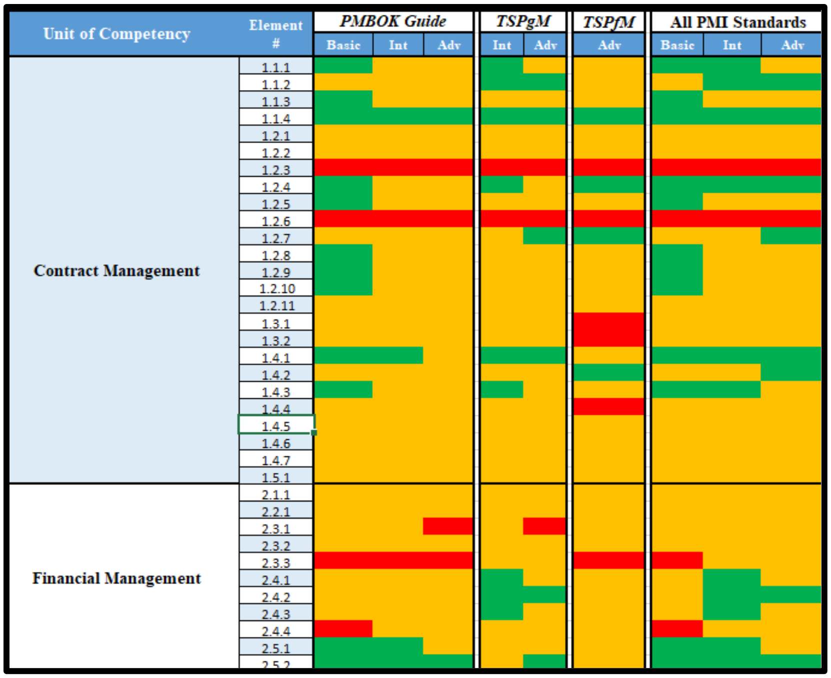


Figure 7. Alignment of Business Management DoD PM Category by PMI Standard

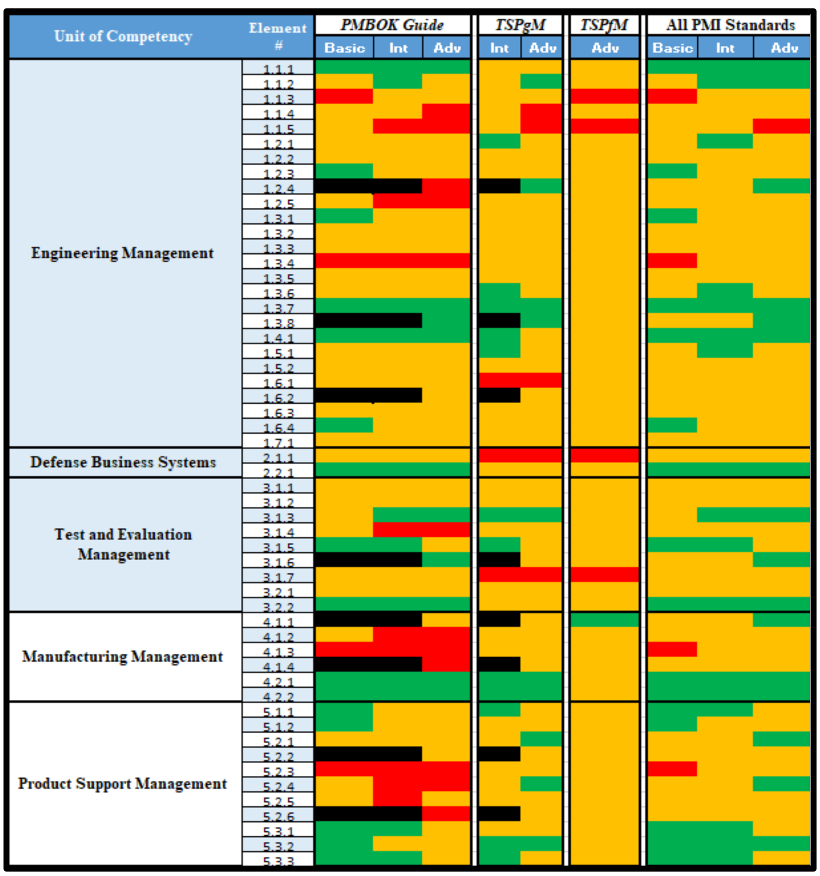


Figure 8. Alignment of Technical Management DoD PM Category by PMI Standard

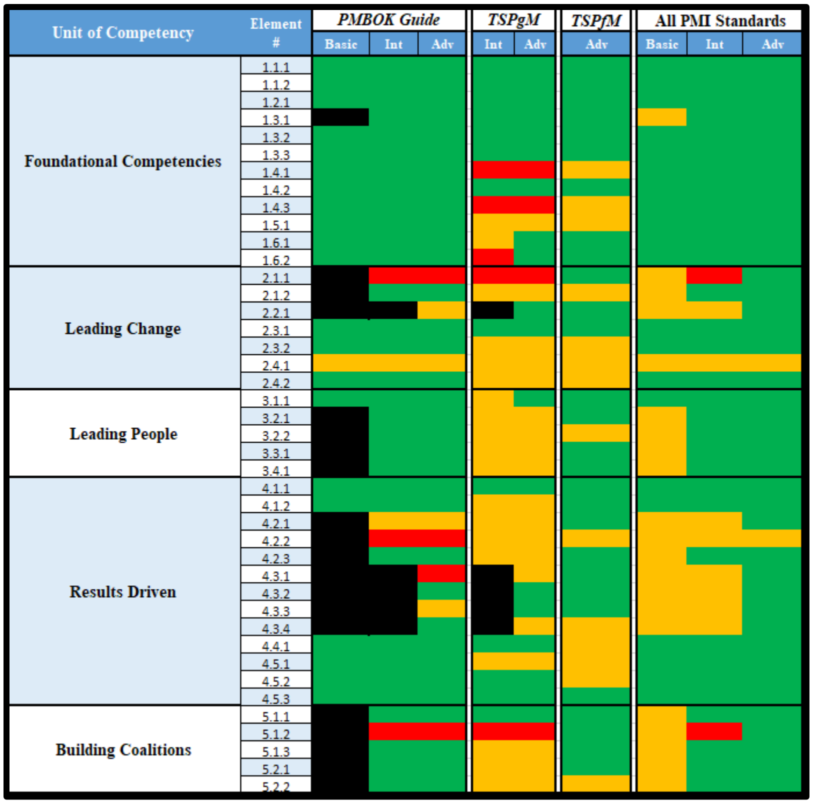


Figure 9. Alignment of Executive Leadership DoD PM Management Category by PMI Standard

The visualizations in these figures demonstrate the alignment improvement of incorporating all three PMI standards to the DoD PM categories. The visualizations also provide a clear view of which DoD PM category is least aligned with the PMI standards. The Acquisition Management DoD PM category from Figure 6 contains the two DoD PM units of competency that are the least aligned across all three PMI standards. They include Acquisition Law and Policy (0% aligned, 33% somewhat aligned, and 67% completely unaligned) and the International Acquisition and Exportability (0% aligned, 74% somewhat aligned, and 26% completely unaligned) units of competency. This is not surprising since these two units of competency are mostly exclusive to the DoD's nature of work and would not contain lexicon that would be commonplace in an industry-wide standard. Therefore, courses in these two units of competency would need to augment acquisition/PM training if the DoD adopted PMI certification standards.

The next section provides a breakdown of the competency mapping by the *PMBOK Guide* project management knowledge areas, *TSPgM* program management performance domains, and *TSPfM* portfolio management performance domains to answer the question, *What PMI knowledge areas and performance domains are most aligned and least aligned with the DoD program management functional career field competency elements?* Analyzing the level of alignment between the DoD's PM functional career field competencies and the PMI standards at this level enables DoD officials to see which knowledge areas and domains are not being applied in the DoD's PM competencies.

This analysis required the approach of mapping the DoD's PM competency elements to the PMI knowledge areas and performance domains by determining the DoD PM competency elements that aligned (both completely and somewhat) with the PMI's knowledge areas and performance domains. This process enabled the tallying of each knowledge area and performance domain that aligned with the DoD PM competency elements. Figure 10 demonstrates the extent to which each of the *PMBOK Guide*'s 10 knowledge areas align with the DoD PM competency elements. This analysis enables DoD stakeholders like the DAU to adjust training objectives to appropriately integrate the *PMBOK Guide* project management knowledge areas into PM certification curriculum.

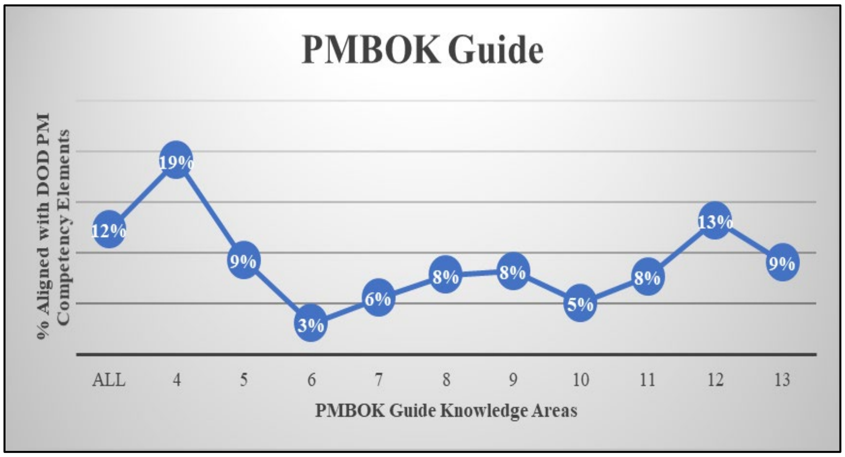


Figure 10. Alignment of the *PMBOK Guide* Project Management Knowledge Areas to DoD Competency Elements

The knowledge areas that exhibited the greatest level of alignment include 4 – Project Integration Management, and 12 – Project Procurement Management.

- **4 – Project Integration Management:** This knowledge area made up 19% of all the aligned and somewhat aligned DoD PM competency elements—more than any other section. Project Integration Management includes the coordination of processes across every *PMBOK Guide* process group (initiating, planning, executing, monitoring and controlling, and closing).
- **12 – Project Procurement Management:** This knowledge area made up 13% of all the aligned and somewhat aligned elements.

The knowledge areas that exhibited the lowest level of alignment include 6 – Project Schedule Management, 10 – Project Communications Management, and 7 – Project Cost Management.

- **6 – Project Schedule Management:** This knowledge area made up only 3% of the aligned and somewhat aligned DoD PM competency elements. This deficiency in alignment is concerning because managing schedule is one of the three project management tenets that make up the triple constraint of project management (Atkinson, 1999).
- **7 – Project Cost Management:** This knowledge area made up 6% of the aligned and somewhat aligned DoD PM competency elements. As stated, cost management is one of the three components of the triple constraint and is therefore critical in project management.
- **10 – Project Communications Management:** This knowledge area made up only 5% of the aligned and somewhat aligned DoD PM competency elements. The impact that communications management can have on a project cannot be overstated. Mortlock (2016) opined that including some form of communications document (e.g.,

a strategic communication [STRATCOM] plan) that conveys a project's or program's desired impact and synchronizes its implementation and execution plans has proven valuable to program success.

To summarize, the least aligned *PMBOK Guide* knowledge areas include project cost, schedule, and communications management. Two of these three are related to the triple constraint, which—if not well-managed—can significantly impact project outcomes. The fact that the DoD PM competencies do not align well with these *PMBOK Guide* sections may be cause for concern because it is an indicator that the DoD is not adequately training their PMs on the importance of managing schedule, cost, and communications—at least in the realm of formal acquisition training.

This section demonstrates the extent to which each of *TSPgM*'s program management performance domains—and elements across all domains—align with the intermediate and advanced DoD PM competency elements (see Figure 11). This analysis enables DoD stakeholders to focus on the most relevant *TSPgM* program management performance domains when restructuring their certification curriculum.

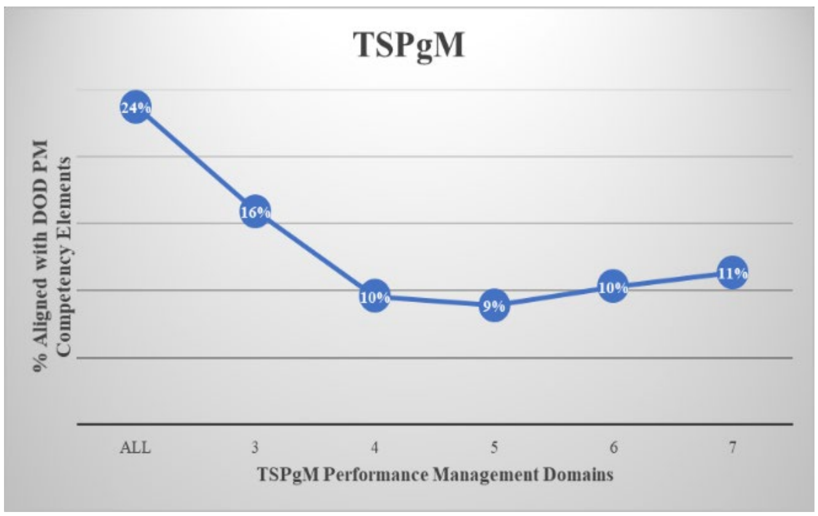


Figure 11. Alignment of *TSPgM* Program Management Performance Domains to Intermediate and Advanced DoD Competency Elements

The program management performance domains that exhibited the greatest level of alignment include All – Elements Across All Knowledge Areas and 3 – Program Strategy Alignment. The remaining four performance domains exhibited mostly similar levels of alignment (9%–11%).

This section demonstrates the extent to which each of *TSPfM*'s portfolio management performance domains—and elements across all domains—align with the advanced DoD PM competency elements (see Figure 12). This analysis enables DoD stakeholders to focus on the most relevant *TSPfM* program management performance domains when restructuring their certification curriculum.

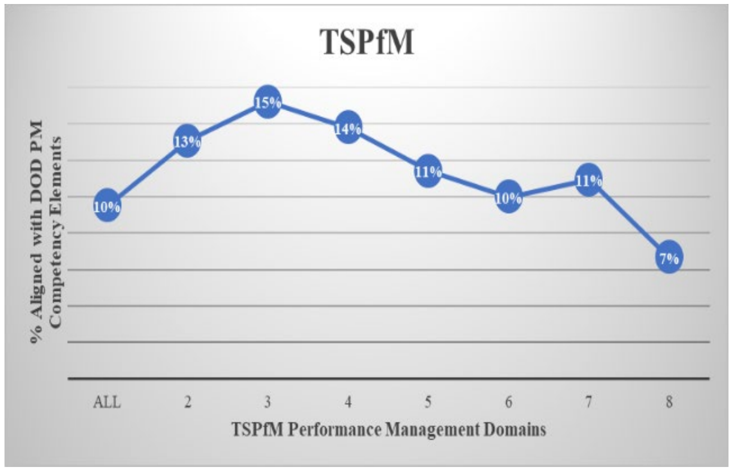


Figure 12. Alignment of TSPfM Portfolio Management Performance Domains to Advanced DoD Competency Elements

The portfolio management performance domains that exhibited the greatest level of alignment include 2 – The Portfolio Life Cycle, 3 – Portfolio Strategic Management, and 4 – Portfolio Governance.

- **2 – The Portfolio Life Cycle:** Just as the *PMBOK Guide* Project Integration Management knowledge area was highly aligned with the DoD PM competencies, so too is this performance domain (13%). Project Integration Management and Portfolio Life Cycle Management heavily rely on information systems that enable effective communication and support seamless and timely transitions between project and life-cycle phases (PMI, 2017a, 2017b).
- **3 – Portfolio Strategic Management:** This performance domain makes up 15% of the aligned DoD PM competencies. Decisions relying on strategic alignment are made at the executive level.
- **4 – Portfolio Governance:** This performance domain makes up 14% of the aligned DoD PM competency elements. The effective implementation of Portfolio Governance aids an organization in becoming auditable (Rendon & Rendon, 2015).

The performance domain that exhibited the lowest level of alignment was 8 – Portfolio Risk Management.

- **8 – Portfolio Risk Management:** This domain made up the lowest number of aligned DoD PM elements. This indicates that the current DoD PM competency elements do not include many elements related to risk management at the advanced level. The DoD should consider addressing this training gap to improve their PMs' ability to identify, analyze, and manage risks at the portfolio level.

Conclusions

This research provided the DoD with information and insight necessary to effectively respond to the Fiscal Year 2020 NDAA's (2019) mandate to base acquisition workforce certification requirements on nationally or internationally recognized third-party standards. The goal of the NDAA's mandate is to improve the quality of the DoD's program management workforce through effective training. As globally recognized standards, PMI's *PMBOK Guide*, *TSPgM*, and *TSPfM* serve as excellent foundations on which to base the DoD's program management certification requirements. The researchers investigated the degree to which the DoD's PM competencies align with the standards of the PMI's *PMBOK Guide, TSPgM,* and *TSPfM.* Analyzing and defining the level of alignment between the two standards enables training organizations to provide more comprehensive training to the acquisition workforce that leverages internationally recognized PM standards.

From a high-level perspective, the *PMBOK Guide* proved to be the most aligned, *TSPgM* is the second most aligned, and *TSPfM* is the least aligned with DoD PM competencies. The knowledge areas and performance domains that were most aligned with the DoD's PM competency elements included concepts for strategic management and life-cycle management. The most concerning finding from this research was the discovery of the relatively low level of alignment of the schedule and cost management knowledge areas across DoD PM competencies.

Table 2. Summary of Research Findings

<table>
<tr><td colspan="4">1. To what extent are the DoD's 2016 program management competency elements aligned with the PMI's PMBOK Guide, TSPgM, and TSPfM? Which PMI standard is the most aligned?</td></tr>
<tr><th>PMBOK Guide</th><th>TSPgM</th><th>TSPfM</th><th>All PMI Standards</th></tr>
<tr><td>34% Aligned (Most Aligned)</td><td>26% Aligned</td><td>25% Aligned</td><td>61% Aligned</td></tr>
<tr><td colspan="4"></td></tr>
<tr><td colspan="4">5. What PMI knowledge areas and performance domains are most and least aligned with the DoD program management functional career field competency elements?</td></tr>
<tr><th colspan="4">PMBOK Guide Knowledge Areas</th></tr>
<tr><th colspan="2">Most Aligned</th><th colspan="2">Least Aligned</th></tr>
<tr><td colspan="2">All – Elements Across All Knowledge Areas</td><td colspan="2">6 – Project Schedule Management</td></tr>
<tr><td colspan="2">4 – Project Integration Management</td><td colspan="2">7 – Project Cost Management</td></tr>
<tr><td colspan="2">13 – Project Stakeholder Management</td><td colspan="2">10 – Project Communications Management</td></tr>
<tr><th colspan="4">TSPgM Performance Domains</th></tr>
<tr><th colspan="2">Most Aligned</th><th colspan="2">Least Aligned</th></tr>
<tr><td colspan="2">All – Elements Across All Performance Domains</td><td colspan="2">N/A</td></tr>
<tr><td colspan="2">3 – Program Strategy Alignment</td><td colspan="2"></td></tr>
<tr><th colspan="4">TSPfM Performance Domains</th></tr>
<tr><th colspan="2">Most Aligned</th><th colspan="2">Least Aligned</th></tr>
<tr><td colspan="2">2 – The Portfolio Life Cycle</td><td colspan="2">8 – Portfolio Risk Management</td></tr>
<tr><td colspan="2">3 – Portfolio Strategic Management</td><td colspan="2">4 – Portfolio Governance</td></tr>
<tr><td colspan="2">4 – Portfolio Governance</td><td colspan="2"></td></tr>
</table>

The following are recommendations based on this research.

1. **Base the new DAWIA PM training certification requirements on the *PMBOK Guide*, *TSPgM*, and *TSPfM.***

A review of the literature and the analysis of the mappings between the DoD's PM functional career field competencies and the PMI standards have led the researchers to believe that the DoD should base their new certification requirements on all three PMI standards. The progressive complexity and scope of the DAWIA certifications "correlate to the complexity and responsibilities required for designated positions and different types of assignments in weapon systems, services, business management systems and information technology, and international acquisitions" (Redshaw, 2011, p. 55). Because the *PMBOK Guide* is exclusively aimed towards individuals charged with managing temporary endeavors (projects), it would not suffice as the sole source of training for the DoD's program management workforce. For example, many PMs lead complex, decades-long programs and manage portfolios that contain a multitude of different projects and programs. Such endeavors require a higher-level managerial perspective and scope of control than the *PMBOK Guide* provides. Therefore, the *PMBOK Guide* would not be able to meet the progressive complexities of the DAWIA certifications and operational responsibilities that are reflected in the DoD's acquisition workforce. By adding *TSPgM* and *TSPfM* to the certification framework of their PMs, the DoD can account for the increase in managerial scope that PMs will see as they progress in their careers.

2. Maintain the three-tiered certification model.

The DAWIA three-tiered certification model consists of Level I (basic), Level II (intermediate), and Level III (advanced). This progressive education model enables PMs to be trained on relevant subject matter that align with required responsibilities and prevents them from learning out-of-scope material too early in their career. For example, a DoD project manager would rarely require training on portfolio life-cycle management when the scope of their responsibilities is to manage small projects. On the other hand, DoD program executive officers, who primarily manage portfolios, require training on basic project management practices because project and program management fundamentals form the basis of portfolio governance and strategic alignment across projects, programs, and portfolios. To guide PMs from project management to being capable of leading vast programs and portfolios, the DoD must establish training that gradually increases in scope in correlation with the scope of the PM's current job responsibilities. This can be accomplished by establishing certification standards based on the following model:

- DAWIA Level I (basic/project managers) – PMP certification based on the *PMBOK Guide*
- DAWIA Level II (intermediate/PMs) – PgMP certification based on *TSPgM*
- DAWIA Level III (advanced/program and portfolio managers) – PfMP certification based on *TSPfM*

This would allow for a gradual increase in program management knowledge and application and align experience to training certifications. To improve upon this model, the DoD should enable cross-sectioning of the three PMI standards into each certification level. As mentioned, the *PMBOK Guide* serves as the foundation for both *TSPgM* and *TSPfM* and therefore holds valuable information that should be used in the training of managers of programs and portfolios. Likewise, including sections of *TSPgM* and *TSPfM* with the Level I education allows inexperienced DoD PMs to understand the larger picture of their projects and how they fit into programs and portfolios.

3. Augment professional certifications with DoD-specific PM training.

As this research has demonstrated, the three PMI standards alone do not cover all the DoD PM competencies. For example, if the PMP certification is adopted for DAWIA PM Level I (basic), *TSPgM* certification is adopted for DAWIA PM Level II (intermediate), and *TSPfM* certification is adopted for DAWAI PM Level III (advanced), additional DAU training courses would need to focus on the areas least aligned, like Acquisition Law and Policy and International Acquisition and Exportability. Additional DAU training would be required in the areas not covered by PMI standards sufficiently, including the following:

- Acquisition Management
    - Acquisition Policy and Law
    - International Acquisition and Exportability
- Business Management
    - Contract Management, specifically in pre-solicitation planning and execution
- Technical Management

- Engineering Management, specifically technical planning in understanding, applying, and ensuring program protection, cybersecurity, and counterintelligence

Considering that 190 DoD PM competencies exist, the fact that PMI standards aligned reasonably well reinforces the recommendation to adopt the PMI standards.

4. **Consider all three components of auditability.**

In conjunction with the modification to its PM certification requirements, the DoD should consider the research of Eckerd and Snider (2017) and Rendon and Rendon (2015). Both sets of research emphasize the importance of ensuring capable processes and effective internal controls. While this research exclusively considered the development of competent personnel through an analysis of training standards, the DoD should ensure that correct measures are being taken in modifying training certifications and in developing effective processes to transition the workforce and the training staff to the new standards.

5. **Revitalize the *U.S. Department of Defense Extension to: A Guide to the Project Management Body of Knowledge*.**

To fill competency gaps that are not covered by PMI standards, the DoD should look to the *U.S. Department of Defense Extension to: A Guide to the Project Management Body of Knowledge* (*PMBOK Guide*; DoD & DAU, 2003).

References

Alie, S. S. (2016). *Interlocking program and project governance with PMI's process groups: How to form the proper baseline and tailor project/program governance* [Paper presentation]. PMI Global Congress 2016, EMEA, Barcelona, Spain. https://www.PMI.org/learning/library/interlocking-program-project-governance-PMI-process-group-10175

Atkinson, R. (1999). Project management: Cost, time and quality, two best guesses and a phenomenon, its time to accept other success criteria. *International Journal of Project Management*, *17*(6), 337–342. https://doi.org/10.1016/S0263-7863(98)00069-6

Bernard, R. H. (1996). Qualitative data, quantitative analysis. *Journal of Computational and Applied Mathematics*, *8*(1), 9–11. https://doi.org/10.1177/1525822X960080010401

Bond, D. P., Davis, S. M., & Pearsall, A. D. (2016). *The Goldwater–Nichols Act of 1986: 30 years of acquisition reform* [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun. https://calhoun.nps.edu/handle/10945/51649

Choi, D. (2009). *A study on improving defense acquisition through the application of Defense Acquisition Workforce Improvement Act (DAWIA) concept to defense industry workforce* [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun. http://hdl.handle.net/10945/4571

Defense Acquisition Workforce Improvement Act, H.R. 5211, 101st Cong. (1990). https://www.congress.gov/bill/101st-congress/house-bill/5211

DoD & Defense Acquisition University. (n.d.). *Certification standards & core plus development guide: Program management level 1*. DAU. https://icatalog.dau.edu/onlinecatalog/CareerLvl.aspx?lvl=1&cfld=9

DoD & Defense Acquisition University. (2003). *U.S. Department of Defense Extension to: A guide to the project management body of knowledge (PMBOK guide).* Defense Acquisition University Press. http://www.risk-services.com/DoDExtPMBOKJune2003.pdf

Eckerd, A., & Snider, K. (2017). Does the program manager matter? New public management and defense acquisition. *The American Review of Public Administration*, *47*(1), 36–57. http://hdl.handle.net/10945/62045

Gansler, J. S., Berteau, D. J., Maddox, D. M., Oliver, D. R., Salomon, L. E., & Singley, G. T. (2007). *Urgent reform required: Army expeditionary contracting*. *Report of the Commission on Army Acquisition and Program Management in Expeditionary Operations.* https://ogc.altess.army.mil/Documentation/EandF/Guidance/Gansler%20Commission%20Report_Final%20Report_10-31-07.pdf

GAO. (1995). *High-risk series: An overview* (GAO/HR-95-1).

GAO. (2010). *Defense acquisition workforce: DoD's training program demonstrates many attributes of effectiveness, but improvement is needed* (GAO-11-22).

GAO. (2018a). *Defense acquisition workforce: Opportunities exist to improve practices for developing program managers* (GAO-18-217).

GAO. (2018b). *F-35 Joint Strike Fighter: Development is nearly complete, but deficiencies found in testing need to be resolved* (GAO-18-321).

GAO. (2019a). *High-risk series: Substantial efforts needed to achieve greater progress on high-risk areas* (GAO-19-157SP).

GAO. (2019b). *Improving program management: Key actions taken, but further efforts needed to strengthen standards, expand reviews, and address high-risk areas* (GAO-20-44).

Holtzman, J. (1999). Getting up to standard. *PM Network, 13*(12), 44–46. https://www.PMI.org/learning/library/ansi-standard-5057

Kupec, C. (2013). *Project management professional training needs for defense industry projects* [Doctoral dissertation, Old Dominion University]. ODU Digital Commons. https://digitalcommons.odu.edu/stemps_etds/64

MacStravic, J. A. (2016, September 6). *Program management functional career field competencies* [Memorandum]. DoD. https://www.dau.edu/training/career-development/program-management/Lists/Blog/Attachments/29/Prog%20Mgmt%20Funct%20Competencies_160906%20(003).pdf

Mortlock, R. F. (2016, April–June). Stratcom strong. *Army AL&T Magazine*, 126–129. https://asc.army.mil/docs/magazine2/armyalt-apr-jun2016.pdf

Mortlock, R. F. (2020, Spring). Been there, done that: The profession of acquisition. *Army AL&T*, 72–77. https://asc.army.mil/armyalt/Spring2020/html/7.html

National Defense Authorization Act for Fiscal Year 2020, Pub. L. No. 116–92 § 861 (2019). https://www.congress.gov/bill/116th-congress/senate-bill/1790/text

Office of the Under Secretary of Defense for Acquisition and Sustainment. (2020a). *Operation of the adaptive acquisition framework* (DoDI 5000.02). DoD. http://acqnotes.com/wp-content/uploads/2014/09/DoD-Instruction-5000.2-Operation-of-the-Adaptive-Acquisition-Framework-23-Jan-2020.pdf

Office of the Under Secretary of Defense for Acquisition and Sustainment. (2020b). *The defense acquisition system* (DoDD 5000.01). DoD. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/500001p.pdf?ver=2020-09-09-160307-310

Pernin, C., Axelband, E., Drezner, J., Dille, B., Gordon IV, J., Held, B., McMahon, S., Perry, W. Rizzi, C., Shah, A., Sollinger, J., & Wilson, P. (2012). *Lessons from the Army's Future Combat Systems program* (Report No. HQD105725). RAND. https://www.rand.org/content/dam/rand/pubs/monographs/2012/RAND_MG1206.pdf

President's Blue Ribbon Commission on Defense Management. (1986). *A quest for excellence: Final report to the president*. Government Printing Office.

Project Management Institute. (2017a). *A guide to the Project Management Body of Knowledge* (6th ed.).

Project Management Institute. (2017b). *The standard for portfolio management* (4th ed.).

Project Management Institute. (2017c). *The standard for program management* (4th ed.).

Project Management Institute. (2020). *Project Management Institute certifications*. https://www.PMI.org/certifications/types

Redshaw, M. (2011). *Developing competencies required for directing major defense acquisition programs: Implications for leadership* [Doctoral dissertation, University of Phoenix]. ProQuest Dissertations Publishing. http://search.proquest.com/openview/3c8ebde08bf348e5c110f4ce06cc3319/1?pq-origsite=gscholar&cbl=18750&diss=y

Rendon, R. (2010). *Professionalization of the U.S. defense acquisition workforce: Progress, problems and future directions* [Faculty publication, Naval Postgraduate School]. NPS Archive: Calhoun. http://hdl.handle.net/10945/40376

Rendon, R. (2019). Enhancing professional and technical excellence: Analysis of contract management competency models. *Proceedings of the 16th Annual Acquisition Research Symposium*, *1*, 86–112. http://hdl.handle.net/10945/63019

Rendon, R., & Rendon, J. (2015). Auditability in public procurement: An analysis of internal controls and fraud vulnerability. *International Journal of Procurement Management*, *8*(6), 710–730.

Ross, D. W., & Shaltry, P. E. (2006). *The new PMI standard for program management* [Paper presentation]. PMI Global Congress 2006, EMEA, Seattle, WA.

Schwartz, M. (2010). *The Nunn–McCurdy Act: Background, analysis, and issues for Congress* (CRS Report No. R41293). Congressional Research Service. https://apps.dtic.mil/sti/citations/ADA524187

Shenhar, A., & Dvir, D. (2004, July 14). *Project management evolution: Past history and future research directions* [Paper presentation]. PMI Research Conference, London, England. https://www.PMI.org/learning/library/project-management-evolution-research-directions-8348

Woolsey, J. (2019). *2019 annual report: Transforming acquisition training*. DAU. https://www.dau.edu/about/_layouts/15/WopiFrame.aspx?sourcedoc=/about/Documents/2019%20Annual%20Report.pdf&action=default

Turnover Among the Civil Service Components of the Department of Defense Acquisition and Medical Workforces

Spencer T. Brien—Naval Postgraduate School [stbrien@nps.edu]

Abstract

This study is an analysis of workforce turnover behavior among two segments of the Department of Defense (DoD) civilian workforce. Its research objective is to identify turnover trends among the civil service medical healthcare workforce and the civilian component of the defense acquisition workforce. The study employs a logistic regression model to estimate the relationship between employee personnel characteristics and the likelihood of separating from the civil service. In both groups, the overall pattern of attrition was consistent with a life-cycle model of turnover behavior. The results highlight specific employment categories where improvement in recruiting a more diverse workforce is needed.

Introduction

This study is an analysis of workforce turnover behavior among two segments of the Department of Defense (DoD) civilian workforce. Its research objective is to identify turnover trends among the civil service medical health care workforce and the civilian component of the defense acquisition workforce. The focus on the civilian medical workforce is notable because the Defense Health Agency (DHA) is currently undergoing a significant reorganization. Establishing baseline turnover patterns prior to the completion of the reorganization will facilitate future program evaluations of the impact of the reorganization on the medical workforce.

In both groups, the overall pattern of attrition was consistent with the life-cycle model of turnover behavior. Employees that are not yet eligible for retirement had higher propensities to separate at early stages of their careers and relatively low attrition rates as they approach retirement eligibility. Retirement-eligible personnel conversely separate at an accelerating rate as they move past their earliest eligibility date. The preliminary statistical analysis illustrated how these two overall patterns hold for both medical and acquisition employees but also highlighted how the acquisition workforce benefits much more from mid-career appointments. This surge of older appointees flowing into the workforce from uniformed services also affects the outflow years later as these employees retire. Even if medical and acquisition employees respond relatively similarly to advancement in age and tenure, the difference in the composition of the workforce will change the recruitment, training, and succession planning strategies needed to manage the two types of employees.

Another finding from this study is that the DoD can do much better in hiring racial minorities into technical and white-collar positions. This analysis of the medical workforce found that personnel identifying as Black, Native American, and Hawaiian or Pacific Islander were almost twice as likely to be employed in blue-collar work, while employees identifying as White or Asian were nearly evenly distributed between the two labor categories. Pursuing diversity and representation within occupational categories, and not just within the workforce as a whole, is essential to achieving a public workforce that aligns with the American public. Furthermore, our attrition analysis found that employees identifying as Black, Asian, and Hawaiian and Pacific Islander had lower likelihoods of voluntarily separating when compared to white employees. Attaining diversity within these technical fields will also bring a more stable workforce.

For the acquisition workforce, the study employed command-level employee satisfaction survey results to estimate whether organizational climate affects turnover behavior. On this point, the model output was inconclusive, and no statistically significant relationship between turnover and employee satisfaction was identified. Sensitivity analysis revealed a similar life-cycle trend for employees in both the medical and acquisition workforces. The primary difference in turnover incidence between the two segments of the workforce is the much larger intake of employees at the mid-career level of experience within the acquisition workforce. This difference in the inflow of acquisition personnel leads to a different demographic profile of the workforce when compared to the medical professionals.

Background

Turnover Behavior in the DoD

Understanding turnover behavior among the DoD civilian workforce is an important component of strategic workforce management. Civilian personnel perform a variety of functions that are crucial to the operations of the DoD. This workforce, however, has for decades been subject to a variety of adverse environmental conditions, such as pay freezes, furloughs, and other factors that can contribute to increased rates of voluntary separation (Fernandez et al., 1985). While turnover among uniformed military personnel has been a topic of focused study for many years, the civilian DoD workforce has received less attention (Baldor, 2018; Buddin, 1984; Gebicke, 1998; GAO, 1997; Rabkin, 2000).

This study explicitly builds off of prior research by Brien (2019) and Buttrey et al. (2018). These studies, conducted in collaboration with TRAC–Monterey and the DoD Office of People Analytics, include nonparametric survival analyses of turnover among civilian employees across the entire DoD. Collectively, the studies examined gender-based turnover differentials in a variety of different contexts, such as the variation in gender-based turnover differentials across different military branches. Their work also explored gender differentials across different categories of science, technology, engineering, and mathematics (STEM)-affiliated occupation statuses. The key findings of those studies centered on the lack of a gender differential in turnover for women filling STEM occupations, while a significant gender gap persists in non-STEM job categories.

This study advances research on turnover among the civilian DoD workforce by examining two subsets of the workforce in greater detail. The healthcare workforce and the acquisition workforce constitute two distinct subsets within the DoD. The nature of their work and their skillsets are crucial to continued DoD operations. Analyzing the trends associated with turnover in each of these communities separately may help produce insights that improve the administration of these populations.

The Medical Civilian Workforce

Understanding the institutional background of the ongoing reorganization of the DoD civilian medical workforce provides context for current turnover trends within this subpopulation. In a thesis completed in collaboration with this research project, Paone (2020) described the motivation for this reorganization:

In Fiscal Year (FY) 2018, the DHA began a four-year transition plan to assume authority of the DoD's Military Healthcare System (MHS) including over 400 military treatment facilities (MTFs) that serve over 9.4 million patients worldwide (DHA, 2019a). The U.S. Congress mandated the formation of a centralized governance of the MHS in 2013, and subsequently, the deputy secretary of defense directed the official establishment of the DHA as a Combat Support Agency (CSA) in March of the same year (DoD Directive:

3000.06, 2013). However, the proposal to integrate military medical services into a joint system to increase quality of care and efficiency is not a new concept and is one that experts have been proposing repeatedly since the 1980s. The most recent push and subsequent creation of the DHA followed a 2011 DoD special task force report on MHS governance that cited inconsistencies and inefficiencies in the delivery system (DHA, 2017). Under the previous system, each of the armed services (U.S. Navy, Army, and Air Force), respectively, managed its own medical treatment facilities in a jointly supportive but non-integrated system.

The purpose of the DHA is to ensure the standardization of quality of care to all its military healthcare beneficiaries as part of a larger effort to provide greater integration of services and a more centralized control over purchased care. The DHA reports its mission is to achieve greater integration of direct and purchased care delivery systems, so it can accomplish a "4-point aim"—achieve medical readiness, improve the health of its beneficiaries, enhance the experience of care, and lower healthcare costs (DHA, 2019a). The DHA further establishes its agency goals as follows:

> Empower and Care for Our People
>
> Optimize Operations Across the Military Health System
>
> Co-create Optimal Outcomes for Health, Well-Being and Readiness
>
> Deliver Solutions to Combatant Commands (DHA, 2019a, p. 1)

The DHA is expected to complete a full transition of operations governance before FY2022 across all armed services and their respective MTFs. At this time, military active duty manpower will fall under the command of a joint readiness center environment, which will share a commanding officer with the MTF during the transition. The commanding officer will be accountable to both the DHA and combatant commanders. This proposed organizational structure has remained fluid with evolving strategy, and it may continue to change with strategy implementation as this landmark organizational transition continues (Paone, 2020, pp. 1–2).

Examining the healthcare workforce separately from the rest of the civilian workforce may help reveal the ways that the unique aspects of healthcare work impact turnover decisions. Healthcare workers experience unique training, labor hours, and environmental conditions relative to other DoD employees. Recent research on turnover among the broader healthcare community highlights the role of burnout in increasing employees' intention to voluntarily separate from medical employment (Aiken et al., 2002; Gesesew et al., 2016; Rambur et al., 2008).

Although medical professionals endure many years of specialized training to become doctors, nurses, physician's assistants, and the other various forms of medical professionals, recent research has identified relatively high turnover rates among younger professionals. McGrail et al. (2017) examine physician turnover rates in both rural and urban communities. Their analysis revealed that, during the 2000–2014 period, physicians under the age of 45 had double the turnover rate of their older colleagues. The authors identify institutional sorting as one problem; it led to poorer and more rural areas losing their physicians at a higher rate.

The institutional factors that McGrail et al. (2017) identify as having an impact on turnover behavior may manifest in the DoD medical workforce. As organizations undergo periods of transition, it is more complicated and challenging to offer the institutional support that younger professionals need to keep them in the profession. Additionally, adverse institutional conditions may cause older professionals to accelerate their retirement plans.

Understanding the baseline turnover prior to the organizational changes will enable future researchers to perform a program evaluation of the impact of DHA's transition plan on employee retention.

The Defense Acquisition Workforce

Ensuring that the Defense Acquisition Workforce is sufficiently staffed, both in terms of size and breadth of skills, is essential to maintain the pace of defense procurement. McCauley's (2020) thesis, written in collaboration with this research project, describes the challenge of monitoring and stabilizing turnover within this workforce:

> The U.S. government annually spends hundreds of billions of dollars on national defense; although this is a tremendous amount of money, every cent of that is needed to ensure that the United States remains the world's strongest military power. The DoD Acquisition Workforce (AWF) oversees the utilization and spending of large amounts of this budget. These General Schedule (GS) employees are responsible for the "development, acquisition and sustainment of warfighting capabilities, systems and services" (Secretary of the Navy [SECNAV], 2019, p. 1). In keeping with the Defense Acquisition Workforce Improvement Act (DAWIA) of 1990, the AWF is made up of highly skilled and qualified federal employees. For example, most of the Department of the Navy's acquisition employees are a GS-13, which is two below the max rank for the General Schedule system; their military counterparts are usually an O-4 ([SECNAV], 2019). These are relatively senior positions that require a great deal of training and responsibility.
>
> However, due to contractual differences between military and civilian personnel, civilian employees can be much harder to maintain and replace. Civilians can decide to terminate from a position whenever they choose, regardless of whether there is a scheduled replacement or not. Acquisition employees also have incredibly valuable skills that can easily transfer to high paying, private sector positions. The very jobs and contracts that the AWF oversees provides the expert experience needed for the lucrative civilian market. For this exact reason, it is critical that the DoD does its absolute best to retain the talent it already has. Hiring new employees within the GS system is incredibly difficult as well. There is a great deal of litigation that goes along with the process that keeps it from being quick and timely. This same process makes it equally difficult to fire or terminate undesirable employees. That is why it's so important to acquire and retain qualified workers the first time around. If the AWF is experiencing high attrition rates, the entire acquisition process can sometimes come to a screeching halt. This is not only a waste of millions—and sometimes billions—of taxpayer dollars, but in some instances, it can cost the lives of U.S. military members. It is always of the utmost importance that we get safe, reliable, and superior equipment to our fighting forces as fast as humanly possible. Therefore, every action possible must be taken to mitigate civilian acquisition attrition and ensure that the AWF always employs the best and brightest people (McCauley, 2020, pp. 1–2).

A series of studies published by RAND have sought to document changes in the inventory of the defense acquisition workforce over the last 2 decades (Gates et al., 2008, 2009, 2013, 2018). These reports provide a detailed descriptive analysis of various characteristics of acquisition employees, such as the distribution of new acquisition hires across the military service branches, the relative level of education of employees, and the pace of retirements among this workforce. Key findings include observations that 1. new hires into acquisition job categories have a higher level of education in 2017 than they did a decade prior, and 2. the civilian acquisition workforce has a lower turnover rate that the rest of the civilian DoD workforce (Gates et al., 2018).

The authors of the RAND studies note that a key limitation of their work is the lack of data on contracted workers performing acquisition tasks. This is a near universal constraint on studies relying on extracts from the Defense Manpower Data Center, which only contains records of direct employees of the DoD. Another aspect of the RAND studies that is not a limitation, but a feature that differs from this analysis, is that they do not engage in statistical modeling to identify the relationships between employee characteristics and turnover behavior. The studies provide deep and detailed historical descriptions of trends within the acquisition workforce, but they do not seek to identify causal relationships between different features of the workforce.

Methodology and Data

The data for this analysis are obtained from the civilian master files maintained by the Defense Manpower Data Center. These data were accessed through the PDE operated by the AAG. The data used in this study consist of two subsets of the full civilian workforce data set used in Brien (2019) and Buttrey et al. (2018). As described in the prior report,

> the civilian master files contain demographic and detailed information found in an employee's personnel file. We merge the civilian transaction files with the civilian master files to catch all the data transactions that take place within each employee's record. These transactions are changes in the employee's career, which include salary changes, changes of appointment, and, most importantly, separation. We flag those employees with separation transactions to determine which employees attrite during the eight years of our study. We also classify employees as "disappeared" who do not have a separation transaction file, but their master file quarterly snapshots end during the research period, indicating they are no longer employed. (Buttrey et al., 2018, p. 30)

After the separations were identified in the data, we limited the data set to a cohort of civilian DoD employees who were appointed to their positions in 2009. This reduced the total count of individuals tracked in our study to 102,009. Another 4,355 records were dropped due to missing or inconsistent data. The total number of individuals tracked in the 2009 employee cohort was 97,654. These individuals were observed over an 8-year period. As time passed, we were able to observe members of the 2009 cohort separate from federal employment and then conduct statistical analyses of the employee attributes that were associated with higher turnover rates.

Limiting the analysis to the 2009 cohort provides a comparable sample of individuals for an analysis of separation behavior. Members of the cohort were subject to the same macroeconomic effects that influence turnover behavior. Although cohort members were all newly appointed to federal service in 2009, the sample contains considerable variation in the age at entry, level of education, and employment history characteristics.[17] This study examines two subsets of this cohort: the acquisition and medical workforces.

Although Brien (2019) and Buttrey et al. (2018) employed nonparametric survival analysis methods to examine turnover behavior, this study ultimately used parametric logistic regression methods to model relationships between employee and environmental characteristics and turnover. This change in the methodology from what was initially

---

[17]See Brien (2019) for descriptive statistics of the entire 2009 cohort.

intended in the project proposal was made for two compelling reasons. First, peer reviewer feedback received on an academic manuscript that was derived from the Brien (2019) report highly recommended using a parametric approach that is more familiar to the academic audience of the public human resource management literature. As noted in Brien (2019), the public human resource management literature contains relatively few studies of turnover behavior, and those that do exist, such as Cho and Lewis (2012), do not have access to individual-level behavior tracked over multiple consecutive years. The relative novelty of the data set used in this series of studies makes replicating some of the earlier estimates in the broader literature of the marginal effects of individual characteristics on turnover behavior a highly desirable contribution. Given this external feedback, the methodology of this study has been adjusted to have greater external impact. A second reason motivating the change in methodology is that it allowed the thesis students who assisted in this research, Paone and McCauley, to participate to a much greater extent because the methods aligned with those presented in their NPS coursework.

Shifting from a survival analysis approach to a logistic regression strategy necessitated a narrowing of the analytic focus of the analysis. Instead of measuring turnover outcomes of the entire cohort from 2009 through 2017, the study instead examines turnover within the cohort that occurred within a single year—2014. This allows the creation of a single binary outcome variable measuring whether the individual separated from employment in 2014.

Two subsets of the data are created: one for the healthcare workforce and the other for the acquisition workforce. The healthcare workforce is identified using the Job Family Standard (JFS) series code (U.S. Office of Personnel Management, 2017). The 0600 classification identifies jobs in the Medical and Healthcare Group. This includes physicians, nurses, physical therapists, and other categories of medical professionals. Individuals in the 0600 group whose occupation is associated with custodial work or dental care were excluded from the subset. Additionally, individuals whose listed age at time of entry to employment was below 18 or over 80 were dropped from the analysis. The final count of individuals taken from the 2009 cohort that were employed for at least part of 2014 consists of 50,946 unique people.

The acquisition workforce was identified by selecting 20 job series codes that are associated with acquisition work. After limiting the 2009 cohort to just these job series and following the same data cleaning process as used with the medical data, a subset of 105,940 individuals were identified.

The personnel data obtained from the PDE environment contain a variety of individual characteristics that may impact turnover behavior. Individual demographic characteristics such as age, sex, and race are each included in the data. The medical workforce subset includes codes that distinguish between nine different occupational categories. These allow differentiation between occupations that require higher-education degrees versus those that do not. Additionally, some professions, such as nurses, require regular shift work, which may incur a higher physical toll on employees over time. Differentiating between nurses and medical professionals that do not commonly engage in shift work, such as physical therapists, may help further identify differences in turnover behavior within the medical workforce.

Prior research stemming from this multiyear project focused on turnover differentials between blue-collar and white-collar employees (Morgan, 2019; Urech, 2019). Their studies sought to determine whether employees engaged in different types of work exhibit different turnover patterns. Copeland’s (2008) review of transformational trends in the composition of the federal workforce highlighted the increase in the number of white-collar jobs and the underrepresentation of women in white-collar jobs, particularly among the Senior Executive Service. More recently, some career areas that have been traditionally filled by women, such as nursing, have been described as “pink-collar” (Howe, 2017). These professions traditionally require significant education and credentialing but do not operate in an office environment.

As shown in Figure 1, we divide the medical workforce into four categories: (1) healthcare tech or clerical assistant, (2) medical officer, (3) healthcare professional, and (4) nurse. We create indicator variables for each category and include them in the model to estimate whether the medical occupation class is associated with differential turnover behavior. Understanding these behavioral differentials may help guide strategic hiring to replace losses to voluntary attrition. It may also help identify subpopulations of the medical workforce that could be targeted for retention policies. The first category, healthcare tech or clerical assistant, is omitted from the model and is treated as the baseline subpopulation. The estimates of the other categories are the differential turnover effect of being in those other categories relative to the base subpopulation’s behavior.

The fields from the personnel records that describe the individual’s federal work history are especially important to this analysis. The appointment date field, in combination with the employee’s age, is used to compute retirement eligibility dates. These dates are also adjusted using records detailing prior years of federal service. From the retirement eligibility dates, we generate indicator variables to identify whether the employee is eligible to retire in 2014. The level of educational attainment is also reported in the civilian records. We create a series of indicator variables to denote different education levels.

The personnel records include fields describing the organizational location of the employee within the DoD. These codes identify both the service branch in the U.S. military and, for the acquisition employees, an organizational code that identifies the naval command employing Navy personnel. Controlling for these organizational characteristics may help capture some unobserved features of organizational climate and culture that influence employee satisfaction and, subsequently, turnover behavior.

The acquisition data are merged with the survey results from a Director for Acquisition Career Management (DACM) employment satisfaction survey. The results from this survey were used in two NPS theses assessing acquisition command climate (Collins & Garcia, 2018; McKeithen, 2016). In the survey, employees were asked to report on 11 dimensions of job satisfaction. The questions related to job satisfaction, supervisor-related commitment, their job characteristics, job role ambiguity, job stress, commute strain/safety, work/family conflict, organizational justice, job fit, workplace values, and high-quality relationships. These questions were designed from work on organizational theory by Fields (2002). We computed average satisfaction scores by Navy command group and then use the command average for each measure as a control for the overall climate on that measure for all employees in the PDE cohort sample.

There are several limitations to this approach. First, the respondents to the survey are not the same individuals in the 2009 cohort. The survey was administered to both uniformed and civilian personnel, and it is not possible to separate them. Second, the survey was administered on a relatively small sample of acquisition employees. After cleaning the data, complete results that are usable for the analysis were only obtained from 672 respondents. Examining the distribution of responses across the Navy commands, some commands only had a handful of usable results. Table 2 displays the distribution of responses by command and sex.

Table 1. Survey Observations by Command (McCauley, 2020)

Observations by Command and Gender	Code	Male	Female	Total
NAVAIR (Naval Air Systems Command)	1	46	79	125
NAVSEA (Naval Sea Systems Command)	2	66	73	139
NAVSUP (Naval Supply Systems Command)	3	74	68	142
SPAWAR (Space and Naval Warfare Systems Command)	4	14	26	40
MARCORSYSCOM (Marine Corps Systems Command)	5	1	2	3
MSC (Military Sea Lift Command)	6	12	6	18
SSP (Strategic Systems Programs)	7	1	1	2
BUMED (Bureau of Medicine and Surgery)	8	1	0	1
OPNAV (Office of the Chief of Naval Operations)	9	2	4	6
OSBP (Office of Small Business Programs)	10	2	2	4
MARCOR I&L (Marine Corps Installations and Logistics Command)	11	2	2	4
NAVFAC (Naval Facilities Engineering Command)	12	78	93	171
ONR (Office of Naval Research)	13	8	9	17
TOTAL		**307**	**365**	**672**

Due to the limited number of responses in some command units, we exclude those commands from the analysis of acquisition turnover. Only five commands, NAVAIR, NAVSEA, NAVSUP, SPAWAR, and NAVFAC are included in the analysis. Making this restriction reduces the subsample size of the acquisition workforce to 5,541 individuals. Although this is costly in terms of data, the scientific benefits of conducting a turnover analysis with both measures of actual behavior and employee satisfaction justify this limitation. As discussed in Brien (2019), studies of public sector turnover have had very limited access to actual turnover data, and those that have had it relied on agency-level figures of turnover rates. The individual-level data this team has accessed in the PDE environment is still relatively unique in this field. Integrating it with employment data to test a more complete model of turnover behavior that incorporates both economic and psychological factors relating to turnover is a significant advancement to the study of workforce attrition.

Model of Turnover Behavior

This study examines turnover behavior among the DoD civilian workforce using a life-cycle model of turnover behavior. This model, described by Cho and Lewis (2012), explores the economic rationale behind the employee's decision to voluntarily separate from their job. This economic model predicts that employees will choose to leave their job when the expected benefits of other employment or retirement exceed their current compensation. Tangible factors such as pay, commuting costs, and health and retirement benefits are key factors in this decision. Other nontangible factors also included in the model are the organizational environment and employee job satisfaction (Pitts et al., 2011).

This analysis employs a logistic regression model. The dependent variable is a binary indicator of whether the individual voluntarily separated from public service in 2014. The model is designed to obtain estimates of the marginal effect of changes in the explanatory variables on the odds that an attrition event will occur.

The estimates generated from a logistic regression are the log of the odds ratio, often referred to as the log-odds ratio. Although the log-odds ratio can indicate the sign and statistical significance of the marginal effect of a variable on the odds of an attrition event, it has relatively little interpretative value. Usually, the log-odds is converted to an odds ratio by exponentiating it. Additionally, because the logistic function is nonlinear, the marginal effects of the odds ratio is calculated at different values of the explanatory variables, especially if they are discrete categorical variables, such as sex, education level, or profession. For continuous variables, the marginal effects are typically calculated at the mean value.

Findings

Figures 2a and 2b present, for the nonretirement eligible population, employee age in 2014 for the medical and acquisition workforces, respectively. The two lines in each graph separate the age distributions of the employees that did separate from federal employment from those that remained in employed status. Figure 2a illustrates that among the nonretirement eligible medical employees, a higher percentage of those that separated fell within the ages of 23 to 40. For those between the ages of 40 and 61, a higher percentage did not experience turnover. Among the acquisition employees depicted in Figure 2b, the younger workforce between 23 and 40 does not appear to show a difference in turnover behavior, while those aged between 40 and 55 show a relatively lower propensity to separate from federal employment. This indicates that retaining medical employees at earlier stages in their careers may be a greater challenge than retaining acquisition workers.

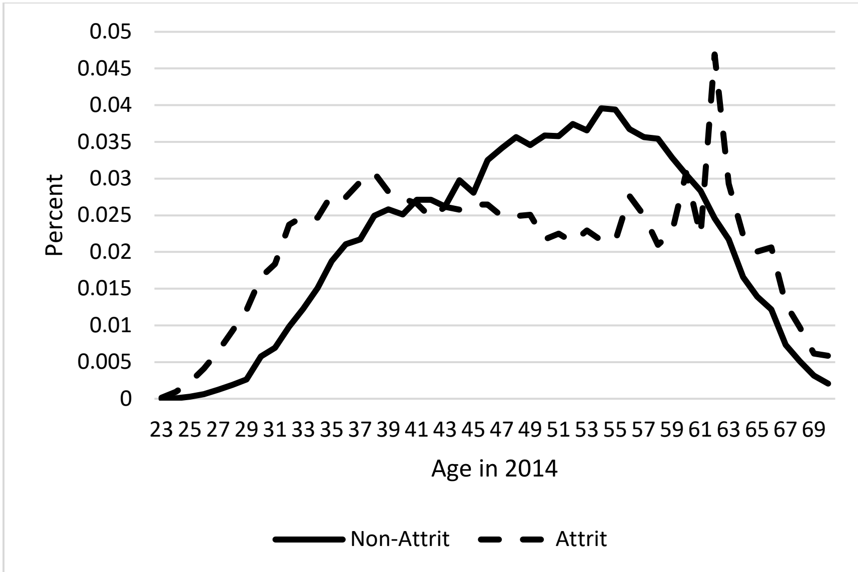


Figure 2a. Nonretirement Eligible Medical Employees' Age by Turnover Outcome

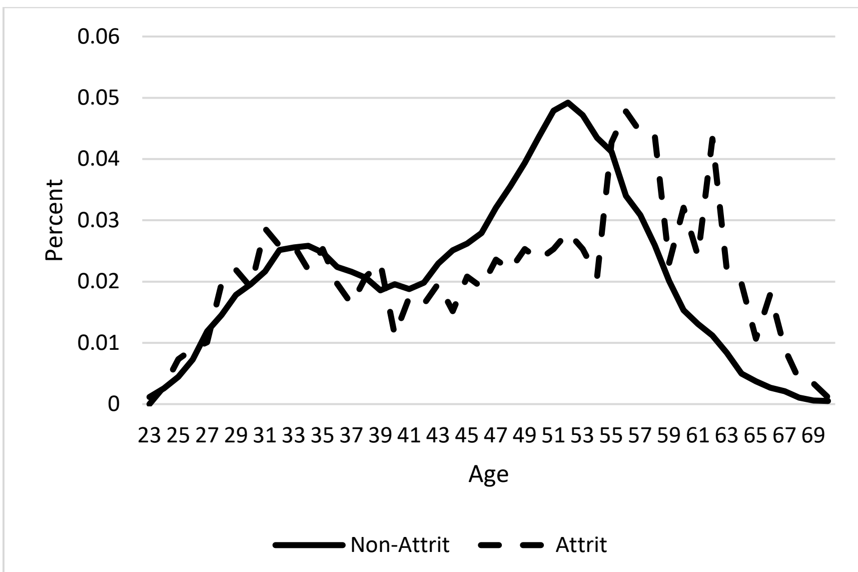


Figure 2b. Nonretirement Eligible Acquisition Employees' Age by Turnover Outcome

The retirement-eligible workforce depicted in Figures 3a and 3b exhibits a sharply different age distribution than the nonretirement eligible workers. The age distribution of the retirement-eligible medical workers show that relatively few are younger than 60 years of age. Additonally, the age distribution of those that separate and those that remain is relatively similar. There is a slightly higher percentage of employees older than 65 among those that do turn over than those that remain within the retirement-eligible medical workforce.

The age distribution of the acquisition workforce is much more lumpy than the medical workforce, which had a single peak in the mid-1960s. The prior reports issued from this research project by Brien (2019) and Buttrey et al. (2018) showed that the relatively large number of civilians entering employment with the DoD in their 50s is driven by retirements among uniformed personnel and their subsequent rehiring in civilian positions. The difference in the age profiles between the medical and acquisition employees illustrates that the medical workforce does not receive the same level of intake of former uniformed personnel as the DoD at large. In contrast, the acquisition workforce appears highly reliant on military rehires. Age does not appear to be associated with a turnover differential among acquisition workers between the ages of 58 and 64. Workers younger than 58 appear to have a lower rate of turnover, while those above 64 have a higher rate of turnover. This observation is consistent with a life-cycle model of employee turnover.

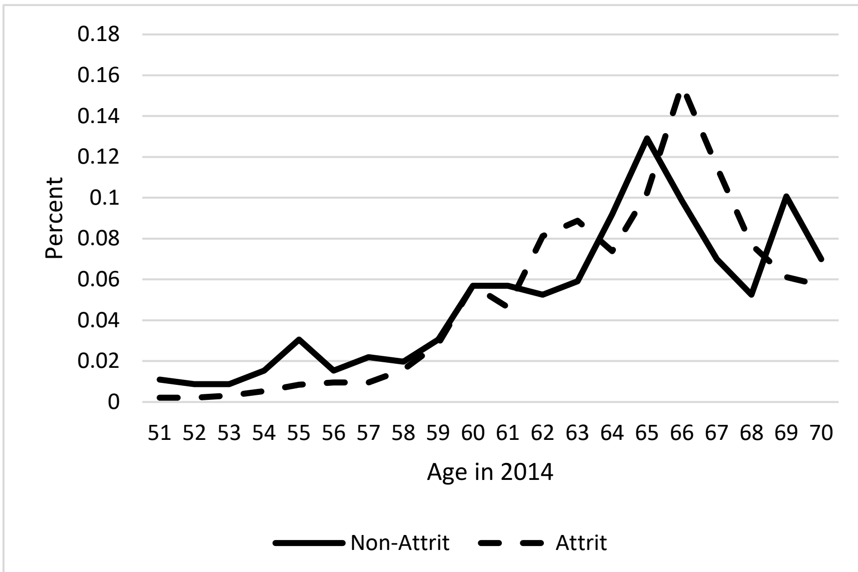


Figure 3a. Retirement Eligible Medical Employees' Age by Turnover Outcome

Figure 3b. Retirement Eligible Acquisition Employees' Age by Turnover Outcome

Another key finding of the preliminary analysis relates to the representativeness of racial groups across blue- and white-collar occupations within the medical workforce. Table

4 depicts the blue/white collar distribution for different racial groups identified within the medical workforce sample. The same statistics are visualized in Figure 6. Whites are distributed almost evenly across blue- and white-collar occupation types, with a slightly higher proportion in white-collar work. All other racial categories had a larger share employed in blue-collar work. Employees that identified as Black, Hawaiian or Pacific Islander, or Native American were highly concentrated in blue-collar occupations. Targeting recruitment to improve representation of racial minorities within while-collar occupations should be an immediate goal of the DoD.

Table 2. Racial Group Representation Across Occupation Type in the Medical Workforce

Racial Group	Total Count	Occupational Type	Counts by Occupation Type	Percent by Occupation Type
Asian	3,475	Blue Collar	1,830	52.66%
		White Collar	1,645	47.34%
Black	9,972	Blue Collar	6,886	69.05%
		White Collar	3,086	30.95%
Hawaiian or Pacific Islander	497	Blue Collar	354	71.23%
		White Collar	143	28.77%
Native American	428	Blue Collar	268	62.62%
		White Collar	160	37.38%
White	28,918	Blue Collar	14,027	48.51%
		White Collar	14,891	51.49%

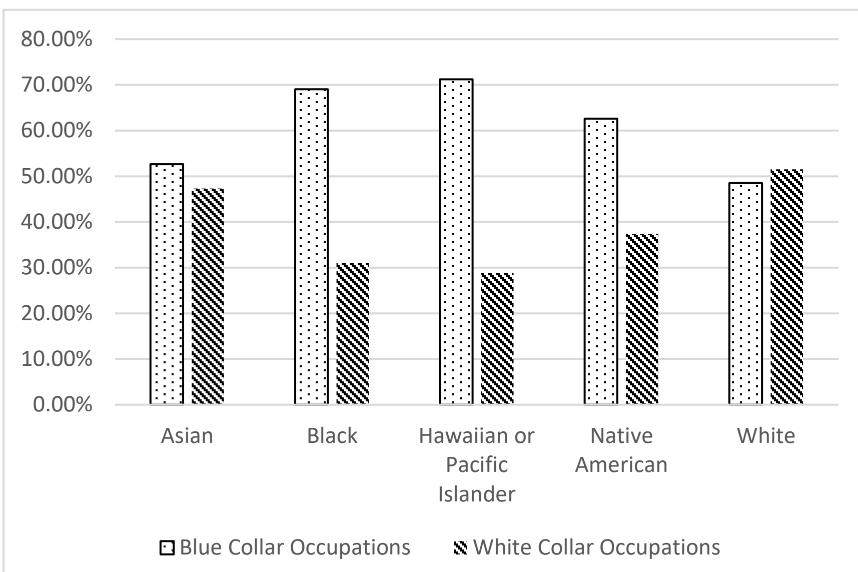


Figure 6. Racial Group Representation Across Occupation Type in the Medical Workforce

Logistic Regression Estimates

The odds ratio estimates obtained from the logistic regressions are presented in Tables 5 and 6. Table 5 shows the results from the medical workforce. After controlling for years of service, age, professional classification, and race, we find that gender has no impact on turnover among medical employees. This is consistent with findings by Lewis and Park (1989) in their study of the determinants of turnover in the federal workforce.

Table 3. Logistic Regression Results of Medical Workforce 2014 Attrition

Variables	Nonretirement Eligible	Retirement Eligible
Sex	0.997	1.127
	(0.024)	(0.157)
Years of Service	0.863***	1.168***
	(0.005)	(0.050)
Year of Service^2	1.004***	0.997***
	(0)	(0.001)
Age	0.710***	1.097
	(0.007)	(0.103)
Age^2	1.003***	0.999
	(0.000)	(0.001)
Graduate Education	0.894***	0.617**
	(0.026)	(0.100)
College Graduate	0.892***	0.716**
	(0.024)	(0.117)
Less Than High School	1.515	0.880
	(0.446)	(0.973)
Medical Officer	1.226***	0.550***
	(0.067)	(0.114)
Administrative Health Professional	0.873***	1.474*
	(0.035)	(0.348)
Nurse	1.033	1.161
	(0.028)	(0.182)
Black	0.929***	0.638***
	(0.025)	(0.097)
Asian	0.698***	1.258
	(0.030)	(0.262)
Native American	0.908	0.843
	(0.101)	(0.488)
Hawaii or Pacific Islander	0.740***	0.587
	(0.078)	(0.368)
Constant	4,902.237***	0.010*
	(1334.304)	(0.025)
N	45,151	1,475
Pseudo R-Squared	0.049	0.089

All estimates have been converted to odds ratios.
Robust standard errors are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Years of service is modeled as having a quadratic relationship with turnover. The base estimate of 0.866 and the estimate of 1.004 for the squared term collectively indicate that at relatively low years of service, a marginal increase in tenure reduces the likelihood of separating. This marginal effect is diminishing, however. The marginal effects of the quadric relationship were calculated and displayed in Figure 7. These lines represent the first differences of the quadratic relationship between increasing years of service and the percent change in the odds of separating from federal service. For the nonretirement eligible group, the left-hand side of the line is below zero, starting at approximately -2.5%. This is interpreted to indicate that an employee with 5 years of service is 2.5% less likely to separate than an employee with 4 years of service, on average. Moving to the right, however, the effect diminishes. Where the line crosses the axis at 20 years of service, there is essentially no statistically significant marginal effect of another year of service.

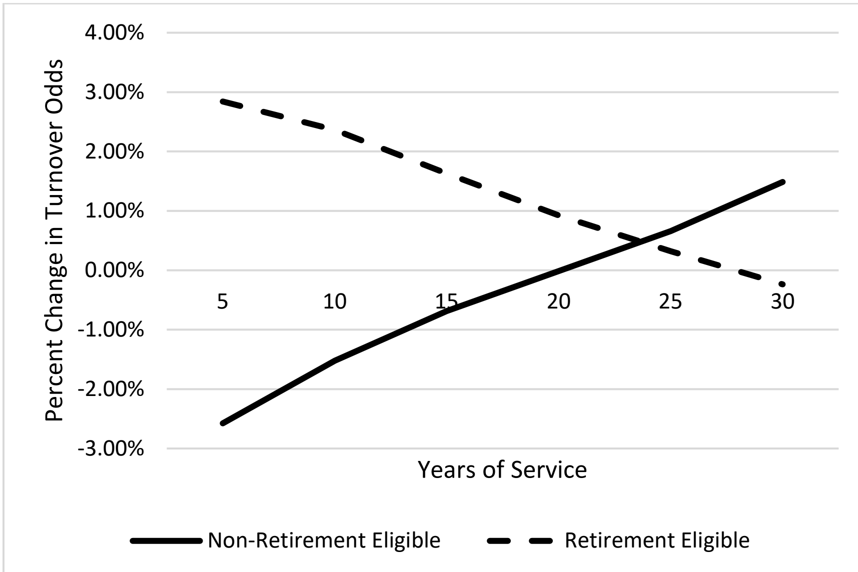


Figure 7. Marginal Effect of Years of Service on Turnover Odds for Medical Civilians

This marginal relationship is inverted for the retirement eligible community. For medical personnel eligible for retirement with relatively few years of service, adding 1 more year to their tenure increases their odds of separating by over 2%. As tenure increases, this effect declines. Once tenure reaches approximately 25 years of service, the marginal effect is not statistically distinguishable from zero, meaning that there is no additional impact on the odds of separating for further years of service. Once employees reach a late career status, length of service has no additional effect.

The age variable exhibits a similar quadratic relationship, as years of service does for the nonretirement eligible population. For younger employees, growing older is associated with a lower retirement likelihood, but at a diminishing rate. In contrast, age does not have a statistically significant relationship with turnover within the retirement-eligible medical workforce. This is consistent with the years of service estimate that once employees reach retirement eligibility there is an overall shift in turnover behavior, but the marginal effects of additional years of service or age are close to zero.

The occupational classifications show that different medical professions appear to experience significantly different turnover rates within the nonretirement eligible population. In comparison to blue-collar healthcare technicians, medical officers have higher odds (approximately 22.6%) of separating during preretirement eligibility. Nurses, however, appear to have a similar turnover rate to the blue-collar medical workers. Other healthcare professionals, which includes administrators and pharmacists, have a lower turnover rate in comparison to the blue-collar medical workers. Their attrition differential is estimated to be approximately 12.7 percentage points lower (1−0.873). This result invites further studies of the medical workforce to understand the environmental conditions and work schedules of medical employees and how that factors into their turnover decisions.

The results of the regressions using the acquisition data are depicted in Table 6. This table depicts estimates of a model of turnover among a subset of Navy acquisition employees. As discussed earlier, the personnel records are merged with average command-level employee satisfaction survey results. The intent of including these variables was to capture organizational differences in different measures of satisfaction. All 11 measures were tested for fitness in the model with a series of sensitivity tests. Ultimately, none of the satisfaction measures were statistically significant in the turnover model. Sensitivity testing revealed that there was insufficient variation in the satisfaction scores across organizational groups to reveal a relationship between employee satisfaction and turnover behavior. Table 3 displays one of the permutations of the sensitivity tests that included two of the most salient measures: job satisfaction and job stress.

Table 4. Logistic Regression Results of Acquisition Workforce 2014 Attrition

Variables	Nonretirement Eligible	Retirement Eligible
Sex	0.759	1.229
	(0.216)	(0.341)
Years of Service	1.038	1.329
	(0.059)	(0.244)
Years of Service^2	1.000	0.996
	(0.002)	(0.003)
Age	0.637***	3.14**
	(0.053)	(1.721)
Age^2	1.005***	0.992*
	(0.001)	(0.004)
Job Satisfaction Average	0.266	0.636
	(0.327)	(0.922)
Job Stress Average	0.382	1.652
	(0.401)	(2.066)
Grade	1.050	1.059
	(0.05)	(0.069)
College	0.895	1.184
	(0.24)	(0.396)
Graduate School	1.026	0.310
	(0.343)	(0.225)
Constant	870,379.8*	0**
	(6960931)	(0)
N	4,839	702
Pseudo R-Squared	0.056	0.065

All estimates have been converted to odds ratios.
Robust standard errors are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

The estimates of the two main components of the life-cycle model, age and years of service, fared differently in the acquisition data than in the medical data. The age variable was modeled with a quadratic term, similarly to the medical data, and produced the expected results for both the retirement eligible and ineligible groups. The years of service variable, however, was not statistically significant in the acquisition data. Additionally, none of the other control variables, such as grade and education, were statistically significant in the acquisition data.

The lack of statistical significance on any of the variables except for age raises the question of whether dropping so many records to be able to include the satisfaction survey fields reduced the power of the model to measure differences in turnover behavior. Since the employee satisfaction fields were not providing useful estimates, we decided to drop them from the model and add back in the acquisition employees from other branches of the Navy as well as other service branches of the DoD. The model depicted in Table 6 was rerun on the larger population of acquisition personnel. The results from this sensitivity test showed that years of service continued to be statistically insignificant, while grade and education indicators each became significant. Their estimates were similar in sign and magnitude to those estimates obtained from the medical data.

Discussion

This analysis of two very different segments of the DoD workforce has helped reveal differences in the composition and behavior of civilian personnel. In both groups, the overall pattern of attrition was consistent with the life-cycle model of turnover behavior. Employees that are not yet eligible for retirement had higher propensities to separate at early stages of their careers and relatively low attrition rates as they approach retirement eligibility. Retirement-eligible personnel conversely separate at an accelerating rate as they move past their earliest eligibility date. The preliminary statistical analysis illustrated how these two overall patterns hold for both medical and acquisition employees but also highlighted how the acquisition workforce benefits much more from mid-career appointments. This surge of older appointees flowing into the workforce from uniformed services also affects the outflow years later as these employees retire. Even if medical and acquisition employees respond relatively similarly to advancement in age and tenure, the difference in the composition of the workforce will change the recruitment, training, and succession planning strategies needed to manage the two types of employees.

Future use of these estimates for strategic workforce management may help defense budgeting. Conducting fiscal impact simulations of different turnover scenarios can prepare the DoD for future fiscal stress (Brien et al., 2020; Hansen et al., 2018). Defense budget cuts or other adverse fiscal conditions may result in hiring freezes or even force reductions. Understanding the pace of attrition behavior may help DoD officials understand the expected losses to the civilian workforce during such a period. Developing a forward-looking fiscal plan that takes into account potential actions by Congress fits within a broader goal of making strategic fiscal policies that are intentionally crafted in response to other levels of government (Brien, 2017; Brien & Sjoquist, 2014; Brien et al., 2017; Brien & Yan, 2020). Additionally, turnover simulations may provide guidance on where to target recruitment to avoid future gaps in workforce capability.

Another finding from this study is that the DoD can do much better in hiring racial minorities into technical and white-collar positions. This analysis of the medical workforce found that personnel identifying as Black, Native American, and Hawaiian or Pacific Islander were almost twice as likely to be employed in blue-collar work, while employees identifying as White or Asian were nearly evenly distributed between the two labor categories. Pursuing diversity and representation within occupational categories and not just within the workforce as a whole is essential to achieving a public workforce that aligns with the American public.

Furthermore, our attrition analysis found that employees identifying as Black, Asian, and Hawaiian or Pacific Islander had lower likelihoods of voluntarily separating when compared to white employees. Attaining diversity within these technical fields will also bring a more stable workforce.

The lack of empirical results with respect to the employee satisfaction survey data obtained from the acquisition workforce was a disappointing outcome, but this null result points towards methodological improvements that may be made in future studies. Using average satisfaction rates at the command level was found to be too aggregated at the organizational level. Satisfaction survey results obtained from organizational units lower down within the DoD and naval organizational structures will be necessary to effectively test the relationship between employee satisfaction and turnover behavior. This remains a pressing goal within the public human resource management literature. While there is a broad body of research examining the relationship between reported turnover intention and employee satisfaction, a growing body of research has questioned turnover intention as a valid proxy for actual behavior (Cohen et al., 2016; Dalton et al., 1999). Finding new sources of employee satisfaction data that can be integrated into the PDE environment for secure and PII-protected research into turnover behavior will be part of the next step of this project.

Understanding the patterns of turnover behavior within different segments of the DoD workforce is one part of strategic workforce management. This study, which used a parametric approach to modeling turnover, was able to contrast the break points in life-cycle behavior as employees approach and move into retirement eligibility. One of the primary benefits of estimating these relationships is that it will provide a baseline for turnover behavior prior to the ongoing reorganization of DoD medical operations. In coming years there will be a pressing need to perform evaluations of the impact of this transformation on the civilian medical workforce. The changes to the nature of the work performed by civilian medical employees may have significant impacts on retention and recruitment. This study lays an important groundwork for the pre–post comparisons that will be conducted.

References

Aiken, L. H., Clarke, S. P., Sloane, D. M., Sochalski, J., & Silber, J. H. (2002). Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *JAMA*, *288*(16), 1987–1993.

Baldor, L. C. (2018). Army lowers recruiting goal; more soldiers staying on. *The Associated Press*.

Buddin, R. (1984). *Analysis of early military attrition behavior.* RAND.

Fernandez, R. L., Gotz, G. A., & Bell, R. M. (1985). *The dynamic retention model.* RAND.

GAO. (1997). *Military attrition: DOD could save millions by better screening enlisted personnel* (GAO/NSIAD-97-39).

Gates, S. M., Keating, E. G., Jewell, A. D., Daugherty, L., Tysinger, B., Robbert, A. A., & Masi, R. (2008). *The defense acquisition workforce: An analysis of personnel trends relevant to policy, 1993–2006* [Report]. RAND. https://dair.nps.edu/handle/123456789/3719

Gates, S. M., Keating, E. G., Tysinger, B., Jewell, A. D., Daugherty, L., & Masi, R. (2009). *The Department of the Navy's civilian acquisition workforce: An analysis of recent trends*. RAND. https://apps.dtic.mil/sti/citations/ADA493748

Gates, S. M., Phillips, B., Powell, M. H., Roth, E., & Marks, J. S. (2018). *Analyses of the Department of Defense acquisition workforce: Update to methods and results through FY2017*. RAND.

Gates, S. M., Roth, E., Srinivasan, S., & Daugherty, L. (2013). *Analyses of the Department of Defense acquisition workforce: Update to methods and results through FY2011*. RAND. https://apps.dtic.mil/sti/citations/ADA585148

Gebicke, M. E. (1998). *Military attrition: DoD needs to better understand reasons for separation and improve recruiting systems*. GAO.

Gesesew, H. A., Tebeje, B., Alemseged, F., & Beyene, W. (2016). Health workforce acquisition, retention and turnover in southwest Ethiopian health institutions. *Ethiopian Journal of Health Sciences*, *26*(4), 331–340.

McCauley, J. W. (2020). *Characteristics of attrition among Navy acquisition employees and the significance of employee satisfaction responses*. Naval Postgraduate School.

McGrail, M. R., Wingrove, P. M., Petterson, S. M., & Bazemore, A. W. (2017). Mobility of U.S. rural primary care physicians during 2000–2014. *The Annals of Family Medicine*, *15*(4), 322–328.

Paone, S. M. (2020). *Characteristics of the current civil service military healthcare workforce and determinants of workforce turnover and attrition*. Naval Postgraduate School.

Rabkin, N. J. (2000). *Military personnel: First-term recruiting and attrition continue to require focused attention*. GAO.

Rambur, B., Val Palumbo, M., McIntosh, B., & Thomas, C. (2008). A cross-disciplinary statewide healthcare workforce analysis. *Journal of Allied Health*, *37*(2), 105–109.

PANEL 5. BLOCKCHAIN AND PREDICTIVE ANALYTICS IN THE SUPPLY CHAIN

Tuesday, May 11, 2021	
10:45 a.m. – 12:00 p.m.	**Chair: Rear Admiral Peter Stamatopoulos, USN,** Commander, Naval Supply Systems Command/Chief of Supply Corps ***Blockchain for a Resilient, Efficient, and Effective Supply Chain: Evidence from Cases*** Adrian Gheorghe, Old Dominion University Unal Tatar, University at Albany – SUNY Farinaz Sab Ali Pour, Old Dominion University Omer Keskin, Old Dominion University ***Blockchain Mergence for Distributed Ledgers Supporting Fleet Logistics and Maintenance*** Britta Hale, Naval Postgraduate School Terry Norbraten, Naval Postgraduate School Jonathan Culbert, LCDR USN, Naval Postgraduate School Don Brutzman, Naval Postgraduate School ***Blockchain Data Management Benefits by Increasing Confidence in Datasets Supporting Artificial Intelligence (AI) and Analytical Tools using Supply Chain Examples*** Tony Kendall, Naval Postgraduate School Bruce Nagy, NAWCWD Avantika Ghosh, University of California Berkeley Arijit Das, Naval Postgraduate School ***Product Supportability Through Lifecycle Modeling and Simulation*** Magnus Ansersson, Systecon Justin Woulfe, Systecon North America

Rear Admiral Peter Stamatopoulos, USN—is a native of San Diego. He earned a bachelor's degree in Business Administration from the University of San Diego and received his commission in 1988 through the Navy Reserve Officer Training Corps. He holds a master's degree from the Naval Post Graduate School and is a graduate of the Columbia University Senior Executive Program.

As a Navy supply corps officer and joint logistician, he has deployed across the globe in submarines, amphibious assault ships, aircraft carriers and operational staffs. In his most recent assignment, he served as director of logistics (J4), U.S. European Command.

His staff assignments include assistant chief of staff, Logistics and Ordnance, Commander, Naval Surface Forces Pacific; commanding officer, Naval Supply Systems Command (NAVSUP) Fleet Logistics Center, San Diego; chief of staff, NAVSUP Global Logistics Support; Logistics Services Division chief, Joint Chiefs of Staff J4, Washington, D.C.; head Program Objective

Memorandum Development Section, Office of the Chief of Naval Operations N80; executive assistant to the vice commander, NAVSUP; and supply officer, Fighter Wing U.S. Pacific Fleet and Fighter Squadron (VF) 124.

As a flag officer, he served as director, supply, ordnance and logistics operations, Office of the Chief of Naval Operations N41, Washington, D.C.; and as fleet supply officer on the staff of Commander, U.S. Fleet Forces Command.

Stamatopoulos has served on teams that have been recognized with numerous awards and is the recipient of the Adm. Stan Arthur Logistician of the Year; Adm. Ben Moreell Award for Logistics Competence; Armed Forces Communications and Electronic Association and the Naval Institute's Copernicus Award for C4I; and the 2006 Adm. Stan Arthur Logistics Team of the Year.

He is entitled to wear the Defense Superior Service Medal, Legion of Merit and Meritorious Service Medal along with various other personal, unit and service awards. He is also a qualified Naval Aviation Supply officer, Submarine Warfare Supply Corps officer, Joint Duty qualified officer and member of the Defense Acquisition Corps.

Blockchain for a Resilient, Efficient, and Effective Supply Chain: Evidence from Cases

Adrian V. Gheorghe—has been a systems engineering professor for 14 years and has more than 40 years of research, academic, and practical experience in the areas of systems engineering, decision-making in engineered systems, and analysis of extreme and rare events—including acquisition. Gheorghe has published five books on systems engineering, critical infrastructure, and derived supply chain/blockchain monography, along with several articles on blockchain governance and cybersecurity. [agheorgh@odu.edu]

Farinaz Sabz Ali Pour—is a PhD candidate in systems engineering and engineering management. She has experience applying blockchain technology in multiple fields, developing agent-based model supply chain systems, and designing resilient systems. She has published several articles on blockchain technology and supply chains. She has experience in resilient system design, blockchain-based model, and model-based systems engineering. [fsabz001@odu.edu]

Unal Tatar—is an Assistant Professor of Cybersecurity at the College of Emergency Preparedness, Homeland Security, and Cybersecurity, University at Albany. He has more than 16 years of cybersecurity experience in government, industry, and academia. He is the former coordinator of the National Computer Emergency Response Team of Turkey. Tatar's research is funded by the NSF, NSA, DoD, NATO, and Society of Actuaries. He holds a BS degree in computer engineering, an MS degree in cryptography, and a PhD in engineering management and systems engineering. His main topics of interest are information/cybersecurity risk management, cyber resiliency, cyber insurance, and blockchain. [utatar@albany.edu]

Omer Faruk Keskin—is a PhD candidate in the Engineering Management and Systems Engineering Department at Old Dominion University. He holds a master's degree in engineering management and a bachelor's degree in systems engineering. He is a graduate research assistant and has worked on several projects, including the grant proposal writing phase. His main fields of research include enterprise cyber risk management and risk quantification, modeling, and simulation. [okesk001@odu.edu]

Abstract

In the modern acquisition, it is unrealistic to consider single entities as producing and delivering a product independently. Acquisitions usually take place through supply networks. Resiliency, efficiency, and effectiveness of supply networks directly contribute to the acquisition system's resiliency, efficiency, and effectiveness. All the involved firms form a part of a supply network essential to producing the product or service. The decision-makers have to look for new methodologies for supply chain management. Blockchain technology introduces new methods of decentralization and delegation of services, which can transform supply chains and result in a more resilient, efficient, and effective supply chain.

This research aims to review and analyze the selected current blockchain technology adoptions to enhance the resiliency of supply network management by facilitating collaboration and communication among suppliers and support the decision-making process. In the first part of this study, we discuss the limitations and challenges of the supply chain system that can be addressed by integrating blockchain technology. In the final part, we analyze multiple blockchain-based supply chain use cases to identify how the main features of blockchain are suited best for supply network management.

Keywords: supply chain, blockchain, supply network, resilience, acquisition

Introduction

The Department of Defense (DOD) spending on goods and services has grown significantly since the Fiscal Year (FY) 2000 to well over $250 billion annually (Walker, 2006). The process of DoD supply chain includes all government and private-sector organizations, processes, and systems that individually or collectively play a role in planning for, acquiring, maintaining, or delivering material resources for military or other operations conducted in support of U.S. national defense interests (Reay, 2000). The supply chain complexities, which create significant challenges throughout the networks, arise from various factors, such as changes in customer expectations, multiple market channels, and international markets. Access to the latest technologies in various fields can be a great support in supply chain management. Innovations, including digitalization and industry 4.0, have developed new paradigms, principles, and models in supply network management. Through the literature review by (Ivanov et al., 2019), the digital technologies include big data analytics, advanced manufacturing technologies with sensors, decentralized agent-driven control, advanced robotics, augmented reality, advanced tracking and tracing technologies, and additive manufacturing. The development of the digital supply chain and smart operations are facilitated using Internet of Things (IoT), cyber-physical systems, and smart products. Blockchain technology is attracting a rising level of interest reflected by Google trends that returned 21.6 million Google queries for blockchain released on January 10, 2017 (Fosso Wamba et al., 2018). However, there is a need for novel models to support supply chain management in the future (Ivanov et al., 2019).

Resiliency, a vital feature for the viability of supply chains, is defined to be a measure of the persistence of systems and of their ability to absorb change and disturbance and still maintain the same relationships between populations or state variables (Holling, 1973). The common resiliency metrics for the supply network system are availability, connectivity, and accessibility, which can be improved by the decentralized, distributed, and fault-tolerant features of blockchain technology. The highly resilient architecture and distributed nature of blockchain technology make it an interesting platform to defend against attacks and preserve the integrity of the identity network (Shrier et al., 2016).

This study proposes applying blockchain technology, which can address many of the mentioned technological challenges and enhance supply network management. This study includes a systematic literature review of the current most critical challenges of supply chain and supply networks in the Literature Review section. The Block Chain in Supply Chain section covers the solutions that can be provided by utilizing blockchain technology for the supply network challenges identified from the academic and grey literature. Several use cases that applied blockchain technology in supply chain management are analyzed in the Case Studies section, followed by a conclusion in the last section.

Literature Review

In this section, the limitations and challenges of the supply chain system that can be addressed by integrating blockchain technology are identified through the literature, and the blockchain-based solutions are mapped to those challenges.

A systematic literature review is an efficient tool for summarizing the results of existing studies and assessing consistency among previous studies. It provides a systematized approach to identify current challenges, new methodologies, and research avenues (Queiroz et al., 2019). This study provides a systematic literature review on the current supply chain challenges and limitations in a supply network. For this purpose, the Web of Science database is used. Table 1 shows the details of the papers that were extracted by each keyword. In the next step, papers that were relevant to the topic were selected for full paper reading in addition to relevant papers that were selected from Google Scholar, ISI journals, and conference papers. The third step included assessing the quality of the selected paper, and lastly, final papers were selected as the references for this report.

The protocols followed for this systematic literature review include 1. determining Web of Science and Google Scholar as the main research databases, and 2. only considering the English language journal publications for this review. The data extractions details are defined and can be seen in Table 1.

Table 1. Literature Review Structure

Database	**Total # of results**	**Number Selected of papers**
Web of Science	3217 + 338	29
Google Scholar	132	127
Total	3687	156

The three keywords that were used for the Web of Science database are *supply network challenge*, *blockchain technology*, and the combination of *supply chain challenges* AND *blockchain*, as shown in Figure 1.

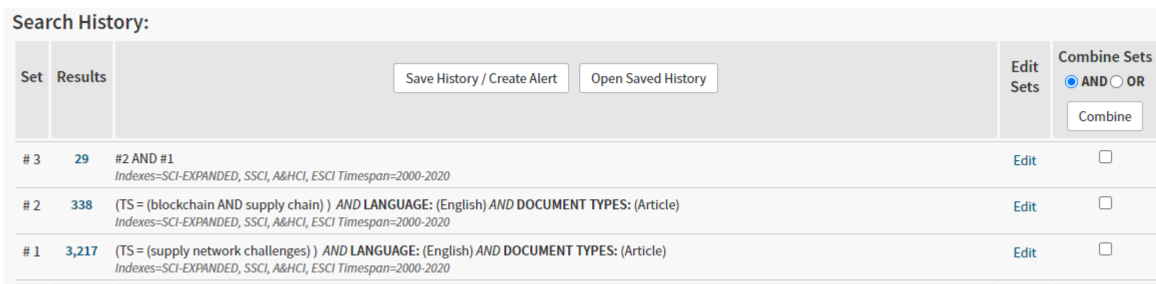


Figure 1. Keyword Search History

The articles in the supply chain field include a broad area of applications and theories. Hence, this study has narrowed down the content search to three main categories of technologies, theories, and applications (see Table 2).

Table 2. Search Content Categories

Content Categories	
Main technologies	Blockchain technology, smart contracts, modeling, & simulation
Main theories	Conceptual, reviews, frameworks, and case studies
Blockchain application area	Supply chain network

The study retained 29 articles out of the combination of 3,217 papers with the topic of supply network challenges, and 338 papers with both topics of blockchain and supply chain up to summer in the two decades of 2000–2020 (see Figure 1). The report categorizes the blockchain applicability in supply network challenges into four areas: communication, transparency, data and information, and performance.

The region-based literature analysis revealed that *supply network challenges* is a popular topic in different regions across the world. Figure 2 shows the distribution of papers in different countries within the past two decades; most articles were published in the United States, followed by China and England.

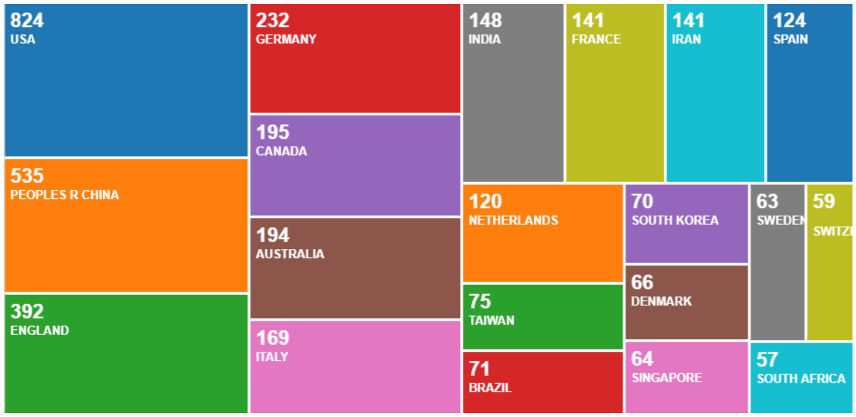


Figure 2. Supply Network Challenges Paper Region Distribution

On the other hand, the systematic review shows that the interest in research in the field of supply networks has gained more attention recently, as depicted in Figure 3. In 2019, there were around 500 articles published related to supply network challenges.

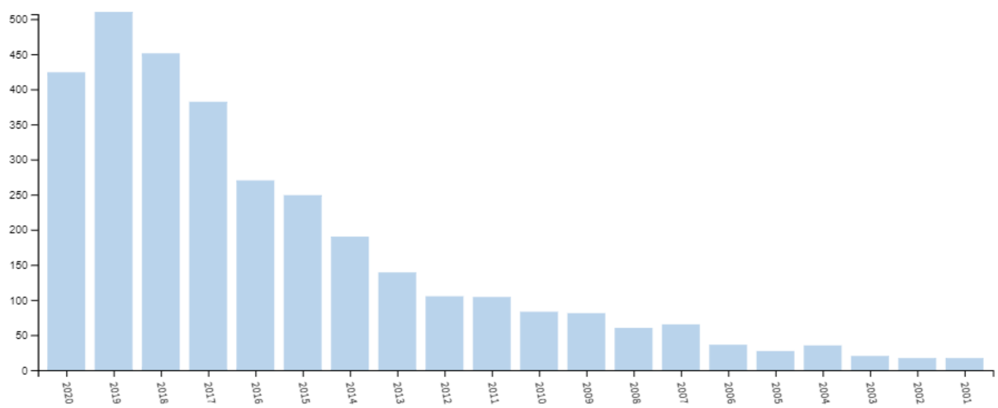


Figure 3. Supply Network Challenges Publication (2001–2020)

Supply Chain Network Management

A supply chain is a network of multiple businesses and relationships (Lambert & Cooper, 2000) considered as a complex system due to having multiple levels and numerous facilities at each level (Beamon, 1999). Supply chain management is a system capable of rational planning, management, and control of the supply chain and the logistics. It enables the stakeholders to accurately monitor and provide real-time responses to the supply chain issues (Yoo & Won, 2018). The evolution of supply chain management was studied by Fawcett and Magnan (2002) and is shown in Table 3. It is rooted back to the 1950s and 1960s, where the focus was on minimizing the production cost. In the 1970s, material requirement planning was developed. In the 1980s, the global competition developed management programs such as Just-In-Time that forced firms to offer low-cost, high-quality, and reliable products with greater flexibility in the design. In the 1990s, the outsourcing of non-core operations was introduced as a solution to reduce or transfer some of the risks, which developed more interactions and integrations among different involved parties in a supply network. The changes shift the focus to more collaborations and performance management of a supply network for the success of a firm. Lastly, with the advent of new technologies, the concentration is more on creativity, improving collaboration, and communication among the stakeholders.

Table 3. Supply Chain Management Evolution Era (Fawcett & Magnan, 2002)

Era	Description
Creation Era	In the 1980s, the supply chain term was coined by an American industry consultant.
Integration Era	Through the 1960s–1990s, the electronic data interchange system was developed and led to enterprise resource planning systems.
Globalization Era	The objective of organizations changed to competitive advantage, creating value-added, and reducing costs through global sourcing
Specialization Era – Phase One	In the 1990s, the focus became more on core competencies and adopted a specialization model. Firms outsourced non-core operations to the other companies and abandoned vertical integration.
Specialization Era – Phase Two	New aspects of supply planning, collaboration, execution, and performance management were adapted to the supply chain.
Supply Chain Management	The use of the World Wide Web (Web 2.0) led to more creativity, information sharing, and collaboration among users.

Harland (1996) described *supply network* as a dynamic, interconnected, complex, interdependent network of suppliers, manufacturing facilities, and linking multiple organizations (Bales et al., 2004). The structure of the supply network includes the member companies and the links between them. Lambert and Cooper (2000) suggest that three primary aspects of the company network structure include:

1. Members of the supply chain
2. Structural dimensions of the network
3. Different types of process links across the supply chain

Based on the characteristics of a supply network and the challenges that might arise managing such supply chains, this study applies blockchain technology to provide a decision-making framework that can capture the emerging phenomena of complex supply network challenges.

- **Blockchain Technology**

Blockchain technology created by Nakamoto (2008) is an emerging information technology that provides new opportunities for decentralized market design with transparent and user-friendly applications that allow consumers to participate in the decision-making (Mengelkamp et al., 2018). Based on the World Economic Forum report in 2015, blockchain has been considered as one of the megatrends that are going to change the world in the next decade (Kshetri, 2018).

Blockchain consists of nodes within a communication network that contain a common communication protocol. Each node stores a copy of the blockchain on the network, and a consensus function verifies transactions to preserve the immutability of the chain (Wang et al., 2019). Each block is identified through its cryptographic hash, and each block is referred to as the hash of the previous block, which creates a link between blocks to form a blockchain. The transactions of each block are hashed in a Merkle tree. The root hash and the hash of the previous block are recorded in the block header. Blockchain provides interaction between users by a pair of public and private keys (Casado-Vara et al., 2018). The hashing process transforms assets into a digitally encoded token that can be registered, tracked, and traded with a private key on the blockchain (Ivanov et al., 2019).

▪ **Supply Chain Challenges and Adaptation With Blockchain Technology**

In a supply network, flows of data have different forms and satisfy different needs. This leads to a complex course of controlling, ensuring immutability, and security transparency. Therefore, an efficient mechanism is required that enhances immutability and ensures confidentiality of transactions in the supply network.

This study aims to explore the feasibility of using blockchain technology to address current limitations, efficiently manage the process, and enhance the resiliency of supply chain systems. We grouped the current supply network systems limitations (Apte & Petrovsky, 2016; Fawcett & Magnan, 2002; Kleindorfer & Saad, 2005; Kshetri, 2018; Wilding et al., 2012) to be addressed under four main categories per the blockchain features:

- *Network Communication and Information Flow:* (1) Analog gaps between customer and supplier, (2) lack of information sharing among all involved stakeholders, (3) lack of an integrated global view concerning increasingly dynamic supply chains
- *Transparency:* (1) Lack of traceability of failures in the flow of the process, (2) limited visibility concerning how and where products are sourced, made, and stored
- *Data and Information Management:* (1) Disparate record keeping, (2) lack of accurate and reliable data for analytics, (3) excessive redundancy and crosschecking, (4) long and costly audit processes
- *Performance measurement:* (1) high cost of managing the network, (2) decreased speed due to network arrangement and communication, (3) lack of flexibility due to various policies and structure in the network

▪ ***Categories of Supply Network Challenges and Blockchain-Related Solutions***

The supply network's complexity consists of multiple elements, including raw material suppliers, distributors, manufacturers, retailers, and end consumers (Francisco & Swanson, 2018). The major complexity of supply chain processes has led to challenges in the supply network and conflicts that are raised from local objective versus network strategies (Terzi & Cavalieri, 2004). Defining boundaries among the multiple interconnections of a network is one of the methodological challenges in studying the supply network (Park et al., 2013). One of the other challenges is the coordination of complex influx and outflow of materials (Park et al., 2013).

This study identifies the main challenges and limitations of supply networks through the literature and aims to provide related blockchain empowered solutions to improve the current challenges. Table 4 describes the taxonomy of challenges in supply chain networks and the related blockchain features identified as a solution in the literature. The blockchain features that can improve those challenges are linked to each category of supply network challenges as the recommended solution to tackle those challenges.

Table 4. Taxonomy of Challenges in the Supply Networks and Adaptation With Blockchain Technology

Category	Supply Chain Network Challenges	Blockchain Empowered Solution
Network Communication and Information Flow	Miscommunication between suppliers and retailer (Ludema, 2002)	Value chain visibility for all parties (Kshetri, 2018)
	Inappropriate use of power and opportunistic behavior (Dani et al., 2003)	Seamless networks, visibility, and symmetric information to all actors (Wang et al., 2019)
	Lack of effective collaboration, communication, and partnership (Fawcett & Magnan, 2002; Saberi et al., 2019; Terzi & Cavalieri, 2004; Wang et al., 2019) Conflicts in local versus global interests Strong reluctance of sharing common information (Fawcett & Magnan, 2002; Terzi & Cavalieri, 2004)	
	Risks and disruptions from natural disasters or conflicts (Ivanov et al., 2019; Park et al., 2013) Defining boundaries among the interconnections of a network (Park et al., 2013) Coordination of complex influx and outflow of materials (Park et al., 2013)	
	Contradictory operational objectives and priorities; Different culture and geographical disperse of the partners (Saberi et al., 2019; Wang et al., 2019)	
	Complicated distribution structure and lack of information about the margins of products for the customers (Yoo & Won, 2018)	

Category	Supply Chain Network Challenges	Blockchain Empowered Solution
Transparency	Lack of common purpose; power imbalances; culture and procedures; autonomy (Fawcett & Magnan, 2002)	Immutable Ledger (Abdirad & Krishnan, 2020; Chen et al., 2017; Francisco & Swanson, 2018; Kshetri,

		2017, 2018; Queiroz et al., 2019)
	Lack of accountability (Fawcett & Magnan, 2002; Kshetri, 2017)	Decentralized platform (Kshetri, 2017, 2018) Real-time basis data and tracking (Kshetri, 2018; Wang et al., 2019)
	Traceability disruptions (Queiroz et al., 2019)	Smart contracts improve responsiveness; reduce lead time; decrease transaction and monitoring cost; enhance visibility, trust, security, and transparency (Queiroz et al., 2019); smart contracts ensure the participation of the consumers (Kshetri, 2017)
	Information privacy of customers; lack of audit trails (Kshetri, 2017; Tatar et al., n.d.)	Encrypted data with hash functions; no single point of failure; secure messaging between devices; audit trail to ensure accountability (Kshetri, 2017, 2018)
	Lack of transparency (Francisco & Swanson, 2018; Yoo & Won, 2018)	Data is controlled with private & public keys (Kshetri, 2017) Owner choose the information that is released (Kshetri, 2017)
	Lack of trust in information legitimacy (Chen et al., 2017; Wang et al., 2019); Fraud, corruption, tampering, and falsifying the information as trust problems (Tian, 2017)	Transparency (Francisco & Swanson, 2018; Saberi et al., 2019); Traceability (Chen et al., 2017; Francisco & Swanson, 2018; Ivanov, Dolgui, & Sokolov, 2019; Mengelkamp et al., 2018; Saberi et al., 2019)
	Unstable distribution prices (Yoo & Won, 2018)	Authenticity and legitimacy (Wang et al., 2019) Accountability (Fosso Wamba et al., 2018)
		Openness, neutrality, reliability, and security for all members of the supply chain (Tian, 2017)

		Trust (Folkinshteyn & Lennon, 2016; Kiviat, 2015)
		Transactions are viewable by the whole network that protects against double-spending (Yoo & Won, 2018)

Category	**Supply Chain Network Challenges**	**Blockchain Empowered Solution**
Data and Information Management	Inefficient transactions, fraud, pilferage, centralized and stand-alone information management system (Saberi et al., 2019)	Disintermediation (Saberi et al., 2019) Transparency (Francisco & Swanson, 2018; Saberi et al., 2019; Tian, 2017) Traceability (Chen et al., 2017; Francisco & Swanson, 2018; Ivanov, Dolgui, & Sokolov, 2019; Saberi et al., 2019)
	Lack of information about the origin of the product (Casado-Vara et al., 2018)	Authentication & privacy (Abdirad & Krishnan, 2020; Kshetri, 2017)

Category	**Supply Chain Network Challenges**	**Blockchain Empowered Solution**
Performance	Cost	Paper records elimination Reduce regulatory compliance costs Tracking processes with IoT Identify the defective products easily from the source Track the quality and counterfeit of the ingredients from the partners; Provide meaningful data to assess the quality No costly regulation and overhead (Yoo & Won, 2018) Cost reduction due to disintermediaries (Folkinshteyn & Lennon,

		2016; Tapscott & Tapscott, 2017)
	Speed	Digitalizing physical process and reduce interactions and communications time (Kiviat, 2015; Kshetri, 2018)
	Delays and defaults in the delivery of goods (Casado-Vara et al., 2018)	
	Dependency	Partners should be more responsible and accountable for their actions Digital certification Audit trail (Kshetri, 2018)
	Risk Reduction	Address the holistic source of risk by verifying provenance Network only permits mutually accepted parties to engage in transactions (Kshetri, 2018)
	Sustainability	Validation of participants' identity (Kshetri, 2018)
	Flexibility	Address consumer's concern about the products Higher level of impact with IoT integration (Kshetri, 2018)

Blockchain in Supply Chain

Blockchain has the potential for supply chain improvements. Based on the features of blockchain technology, Kshetri (2018) claims that blockchain has the potential to help achieve supply chain critical objectives. There are some pilot practices of blockchain technology in a supply chain with no evidence of large-scale adoption (Wang et al., 2019). The supply chain has been expected to be one of the most promising non-finance application domains of blockchain (Kshetri, 2018). There is limited empirical evidence of the advantages of blockchain on the existing supply chain. Supply chain as a complex workflow has been identified as one of the main potential areas of blockchain application to deliver a real rate of interest (Kshetri, 2017). Wang et al. (2019) categorize the current literature of blockchain in the supply chain into four types: descriptive, conceptual, predictive, and prescriptive. The sources of insecurity can be tracked within a supply network. Blockchain can facilitate managing crisis situations regarding security vulnerabilities. Blockchain can be applied to register time, location, price, involved parties, and the related information while the ownership of an item is changing (Kshetri, 2017). Trust enhancement, accurate information sharing, and verifiability are crucial because of current challenges such as inefficient transactions, fraud, pilferage, and poor performance in the supply network (Saberi et al., 2019). The technological developments and applications of blockchain technology can improve supply chain transparency, security, durability, and process integrity, which results in more organizational, technological, and economic feasibility (Saberi et al., 2019). As the supply networks contain large numbers of stakeholders, process tracking would be more difficult.

Smart contracts can automate the processes. The agreed contracts can be delivered to the specified parties for digital execution. Programs can be updated based on agreed verifications, and copyright documents can be released to the relevant parties. Smart contract adoption can fundamentally change the supply chain structures and governance (Wang et al., 2019). The governance and process rules of smart contract in a blockchain-based supply chain provides actor certification and approval and the processes that are permitted to be accessed for execution (Saberi et al., 2019; Sabz Ali Pour et al., 2018).

A blockchain-based supply chain management system can improve the system in several ways. First, it provides the ability to record, provide, and share prices. Second, companies can deliver honest information to consumers. Third, the purchase intentions of buyer information can be obtained. Fourth, marketing operations for exploiting customers' propensity with no personal information can be included. Fifth, the trading contracts process can be automated using smart contracts (Yoo & Won, 2018). The supply chain management processes can be improved by blockchain monitoring, which provides efficient customer service management and convenient demand management (Yoo & Won, 2018).

The literature on blockchain technology in the supply chain is still in its early stages. The literature mostly describes blockchain as a distributed ledger technology because it is data-management technology that consists of a chain of decentralized computer terminals and a network software protocol on the base of a peer-to-peer node's network (Fosso Wamba et al., 2018). Several systematic literature reviews studied the blockchain applications in supply chain management (Denyer & Transfield, 2009; Transfield et al., 2003) and offered a more in-depth understanding of the technology (Queiroz et al., 2019). A list of literature review articles in the field of supply chain and blockchain is presented in Table 5.

Table 5. Supply Chain and Blockchain Literature Review Studies

Author(s)	Objective(s)	Outcome(s)
Fosso Wamba et al. (2018)	A systematic review of supply chain cases over the knowledge gap in bitcoin, blockchain, and financial technology	Illustrates technology evolvement and adaptation of organizations to apply the advantages of blockchain technology.
Wang et al. (2019)	Systematic academic and practitioner literature review on understanding blockchain technology for future supply chain	Retained 24 articles out of 227 papers in 2017 and categorizes the blockchain applicability in the supply chain into four areas: visibility and traceability, supply chain digitalization and disintermediation, improved data security, and smart contracts. The study identified the main drivers of blockchain development within supply chains as trust, product safety, authenticity and legitimacy, public safety and anti-corruption, and supply chain disconnections and complexities.
Queiroz et al. (2019)	Systematic literature review on blockchain supply chain management integration	Twenty-seven papers were identified in the past decade that address it, with the main theoretical approach of conceptual and framework. It shows essential implications for managers, practitioners, consultants, and decision-makers in the field (tracking enhancement, real-time visibility, decentralized operation, smart contracts, improving securities, reduced transaction costs). Also, the study identified a vital gap in the literature relate to blockchain–supply chain management integration in emerging economies and developed empirical studies.
Saberi et al. (2019)	Literature review on the application of blockchain and smart contract to overcome the potential barriers in supply chain	Introduced four categories of barriers for blockchain technology adaptation, including inter-organizational (new rules, responsibilities, policies, and expertise), intra-organizational (relationships among parties and their privacy policies related to information and data usage), technical (technology access limitation to get real-time information, data manipulation, and information immutability) and external barriers (pressures, lack of proper governmental and industry policy).
Ivanov et al. (2019)	Conceptual model	Adoption and application of blockchain technologies applied to supply chain traceability and introduced the behavioral theory as the lens for this framework on theoretical guidance of Unified Theory of Acceptance and Use of Technology (UTAUT).
Kim & Laskowski (2018)	Platform development	Smart contracts on the Ethereum blockchain platform that execute a provenance trace and enforce traceability limitations.
Chen et al. (2017)	Conceptual model	Adoption of blockchain technology to improve supply chain quality management and develop a blockchain-based supply chain quality management framework.
Gausdal et al. (2018)	Theoretical framework	Identify the key elements and barriers to digital innovation. The main identified barriers include high cost of implementation, technology-oriented culture, lack of investment initiatives, low level of blockchain diffusion through the supply chain, and risk aversion.
Yoo & Won (2018)	Platform Development	Applied blockchain and smart contracts for price-tracking that improve the transparency of the product distribution structure.

Case Studies

In this section, three blockchain-based supply chain use cases are analyzed to explore how the main features of blockchain are suited best for supply network management.

Walmart

Partners: Walmart, the largest grocery retailer in the United States, has partnered with IBM, JD.com, and Tsinghua University to conduct studies on the adoption of blockchain technology in the food supply chain (JD, 2017).

Purpose of Blockchain: Walmart's ultimate goal of using blockchain in their supply chain is to enhance transparency (IBM Blockchain, 2017). The partners established the Blockchain Food Safety Alliance to design blockchain solutions for food tracking, traceability, and safety (JD, 2017).

Benefits of Blockchain: Traceability includes tracking and tracing the products throughout both directions of product flow within the supply chain. Products can be tracked from their origins to the stores, and they can be traced back from the shelves to the farms. With blockchain, identifying the sources of foodborne illnesses and tracing back to the farms/origins can be reduced from days to seconds (JD, 2017). Such abilities not only provide benefits for public health but also reduce the economic impact for Walmart since only the contaminated products would be discarded, rather than all similar products (Tan et al., 2018).

Other benefits of blockchain for Walmart are improved security and trust. Customers can learn more about the products they consume, which results in higher confidence. In the food supply chain, most of the data is still processed on paper or in systems that cannot talk to each other (IBM Blockchain, 2017). Blockchain adoption provides immutability that avoids any alteration and transparency that provides everyone to access the ledger. This can effectively reduce the chance of food fraud and human errors (Tan et al., 2018).

Another benefit of the adoption of blockchain technology is to reduce waste by providing faster routes for perishable items, eventually leading to more sustainable operations. Still, a large portion of food is spoiled before arriving at the stores. Decreased delivery times can reduce waste by applying blockchain technology (Tan et al., 2018).

Method: Walmart conducted two pilot projects to test the effectiveness of the developed blockchain application. A pilot study was conducted in China on the pork supply chain, and the other pilot study was conducted in the United States on the mango supply chain (Tan et al., 2018). Both studies were successful in improving food safety, increasing recall speed, building higher trust for customers, and decreasing costs (Tan et al., 2018).

Challenges: The adoption pilot studies showed that adoption of blockchain technology is achievable; however, the mass adoption would introduce more challenges. Walmart has hundreds of thousands of suppliers worldwide, more than 4,000 retail stores in the United States, and more than 6,000 stores in 23 other countries (Walmart, 2021). Within its huge supply chain, most small- and medium-size enterprises do not have the technological infrastructure or training to adopt blockchain. Moreover, broad adoption of blockchain requires a high cost of implementation.

Maersk

Partners: Maersk, the world's largest integrated shipping company, has partnered with IBM to develop the TradeLens platform utilizing blockchain technology in the global supply chain in 2018 (Gausdal et al., 2018). TradeLens brings a diverse set of stakeholders together in a platform, including shippers/cargo owners, freight forwarders, intermodal operators, ocean carriers, ports, terminal operators, customs authorities, and financial service providers (TradeLens, 2018). Similar supply chain companies, including Hapag-Lloyd, Ocean Network Express (ONE), CMA CGM, and Mediterranean Shipping Company (MSC), joined TradeLens, extending the scope of the consortium to include more than half of the global ocean container carrier industry (Maersk, 2019).

Purpose of Blockchain: Maersk's ultimate goal of using blockchain technology is to improve collaboration and trust across the partners of the global supply chain (Maersk, 2019). It is expected to increase the efficiency of the supply chain that mostly depends on manual processes in current technology (Kralingen, 2018).

Benefits of Blockchain: Senior Vice President of IBM states that "blockchain for the enterprise is solving previously unsolvable problems" (Maersk, 2019). TradeLens provided supply chain visibility, ease of documentation, and the ability to add new features on top of the platform (Kralingen, 2018). The built-in security feature of blockchain that makes it immutable prevents any alteration in the history of the transactions or smart contracts. This enables trust among the partners and keeps the records so that partners can keep track of the documentation digitally rather than undertaking all the processes manually on paper.

Method: Maersk and IBM utilized the open-source Hyperledger technology program by the Linux Foundation, contributed by a couple of hundred developer enterprises to develop TradeLens (Kralingen, 2018; TradeLens, 2018). The platform is governed transparently by the partners, enabling the trust that brings the partners together. The platform supports innovation with its structure that eases adding new features and applications to serve the diverse needs of different types of stakeholders of the supply chain (TradeLens, 2018). All communication among the blockchain nodes is end-to-end encrypted, increasing its security (Kralingen, 2018). Only the partners of the permissioned blockchain platform can access the data. The partners participate in consensus for transaction validation and data hosting.

Initial Phases: A pilot project involved shipment with Saudi Customs demonstrated immutability, auditability, and transparency features of the platform, in addition to reducing costs and processing time (Madsen, 2019).

The initial phase of the developed platform implemented the processing of Bill of Lading among the supply chain partners. It resulted in a significant decrease in the administrative costs—up to 15% of the cargo value based on the initial tests. Considering the industry covers almost 60% of the world's GDP, the efficiency increase is considered astounding (Gausdal et al., 2018).

As of September 2019, the utilization of TradeLens included more than 100 organizations, five out of the world's six largest shipping companies, 55 ports, and almost a dozen customs authorities, with more than 10 million weekly shipping events (Madsen, 2019).

DHL and Accenture for Pharmaceutical Industry

Partners: DHL, an international courier, package delivery, and express mail service, and Accenture, a multinational company selling consulting and processing services, cooperated to develop a blockchain application for the healthcare and life sciences industry (Accenture, 2018). The blockchain application is suitable to establish communication among various stakeholders, including manufacturers, storage facilities, distributors, hospitals, pharmacies, and healthcare providers (Accenture, 2018).

Purpose of Blockchain: DHL and Accenture's ultimate goal of using blockchain technology is to fight against counterfeit medications. They aim to reach this goal by implementing serialization, tracking, and tracing features in a blockchain platform (Heutger et al., 2018).

Benefits of Blockchain: The pharmaceutical industry is under the threat of counterfeit drugs. According to Interpol, more than one million deaths are related to counterfeit drugs every year (Aces & Kleeberger, 2018). The developed blockchain platform can provide the ability to verify the point of origin of the drug and whether it is genuine or counterfeit, helping to save lives.

It helps pharmaceutical supply chain companies to maintain their reputation by giving them the ability to track every step of drugs in all parts of their life cycle. It enables better management of drug inventory at any part of the supply chain by determining faster delivery routes, handling and storage conditions, and tracking expiration dates (Heutger et al., 2018).

Another benefit of the blockchain platform is to keep drug quality high. When a drug is detected as non-compliant, it can immediately be traced back to the origin. All drugs manufactured under the same conditions can seamlessly be located and recalled. This process can take weeks with a paper-based supply chain. However, it can be completed in seconds using blockchain technology.

Method: The partners developed a working prototype of the blockchain platform. After working on the proof of concept, they developed the blockchain-based serialization prototype with supply chain partners in six locations to track the life cycle of drugs. Simulations demonstrated that the blockchain platform for genuine medicine could process 7 billion new serial numbers and 1,500 new transactions per second (Accenture, 2018).

The events related to the drugs in each step of the supply chain are recorded in the blockchain, and since it is immutable, they cannot be changed. A serial number is given to each sealed unit of drugs. Information including manufacturer, plant ID, and the expiration date is associated with the serial number and stored in the blockchain. While each unit is aggregated into cases and pallets, shipped to distributors, and eventually placed on the shelves, blockchain can track the exact location of each drug unit. Pharmacies, healthcare providers, and patients can trace back the drugs to see the origin of the drugs and whether it is counterfeit (Accenture, 2018; Aces & Kleeberger, 2018; Alla et al., 2018).

Conclusion

This research reviewed current blockchain technology adoptions aiming to enhance the resiliency of the supply network. Most of the blockchain adoption efforts facilitate collaboration and communication among suppliers to support the decision-making process. The limitations and challenges of the supply chain system were addressed by integrating various applications of blockchain technology.

In this study, we analyzed three blockchain-based supply chain use cases to identify how the main features of blockchain are suited best for supply network management. Immutability, traceability, tracking, and security by encryption are the features of blockchain technology utilized by most of the applications. Transforming from traditional paper-based manual supply chain management procedures to digital, immutable, and rapidly processing characteristics of blockchain technology helps enterprises provide solutions to improve public health, prevent fraud, significantly reduce costs and processing times, and ensure trust among partners.

Blockchain brings some benefits for the supply chain; however, it also presents some challenges to implement. In particular, some of the case studies we explored showed that the supply chain architecture and characteristics, stakeholder relations, and technological infrastructure of the organization and its stakeholders are important parameters for blockchain adoption. Qualitative studies have been conducted in the literature so far; however, there are still benefits of developing quantitative studies that enable researchers to test various scenarios on a well-informed blockchain adoption decision support system that uses modeling and simulations techniques.

References

Abdirad, M., & Krishnan, K. (2020). Industry 4.0 in logistics and supply chain management: A systematic literature review. *Engineering Management Journal*, 1–15.

Accenture. (2018, March 12). *DHL and Accenture unlock the power of Blockchain in logistics*. Accenture Newsroom. https://newsroom.accenture.com/news/dhl-and-accenture-unlock-the-power-of-blockchain-in-logistics.htm

Aces, C. G., & Kleeberger, M. (2018, June). Revolutionizing chains—The impact of blockchain. *Delivered*. DHL. https://www.dhl.com/global-en/home/about-us/delivered-magazine/articles/2018/issue-3-2018/revolutionizing-chains-the-impact-of-blockchain-on-the-supply-chain.html

Alla, S., Soltanisehat, L., Tatar, U., & Keskin, O. (2018). Blockchain technology in electronic healthcare systems. *Proceedings of the 2018 IISE Annual Conference*, 1–6.

Apte, S., & Petrovsky, N. (2016). Will blockchain technology revolutionize excipient supply chain management? *Journal of Excipients and Food Chemicals*, *7*(3), 910.

Bales, R. R., Maull, R. S., & Radnor, Z. (2004). The development of supply chain management within the aerospace manufacturing sector. *Supply Chain Management: An International Journal*.

Beamon, B. M. (1999). Measuring supply chain performance. *International Journal of Operations & Production Management*.

Casado-Vara, R., Prieto, J., De la Prieta, F., & Corchado, J. M. (2018). How blockchain improves the supply chain: Case study alimentary supply chain. *Procedia Computer Science*, *134*, 393–398.

Chen, S., Shi, R., Ren, Z., Yan, J., Shi, Y., & Zhang, J. (2017). A blockchain-based supply chain quality management framework. *2017 IEEE 14th International Conference on E-Business Engineering (ICEBE)*, 172–176.

Dani, S., Burns, N. D., & Backhouse, C. J. (2003). Human aspects of supply chain optimization. *IEMC'03 Proceedings. Managing Technologically Driven Organizations: The Human Side of Innovation and Change*, 350–353.

Fawcett, S. E., & Magnan, G. M. (2002). The rhetoric and reality of supply chain integration. *International Journal of Physical Distribution & Logistics Management*.

Folkinshteyn, D., & Lennon, M. (2016). Braving bitcoin: A technology acceptance model (TAM) analysis. *Journal of Information Technology Case and Application Research*, *18*(4), 220–249.

Fosso Wamba, S., Kala Kamdjoug, J. R., Bawack, R., & Keogh, J. G. (2018). Bitcoin, blockchain, and FinTech: A systematic review and case studies in the supply chain. *Production Planning and Control, Forthcoming*.

Francisco, K., & Swanson, D. (2018). The supply chain has no clothes: Technology adoption of blockchain for supply chain transparency. *Logistics*, *2*(1), 2.

Gausdal, A. H., Czachorowski, K. V., & Solesvik, M. Z. (2018). Applying blockchain technology: Evidence from Norwegian companies. *Sustainability*, *10*(6), 1985.

Harland, C. (1996). Supply network strategies the case of health supplies. *European Journal of Purchasing & Supply Management*, *2*(4), 183–192.

Heutger, M., Kuckelhaus, M., Chung, G., González-Peralta, J., Turner, K., Gockel, B., Acar, T., & Forster, M. (2018). *Blockchain in logistics*. DHL Customer Solutions & Innovation.

https://www.dhl.com/content/dam/dhl/global/core/documents/pdf/glo-core-blockchain-trend-report.pdf

Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics*, *4*(1), 1–23.

IBM Blockchain. (2017, August 22). *Walmart's food safety solution using IBM Food Trust built on the IBM blockchain platform*. https://www.youtube.com/watch?v=SV0KXBxSoio

Ivanov, D., Dolgui, A., Das, A., & Sokolov, B. (2019). Digital supply chain twins: Managing the ripple effect, resilience, and disruption risks by data-driven optimization, simulation, and visibility. In *Handbook of ripple effects in the supply chain* (pp. 309–332). Springer.

Ivanov, D., Dolgui, A., & Sokolov, B. (2019). The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics. *International Journal of Production Research*, *57*(3), 829–846.

JD. (2017, December 14). *Walmart, JD.com, IBM and Tsinghua University launch a blockchain food safety alliance*. JD Corporate Blog. https://jdcorporateblog.com/walmart-jd-com-ibm-and-tsinghua-university-launch-a-blockchain-food-safety-alliance-in-china/

Kim, H. M., & Laskowski, M. (2018). Toward an ontology-driven blockchain design for supply-chain provenance. *Intelligent Systems in Accounting, Finance and Management*, *25*(1), 18–27.

Kiviat, T. I. (2015). Beyond bitcoin: Issues in regulating blockchain transactions. *Duke LJ*, *65*, 569.

Kleindorfer, P. R., & Saad, G. H. (2005). Managing disruption risks in supply chains. *Production and Operations Management*, *14*(1), 53–68.

Kralingen, B. V. (2018, January 16). *IBM and Maersk to use blockchain to improve shipping*. CNBC. https://www.cnbc.com/video/2018/01/16/ibm-and-maersk-to-use-blockchain-to-improve-shipping.html

Kshetri, N. (2017). Blockchain's roles in strengthening cybersecurity and protecting privacy. *Telecommunications Policy*, *41*(10), 1027–1038.

Kshetri, N. (2018). Blockchain's roles in meeting key supply chain management objectives. *International Journal of Information Management*, *39*, 80–89. https://doi.org/10.1016/j.ijinfomgt.2017.12.005

Lambert, D. M., & Cooper, M. C. (2000). Issues in supply chain management. *Industrial Marketing Management*, *29*(1), 65–83. https://doi.org/10.1016/S0019-8501(99)00113-3

Ludema, I. M. W. (2002). Designing a supply chain analysis framework. *INCOSE International Symposium*, *12*(1), 1092–1099.

Madsen, J. T. (2019, September 20). *A game changer for global trade*. Maersk. https://www.maersk.com/news/articles/2019/09/20/a-game-changer-for-global-trade

Maersk. (2019, July 2). *TradeLens blockchain-enabled digital shipping platform continues expansion with addition of major ocean carriers Hapag-Lloyd and Ocean Network Express*. Maersk. https://www.maersk.com/news/articles/2019/07/02/hapag-lloyd-and-ocean-network-express-join-tradelens

Mengelkamp, E., Gärttner, J., Rock, K., Kessler, S., Orsini, L., & Weinhardt, C. (2018). Designing microgrid energy markets: A case study: The Brooklyn Microgrid. *Applied Energy*, *210*, 870–880.

Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system*.

Park, A., Nayyar, G., & Low, P. (2013). Supply chain perspectives and issues. *A Literature Review, WTO and Fung Global Institute*.

Queiroz, M. M., Telles, R., & Bonilla, S. H. (2019). Blockchain and supply chain management integration: A systematic review of the literature. *Supply Chain Management: An International Journal*, *25*(2), 241–254. https://doi.org/10.1108/SCM-03-2018-0143

Reay, J. H. (2000). *DOD supply chain management implementation guide*. Logistics Management.

Saberi, S., Kouhizadeh, M., Sarkis, J., & Shen, L. (2019). Blockchain technology and its relationships to sustainable supply chain management. *International Journal of Production Research*, *57*(7), 2117–2135. https://doi.org/10.1080/00207543.2018.1533261

Sabz Ali Pour, F., Tatar, U., & Gheorghe, A. (2018). Agent-based model of sand supply governance employing blockchain technology. *Annual Simulation Symposium (ANSS 2018)*. 2018 Spring Simulation Multi-Conference, Baltimore, MD. https://doi.org/10.22360/SpringSim.2018.ANSS.023

Shrier, D., Wu, W., & Pentland, A. (2016). Blockchain & infrastructure (identity, data security). *Massachusetts Institute of Technology-Connection Science*, *1*(3), 1–19.

Tan, B., Yan, J., Chen, S., & Liu, X. (2018). The impact of blockchain on food supply chain: The case of Walmart. *International Conference on Smart Blockchain*, 167–177.

Tapscott, D., & Tapscott, A. (2017). How blockchain will change organizations. *MIT Sloan Management Review*, *58*(2), 10.

Tatar, U., Gokce, Y., & Nussbaum, B. (n.d.). *Law versus technology: Blockchain, GDPR, and tough tradeoffs*. 9.

Terzi, S., & Cavalieri, S. (2004). Simulation in the supply chain context: A survey. *Computers in Industry*, *53*(1), 3–16.

Tian, F. (2017). A supply chain traceability system for food safety based on HACCP, blockchain & internet of things. *2017 International Conference on Service Systems and Service Management*, 1–6.

TradeLens. (2018). *Digitizing global supply chains*. https://www.tradelens.com/

Walker, D. M. (2006). *DoD acquisitions contracting for better outcomes* (GAO-06-800T). GAO. https://www.gao.gov/assets/gao-06-800t.pdf

Walmart. (2021). *Location facts*. https://corporate.walmart.com/our-story/our-locations

Wang, Y., Han, J. H., & Beynon-Davies, P. (2019). Understanding blockchain technology for future supply chains: A systematic literature review and research agenda. *Supply Chain Management: An International Journal*, *24*(1), 62–84. https://doi.org/10.1108/SCM-03-2018-0148

Wilding, R., Wagner, B., Colicchia, C., & Strozzi, F. (2012). Supply chain risk management: A new methodology for a systematic literature review. *Supply Chain Management: An International Journal*.

Yoo, M., & Won, Y. (2018). A study on the transparent price tracing system in supply chain management based on blockchain. *Sustainability*, *10*(11), 4037.

Blockchain Mergence for Distributed Ledgers Supporting Fleet Logistics and Maintenance

Britta Hale—is a Cryptographer and Assistant Professor in Computer Science at the Naval Postgraduate School. Hale has a PhD from the Norwegian University of Science and Technology and a Master of Science in Mathematics of Cryptography and Communications from Royal Holloway University of London. Specialization areas include cryptographic key exchange and authentication protocols, protocol self-healing, post-quantum hybrids, unmanned vehicle security, and secure channels within constrained settings. Hale is currently a member of the Message Layer Security working group of the Internet Engineering Task Force. [britta.hale@nps.edu]

Don Brutzman—is a Computer Scientist and Associate Professor of Applied Science working in the Modeling Virtual Environments Simulation Institute, Undersea Warfare Academic Group, and Information Sciences Department. He leads the Network Optional Warfare project, exploring fleet stealth using efficient messaging, optical signaling, semantic coherence, and ethical control of unmanned systems. His research interests include underwater robotics, real-time visualization using the Extensible 3D Graphics International Standard, artificial intelligence, and networking for distributed large-scale virtual environments. [brutzman@nps.edu]

Terry Norbraten—earned a Master of Science in Modeling, Virtual Environments and Simulation (MOVES) at the Naval Postgraduate School in 2004. Terry retired from active duty in the U.S. Navy in 2005 and has served as a Research Associate with the MOVES Institute by instructing in the Java and JavaScript programming languages and working on various software projects involving stand-alone and web-based scenario generation, discrete event simulation, and training and analysis for fleet and field requirements. [tdnorbra@nps.edu]

Jonathan Culbert—completed his master's degree at Naval Postgraduate School in 2020. His thesis was titled *Toward Understanding the Longitudinal Stability of an IP Geolocation Database*. [jaculber@nps.edu]

Abstract

Blockchain is highly adaptable and enables distributed transaction logging through its cryptographic underpinnings, making it an attractive technology for diverse suppliers and acquisition integrators. Supply chain tracking using blockchain must, however, support updates to item records throughout the life cycle—including repair and carcass tracking within Depot Level Repairable (DLR) and back into operation. Unmanned systems, additive manufacturing of parts, and version-control of software updates are all exemplars related to the supply chain requiring addition, deletion, updating, and mergence of a wide array of records. This raises the question of how to build and integrate an integrity-protected item history record that is updateable regardless of when or where changes may occur. We call this approach to updateable record management *blockchain mergence* and investigate how item tracking can be achieved throughout the full item life cycle, even under intermittent connectivity of deployed assets in combat environments. We demonstrate blockchain mergence through an interweaving of dual chains—an authenticated local history signature chain and a global blockchain—and apply it to an unmanned aerial system repair case. Blockchain mergence offers significant opportunities for distributed decentralized trust among diverse producers and consumers of both materiel and information, ashore and afloat.

Background

Blockchain has been widely researched for applications due to the technology's ability to support consensus among distributed participants. The chain is, by design, required to be a single, forward path of events; if branches appear, the chain consensus ensures that all but one branch is discarded (Zheng et al., 2017). A supply chain, in comparison, particularly on the production side, is a reversed architecture. In this case, small parts are used to build larger parts, hence requiring a form of *mergence* (e.g., a final ready-for-use vehicle is comprised of multiple smaller parts sourced from various vendors, manufacturers, and even countries). Blockchain, according to its current design, fundamentally disallows this. Assuring supply chain integrity and visibility requires an adaptation of blockchain use to allow for the mergence that the original concept was not designed to handle. Such adaptive blockchain solutions provide new capabilities and support the potential for future manufacturing changes.

Approach

We survey existing blockchain solutions for forms of mergence, that is, solutions for merging chains into a single blockchain, such as would be necessary for supply chain assurance.

We analyze potential solutions using partner signatures (where supply chain partners commit to chain addenda by digitally signing new blocks while also committing to the entire previous chain). This requires analysis of security considerations based on different commitment variants. Furthermore, it requires consideration of potential time lines and time line collisions of block production.

The above solutions are evaluated with respect to formal blockchain integration. In particular, we investigate whether mergence of distributed ledgers is possible within exiting blockchain architectures or if it is feasible as a parallel assurance mechanism, such that commitments are uploaded to an existing blockchain. This evaluation will be made on mathematical feasibility as well as use case comparison.

Current Research on Ledger Mergence

Blockchain technology has been often touted as a solution to various challenges since its inception under Bitcoin and cybercurrency. An interesting question is how blockchain technology might benefit Navy logistics. In essence, blockchains are a list of records, or blocks, cryptographically linked as a distributed ledger for recording transactions among parties in a permanent and verifiable way (Zheng et al., 2017). Blockchain might also support “smart contracts,” which may be a way to reduce the administrative friction associated with honoring requirements across a large enterprise.

The hallmarks of a robust Distributed Ledger Technology (DLT) are decentralization between blockchain networks and the individual nodes in those networks, as well as the consensus reached when validating individual blocks when added to a network’s blockchain ledger (Khan, 2019). Khan (2019) also noted that characteristics such as the number of transactions per second (TPS) that a network can process, the network’s scalability, and how a particular network guards against malicious attempts to add false information are also key to a good system.

We focus on authentication of changes at the micro-level, with transparency in a ledger for support of supply chain assurance. Industry is working on a number of efforts involving supply chain logistics and supply chain management (SCM)—such as Hyperledger (2020), Everledger (2020), and Ethereum (2020)—that may have an application to the Navy’s logistical systems and perhaps could contribute to an agile logistical system. The central challenge is applying such efforts beyond acquisitions to the whole life cycle of the supply chain.

A Quick Look at DLT Then and Now

Nakamoto (2009) is considered the conventional originator of the original description of blockchain technology, although it is focused on the financial and Bitcoin applications. One should note, however, that the concepts surrounding blockchain predate this by a decade or so, and there is other research available on distributed ledgers before that time frame. Beyond Bitcoin, DLT and blockchain have been researched for various financial and operational tracking purposes. Zheng et al. (2017) and Natarajan et al. (2017) provided a general and fairly informal introduction into DLT and how it might integrate into mainstream day-to-day operations in the financial, private, and government sectors. Natarajan et al. (2017) also provided a sense of how "decentralized records of flow of commodities and materials across a supply chain by using trusted stakeholders to validate flows and movements" could benefit those stakeholders, lending credence to adopting DLT, which would enhance trust in the supply chain. For an overview of blockchain research, we point to Fosso Wamba & Queiroz (2020), which highlights the benefits of the creation of value in operations and supply chain management (OSCM). Statistics such as the number of published papers by country, topic, keyword summary, and relationships are recorded.

Although not explicitly addressing blockchain technology, Bonanni (2011) discussed supply chain discovery/awareness, concepts and concerns that motivate our work. Bonanni argued for "Radical Transparency" in the context of sustainable (carbon cost) supply chains, carbon-footprint measured supply chains, and product life cycle awareness and optimization. This runs into a similar problem set that Department of Defense (DoD) acquisition may encounter—companies' unwillingness to reveal their supply chain details as trade secrets, or an inability to do so because the companies are unaware of the source of their sources.

There has also been a line of research covering direct application of blockchain to SCM. Korpela et al. (2017) provided an analysis of how blockchain could be used to solve or ameliorate the issues of concern of the major stakeholders involved in a very large supply chain operation. The main contribution of the paper is the proposed elimination of a third party to mediate/handle supply chain inter-business and then address these popular concerns. Meanwhile, Banerjee (2018) provided an overview/summary of the use and benefits of blockchain in supply chain operations, such as

- reduced counterfeiting and origin tracing
- digital product details/life cycle
- custom-built provenance solutions: Software service providers can use the blockchain framework to build provenance solutions for its customers (permission blockchain)

Based on Banerjee's work, custom-built solutions appear to have gained traction within industry. For instance, Infosys has developed a product provenance solution using Oracle Blockchain Cloud Services that is based on Hyperledger Fabric. Infosys has also developed a coffee bean tracking provenance solution for its customers. Such examples point towards a demand for custom-built provenance solutions that can be developed with product- or industry-specific validations. It is important to note that the concept of provenance only functions when all the supply chain stakeholders are part of the blockchain network. The architecture of blockchain inherently traces products as they pass from one supply chain entity to another. These transactions are stored as blocks and chronologically linked according to the physical movement of "the goods." Supporting such tracking technologies motivates our solution (see the section titled A

Ledger Mergence Either in Blockchain or as a Module Approach Leveraging Existing Blockchain Solutions).

Kshetri (2018) provided a theoretical framework related to key objectives of SCM. Kshetri's work covers several corporate case studies of how the Internet of Things (IoT) blockchain SCM can be used by companies with differing levels and areas of interest in supply chain verification/source confidence (see Table 1). Such case studies, including the Chipotle E. coli outbreak ingredient tracing case study, may shed light on potential parallel solution behaviors involving a faulty/compromised hardware component recall in the DoD. Under a similar formal goal, Queiroz and Fosso Wamba (2018) covered blockchain SCM adoption in the United States and India. The study advocated for drawing on emerging literature on blockchain, supply chain and network theory, and technology acceptance models (TAMs). Queiroz and Fosso Wamba (2018) introduced a model based on a slightly altered version of the classical unified theory of acceptance and the use of technology (UTAUT).

Table 1. Cases Selected and Their Classification in Terms of Incorporation of the IoT and Deployment of Blockchain to Validate Individuals' and Assets' Identities (Kshetri, 2018)

The cases selected and their classification in terms of incorporation of the IoT and deployment of blockchain to validate individuals' and assets' identities.

Deg. of incorporation of IoT → / Deg. of deployment of blockchain to validate identities of individuals and assets ↓	High	Low
High	Maersk	Lockheed Martin Everledger
Low	Alibaba Chronicled Modum Walmart Gemalto Intel's solution to track seafood supply chain	Bext360 Provenance

A Method for Adapting Distributed Ledger for Supply Chain Use

Unmanned Aerial Vehicle Use Case: Systems Deployed by U.S. Navy Ships

Littoral combat ships (LCS) have two classes of relatively small surface warships designed for operations near shore by the U.S. Navy ("Littoral combat ship," n.d.). Modern designs allow for flexible mission execution, various mission payloads, and other tasking. Reduced crew complements mean individuals are assigned yet with reduced inventories of spare parts and supplies.

The use of unmanned aerial vehicles (UAVs) help in this regard. These vehicles are employed for scouting and other rapid response detailing that minimize risk to the overall mission, ship, and crew. The ecosystem for a typical UAV consists of four categories of components:

- *hardware:* airframe, sensors, computers
- *software:* communication, guidance and control
- *additive manufacturing (AM):* 3D printed wings, tails and other small parts for ad hoc repair
- *information:* keys, training, repair instructions, feedback, safety

Within these four categories of components, each is different and necessary for aggregation into a complete device, and each has different stakeholders and supply chains feeding ships' supplies. Thus, four parallel supply chains of interest exist, and each is interdependent; therefore, any mergence solution should necessarily support all four aspects, as seen in Figure 1. Note that even with acquisition of a device as a single unit, the nature of updates, potential repairs, and parts reuse between devices imply that it must be possible, for tracking purposes over the device lifetime, to handle the mergence of all four aspects.

System Constraints and Requirements Identification

In this section we explore the various system requirements in the context of the UAV use case.

▪ ***Scenario: UAV Deployment, Repair, and Operations***

Suppose that a ship deploys with stock gear and consists of two distinct yet similar versions of a UAV. The ship must maintain its current pace of operations until return to port or resupply.

Under normal operations, the following issues may affect device history—in that they impact the integrity of the device or its trustworthiness—and, therefore, should be added in an authenticated manner to the device history:

- software updates
- training and safety updates to ship tactics, training, and procedures (TTP) and standard operating procedures (SOP)

Now suppose that a collision occurs during testing between the two UAVs, causing damage to each vehicle. The following may also be important changes to the device history, requiring authenticated changes in device records:

- hardware replacements on board, including classified components
- 3D printing for upgraded tail assemblies
- maintenance feedback to shore commands

Any mergence solution must, therefore, support per minimum such a variety of changes to the item history.

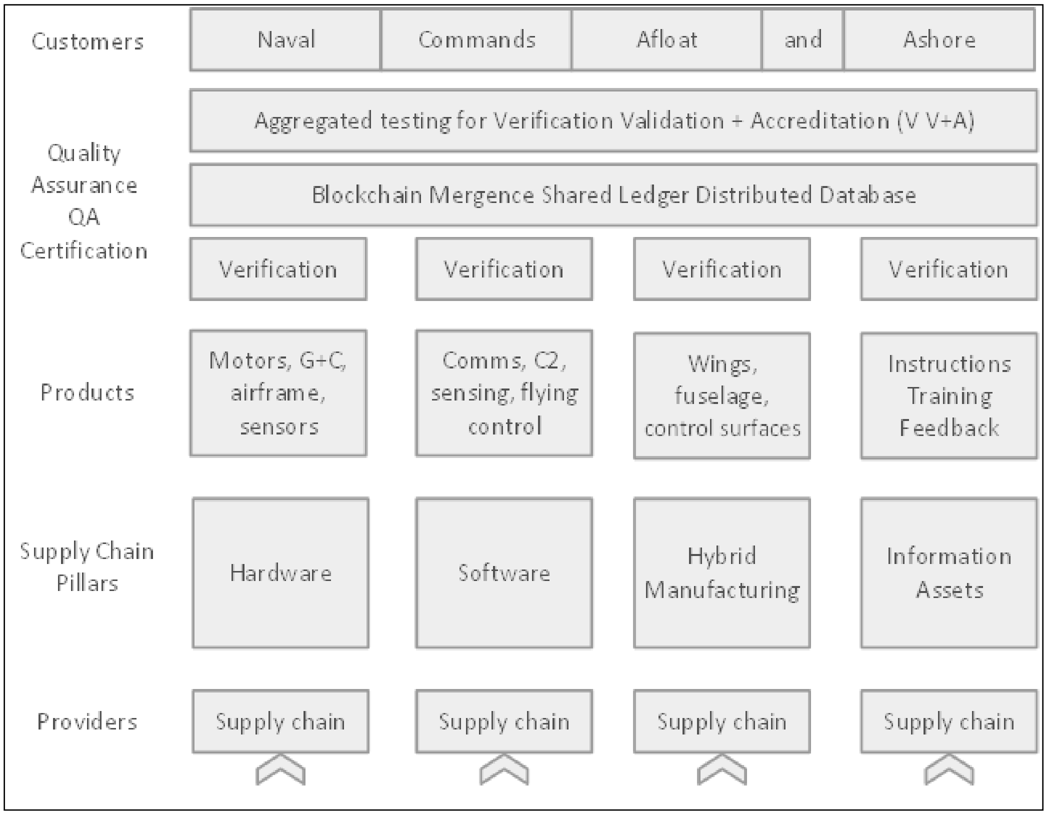


Figure 1. UAV Operational Assembly and Modification

- ***Mergence Requirements Assessment: Verifiability***

In addition to the aforementioned types of item record changes, there are also requirements in how a change is recorded. In particular, the record must be verifiable. In terms of verifiability, the following requirements are also essential and must be supported by a mergence solution:

- conformation that a given component X is on the ship
- conformation of all devices in the inventory that have X as a component
- conformation if and when X has been replaced/repaired and so on within a particular device
- conformation of the change entity—the responsible party to change/split/remove/combine X as a component within devices
- ability to add logs or metadata

The above requirements emerge from use case issues. For example, if a device component is found to be compromised and must be removed, the logged data associated with the device should indicate if it has been removed and by whom. Furthermore, it is important for administration purposes to identify all possible devices containing the compromised component for swift handling and damage mitigation. In these contexts, "components" may refer not only to hardware but also malicious software or poorly executed AM (e.g., 3D printed wings with vulnerable integrity).

- ***Mergence Requirements Assessment: Flexibility***

Finally, we list flexibility requirements associated with mergence. Since mergence solutions must support potential external (industry/non-DoD) supply chain tracking of unpredictable natures, the mergence solution must be fairly adaptable. Furthermore, external supply chain tracking may differ from internal (DoD) tracking, and the potential solution must support one or more blockchains used internally to the DoD. As many acquisition devices may be of a sensitive nature, the mergence solution must furthermore support various classification levels, such that unclassified devices may be administered in unclassified environments, while devices of higher classification levels can also be managed within the same mergence solution without sensitive information leakage. Finally, in addition to all of these, devices transfer hands between organizations, ships, and so on, requiring a flexibility to record management. This leads us to the following final four solution requirements:

- flexibility independent of source/industry in the external supply chain
- flexibility with internal blockchains(s) within the DoD
- flexibility with classification levels
- flexibility for device transfer between organizations, ships, and so on internally

Ledger Mergence Either in Blockchain or as A Modular Approach Leveraging Existing Blockchain Solutions

There is a natural separation between external-DoD and internal-DoD supply chain tracking. This intrinsically leads to a dual solution, with the acquisition boundary denoting a change in authenticity tracking. Even for internal supply chain tracking, satisfying all solution requirements appears, on the outset, to be impossible. Notably, a solution that crosses classification boundaries must be carefully handled, especially for full item records and tracking information. We handle this by further separating out the internal DoD authentication chain into two parts.

- ***External and Internal Chains***

DoD equipment is typically procured via outside commercial manufacturing vendors. The supply chain starts outside of the DoD, where parts and other equipment must be verified and validated before becoming available inside internal supply chains. Conceptually, manufacturers may require supply chain assurance as well, tracking purchased components for integration in building devices. This may take the form of various blockchains (see Figure 2). Minimally, manufacturers may be required to present verification on the types and sources of a device's components. At acquisition, a new item record will be formed, such that the component history of the acquired device is verified and authenticated by the acquisition authority, who registers components under a digitally signed genesis block. Once a genesis block for the internal ledger is formed, tracking may proceed internally.

What is essential at the DoD boundary/component registration step is the actual verification of internal components to a device. Information on processing chips, software, and so on must be recorded. This enables future tracking such that if, for instance, a component is later discovered to be compromised in the manufacturing chain, all devices containing the critical component can be identified. The genesis block thus serves as an initial registration for all components, such that it is only necessary to record changes to that initial list within the device history record.

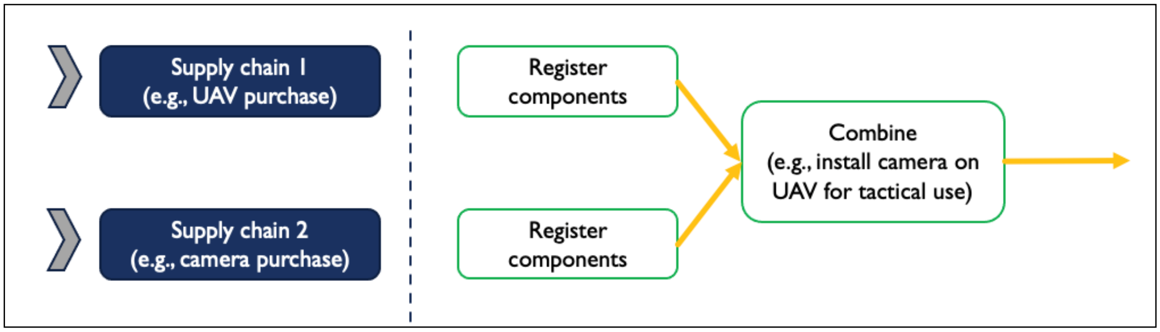


Figure 2. DoD Acquisition: Verification of External Input Chains(s), Registering of Devices to Initiate Internal Chains

- ***Mergence Operations***

We break internal tracking into two further chains to support classification boundaries. The *device chain* handles immediate time history and authenticates records as visible to the admin. Moreover, the device chain is designed to support the current Depot Level Repairable (DLR) system. Within the mergence solution, in addition to the device chain, we also support one or more internal blockchains. These may be organization level blockchains or classification level blockchains. Here we employ the term *blockchain* for a distributed immutable ledger, without specification that restricts to any particular ledger format or consensus method. This in turn meets the flexibility requirements specified in the section of this paper titled System Constraints and Requirements Identification.

We are motivated to show how blockchain technology might be applied to handle the full range of potential fleet resupply maintenance and modification requirements. We define the following device chain operations, in accordance with system requirements, as the fundamental primitive operations that together provide the general coverage necessary for distributed accountability of system modifications:

- *device registration*: adding a new, original item to a ledger. This creates the genesis block for the device chain.
- *device repair*: adding a new component onto an existing device. This differs from *device combine* in that the component being added has no registration history (i.e., no genesis block). This may occur if the repair takes place using AM.
- *device split*: separating components within an existing device. This supports potential reuse or disposal, such as when a component breaks and is removed from the current item record (device history is still maintained). This creates two separate device chains: one for each split component.
- *device combine*: integrating two components into a new combined device. This supports customization of devices after acquisition and parts replacement (e.g., a newly purchased component added to an existing device).

Device split can be employed if a device breaks but components can be reused. For example, suppose that a UAV (UAV1) malfunctions, but certain components can be used to repair another UAV (UAV2). The broken device would then have a device split operation in its item record, creating two new chains: one for the component that will be reused and one for the remaining unusable assembly UAV1. A device combine operation then integrates the split component into UAV2. As such, the item history of UAV1 is now linked to UAV2. If there were relevant repairs to the reused component or if it comes to light that the reused component was compromised during manufacture and must be pulled from use, it will be immediately clear from UAV2's record history that the part now resides within UAV2 instead of the UAV1 device carcass.

Each of the stated operations must be authenticated. For this, we use the public key infrastructure (PKI) already inherent in the DLR system. The operator responsible for the device signs the various operations. The signature covers the current record for the device(s) being operated on as well as what type of operation is performed. The authenticated transcript is stored as part of the device chain. These operations are shown in Figure 3.

The distributed ledger and shared memory exist beyond the immediate device chain history, such that an item record cannot be changed a posteriori. For this we employ a blockchain, which records the signatures from the device chain operations. Note that we only require the signatures, and not the related device information, to be stored on the blockchain, although the latter may be stored also. Storage of further information or metadata may be beneficial for device tracking but could also leak information (such as if the device or its location is sensitive). Instead, we require the minimum information on the blockchain concerning the current signature state.

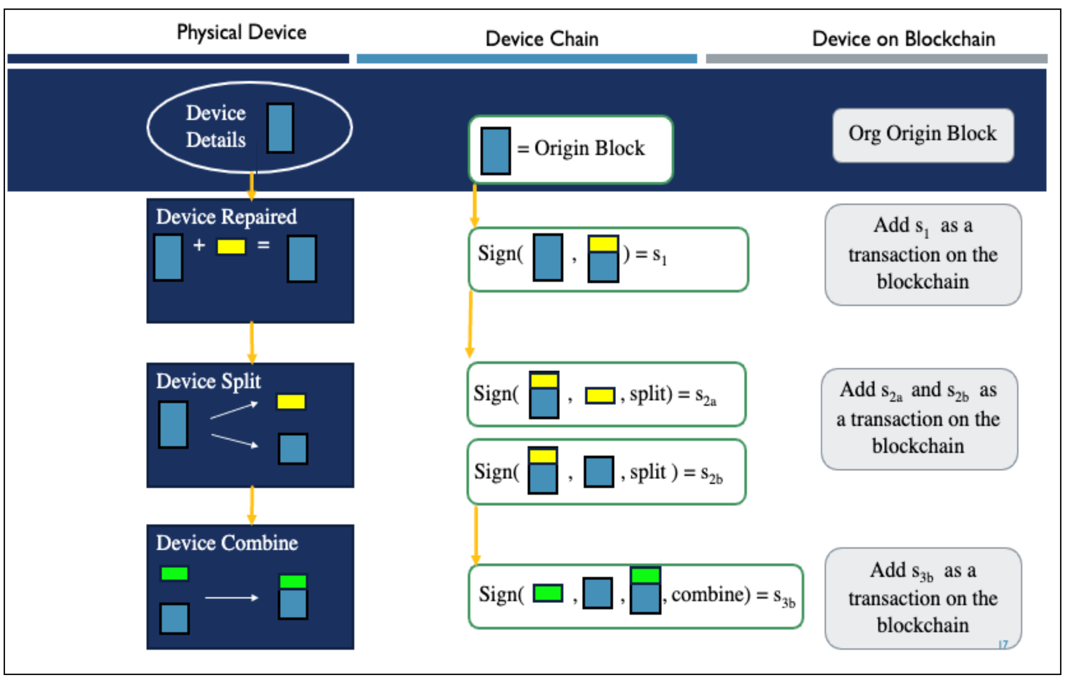


Figure 3. Sequencing Operations for Blockchains

- ***Multilevel Security Classification Agility Considerations***

For hybrid devices used in the fleet, activities may occur and be needed across multiple levels and domains of security, such as UNCLASSIFIED, CONFIDENTIAL, SECRET, and TOP SECRET. For this, we map our solution to the multilevel security (MLS) classification system and demonstrate interoperability.

Note that while a distributed chain contains data blocks of potentially classified information, the information sent to the blockchain is comprised of merely the signature on the data rather than the data itself; any further additions are optional. Even with a time code associated to the signature object representation, there is no intrinsic value to the information outside of the context of the signed data, especially with a plentitude of blockchain transactions. Thus, the blockchain information can be shared across MLS systems since these codes are useless to an attacker without corresponding ledger (database) access.

In addition to the above observation, we can also allow for blockchains operating at different classification levels, such that more relevant device information than merely the signature may in fact be added to the blockchain. This in turn implies that any device may have a record with varying classification levels attached to different aspects of the associated information and that the associated data may be placed on the relevant blockchain. Naturally, higher classification can correlate same- and lower-level data records, but not write to them, per the properties of the MLS system. Figure 4 illustrates this framework in practice.

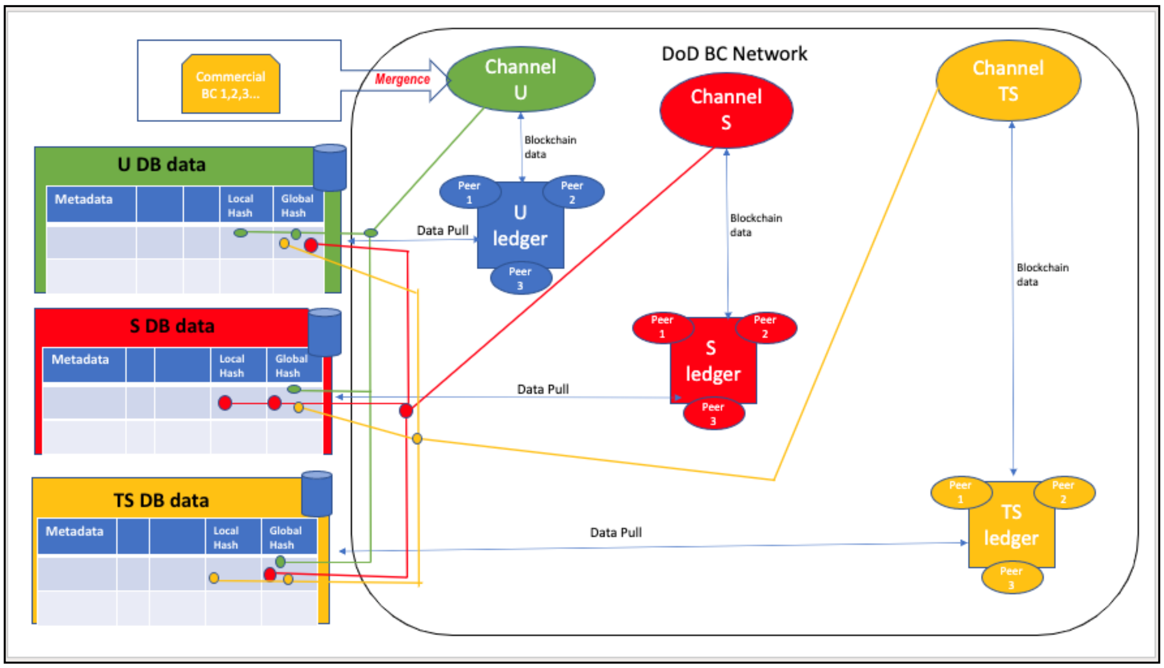


Figure 4. MLS Classification Agility Within a DoD Blockchain Network

To consider how this may work in practice, we walk through the following conceptualized steps for handling the mergence solution of device chains and blockchains within MLS:

1) A new item genesis record for an UNCLASS device is created for the device chain. This correlates to a device origin record with signature information populated to the various blockchains.
2) The device is transferred from an UNCLASS environment to a SECRET environment. The device chain is now maintained at SECRET.
3) The device is repaired, using a combine operation on the device chain. Necessary information is populated to the appropriate and corresponding

SECRET level blockchain, while other chains record signature information only.

If it is later discovered that the device of interest contained a malfunctioning or compromised component (e.g., through manufacturer notification), then an operator might inspect all associated components in the device genesis block, recognizing that the critical component is present. It can then be seen that the device was associated to a different classification (or organization blockchain), and the appropriate authority for that blockchain can be contacted, who can then trace the device's history to identify if the component is still present or has been replaced.

DOD Equipment Repair Transactions Using the Hyperledger Fabric Framework for Distributed Blockchain Ledger Keeping

Using the open-sourced Hyperledger Fabric framework (Hyperledger, 2020) hosted on GitHub (2020), we have constructed a case study exemplar using open-source software and demonstration records that shows how various repair level organizations might use blockchain mergence solutions to record supply chain transactions between participants using blockchain as a distributed ledger technology.[18]

- ***UAV Camera***

For our case study, we have an organizational level (O-level) end user that currently possesses a UAV that houses a camera subcomponent, a DLR that requires repair at the depot level (D-level). The client application transactions that take place and are recorded on the blockchain ledger are

1. O-level issues the nonfunctional DLR camera to D-level for repair
2. D-level accepts and conducts the required repairs for the DLR camera
3. D-level reissues the repaired DLR camera back to O-level

The intent is to show chain of custody for the DLR camera subcomponent, camera metadata (i.e., serial number), and status of repair of the DLR in the supply chain. The blockchain network (N) is comprised of the following consortium organizations, components, and entities (illustrated in Figure 5):

- Organizations R1 (D-level), R2 (O-level), and R4 (blockchain network administrator).
- Client applications A1 (D-level transactor) and A2 (O-level transactor). Client applications conduct transactions on behalf of their respective organizations.
- Certificate authorities CA1, CA2, and CA4. Each organization can prefer their own vetted certificate authority.
- Peers P1 (D-level) and P2 (O-level). Peers maintain local copies of and record blockchain ledger transactions in accordance with agreed upon smart contracts (chaincode) within the consortium.
- Blockchain ledger L1. Each peer maintains and communicates with other network peers to ensure local blockchain ledger copies are kept uniform throughout the network.

[18] Details can be found at https://gitlab.nps.edu/tdnorbra/blockchain-mergence

- Smart contract (chaincode) S5. Peers are able to maintain blockchain ledger uniformity through consortium member agreed upon smart contracts.
- Network ordering service O4. The ordering service serves as the initial administrative gateway between consortium members upon network standup.
- Network configuration NC4. Consortium members R1, R2, and R4 all agree upon the blockchain network configuration policies administered by ordering service O4 via NC4.
- Channel configuration CC1. The channel configuration allows for network peers to accept and distribute blockchain ledger transactions between authorized organizations in accordance with NC4.
- Channel 1. The communications channel where organizational peers accept and record transactions between client applications A1 and A2 on Channel 1.

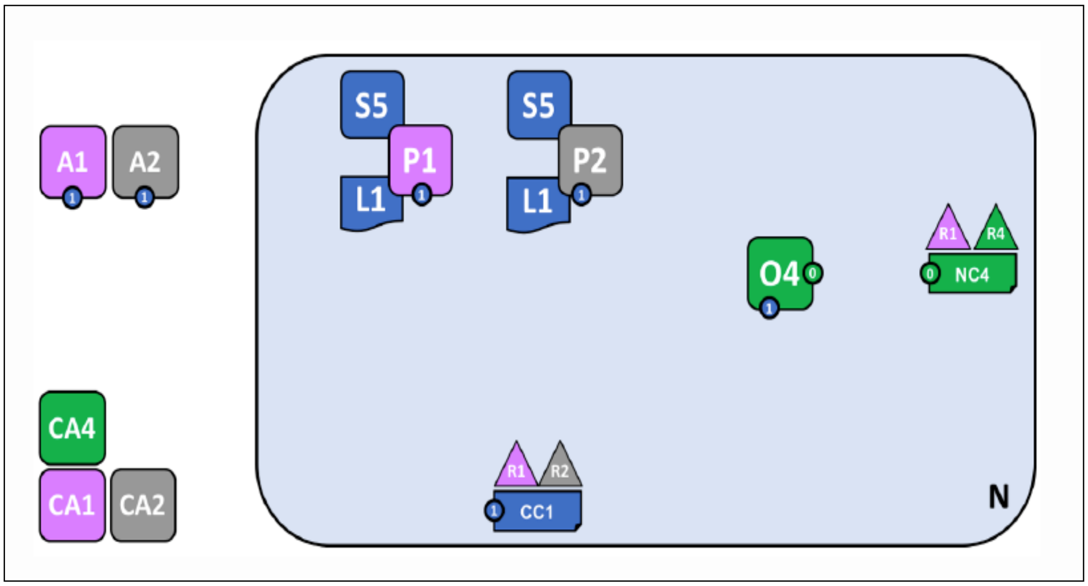


Figure 5. DoD Equipment Repair Blockchain Network (N) Configuration Topology

- ***Using the Fabric Test Network***

We apply the Hyperledger Fabric test network for our case study test benchmarking (Hyperledger, 2020). Depending on the operating system of the developer's choice, a specific environment may first be set up on a local machine to leverage the Hyperledger Fabric framework. Once the prerequisites are complete, the test network is invoked via command line scripts and the demonstration can proceed. In our case, the demonstration code was modified locally to mimic our UAV camera scenario.

- ***DLR Transactions in Action***

The codebase and a demonstration video for this contribution is maintained in open-source version control, located on the Naval Postgraduate School's GitLab server (Norbraten, 2020). In particular, we trialed the following case study stages:

1. Upon network standup/startup, the various organizational peers, admins, and so on are defined, registered, enrolled, and assigned certificate authorities, which then issue authenticating certificates for each respective network node. The agreed upon smart contract (chaincode) is deployed to each organizational peer node and tested. Finally, each organizational peer node is given local custody of the blockchain ledger, which is then readied for acceptance and recording of ledger transactions.
2. The network facilitates transactions. Each client application (authorized organizational transaction entity) submits their authentication data to the network before the network authorizes the transaction to take place.
3. Finally, the network is operational, and authorized organizational entities are recognized. Transactions may begin, but only those that are delineated in the smart contract (an O-level entity submitting a UAV camera to D-level for repair). We run three example transactions, as per the smart contract constructs initiated from each of the O-level and D-level authorized entities. These transactions are initiated from each client gateway interface application that has knowledge of the network from their respective remote locations.

- ***Basic Network Timing Data***

For any system component introduction, computational expense is valuable to assess. We performed distributed ledger testing on commodity hardware, and delays were tolerable once system initialization was complete. Table 2 below annotates various local network timing data benchmarks for tests performed. Experiments were run using a 2015 Apple MacBook Pro laptop with a 3.1 GHz dual-core Intel i7 and 16 GB of RAM running the latest macOS operating system.

Table 2. Network Timing of Various DLR Transactions on Network (N)

Network Action	Time Units
Blockchain Network Startup/Standup	2 min 19 sec
Client Application Authentication	19.4 sec
O-level DLR Issue Transaction	18.3 sec
D-level Acceptance/Receipt ACK of DLR	17.8 sec
D-level Reissue of RFI DLR	18 sec

Conclusions and Recommendations

The span of device life cycle and rapid evolution from design to development, prototype testing, requirements approval, acquisition testing, deployment, casualty response, and system upgrades is immense. Automated authentication and change logging within this life cycle support human-mediated checkpoints. Industry is looking to blockchain solutions for supporting the ecosystem, which opens an opportunity for leveraging the technology for parts tracking within the DoD. Blockchain mergence provides a bridge between local device repairs and blockchain integrity. In essence, it builds on the DLR system with similar fine-grain parts replacement and carcass tracking by means of hard-coded integrity stamps on the blockchain as a defense against adversarial or even undesired but accidental changes to a device's logged history.

We investigated the logistics challenges that a blockchain mergence solution must address, including pre- and post-acquisition concerns, procedural modifications, preventative and corrective maintenance, and field repair. Furthermore, we evaluated potential solutions against fundamental requirements for interoperability, such as with the existing DLR system and MLS. Finally, our initial open-source, Hyperledger Fabric–based simulation of the refined blockchain mergence solution demonstrates proof-of-concept capabilities by applying a widely used industry software library for blockchain configuration, simulation, and confirmation.

Recommendations for Future Work

This proof-of-concept exemplar work is ready for further development to match the full logistics life cycle. A real-world prototype case study emulating multiple participants (ashore hardware suppliers, operational forces, and unmanned systems with hardware-dependent software updates) might further test and demonstrate necessary capabilities and ledger distribution. Using the preexisting CAC infrastructure for individual identification further enables a full, ready-to-test blockchain mergence solution.

System engineering assessments can include potential augmentation of existing systems to better support the increased requirements accompanying the fast-growing deployment of unmanned systems. Use case considerations may also be assessed for trusted deployment and updates to hardware and software, as well as inclusion of additive manufactured parts. Future work might then emulate the full life deployment cycle for a fleet-critical system of interest to explore operational parallels in data-centric security for human–machine tactical deployments.

The blockchain mergence solution is designed for smooth integration with the existing DLR system. Feasibility testing of the combined system comprised of blockchain mergence, with the DLR system, and CAC infrastructure is a logical next step for future work.

References

Banerjee, A. (2018). Blockchain technology: Supply chain insights from ERP. *Advances in Computers*, *111*, 69–98. https://doi.org/10.1016/bs.adcom.2018.03.007

Bonanni, L. (2011). Sourcemap: Eco-design, sustainable supply chains, and radical transparency. *ACM Crossroads*, *17*, 22–26. https://doi.org/10.1145/1961678.1961681

Enterprise resource planning. (n.d.). In *Wikipedia*. Retrieved December 22, 2020, from https://en.wikipedia.org/wiki/Enterprise_resource_planning

Ethereum. (2020). *Ethereum is a global, open-source platform for decentralized applications*. https://ethereum.org/en/

Everledger. (2020). *Meet the Everledger platform*. https://www.everledger.io

Fosso Wamba, S., & Queiroz, M. (2020). Blockchain in the operations and supply chain management: Benefits, challenges and future research opportunities. *International Journal of Information Management*, *52*. https://doi.org/10.1016/j.ijinfomgt.2019.102064

GitHub. (2020). *Repository for the open-sourced Hyperledger Fabric Framework source code*. https://github.com/hyperledger/fabric#releases

Hyperledger. (2020). *Open, proven, enterprise-grade DLT*. https://www.hyperledger.org/wp-content/uploads/2020/03/hyperledger_fabric_whitepaper.pdf

Kahn, F. (2019, February 14). *What are the different types of DLTs & how they work?* Data Driven Investor. https://www.datadriveninvestor.com/2019/02/14/what-are-the-different-types-of-dlts-how-they-work

Korpela, K., Hallikas, J., & Dahlberg, T. (2017). Digital supply chain transformation toward blockchain integration. *Proceedings of the Hawaii 50th International Conference on System Sciences*, 4182–4191. https://doi.org/10.24251/HICSS.2017.506

Kshetri, N. (2018). 1 blockchain's roles in meeting key supply chain management objectives. *International Journal of Information Management*, *39*, 80–89. https://doi.org/10.1016/j.ijinfomgt.2017.12.005

Littoral combat ship. (n.d.). In Wikipedia. Retrieved December 3, 2020, from https://en.wikipedia.org/wiki/Littoral_combat_ship

Nakamoto, S. (2009). *Bitcoin: A peer-to-peer electronic cash system*. Bitcoin. https://bitcoin.org/bitcoin.pdf.

Natarajan, H., Krause, S., & Gradstein, H. (2017). *Distributed ledger technology and blockchain*. World Bank Group. https://doi.org/10.1596/29053

Norbraten, T. (2020). *Developer code hosted on the Naval Postgraduate School's GitLab repository for this NRP project*. NPS GitLab.

Queiroz, M., & Fosso Wamba, S. (2018). Blockchain adoption challenges in supply chain: An empirical investigation of the main drivers in India and the USA. *International Journal of Information Management*, *46*, 70–82. https://doi.org/10.1016/j.ijinfomgt.2018.11.021

Shahaab, A., Lidgey, B., Hewage, C., & Khan, I. (2019). Applicability and appropriateness of distributed ledgers consensus protocols in public and private sectors: A systematic review. *IEEE Access*, *7*, 43622–43636. https://doi.org/10.1109/ACCESS.2019.2904181

Zheng, Z. Xie, S. Dai, H., Chen, X., & Wang, H. (2017). An overview of blockchain technology: Architecture, consensus, and future trends. *2017 IEEE International Congress on Big Data (BigData Congress)*, 557–564. https://doi.org/10.1109/BigDataCongress.2017.85

Blockchain Data Management Benefits by Increasing Confidence in Datasets Supporting Artificial Intelligence (AI) and Analytical Tools using Supply Chain Examples

Anthony Kendal—is a lecturer and researcher at the Naval Postgraduate School where he is in the Information Sciences department researching and teaching in the field of data science including Investigating the emerging "Big Data" and machine learning concepts and technologies on how they enhance analytics for decision-making in DoD areas such as logistics and CTAP (Common Tactical Air Picture). He currently heads a project on using "money ball" and other methods to provide analytics for Navy aviation maintenance as well as a project on investigating how blockchain might assist in Navy logistics processes. Tony graduated from the University of Texas at Arlington. In his earlier career he was a Naval Flight Officer (lieutenant commander) He wrote an award-winning article (USNI) on creativity in the military. He received his Master of Science at the Naval Postgraduate School.

Arijit Das—is a Computer Science research faculty at the Naval Postgraduate school specializing in teaching CS courses and doing applied research trying to find industry solutions for DoD/DoN challenges. His educational background is BECS from MNREC (Allahabad University, India), followed by a MSCS from Oregon State University (Corvallis, Oregon). His other MS degree is an MSEE from UNLV (Las Vegas, Nevada). His work experience is in the Philadelphia tri-state area service industries as a developer using technologies running on Mainframe, Oracle, and UNIX as production servers. Arijit is a frequent speaker in Oracle Database user group events in the Silicon Valley. His current research focus is Big Data, AI/Analytics and Blockchain.

Bruce Nagy—is a Senior Scientist and Systems Engineer at the Naval Air Warfare Center, Weapons Division at China Lake. His research focuses on advanced game theory techniques, artificial intelligence and machine learning applications as applied to naval weapon systems and tactical decision aids. Nagy received an BS in Biology and BA in Mathematics from the Citadel. He went on to receive an BS and MS in Electrical Engineering from the Naval Postgraduate School. He continued his neural network interest with post graduate work in modeling brain stem activities to muscle fibers at UCLA, in cooperation with NIH. Nagy is a former Engineering Duty Office. bruce.nagy@navy.mil

Avantika Ghosh—is a sophomore at University of California, Berkeley studying Computer Science and Economics. She has been a research intern at NPS in 2018 and 2020. In 2018, she conducted a Data Analytics study on benchmarking various machine learning algorithms' to detect anomalous behavior in military aircraft data. Through her current research, she is exploring possible applications of blockchain technology in the Navy. Avantika has been awarded the Bay Area Affiliate Winner award by the National Center for Women in Technology.

Abstract

We describe how Hyperledger Fabric (HLF) blockchain (BC) technology that we previously applied to Navy logistics supply chains can be applied to data supporting artificial intelligence (AI) and software development in terms of system safety and the timely acquisition of data. Data-driven AI/machine learning (ML) requires trusted data for their use in AI functions and requires significant amounts of training data from diverse sources including Internet of Things (IoT) devices/sensors. Unauthorized alterations to data supporting AI/ML could go unnoticed within the AI function build process but surface during operation in hazards affecting unwanted human death or resource destruction. AI/ML controlling hardware usually falls into the two highest software control categories: Levels 1 and 2, risk of death, disability, or resource destroyed.

HLF BC is a tamper-resistant decentralized trusted ledger that provides proof of transaction where trust is implemented through distributed consensus to ensure that only authorized people can modify data and that the modification is traceable and transparent. Distributed ledgers provide system safety through BC provenance, immutability, and policy enforcement through smart contracts.

We show how BC can contribute to the safety of the data and transactions and provide data to the researchers in a timely manner through "smart repositories."

Introduction

Artificial intelligence (AI) software exercises a high degree of control over a particular system function (e.g., movement/guidance of a missile). If the function creates a hazard, this can cause a mishap that has a consequence of a catastrophic and critical event, resulting in death or resources destroyed. There is no redo, reboot, or retraining of an AI function that fails in this scenario. Software safety engineering and test and evaluation efforts to ensure fidelity include data-related elements such as process flow, code level, and data structure analysis. These flows, for example, are similar to the use of blockchain (BC) in supply chains used in our previous research, the Navy supply chain process, which we believe can be adapted for system safety and software integrity.

AI/machine learning (ML), the training sets, algorithms, and associated software supporting weapons systems are targets for increasingly sophisticated adversarial machine learning attacks, which attempt to fool models through malicious input into the system such as AI poisoning and other attacks. Athalye et al. (2017) showed that it is even possible to fool an AI into having it identify a turtle wrongly as a rifle. BC could be used as a countermeasure to prevent such poisoning as well as safeguard system integrity. BC, specifically Hyperledger Fabric (HLF), is a tamper-resistant decentralized trusted ledger that provides proof of transaction where trust is implemented through distributed consensus to ensure that only authorized people can modify the code base, AI algorithm, or training set and that the modification is traceable and transparent. Distributed ledgers provide system safety through BC provenance and policy enforcement through a feature called smart contracts, which imbed logical code. Data in support of AI and software development can also suffer not only from deliberate sabotage or ruse but also from human error.

Machine learning increasingly requires complex data sources from repositories and sensors down to the edge for training sets supporting AI development. Getting the right and accurate data can be a complex process, and error or intentional manipulation is always a concern. The number of sensors and Internet of Things (IoT) devices, such as smart thermometers/oximeters to track COVID-19, has caused an explosion of data generation but not an increase in safeguards to ensure system safety if these edge devices are used to control machines or make life-critical decisions.

Centralized security and authentication controlling IoT devices could lead to a single point of failure, a new target for cyberattack, and cause a bottleneck and high latency (Jia et al., 2020). Typically, an ML project may require diverse data sources and modalities. One example may be drones flying over an urban area, which requires its ML training set data on the region, including crime rate, weather, and road conditions/constraints. For just this simple example, data needed include Naval War College (NWC) wargaming, Naval Postgraduate School (NPS) wargaming, Live Fire Event, Program of Record product performance specifications, contractor specifications, test evaluation results, a diverse set of sensors, IoT devices, and so on. Once an AI is trained, BC can be used to ensure the integrity of the data during operations. BC can be used to find the right data, what is in it, who owns it, and how to get it with quick authorization. Data scientists have long recognized that just getting the right data and permission to use it can be an arduous and long process.

The Problem of System Safety

AI has the potential of creating a technological leap (Eden et al., 2013). That potential leap, especially when dealing with weapon systems, needs scrutiny. This scrutiny focuses on the specificity of the composition and size of the training data algorithm. This research describes how an HLF architecture can be used to increase safety and confidence in the deployment of AI functions. There must be confidence in the data and training sets and the algorithms, and there must be confidence that they are tamper-proof and free from anomalies, intentional or by accident. Acquisition communities cannot identify and certify operational constraints of an ML algorithm for deployment without having confidence in the training data quality, including any negative side effects (Everitt, 2018) that might result from the training process.

The *system safety* concept calls for a risk management strategy based on identification, analysis of hazards, and application of remedial controls using a systems-based approach. This is different from traditional safety strategies (Roland & Moriarty, 1990).

AI safety issues for naval weapon systems usually have not included consideration of adversarial attacks that might affect functional performance. AI adversarial network attacks using techniques like deepfakes, putting an image/video into another image/video for miscategorization (Chauhan, 2018), will be considered within our BC discussion.

When assessing safety, the goal is to identify anything that might be safety-critical. *Safety-critical* is "a term applied to a condition, event, operation, process or item whose mishap severity consequence is either catastrophic or critical (e.g., safety-critical function, safety-critical path, and safety-critical component)" (Defense Standardization Program Office, 2012). Specifically, the publication MILSTD-882E (Defense Standardization Program Office, 2012) helps software engineers determine the level of rigor (LOR), which specifies the depth and breadth of software analysis and verification activities necessary to provide a sufficient level of confidence that a safety-critical or safety-related software function will perform as required. ML/AI usually falls into the system safety two highest software control categories: Level 1 (autonomous) and Level 2 (semiautonomous). We contend that BC could contribute to the analysis and verification of software activities by ensuring data integrity and better accessibility to the data.

Applying Successful BC Techniques to Ensure System Safety of AI Deployed Weapon Systems:

Our previous research used the HLF BC to generate three general use cases for Navy logistics, including financial and inventory transaction audit trails, serial number tracking, and maintenance log integrity. We believe the BC network derived from these three use cases could be adapted for system safety purposes since all our previous demonstrations dealt with the integrity of the data supporting work processes and events. BC tracks food/parts items as assets recorded on ledgers, and training data are assets and also created with similar work processes and events. With HLF you can control who, what, and when and identify those who have access to the logistics data representing assets as well through an immutable ledger containing logistics data that cannot be tampered with. HLF is as transparent as needed but can hide data from those without a need to know.

The data source flows of data and training sets supporting data scientists are similar to previous BC research on Navy supply chains to improve transparency and the safety of the related supply chain data and transactions, but there is a higher level of risk since they are often at Level 1 or Level 2 autonomous systems. In a sense, training sets and analytical data are like the tracking of parts and food since they point to resources represented by the information that needs to be protected and distributed in a friction-free manner. Control of these sources during the integration process to create training data and general analysis is vital to ensure the training sets and AI algorithms are transparent to those who need them, controlled, and their validity supported by an audit trail that BC provides. Training set alterations could go unnoticed within the AI function build process but revealed during operation in hazards affecting unwanted human death or resource destruction. Our previous research demonstrated how BC can provide a needed data management technology through a tamper-resistant decentralized trusted ledger that provides proof of transaction where trust is implemented through distributed consensus. Only authorized people can modify the code base, AI algorithms, or training set modifications that are detectable, traceable, and transparent. Distributed ledgers provide system safety through BC provenance and policy enforcement through smart contracts.

HLF is a consensus-based network that the Department of Defense (DoD) can control and has no “Proof of Work” protocol, which is a wasteful use of computer resources. HLF uses channels to control who can see what data and through consensus; the DoD can control what is allowed to be put on the BC ledger. Such technologies can not only be used in Navy supply and logistics to streamline and improve effectiveness in terms of how workflow can be improved to provide more rapid and secure distribution of material and two-way financial transactions but can also be used on data transactions such as datasets requested by data scientists. Data scientists have long recognized that obtaining “clean data” and the permission to use it has been hampered by administrative friction, which can be caused by data owner’s requirements, trust issues from generated data source transactions, and other administrative processes.

The benefits of BC technology described in this paper support system safety in terms of providing objective quality evidence about data integrity, as well as test and evaluation teams in terms of data management control. We believe elements of BC, such as smart contracts, could contribute to all acquisition groups involved. We will discuss our previous logistics use case as well as new use cases specifically for software safety.

The Hyperledger Fabric Blockchain Solution

HLF provides proof of transaction where trust is implemented through distributed consensus and not centralized policy enforcement. The specific version of BC we used is HLF, which is open-source from the Linux Foundation. HLF is a permissioned, distributed ledger that works on the consensus model that is an integral component of the “trust system” in the BC. Essentially, the Fabric environment provides the “common logging” and service management components on the platform, and the containerized infrastructure allows developers to build a BC network where data is recorded on distributed ledgers where the data written can be trusted, and transactions are immutable and tamper-proof. Smart contracts can embed legal knowledge, laws, and regulations, and enforce Navy data policy. BC/HLF can also provide “provenance” of an item, such as food or a part, and trace back to the source of that part or food item in case of contamination or counterfeit/defective parts as well as other times such as blocks of data in support of AI.

BC can be used for cyber currency such as bitcoin; cyber currency is not a part of this study, and a semiprivate BC in support of data integrity needs a specific set of BC features other than Everledger or Ethereum, which uses an inefficient way to verify blocks called Proof of Work (PoW) instead of the more efficient consensus algorithm such as Proof of Stake.

With our previous research questions—Could BC simplify and enable access and identity management for the Navy supply and logistics systems in a cost-effective manner to reduce this friction? How could BC improve Navy logistics to the last tactical mile?—we demonstrated the feasibility in previous research of using IBM and Oracle versions of HLF to track assets such as food items. Tracking and moving assets could be applied to data assets and adapted for software safety use because in both cases we care about the integrity of the data generated. There have been planned pilot projects in the DoD, usually supply chain scenarios (Simerly & Keenaghan, 2019).

Although HLF is a Linux open-source project, several software companies have adapted HLF as its core BC enterprise solutions and have added additional value through add-ons, cloud support, and company expertise that goes beyond the plain vanilla HLF. This is common with open-source products as you pay for more capability and support. We compared to enterprise versions of HLF, the IBM, and the Oracle HLF BC platforms and evaluated their ability to maintain an efficient, streamlined, and accurate ledger of all shipment transactions during transportation. Additionally, the team developed a ledger serialization function in the smart contracts for synchronized connection on ships and bases to the HLF Framework. The characteristics of enterprise BCs include

- Permissioned architecture
- Highly modular
- Pluggable consensus
- Open smart contract model—flexibility to implement any desired solution model.
- Low latency of finality/confirmation
- Flexible approach to data privacy: data isolation using "channels," or share private data on a need-to-know basis using private data collections
- Multilanguage smart contract support: Go, Java, JavaScript
- Designed for continuous operations, including rolling upgrades and asymmetric version support
- Governance and versioning of smart contracts
- Flexible endorsement model for achieving consensus across required organizations
- Queryable data (key-based queries and JSON queries)
- Uses X.509 public key infrastructure (PKI), which is quite familiar to the DoD for a signed data structure that binds a public key to a person, computer, or organization. Certificates are issued by certification authorities (CAs)
- Cloud support and SaaS (Software as a Service)

Figure 1 is an example of a very simple BC ordering network. A1, A2, and A3 are different "off-chain" applications that could be on IoT devices or web browsers on computers or smartphones. These applications connect the on-chain world with the BC network/database. These client applications represent the "last mile" and could include legacy programs pre-BC. The blue-shaded background represents the BC logical infrastructure layer—not whatever physical layer infrastructures might be used, such as satellite or fiberoptics. O4 is an ordering service. Network configuration (NC4) gives administrative rights to organizations R1 and R4. At the network level, Certificate Authority CA4 (DoD certs can be used) is used to dispense identities to the administrators and network nodes of the R1 and R4 organizations. Certification authorities CA1 and CA4 provide entity validation, as well as other CAs shown in the diagram. In this example, there are two consortiums (common interest parties), represented by R1 and R4 entities who set network configuration policies, seen CC1 and CC4 which set up channels. Channels are ways to decide who gets to see what ledgers. There are three peers: P1, P2, and P3. On the left, P1 has S5, which is a smart contract that provides the rules for the ledger L1. Only those who have access to Channel 1 (C1) have access to the ledger L1. You see that if you have access to A1 or A2 you have access to C1, but the A2 application has access to both C1 and C2 and, therefore, access to ledgers L1 and L2, which is set by configuration control (CCL).

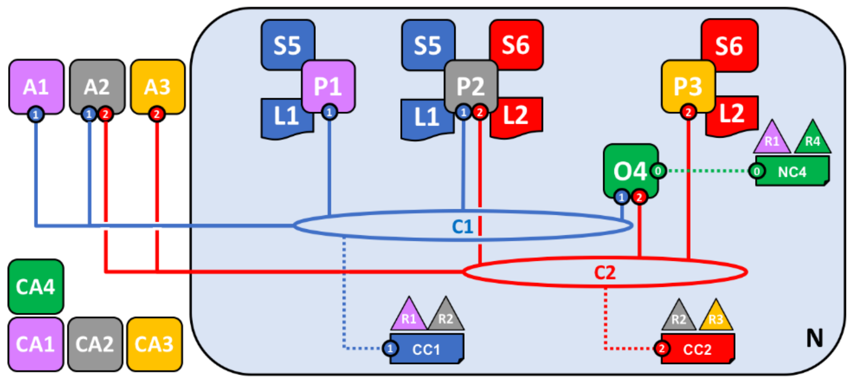


Figure 1. Generic HLF BC Network

Methodology

Our methodology involves two sets of use cases. The first set (original cases) were used in our previous research in Navy logistics, which we believe can also demonstrate BC use for system safety if modified, as both sets of use cases track assets—one tracks food items and the other tracks datasets as assets. The key for repurposing a supply chain for use in software safety support is through the addition of off-chain application programming interface (API), such as Representational State Transfer (REST, or many others), which provides an interface between the BC and the outside world and to what is called "the last mile," which in most of our use cases is a web client. In our first set of use cases (the original use cases) we built two demos (Oracle and IBM cloud versions) illustrating the Navy logistics/supply chain. We demonstrated how BC can document and authenticate transactions along the supply chain, which would be similar to a data supply chain used for data system safety. We worked with both Oracle and IBM enterprise BCs to demonstrate the first set of use cases. In a work in progress, we have an additional set of use cases (labeled new cases) specifically for use with system safety using the open-source version of HLF (https://www.hyperledger.org/).

- **Blockchain Use Case Examples for the Navy Logistics/Supply Chain**

Working with our Navy sponsor (Navy Logistics N4) we looked at three general use cases to apply BC technology using both cloud versions of IBM and Oracle BC platforms: (1) financial and inventory transaction audit trails; (2) serial number tracking, and (3) maintenance log integrity. Maintenance log integrity involves the same issues as AI dataset integrity. The three examples are

- **Original Case 1:** Financial and inventory transaction audit trials. An investigatory inventory and financial transactions via audit trails can be a costly and timely process, and the audit trails could encompass different systems throughout a vast network in such an organization as the Navy. The questions to be answered might include what, where, and who—where a distributed ledger would be able to track "what" through immutable data blocks that make up the ledger. One of the BC strengths is identity verification and management, which would be able to verify and track the "who" in any financial and inventory transactions on the BC.
- **Original Case 2:** Serial number tracking/BC tracking can also be applied to the tracking of specific items in the supply chain, such as serial numbers. Also, the tracking could include a visual identification of the item by an individual, which would automatically be identified as a trusted agent to make that verification along with the where and the when.
- **Original Case 3:** Maintenance log integrity/maintenance repairs—such as on naval aircraft, ground, or ship systems—typically generate data on various transactional databases, which in turn may be sourced to other databases or repositories such as data warehouses' Enterprise Resource Planning (ERP) systems. Our past research on aviation and ground maintenance systems databases shows that there are errors in the databases, and often information is not updated. At the tactical and operational levels, this could have an impact on the effective efforts to ensure maximum mission readiness. Smart contracts, which are integral to HLF, are code that can check, enforce, or flag bad data. Certainly, relational databases can have triggers to check for illogical data entries, but it isn't always being done, and typically several databases and sources may be involved in a maintenance information system to make such error checking costly or not practical. While some minor errors may be acceptable in transactional databases, these errors could have an impact on data analysis and ML/AI if the data in these systems are used as training datasets. BC could use smart contracts to flag errors over a diverse set of data sources and provide basic provenance.

Blockchain Use Case Examples for System Safety

In our second set of use cases, we specifically address three software system safety use cases applied to the open-source HLF:

- **New Case 1:** A researcher/data scientist needs to manage data or training sets for research or ML to process text or binaries (images, RFI signals), structured and unstructured.
- **New Case 2:** A data scientist needs to derive metrics on a dataset but is not allowed to see raw data.
- **New Case 3:** BC is used as a database for relatively small source code.

Figure 2 is a simplified HLF BC network that could support our three scenarios for software safety in the blue background square on the right (see https://www.hyperledger.org/). This is the BC. This BC is supported by a physical network that could be cloud-based and supported by the internet. The "off-chain" applications, IoT, and storage are shown outside of the square. These are applications developed in a normal way and not a new technology. The applications use standard APIs such as REST to interface between the user, databases, and the outside world to connect to the BC. They are called off-chain because while they interface with the BC, they are not part of the BC. From left to right are the identify certificates--CAs such as CA1, CA2, CA3 in our example to identify those who have access. BC is good at leveraging existing technologies, and CA is old technology using X.509 Public Key Infrastructure (PKI), which used to encrypt and sign email. A1, A2, A3, and so on are off-chain client applications that have access to various ledgers (our database) which are controlled through CC1 and CC2 (CCL), which sets up channels and their access. P1 and P2 are peer nodes that in the example host ledgers L1 and L2 for P1, and L3 for P2. Each ledger is supported by smart contracts (S5, S6, S7) that determine the business rules and logic of how the ledger is to be written and who can write on it. C1 and C2 are channels to determine what applications or entities are allowed to see what ledger, which makes Hyperledger very powerful as you can control who sees and changes what—such as Navy personnel and contractors having access to different data.

Off-chain A1 is an application that administers access to the repository and writes to the ledger, which records the metadata in each dataset and provides a digital signature/hash value. CCL provides access to Channels 1 and 2 and, as shown, access to all ledgers. For structured data in the repository (maybe more than the one shown in the diagram), AI would post/write the metadata of a dataset of interest including, if practical, all of the data fields, DTG, and record a hash value or signature. This would be entered either in L1, L2, L3, or other ledgers created. It is not practical to record/post large datasets on a BC ledger, but metadata and pointer/anchors to the data could be provided through URLs. It is possible that through the administrator interfacing with a peer node, the BC could store some small datasets through CouchDB, which would provide the current information/state of an asset such as a dataset.

New Case 1: Figure 2 shows application A2, which could be a customer/client such as a data scientist that is interested in datasets or training sets for an AI project. This customer per the diagram (set up by CC) has access to Channels 1 and 2, which means he can view Ledgers 1 through 3, which would be information about various datasets that can be accessed. In one scenario, the person using A2, the web application, for example, could search for a specific dataset or topic and then request that dataset through the application, which would check the smart contract—let's say for L2—to see if the system allows read access to the repository. Existing off-chain software would complete the task and send an anchor or link (URL) to retrieve that dataset. The customer could later check back and see if the data have changed/been tampered with, or if the data were given to another user. Also, the client would be provided the provenance and metadata and even points of contact, including subject matter experts and the owner of the data. The client can check to see if the dataset has changed and who changed it, since any changes to the repository would be recorded in the appropriate ledger as to who, when, and what. Smart contracts could also provide some prefiltering through smart contracts to reduce unintentional errors. In the past, this has been done pre-analysis but by using smart contracts this would only need to be done once and not by each researcher or customer. This AI system safety idea is similar to the IBM concept (Sarpatwar et al., 2019), where the authors sought a trusted AI environment through provenance with a BC library exposed by REST or Python APIs that provided support for "immutable recording of the AI process, querying for traceability and audit, fair value attribution, etc." We take it a further step to suggest that BC can be part of a smart repository solution that allows clients to search and find trusted datasets and safeguard them. A variation of this use case is a federated learning (FL) scenario that uses a collaborative ML technique whereby the devices collectively train and update a shared ML model while preserving their datasets. Even in a trusted military network using a private BC, some devices on the edge may prove untrustworthy, and ur Rehman et al. (2020) propose a reputation-aware FL that enables trust through BC consensus and trust algorithms through BC smart contracts.

New Case 2: A user wants to compile metrics but is not allowed access to the raw data because of security or cross-domain restrictions. Lampropoulos et al. (2019) proposed a similar scenario, where one Telco A holds private datasets and internally processes a data request by another Telco B, and Telco A only returns the results to Telco B and not the raw sensitive data. The whole process is performed with transparency, ensuring the quality of the results and the privacy of the processed data. A3 in Figure 2 is an application that only has access to Channel 2. The user then picks the dataset to use and looks at the metadata and fields; then the smart contract (S7) executes the query through A1 and post the results in the ledger L3. This use case could also be used for a cross-domain solution setting up rules when a user could have access to a different domain, the raw data, or just the results.

New Case 3: Our last scenario is the data are not stored off-chain but in the BC itself. HLF has the option of using CouchDB that can use standard JSON queries to get the "World" or current state of an asset (like a dataset). Perhaps this use case would apply to IoT devices where you want real-time data from sensors but still want to ensure software safety. The data would be immutable but replicated throughout the network.

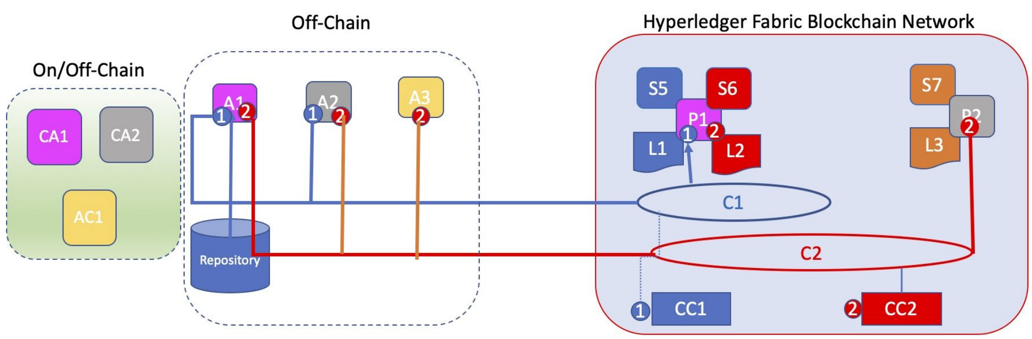


Figure 2. HLF Scenarios

Figure 3 summarizes the flow in our simple scenarios. First, the "customer"—a data scientist or developer—wants to access data such as for training sets in ML, or a developer wants access to code. The customer wants to find the right data quickly, know who owns it, and know that it can be reasonably trusted. In our example, this data resides in a repository that may include both structured data (relational databases) and semistructured and unstructured data such as in the form of .JSON files, text, or graphics. The customer starts a request for the data, and an answer comes back with the metadata, data fields, a date–time group, and a hash value of the set. This information is in a ledger in addition to an encrypted link to access the dataset. The customer can also see the complete history of changes to the data and can verify that the training set, data, or code has not been tampered with through the hash code both in the metadata and the ledger on the BC. Only those authorized can add to the chain, and it is immutable.

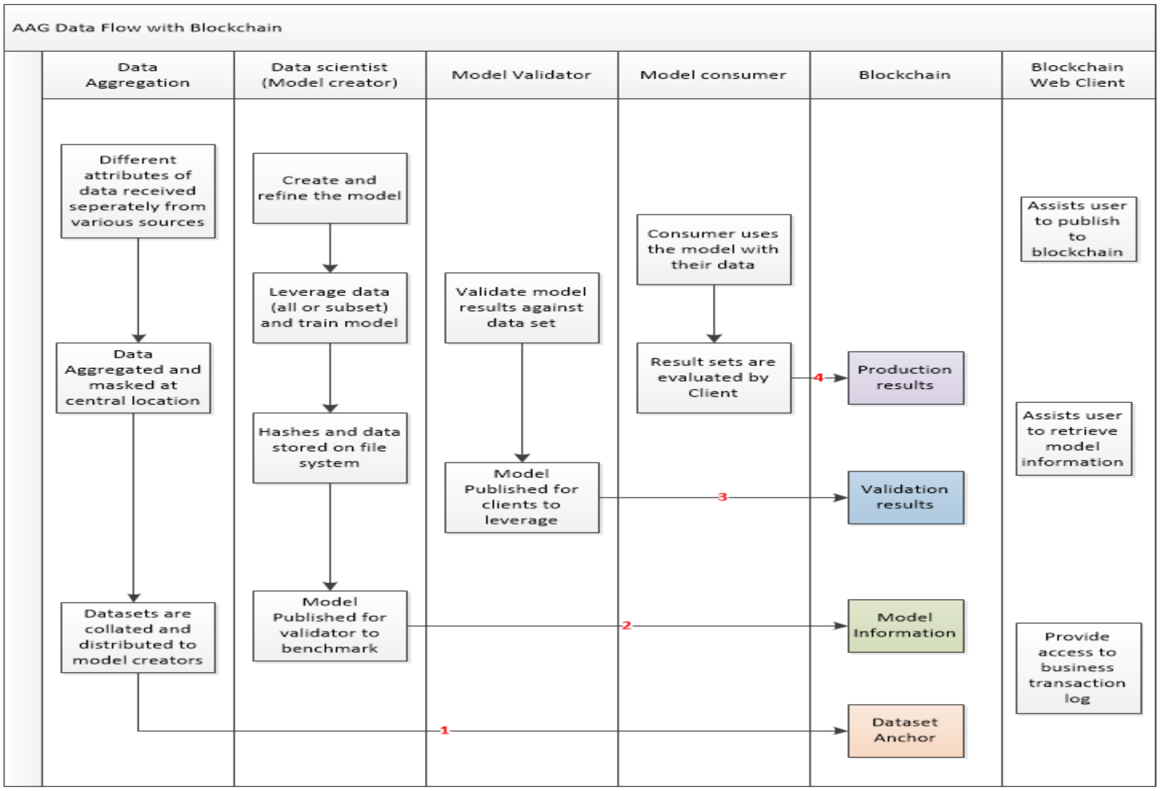


Figure 3. Summary of New Use Cases (adapted from Oracle diagram)

- **Use Cases Using Three Hyperledger Fabric Versions**

We discuss our results using the IBM, Oracle, and Linux Foundation versions of HLF and their application to system safety scenarios. Figure 4 provides a simplistic view of the system safety scenario where the data scientist is looking for training sets or related data.

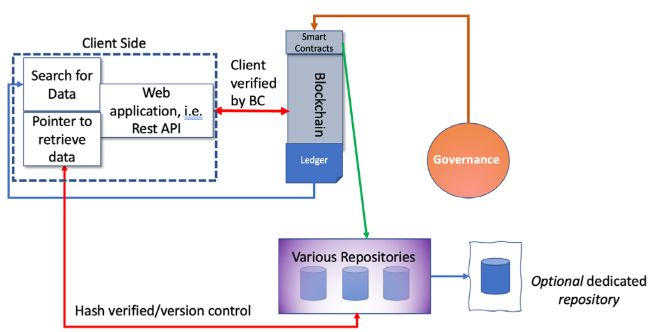
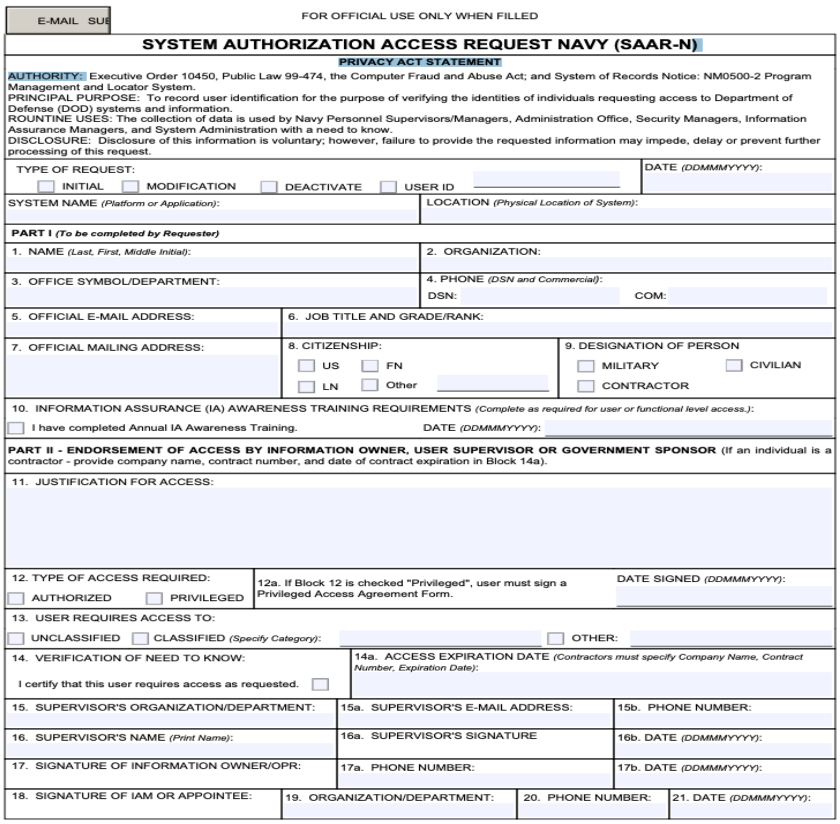


Figure 4. Data Scientist Use Case Example and Smart Repositories

The data scientist (the client) uses a web browser, enabled by Rest API or other development interfaces, and searches for a dataset or training set through the BC which, through certificates (x.509) and smart contracts, knows who the client is. Based on governance, the BC and smart contract will decide if that data scientist has the authority to retrieve the data. If so, the client will be sent a link to access the repository or even an

IoT device or a BC repository with frequently used datasets. Through x.509 certificates, which will have the identity verified by the BC, as well as the sending to the client hash value making sure the dataset hasn't been tampered with. The smart contract may do some initial cleaning up and filtering of the data. What normally takes months to get the data may only take a day and comes with assurance that the data had not been tampered with through an immutable BC. Such forms as in Figure 5 or other members of understanding could be eliminated.



E-MAIL SUE

FOR OFFICIAL USE ONLY WHEN FILLED

SYSTEM AUTHORIZATION ACCESS REQUEST NAVY (SAAR-N)

PRIVACY ACT STATEMENT

AUTHORITY: Executive Order 10450, Public Law 99-474, the Computer Fraud and Abuse Act; and System of Records Notice: NM0500-2 Program Management and Locator System.
PRINCIPAL PURPOSE: To record user identification for the purpose of verifying the identities of individuals requesting access to Department of Defense (DOD) systems and information.
ROUNTINE USES: The collection of data is used by Navy Personnel Supervisors/Managers, Administration Office, Security Managers, Information Assurance Managers, and System Administration with a need to know.
DISCLOSURE: Disclosure of this information is voluntary; however, failure to provide the requested information may impede, delay or prevent further processing of this request.

TYPE OF REQUEST: ☐ INITIAL ☐ MODIFICATION ☐ DEACTIVATE ☐ USER ID | DATE *(DDMMMYYYY)*:

SYSTEM NAME *(Platform or Application)*: | LOCATION *(Physical Location of System)*:

PART I ***(To be completed by Requester)***

1. NAME *(Last, First, Middle Initial)*: | 2. ORGANIZATION:

3. OFFICE SYMBOL/DEPARTMENT: | 4. PHONE *(DSN and Commercial)*: DSN: COM:

5. OFFICIAL E-MAIL ADDRESS: | 6. JOB TITLE AND GRADE/RANK:

7. OFFICIAL MAILING ADDRESS: | 8. CITIZENSHIP: ☐ US ☐ FN ☐ LN ☐ Other | 9. DESIGNATION OF PERSON ☐ MILITARY ☐ CIVILIAN ☐ CONTRACTOR

10. INFORMATION ASSURANCE (IA) AWARENESS TRAINING REQUIREMENTS *(Complete as required for user or functional level access.)*:
☐ I have completed Annual IA Awareness Training. DATE *(DDMMMYYYY)*:

PART II - ENDORSEMENT OF ACCESS BY INFORMATION OWNER, USER SUPERVISOR OR GOVERNMENT SPONSOR (If an individual is a contractor - provide company name, contract number, and date of contract expiration in Block 14a).

11. JUSTIFICATION FOR ACCESS:

12. TYPE OF ACCESS REQUIRED: ☐ AUTHORIZED ☐ PRIVILEGED | 12a. If Block 12 is checked "Privileged", user must sign a Privileged Access Agreement Form. | DATE SIGNED *(DDMMMYYYY)*:

13. USER REQUIRES ACCESS TO: ☐ UNCLASSIFIED ☐ CLASSIFIED *(Specify Category)*: ☐ OTHER:

14. VERIFICATION OF NEED TO KNOW: I certify that this user requires access as requested. ☐ | 14a. ACCESS EXPIRATION DATE *(Contractors must specify Company Name, Contract Number, Expiration Date)*:

15. SUPERVISOR'S ORGANIZATION/DEPARTMENT: | 15a. SUPERVISOR'S E-MAIL ADDRESS: | 15b. PHONE NUMBER:

16. SUPERVISOR'S NAME *(Print Name)*: | 16a. SUPERVISOR'S SIGNATURE | 16b. DATE *(DDMMMYYYY)*:

17. SIGNATURE OF INFORMATION OWNER/OPR: | 17a. PHONE NUMBER: | 17b. DATE *(DDMMMYYYY)*:

18. SIGNATURE OF IAM OR APPOINTEE: | 19. ORGANIZATION/DEPARTMENT: | 20. PHONE NUMBER: | 21. DATE *(DDMMMYYYY)*:

Figure 5. SAAR-N Form

Successful Applications of Blockchain for Naval Supply Chain Tracking

As discussed, our previous research investigated how BC could simplify and enable access and identity management for the Navy supply and logistics systems in a cost-effective manner to reduce administrative friction and how BC could improve Navy logistics to the last tactical mile. In our scenario, the first destination transportation (FDT) refers to the movement and cost of moving shipments from free on board (FOB) points of origin to the location at which the shipment is first received for use or storage. As naval regulations apply, the first checkpoint of where a shipment is received, whether within the United States (CONUS) or outside (OCONUS), begins with a supplier outside of the DoD supply system or industrial activity that creates the shipment. The labor and transportation charges, including freight drayage, cartage, port handling, and other in-transit costs, are processed at the FDT. Freight cartage refers to any inland transit of cargo between locations, which serve as the "checkpoints" in the BC network. When a location is assigned responsibility for "cartage of consignments" to land-based activities, ships, or other transport units, the charges of transportation are given to the location of assigned responsibility, which acts as a peer node checkpoint in the network. At this point, the initial entry in the ledger may be created and committed by the peer node belonging to the FDT and the orderers. It is important to note that FDT does not only include shipments of equipment but also the initial transportation of Navy-owned materials that are provided to a contractor for research. This indicates that the charges of a shipment from a contractor's facility to its final destination point are paid by the government. However, to maintain the legitimacy of a decentralized ledger in this research study, the network for which the ledger is maintained consists of only contractors, supply facilities, and the final base destinations. Essentially, tracking responsibility is passed down from supplier to checkpoint. The checkpoint managers responsible for the charges in a shipment delivery may create and commit the transaction over the BC network, and the next checkpoint manager may agree or disagree about the condition and extraneous details of the shipment that the previous manager signed. Currently, the DON uses service-wide transport (SWT) as a clearinghouse, which is a centralized operations and maintenance manager created to provide transportation funds for naval shipments and mail. Since naval cargo and the movement of mail to bases is not a responsibility of a destination location, the SWT was created to pay for the movement of material, such as aircraft engines, mission module packages, catapult and arresting gear, propellers, shafts, civil engineering support equipment, safety equipment, drones, overseas mail, and Navy Exchange Service Command (NEXCOM) merchandise shipped from within the United States to international locations.

For disconnected operations, to maintain an accurate ledger with the consensus algorithm, the peer nodes must be connected to the Fabric environment unless the peer node decides to save the ledger as a .JSON file and re-upload the ledger as a .CSV file once back online. The ledger is automatically updated after the node reconnects following disconnections due to shipboard communications. The Fabric environment will make BC technology a more viable option for all naval transportation activities.

The Navy requires a multifunctional and secure platform that enables personnel to track multiple shipments from production facilities to bases and a secure ledger of inventory that can only be modified with either an undisputed consensus or access to the smart contract. Once a peer node administrator or user in the network has access to their smart contract, they can modify the transaction protocol that occurs on transactions in the network. However, the network will not instantiate a new version until there is an agreement with the channel creator or the majority of the channel members.

In this simplified logistics BC network, the smart contract contains six methods that carry out the protocol for each transaction on the ledger: foodAssetExists, createFoodAsset, readFoodAsset, updateFoodAsset, trackFoodAsset, and deleteFoodAsset. The method of using names indicates that each shipment is checked to verify if it already exists at a location denoted by a string. After checking for duplication, the asset is created in the ledger using a key-value pair, such as "001: a shipment of supplies." Once the asset is created, it is always a good practice to read the asset's details into the ledger so that users further down the network have a detailed understanding of what a package is supposed to contain. Also, if a shipment is changed—say, a package is redirected to a base that requires supplies urgently—the shipment's location is updated within the ledger and deleted once the shipment arrives.

A multifunctional and secure platform that enables personnel to track multiple shipments from production facilities to bases or ships in transactions involving money, items, material, and history should be trusted, transparent, and traceable back to the origin of the item. These transactions involving information, money, or physical items such as food or parts usually involve the enforcement of policy, technical, or legal requirements that require the enforcement of business rules. BC can maintain a secure ledger of inventory (or transactions involving data or information) that can only be modified with either an undisputed consensus or access to the smart contract, which can enforce business rules and flag “violations.” Once a peer node administrator or user in the network has access to their smart contract, they can modify the transaction protocol that occurs on transactions in the network. However, the network will not instantiate a new version until there’s an agreement with the channel creator or the majority of the channel members.

Based on the above process, we showed how a food or item tracking scenario would work using both IBM and Oracle cloud versions of HLF (see Figure 6). In these BC networks we set up, the smart contracts contain six methods that carry out the protocol for each transaction on the ledger: foodAssetExists, createFoodAsset, readFoodAsset, updateFoodAsset, trackFoodAsset, and deleteFoodAsset. In our food/item tracking scenario, the method of using names indicates that each shipment is checked to verify if it already exists at a location denoted by a string. After checking for duplication, the asset is created in the ledger using a key-value pair, such as “001: a shipment of supplies.” Once the asset is created, it is always a good practice to read the asset’s details into the ledger, so that users further down the network have a detailed understanding of what a package is supposed to contain. Also, if a shipment is changed, say, a package is redirected to a base that requires supplies urgently, the shipment’s location is updated within the ledger and deleted once the shipment arrives.

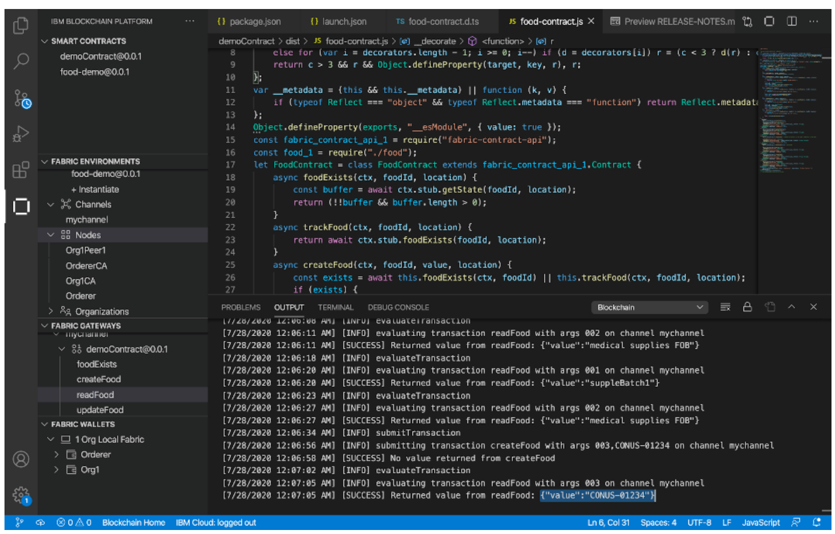


Figure 6. Sample Ledger of Shipments That Are Added and Updated (Contents/Location)

Blockchain Use Case Examples for the Navy Logistics/Supply Chain

Using IBM BC Platform™: To use the IBM BC Platform, users are required to install four vital components: (1) the Virtual Studio Code environment, (2) Node.js, (3) Docker, and (4) Kubernetes. The Virtual Studio Code environment is the offline integrated development environment (IDE), where developers create smart contracts using the open-source programming language Typescript, which was developed by Microsoft.

Smart contracts serve as the fundamental basis of all enterprise BCs because they give certified users the ability to create new transactions and assets, as well as other functions specific to a project. In this project, the team's main goal was to create a consensus network that has the power to create food shipment assets, update or delete them from the ledger when required, and track their location using the "foodId" string, which may be replaced by radio-frequency identification (RFID).

The HLF (from the Linux Foundation) is the basis of both IBM and Oracle platforms. Its components are created in a Kubernetes cluster usually within the IBM Cloud. A Kubernetes cluster contains a set of working machines (nodes) that run containerized applications. The nodes within the cluster host the components of the application workload. Within the cluster, the control plane manages the nodes and workloads that run across multiple machines, as shown in Figure 7:

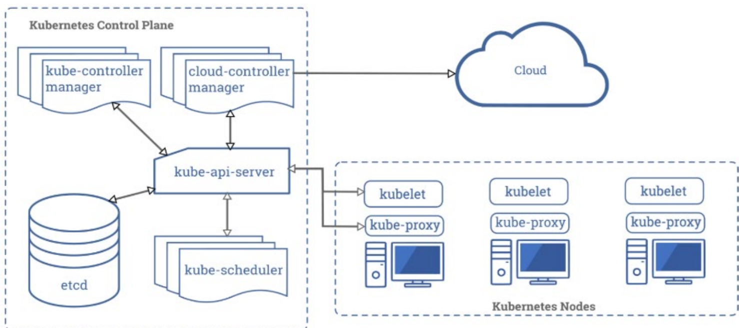


Figure 7. Visual Representation of the Interaction Between Kubernetes and Cloud

Figure 8 illustrates the ordering service. When the Fabric environment is running, you can create the ordering service. The ordering service is a group of orderers that accepts approved transactions endorsed by the peer nodes based on the smart contracts and organizes the transactions in the appropriate order in the ledger blocks based on the consensus algorithm.

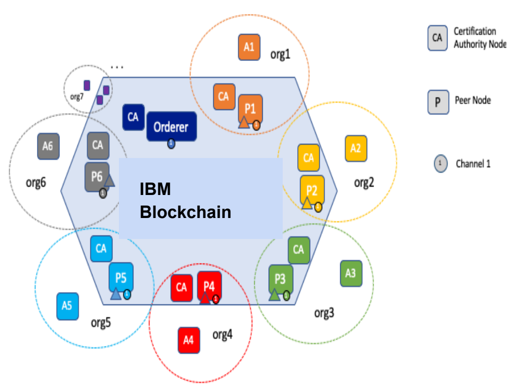


Figure 8. Visual Representation of the Integration of Security and Nodes in a BC Channel

The peer nodes host ledgers and smart contracts—the backbone of the BC network. The smart contract—the transaction protocol—automatically executes, controls, and documents transactions or events occurring on the network.

Like all BC frameworks, the network's integrity is upheld by the consensus algorithm. Each node in the network reviews the entire BC and checks that all previous blocks are valid so that a new transaction may be initiated into the network. However, alternatively, in a permissionless public BC, the consensus algorithm is replaced by the PoW, which creates a hash system of all of the transactions.

In a PoW system, miners constantly attempt to solve the algorithm so that they may mine new blocks and be the first to extend their BC. HLF doesn't use the wasteful PoW but uses a system closer to the "Proof of Stake" as a consensus mechanism. Essentially, decisions are authorized by users who are permitted to join the system and specific channel, as not everyone can join the network. Unlike PoW, computational power is not required, since there are no puzzles needed to obtain "currency." In a "Proof of Stake" system, "validators" are discouraged from creating faulty empty blocks because they have the motivation to incorporate a maximum number of transactions for gains.

To ensure security, the hash must be solved by all the peer nodes in the network so that the new transactions may be approved for the network. While this alternate approach is viable, it is also time-consuming because ensuring that the ledger is tamper-free requires each ledger copy in the nodes to be changed and hashes to be solved.

Developers should install Node.js and Docker unless the developer exports both items into a .JSON file and re-uploads both the files onto the peer nodes as a .CSV file. Docker serves as an OS-level platform to package containers and bundled software, libraries, and configuration files.

Figure 9 shows that using well-defined channels within the software, these containers communicate with each other to allow the user to connect to the Fabric environment and add to or change the ledger. Finally, the Kubernetes system, which was designed by Google and maintained by the Cloud Native Computing Foundation, is the main system that allows the IBM BC Platform to package, install, deploy, and manage the multiple peer nodes in the platform.

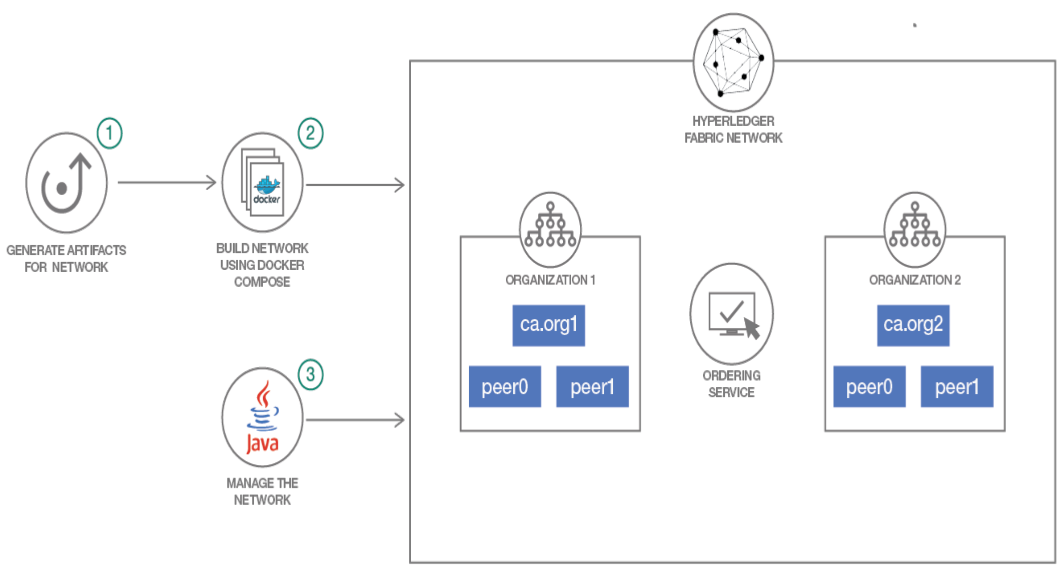


Figure 9. Visual Representation of the Interaction Between the External Software and the HLF Environment

Figure 10 provides an overview of how you would manage the offchain (UI[2]) and the actual BC network consisting of three Fabric components: CA(4), the peer nodes, and the ordering service. Compare Figure 8 to Figures 1 and 2.

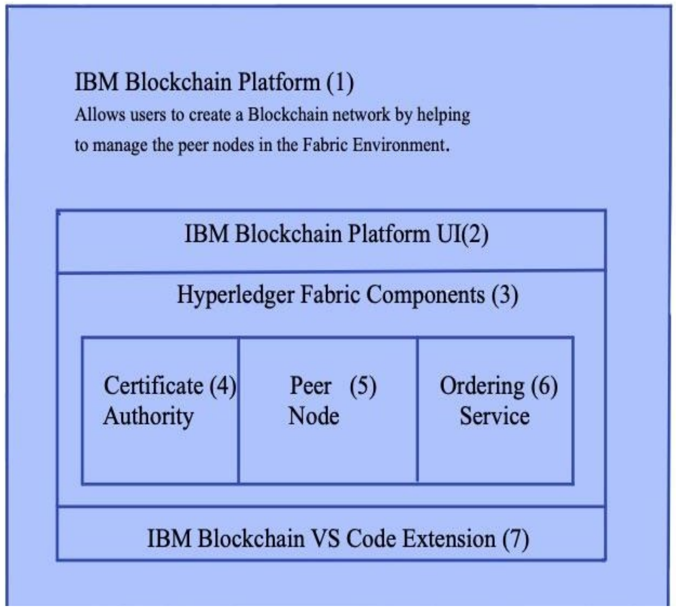


Figure 10: High-Level Representation of IBM Blockchain Platform Architecture

Using the Oracle Blockchain Platform: On both Oracle and IBM platforms, we were able to set up a BC network with peer nodes (stakeholders) with smart contracts that set up the rules for transferring and tracking assets such as food items discussed previously. More work needs to be done on enhancing the network to accurately represent this aspect of the supply chain.

The team set up the network using an Oracle cloud with four peer nodes set up over a single channel and used Oracle Identify Management for role-based access. Separate roles are required for adding users to a role with BC provisioning entitlement, which requires tenancy admin. Additionally, the cloud platform was used instead of the software package due to the amount of storage memory required to host the software appliance VM packages on a local computer. However, the fundamental concepts of using an HLF environment and consensus algorithm remained the same for both platforms to build a BC network.

Oracle Blockchain Platform also provides wizards to simplify joining multiple instances to the network, creating new channels, and deploying chaincodes. Implementation of smart contracts is through Typescript (see Figure 11). These and other DevOps functions are also available via extensive REST APIs for off-chain applications to interface the BC network.

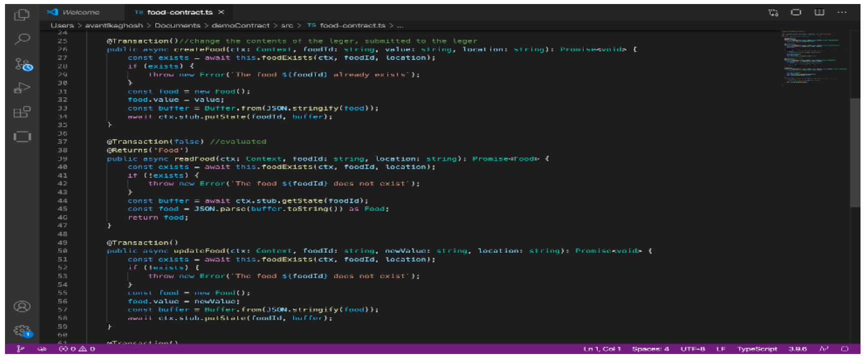


Figure 11. Sample Smart Contract for Tracking Food Shipments (Language: Typescript).

Oracle offers both a managed Cloud version (Oracle Blockchain Platform) of OBP (Blockchain-as-a-Service) and a customer-managed OBP Enterprise Edition for on-premise (or 3rd party cloud) deployment, and nodes can be deployed using both for a hybrid network deployment (see Figure 12). The Cloud SaaS version was used for this project. To access this platform, users must log in with authenticated credentials in Oracle Cloud Infrastructure. Once logged in, users can provision an instance, which comes with a default channel and participant nodes, along with “orderers” that are responsible for maintaining the order of the ledger. An operations Console is provided, and users are not required to download any external software to work with the platform, other than an Integrated Development Environment (i.e., Visual Studio Code) to develop the chaincode and the REST API Testing tool, such as Postman and/or HLF Software Development Kit, which is downloadable from the OBP Console under the Developer Tools tab.

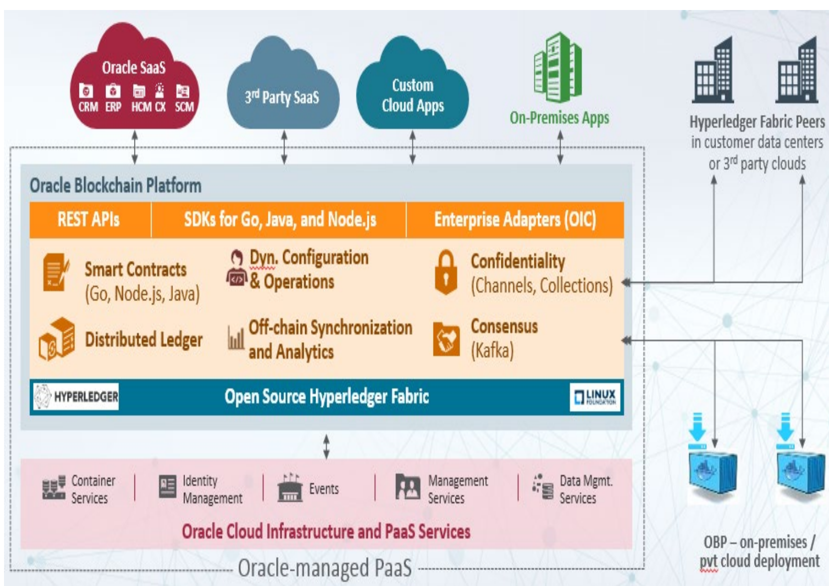


Figure 10. Oracle Blockchain Platform Cloud Service Architecture

The Oracle Blockchain Platform (see Figure 10) comes with an API Gateway that supports REST API so that developers can invoke a transaction, invoke a query, subscribe events with a registered callback, and view the status of a transaction within the ledger as well as a set of DevOps REST APIs for administration, configuration, and monitoring tasks.[19]

Current Use Cases Using Linux version of HLF for System Safety

While both IBM and Oracle are HLF-based, their complete solutions use their respective cloud services are enhanced by their specific products. For our new set of system safety use cases (a work in progress), we installed HLF on a Naval Postgraduate School virtual Red Hat server and installed HLF from the Linux open-source foundation, which provides all the needed images and tools to set up a BC. Unix tools include the Git client, CURL, and Docker with Docker-Compose without Kubernetes, which are key components to build the BC network in a rapid manner. This model suits the researcher who wants to study and test out the concepts before moving to production, at which point a vendor-supported option can better address the challenges. Typical enterprise BC platforms provide dashboards for BC management such as the status and health of the HLF network. In the case of the open-source version, no such tools are provided; instead, everything is done via command; thus one has to have a good idea of Unix command line tools and scripting languages like BASH. Both IBM and Oracle allow you to use an IDE to build the applications. All three platforms offer interfaces via APIs to programming languages like JavaScript, Java, Microsoft Visual Studio, and others. For most production instances, we think a cloud-based BC is usually the right way to go for maintainability, support, ease of use, and security.

For the Linux Foundation version of HLF, the complete install includes commands to set up an HLF network, issue certificates, set up the ledger, create channels, install chaincode, and more. A sample BASH script is provided that goes over all these steps and can be customized for new projects such as for our three system safety use cases. The Docker container-based platform allows one to have several HLF projects to coexist. The test network is shown in Figure 13 with two organizations, R1 and R2. Organization R0 owns the ordering service (O) of channel C1. A copy of the ledger L1 is on all nodes. The root CA issues the certificates CA0, CA1, and CA2 for the three organizations.

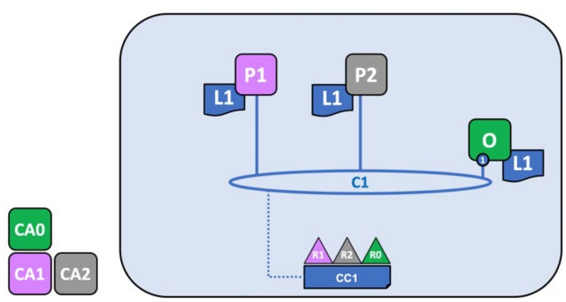


Figure 13. Initial Blockchain Network on Open-Source HLF

[19] The team was given access to the Oracle Cloud Platform thanks to the NPS liaison relationship with the Oracle Blockchain team.

Use Case 1: First a channel is created, and member organizations are added into the channel. The ledger contains the needed URLs to access the different binaries. Using APIs, depending on the requester, the chaincode will craft a unique response to be sent back. When the response is received, the URL and text are preprocessed; this happens in a middleware server outside the HLF, thus a custom webpage is created and served to the end user. This webpage has links to authenticated repositories (database back end). Once authenticated access is granted to the data scientist (data and training sets used in research and stored in a database), an encrypted anchor or URL may be sent to the data scientist to download the dataset.

Use Case 2: When a member organization needs to see a part of the ledger, then channels have to be created. Membership to a channel is restricted to a subgroup of the organizations. Using chaincode, the metrics part of the ledger is provided to members of the metrics channel. There exists another channel where members can see the raw data using queries—again dictated by the chaincode. Membership to the two channels is a different set of organizations. Similar development methods used in Use Case 1 apply to Use Case 2.

Use Case 3: In this scenario again, a channel is created for a certain member organization (not all). These members will be able to access the data in CouchDB via API queries. Chaincode will decide which source code (stored in the CouchDB) is provided as a returned result of the query. The database is replicated on every node, which might be an advantage on the edge and an I/O to IoT devices on the edge. Specific use cases for this capability haven't been developed.

Summary and Conclusions

From our work, the following lessons learned can be applied to protecting datasets such as training sets for AI:

1. Various versions of HLF will work adequately, but due to complexity, we recommend not using open-source but software vendors such as Oracle, IBM, Microsoft Azure BC, and others. BC is not a DoD core competency; therefore, contractor support is needed.
2. HLF or other BCs alone are not the entire solution, since BC is an enabling or general purpose technology (GPT)—so in itself, it is not a solution. You must use a BC protocol within an integrated network infrastructure that also provides for the last mile to bring the data to the user, and this is through APIs. We recommend, ceteris paribus*,* that you consider using the same company that runs your relational databases or ERP, as your team will be familiar with that architecture.

We used a qualitative methodology that included three general logistic use cases: (1) financial and inventory transaction audit trails, (2) serial number tracking, and (3) maintenance log integrity. These were used in consultation with the topic sponsor. We created simple scenarios where items were tracked through a BC network, and smart contracts would check for certain conditions that would simulate quality control and tracking. We selected two enterprise HLF platforms, Oracle and IBM, and evaluated them in terms of functionality, development ease, and security.

We found that both the IBM and Oracle BC platforms may be used to create a secure network of peer nodes and a consensus for the legitimacy of the shipment ledger, which can only be modified using smart contracts. A special concern with Navy logistics is the possibility of unreliable networks, especially from shore to ship. The BC protocol creates a multitude of copies of the blocks (the public ledger) and if connectivity is lost, the blocks will be updated once the network node communications are reestablished. Both IBM and Oracle BC platforms were accessed through the cloud, but the option is for the Navy to put either platform on its implementation of the cloud or servers.

There were differences between IBM and Oracle implementation of HLF, such as how the whole network infrastructure was implemented, user interfaces, the developer tools and application programming interfaces provided, and how the implementation would connect to the Navy's legacy systems to reach the last mile—such as on the ship. These were real value-added capabilities since HLF alone cannot make an enterprise BC system that supports the existing logistics information system.

BC technologies offer the potential to reduce costs and logistical friction by providing a trusted ledger in support of logistic transactions and processes. Errors can be reduced through smart contracts, as demonstrated in both IBM and Oracle BC platforms. BC tracks assets, and therefore, BC can track data assets just as well as a partial solution to software safety.

Intermittent Communications

The Navy primarily operates at sea, which means the communications infrastructure supporting the BC network may not always be available or reliable, or provide bandwidth. A significant concern when implementing BC technology in cargo shipments is its dependence on a continuous connection to the Fabric environment. However, HLF is a robust distributed database (ledger) that has many copies of itself.

The BC platform does require you to be connected to the Fabric environment at all times or to consistently re-upload the ledger to the peer nodes to have a constant accurate ledger. BC provides an update method that if a node is offline, it will have an update of its BC once reliable network is reestablished.

The Issue of Governance

Figure 4 showed a simple notational circle labeled "Governance," but this issue is far from simple and is the key to any implementation of BC in support of data. While a detailed discussion of governance is beyond the scope of this paper, Gaur and Gaur (2018) presented a variety of frameworks, some of which would apply to permissioned BC networks. Previous discussions of BC governance tended to be about public BCs supporting cyber currencies. They noted that while BC is about decentralization, there will have to be some aspects of centralized governance—especially ones involving policy and legal aspects in the storage and use of data. For example, governance could include safeguards through smart contracts that could flag possible AI bias, especially ones used for human resources. Governance can consist of different layers, and one classification recognizes the different levels the data serves and classified as strategic, operational, and tactical governance. Since BC is decentralized by nature, the governance should be at the lowest level if diversity and flexibility are important. Ziolkowski et al. (2020) looked at governance that includes demand and data management, system architecture design and development, membership, and data ownership. Each one represents a possible off- or on-chain solution that involves technical and policy considerations—both of which may include smart contracts as solutions and resources (Feagan, 2020). System architecture design and development are not trivial tasks and are based to a great extent on governance and policies. To resolve this, IT network engineers must work as part of a consortium to determine the appropriate way to expose their peers to other organizations to receive transaction endorsement proposal/simulation requests while minimizing an attacker's ability to gain access to sensitive information stored in the simulating peer's database (Feagan, 2020).

The level of rigor is ultimately determined by the policies derived by governance. Data accessibility is also a key, so governance should have policies that allow scientists working for the DoD to find data not through randomness but structure, without undue delay, and data that complies with software safety. BC supporting “smart repositories” may facilitate this goal. The default should be to allow our data scientists and analysts timely access to data unless there is a good reason not to. Our adversaries work for AI superiority, and withholding data from their researchers is something they avoid. We refer to unclassified and non-PII/medical data.

Findings

We demonstrated through IBM and Oracle examples that HLF could meet logistics/audit and security requirements through smart contracts and the inherent trust systems with embedded certificates. Data entry errors could be reduced through smart contracts, which is an inherent feature of HLF. We believe a consortium BC through HLF would be a way to go to be able to share information (through the ledger) with suppliers and other third parties but also have the capability not to share when appropriate. BC could add the capability for secure transactions through certificates and the immutability of the transactions on the BC. The additional capability of BC on Navy logistics and supply would be able to catch some data entry errors, to trace back to the source, and basically to better know the what, the who (verified), and the where of various transactions generated by the supply chain.

We found that both the IBM and Oracle BC platforms may be used to create a secure network of peer nodes or naval hotspots that can generate a consensus for the legitimacy of the shipment ledger, which can only be modified using smart contracts. Since a key component of both platforms is maintaining accuracy and security of the ledger, all users must consistently export and import the smart contracts and ledgers onto their respective peer nodes every time an update is made on the ledger or if the transaction protocol on the smart contract is changed. A special concern with Navy logistics is the possibility of unreliable networks, especially from shore to ship. The BC protocol creates a multitude of copies of the blocks (the public ledger), and if connectivity is lost, the blocks will be updated once the network node communications are reestablished. Both IBM and Oracle BC platforms were accessed through the cloud, but the option is for the Navy to put either platform on its implementation of the cloud or on servers.

There were differences between IBM and Oracle implementation of HLF—such as how the whole network infrastructure was implemented, user interfaces, the developer tools and application programming interfaces provided, and how the implementation would connect to the Navy’s legacy systems to reach the last mile, such as on the ship. These were real value-added capabilities, since HLF alone cannot make an enterprise BC system that supports the existing logistics information system.

We found a “consortium BC” with a BC consensus network to be the best fit for the use cases. A consortium allows both private and public users to use the BC while control is maintained by the private users (the Navy) through a consensus network, which means by the consensus of trusted Navy entities. This is contrasted by PoW BC networks used in cyber currency, which are inefficient and not appropriate for a government entity. BC technology has the potential for revolutionizing the logistics process by ensuring the quality and trustworthiness of logistical generated data as well as providing provenance of parts and food, but it is new and risky.

The team also compared the IBM and Oracle BC platforms on efficiency and maintainability of a ledger of shipments and discovered that it was easier to use the IBM platform to create and export smart contracts and ledger; however, in September 2021, Oracle will provide similar capabilities for developing and deploying smart contracts. The IBM platform required users to develop their smart contract on the Visual Studio Code environment, export the contract as a .JSON file, log in to the online BC network, and import the contract and ledger as a .CSV file using a converter.

The Oracle Blockchain Platform, on the other hand, allowed users greater flexibility to join ledgers more cohesively. The Oracle platform allowed users to log in to the Oracle cloud after they were approved by an administrator and used simple software like IDE and the Software Development Kit. Furthermore, the Oracle Blockchain Platform employed chaincode as a smart contract for transactional protocols in the network. A chaincode is written in either Java, Node.js, or Go and packaged into a ZIP file, which can be installed on the network. This is similar to how smart contracts are exported as .JSON files and uploaded on the IBM network as .CSV files. More specifically, chaincodes outline the structure of the ledger, initialize it, create updates (such as reading or updating entries), and respond to queries.

Should HLF be used for software safety for ML and AI development? BC is general purpose technology (GPT) like the Internet, so BC isn't a solution in and of itself, but it acts as an enabler that provides a trusted, distributed ledger that could be used for smart repositories and software safety. If other technologies are better, then why aren't they commonplace? BC isn't *the* solution but, along with off-chain technology, may be a technology that enhances existing business processes.

References

Athalye, A., Engstrom, L., Ilyas, A., & Kwok, K. (2017). *Synthesizing robust adversarial examples*.

Chauhan, G. (2018, September 14). *AI safety*. Towards Data Science. https://towardsdatascience.com/ai-safety-9aeb9ca42907#:~:text=AI%20Safety%20is%20collective%20termed,of%20real%2Dworld%20AI%20systems

Defense Standardization Program Office. (2012). *System safety* (MIL-STD 882E). Pentagon.

Eden, A. H., Moor, J. H., Soraker, J. H., & Steinhart, E. (2013). *Singularity hypotheses: A scientific and philosophical assessment*. Springer.

Everitt, T. (2018). *Towards safe artificial general intelligence* [Doctoral thesis, Australian National University]. https://www.tomeveritt.se/papers/2018-thesis.pdf

Feagan, L. (2020, April). *Hyperledger fabric myths and reality*. Object Computing. https://objectcomputing.com/resources/publications/sett/april-2020-hyperledger-fabric-myths-and-reality

Gaur, N., & Gaur, N. (2018). *Hands-on blockchain with Hyperledger: Building decentralized applications with Hyperledger fabric and composer* (1st ed.). Packt Publishing.

Jia, X., Hu, N., Yin, S., Zhao, Y., Zhang, C., & Cheng, X. (2020). A2 chain: A blockchain-based decentralized authentication scheme for 5G-enabled IoT. *Mobile Information Systems*, *2020*. https://doi.org/10.1155/2020/8889192

Lampropoulos, K., Georgakakos, G., & Ioannidis, S. (2019). Using blockchains to enable big data analysis of private information. *IEEE 24th International Workshop on Computer-Aided Modeling and Design of Communication Links and Networks*, *2019*, 1–6, https://doi.org/10.1109/CAMAD.2019.8858468

Roland, H., & Moriarty, B. (1990). *System safety engineering and management* (2nd ed.). Wiley.

Sarpatwar, K., Vaculin, R., Min, H., Su, G., Heath, T., Ganapavarapu, G., & Dillenberger, D. (2019). Towards enabling trusted artificial intelligence via blockchain. In S. Calo, B. Seraphin, & D. Verma (Eds.), *Policy-based autonomic data governance* (1st ed., pp. 137–153). Springer International Publishing. https://doi.org/10.1007/978-3-030-17277-0_8

Simerly, M. T., & Keenaghan, D. J. (2019). Blockchain for military logistics. *Army Sustainment*, *51*(4), 48–49.

ur Rehman, M., Salah, K., Damiani, E., & Svetinovic, D. (2020). Towards blockchain-based reputation-aware federated learning. *IEEE Conference on Computer Communications Workshops*, 183–188. https://doi.org/10.1109/INFOCOMWKSHPS50562.2020.9163027

Ziolkowski, R., Miscione, G., & Schwabe, G. (2020). Decision problems in blockchain governance: Old wine in new bottles or walking in someone else's shoes? *Journal of Management Information Systems*, *37*(2), 316–348. https://doi.org/10.1080/07421222.2020.1759974

Product Supportability Through Lifecycle Modeling and Simulation

Magnus Andersson is a director at Systecon. Andersson started his career at the Swedish utility Vattenfall, where he worked on various wind power research and development topics. In 2011, he joined Systecon, where he has worked as a consultant within the fields of life asset cycle management, O&M, and lifecycle economics. Today he splits his time between Systecon's Consultant Department and business development. Magnus holds an MSc in Energy Systems Engineering from Uppsala University in Stockholm, Sweden.

Justin Woulfe is a principal and cofounder of Systecon North America, with expertise in systems, logistics, and cost optimization. For the past decade, he has focused on balancing cost and capability within the aerospace and defense industry. Justin has a BS in Electrical Engineering from Virginia Military Institute, an MS in Engineering Management from Drexel University, and an MS in Supply Chain Management from Syracuse University through the DoD LOGTECH program. His research and analysis has resulted in billions of dollars in savings and increases in readiness on large, complex acquisition and sustainment programs. His work in model-based systems engineering, optimization, and readiness analysis is widely published and taught in both university and DoD programs.

Abstract

Current changes in Department of Defense (DoD) budgeting processes and in the constraints on available funding have resulted in inadequate support for the warfighter's needs. The decision environment evolves into a key question impacting warfighter capabilities: *How should the funding be distributed to achieve the optimal balance between readiness, performance, and cost?*

This paper outlines the fundamentals of successful Product Lifecycle Management (PLM), a method to monitor systems towards fulfilling the operational needs at the lowest possible total ownership cost (TOC). The paper discusses critical decision points in different phases of the system's lifecycle and suggests an approach to use modeling and simulation tools to answer key questions and provide the required decision support.

Introduction

Today's constraints on funding the acquisition of systems and their associated lifecycle support costs require a rigorous and consistent analytical process to ensure the systems and supporting processes provide capabilities that are worth the expenditures. These funding constraints come at a time when many of our systems are very mature and "war-weary." This fact exacerbates an already complex decision environment. The decision environment evolves into a key question impacting our warfighter capabilities: *How should the funding be distributed to achieve the optimal balance between readiness, performance, and cost?*

Key Points: Recent Department of Defense (DoD) policies and guidance make significant strides towards identifying and promoting broad-based Product Lifecycle Management (PLM) strategies to design, field, and sustain more affordable and ready warfighting capabilities. The practical implementation and institutionalization of these strategies, however, has not kept pace with available analysis capabilities. The most significant barriers to attaining the desired implementation and institutionalization of these strategies are

- The deep-rooted divisions between systems engineering, lifecycle product support, and programmatic and cost functions;
- Divergence between policy requirements and organizational business

strategies/investments in enterprise-wide lifecycle process and knowledge management;

- Sustainment data from the many “stovepiped” information sources within each of the services/organizations that needs to be extracted, transformed, and loaded into a common information analytics data warehouse with other PLM data sources and capabilities;
- The need for developing and employing a comprehensive “Big Data” strategy to use effectively the large volume of sustainment data and resolve the complexities involved with effective integration of this data;
- A scarcity of competency and proficiency in structured analytics, business intelligence, lifecycle product support package design, PLM technologies, and reliability, availability, maintainability, and cost (RAM-C) trade studies.

In addition, the complexity of the decision environment is increased by:

- The potential cost growth of continuing to operate systems that have been significantly degraded by war fatigue or have had their original operational life extended many times;
- The decreased budgets and increased costs to maintain systems ultimately leads to a realization that spreading budget cuts across every program is probably no longer a viable solution;
- Early decisions regarding concepts, requirements, and choice of supplier will impact the total ownership cost (TOC) more than anything.

This paper outlines the fundamentals of successful PLM, a method to monitor systems towards fulfilling the operational needs at the lowest possible TOC. The paper discusses critical decision points in different phases of the systems lifecycle and suggests an approach to use modeling and simulation tools to answer key questions and provide the required decision support.

Advances in lifecycle modeling and simulation technologies have provided a significant opportunity for the DoD to address these complex issues. Lifecycle management (LCM) simulation tools and techniques have been developed to automate and modernize the collection, aggregation, measurement, and visualization of system and platform performance from the in-service engineering agent’s (ISEA’s) perspective, with potential for providing valuable information to the service components and to the acquisition community. These new technologies assist with the capture, retention, translation, and aggregation of numerous forms of structured data. There are numerous databases being used that perform just as many tasks, and the primary purpose is to aggregate their data. In some cases, tools can translate database data elements so that they are compatible with other databases’ data elements. Data translation then paves the way for data integration. Data aggregation and integration reveal data relationships not otherwise known to program managers and subject matter experts.

Additionally, early decisions regarding concepts, requirements, and choice of supplier will impact the TOC more than anything else. Unfortunately, these decisions need to be made without exact knowledge about all influencing parameters. To make these kinds of decisions under major uncertainties calls for an efficient and systematic decision-making process, using modeling and simulation tools to analyze the consequences of the decisions. Figure 1 shows the basic data modeling and analysis process.

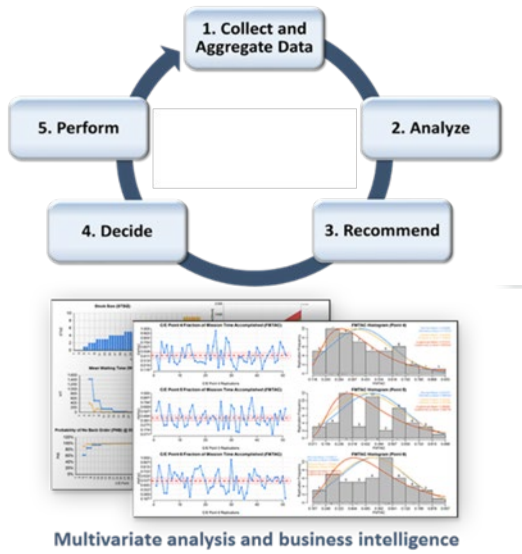


Figure 11 Data Modeling and Analysis Process

Supportability in the Design and Acquisition Phases

From a customer and owner perspective, any system typically goes through several phases—starting with concept definition, specification, and acquisition; continuing with system design and development, production, entry to service, operations, and maintenance; and finally disposal. All through the lifecycle, a program or product manager needs to make a lot of decisions regarding the technical system, its operations and maintenance, and the logistic support. The important point here is that consequences of decisions made will not come in daylight until many years after a decision is made. That is the background to the classic characteristics of a lifecycle cost curve (LCC), shown in Figure 2.

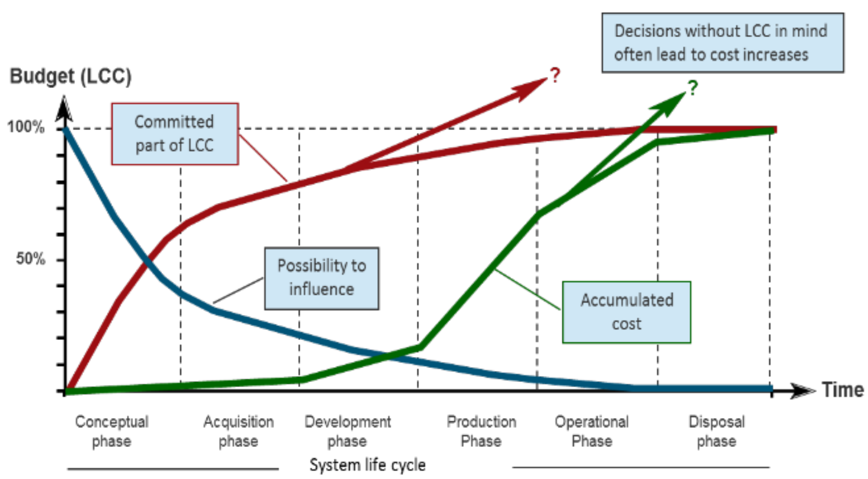


Figure 12 Characteristics of a Lifecycle Cost Curve (LCC; Woulfe, n.d.)

The green curve shows the actual expenditures (both CapEx and OpEx) for a system throughout its lifecycle. The red curve, however, describes when stakeholder(s') decisions make them commit to the costs, which usually occur long before the actual expenditures.

Thus, their possibility to influence the TOC will decrease during the system´s lifecycle according to the blue curve.

It is also important to point out that if decisions are made in later phases without analyzing the potential consequences on operational performance and lifecycle cost, there is a great risk that you commit to future cost increase.

Cost–Benefit Assessment During Product Lifecycle

When should replacement of fleet of systems take place? What requirements should be put on a new system? Which systems should be purchased? What investments in logistic support, spares, and other resources should be chosen? What improvements are most cost-effective to make to enhance my operations?

These are some examples of major questions for a system manager. They all require an understanding of what the consequences of the choices at hand will be on operational performance and total cost of ownership. The questions are complicated to answer since there are so many influencing parameters. Figure 3. illustrates the decision problem and the three main influencing domains.

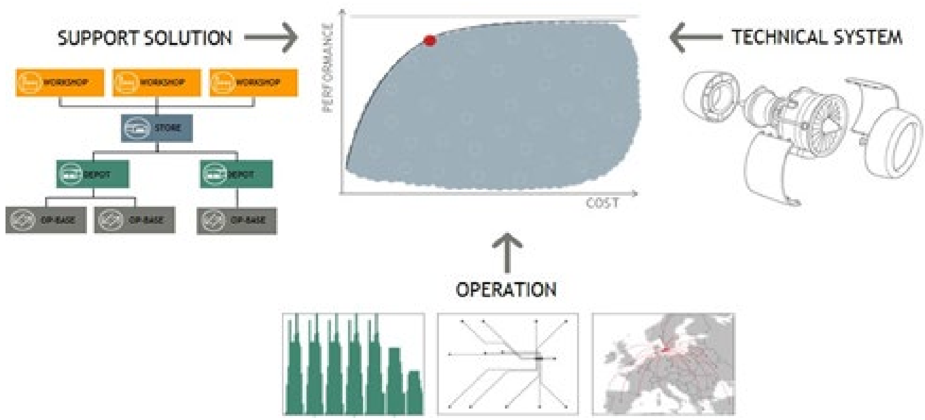


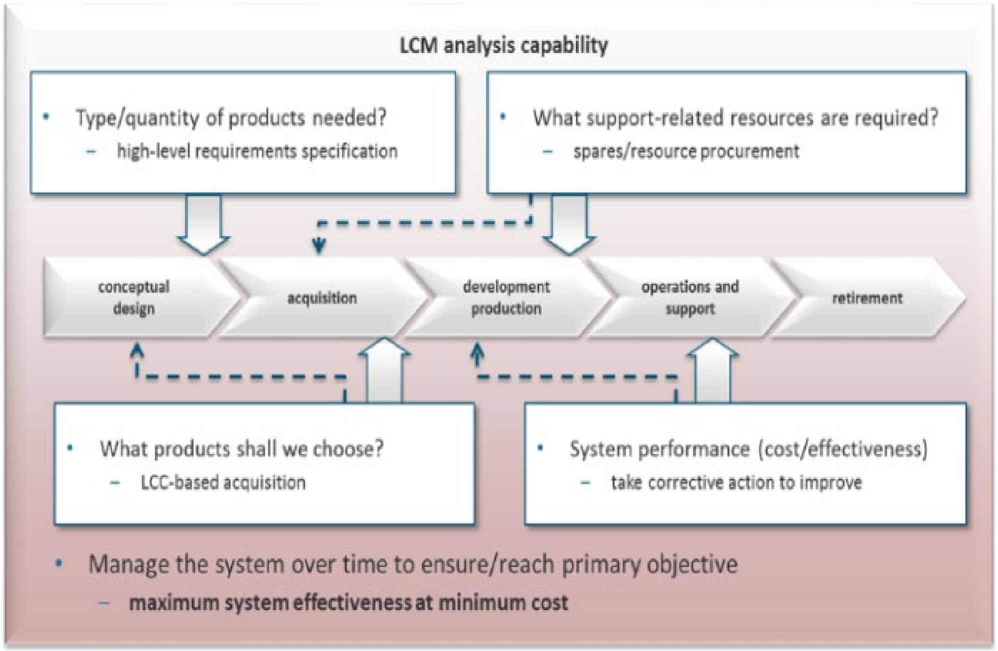
Figure 13 The Dimensions That Influence the Relationship Between Cost and Availability (Woulfe, n.d.)

To be able to assess consequences of alternative solutions in a systematic and consistent way throughout the system's lifecycle, there is a need to use an analytical approach supported by efficient decision support models—a combination of tools to assess different aspects of a decision. Typically, an optimization tool is used to identify the best logistic support solution from a cost effectiveness perspective and to optimize the spares assortment. A simulation tool is used to validate sustainability and ability to handle different scenarios and to dimension fleet size, personnel, repair equipment, and other resources. A cost calculation tool is used for LCC comparisons, identification of cost drivers, budgeting, and cost analysis. These tools work together as a suite to provide decision support for each type of decision and to help find the optimal trade-off between cost and availability.

A general approach when working with LCM analyses includes the following:

- Define a system and scope, the decision at hand, and the alternative solutions;
- Define prerequisites and limitations for operations and maintenance;
- Define influencing parameters and create a model;
- Acquire input data, beginning with a rough data model;

- Validate the model and the data quality, and improve data that have significant impact on the decision at hand;
- Perform analyses and evaluate the results;
- Perform sensitivity analysis, identify drivers of cost and effectiveness, and iterate to find the best solution.



As per Figure 4, in the early phases, stakeholder(s) make the major decisions, which

Figure 14 Lifecycle Maintenance Analysis Capability (Woulfe, n.d.)

will commit most of the future lifecycle costs. This means that it is in the early phases that stakeholder(s) need to put in most of the effort. Nevertheless, to achieve the availability performance and the lifecycle cost that the early decisions have made possible, stakeholder(s) need to carry on making decisions in a systematic way throughout the rest of the system's lifecycle. Otherwise, there is a great risk that stakeholder(s) will suffer from uncontrollable increasing costs or poor availability performance.

Managing decisions over the lifecycle with overall requirements and goals on macro level in focus, modeling detailed data on micro level is a true lifecycle management challenge.

Sample Test Cases

Case 1 Objective

A power utility company wants to investigate and analyze if it would be cost effective to invest in the procurement of spare transformers. Additionally, they need to determine the storage location for each of the transformers to optimize operational availability (Ao) of the power plant and operational costs.

- **Case 1 Sample Data**

  The power utility company used the data in Figure 5:

Parameter	Description
Power Plant	Name of power plant
Manufacture	Manufacturer of transformer
Apparent Power	The magnitude of the complex power [VA]
Voltage Ratio Max/Min	Ratio between LV and HV side
Vector Group	Winding configuration of
Existing Spare Transformer	If spare units exist and its location
Quality/Reliability	Reliability of transformer
Transformer Price	Price of transformer [EUR]
Downtime in case of spare	Time duration required to replace if spares
Downtime in case of no spare	Time duration required to replace if no
Expected annual gross margin of	Expected gross margin per annum if no

Figure 15 Available Transformer Data

The data concerning downtimes with and without spare units, and the data concerning the expected margin, enabled the utility to assess what possible downtimes would imply in terms of lost profit. Together with the reliability data and the price of each transformer, the risk of losing profit could be evaluated against the risk mitigation of investing in spare units.

Case 1 Methodology

The utility used a spare part and logistic support optimization tool to model and analyze their transformer case. The basics of the methodology is depicted in Figure 6.

This tool uses turnaround times, reliability, and price data together with other logistics, maintenance, and technical data to calculate the optimal assortment and allocation of spares from a system cost-efficiency perspective.

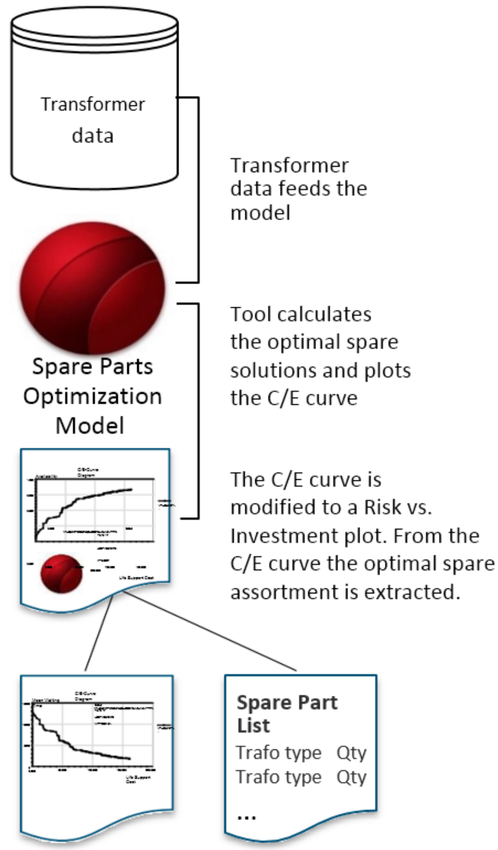


Figure 16 Overview of the Analysis Method (Andersson, 2015)

The spare part and logistic support optimization tool generates a cost/effectiveness (C/E) curve that plots the spares investment against the availability of the whole system (i.e., the average availability of all transformers). Each point on the C/E curve represents the optimal sparing solution for a specific budget frame, and as one progresses to the right in the C/E curve, the spares investment increases as power utility company invests in more transformers. As a consequence of the larger spares investment, the resulting availability also increases.

As the value of availability can differ between transformers in this case, the utility took advantage of the possibility to prioritize the plants in the model and used the expected annual gross margin as the priority factor in the input model.

Once the C/E curve had been established, the utility extracted the availability for each transformer in the case and for each point on the curve. Together with the information about the expected annual gross margin, the C/E curve was modified to a risk versus investment curve.

Case 1 Results

Figure 7 shows how the investments in spares influence the lost profits due to downtime caused by transformer failures. Naturally, lost production, and hence lost revenues, decreases with higher investment levels in spare transformers.

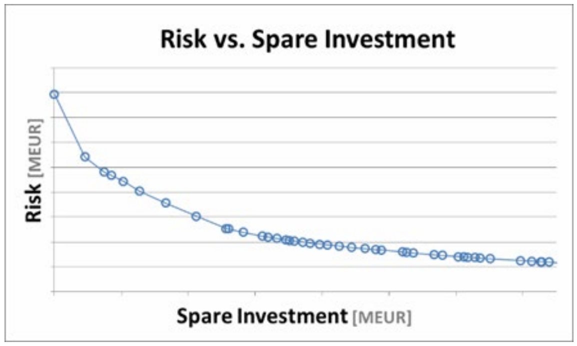


Figure 17 Risk Versus Spares Investment (Andersson, 2015)

The power utility company was interested in evaluating how many and which transformers could be economically motivated to purchase as spares. Therefore, the delta risk reduction was divided with each respective spares investment to create Figure 8.

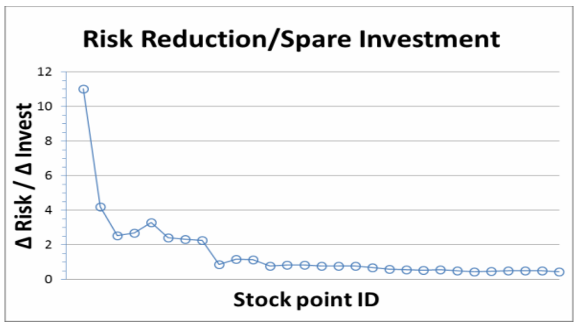


Figure 18 Delta Risk/Delta Investment (Andersson, 2015)

In the plot in Figure 8, the dimensionless ratio between risk reduction in dollars and investment in dollars is depicted. If this ratio is below 1, the investment is inevitably not profitable. However, all ratios above 1 will not necessarily prove themselves profitable since there are some uncertainties built into the risk value.

The power utility company opted to vary different input parameters (e.g. the failure frequencies of the transformers), in order to study the sensitivity of the results. Results from three scenarios with different failure rates are shown in Figure 9.

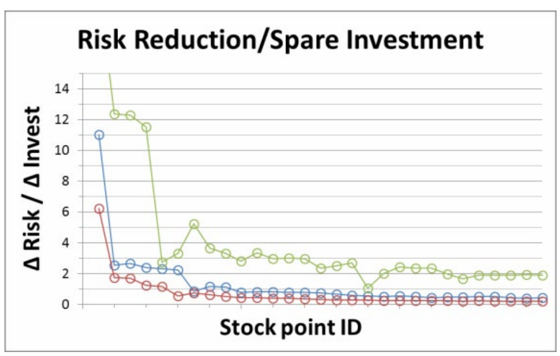


Figure 19 Delta Risk/Delta Investment at Different Failure Scenarios (Andersson, 2015)

Properly investigating the sensitivity of the results was an integral part of the analysis. To find the absolute availability level was not the priority of the analysis, more so was formulating a short list of transformers to invest in. After evaluating the case in different scenarios, the power utility company could select a ratio between risk reduction and spare investment with good judgment and formulate a short list of transformers for their investment program.

Case 2 Objective

Navy type commanders (TYCOM) want to make sure that all the ships pass their Board of Inspection and Survey (INSURV) inspections. Ships are typically notified 1 year prior to the conduct of this upcoming INSURV. What can the TYCOM do to mitigate the risks to the ships of failing an INSURV, and where should they focus their limited resources? Develop a statistical model to prioritize ship departments for focus of upcoming INSURV inspections.

Case 2 Sample Data

The TYCOM used the following data (see Table 1):

Table 5 INSURV Data

Parameter	Description
INSURV	Material Inspection (MI) Data
3-M	Maintenance Material Management Data
Training Sets	Prior INSURV MI data

Case 2 Methodology

Develop a statistical inspection model using binomial logistic regression using the following parameters:

- Formula

$$D = x_R + x_{Am} + x_i + x_{Av} - 1$$

where

$$D = Discrepancy\ (binary)$$

$$x_R = Root\ Cause\ Code$$

$$x_{Am} = Ship\ Age\ (months)$$

$$x_i = i^{th} Inspection$$

$$x_{Av} = number\ of\ Availabilities$$

$$-1 = No\ intercept$$

- Training Set = InspectionDate ≤ 2016 (90 Inspections)
- Test Set = InspectionDate > 2017(24 Inspections)
- There is no equivalent R^2 for logistic regression
- McFadden R^2 index (0.2–0.4 = excellent fit)
- Receiver operating characteristic (ROC) area under curve (AUC) is a (preferred?) binary classifier performance measurement (1.0 is ideal)

Case 2 Results

Figure 10 shows approximately 9 times out of 10 that the model correctly identified that a specific discrepancy will occur within this Anti-Submarine (AS) Department with a root cause (i.e., model is a realistic representation of predicting root causes).

- R^2 = 0.3534082
- Fit versus actual accuracy = 0.888888888889
- AUC = 0.8476919

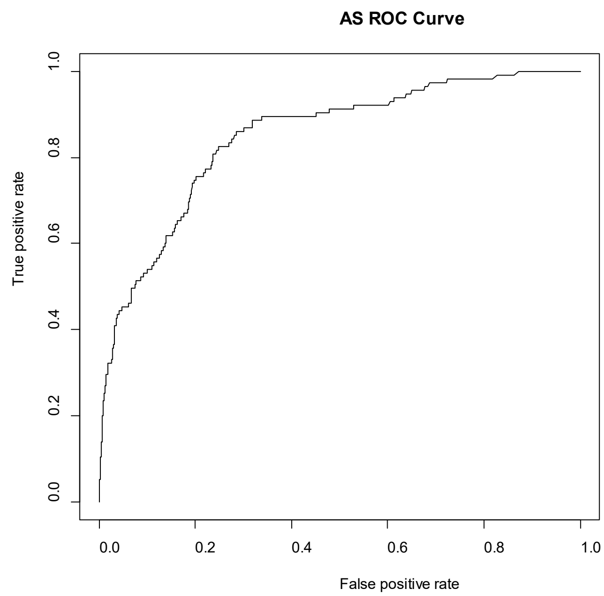


Figure 20 Model Fidelity Curve

Figure 11 shows the probability the defect (Pd) will occur for a particular area on the ship. ELEX/CCA/MODULE component failure is rated the highest probable defect in the reliability area (A). This provides a heads-up to the TYCOM team for a particular discrepancy area prior to the actual inspection. They may ask the ship to conduct additional preventive maintenance in order to mitigate these issues.

	P(d)	Discrepancies per Inspection	Avg EOC	Average EOC +/- 1 Standard Deviation (68%)
A-RELIABILITY				
1-MANUFACTURING DEFECT	17.48%	1	x-1σ 0.26 x 0.50 x+1σ 0.73	
2-INSTALLATION DEFECT	23.30%	28	x-1σ 0.24 x 0.48 x+1σ 0.71	
3-INADEQUATE DESIGN	12.62%	1	x-1σ 0.34 x 0.56 x+1σ 0.78	
4-ELEX/CCA/MODULE COMPONENT FAILURE	64.08%	7	x-1σ 0.41 x 0.63 x+1σ 0.85	
5-COMPONENT FAILURE	41.75%	3	x-1σ 0.39 x 0.61 x+1σ 0.83	
6-SEAL FIGURE	28.16%	4	x-1σ 0.38 x 0.63 x+1σ 0.89	

Figure 21 Probability of Discrepancy per INSURV Area

Conclusion

This paper has presented a tool-based methodology to enhance supportability. These models can be used for optimizing spares and predicting areas where failures can occur.

By conducting the analysis, the customers will be better prepared to provide informed decisions. The methodology quantifies the risks.

Moreover, the case presented in this paper shows how logistics modeling tools can be successfully employed and deliver fact-based results, also in cases with low failure frequency systems.

References

Andersson, M. (2015, June). *Balancing spares investment to minimize revenue loss.* Systecon. http://www.systecon.kr/download/paper_-_spare_transformer_concept.pdf

Woulfe, J. (n.d.). *Decision making in life cycle management: An analytical approach.* Systecon. Retrieved March 17, 2021 from https://www.systecongroup.com/sites/default/files/2020-08/decision_making_in_life_cycle_management_1.pdf

PANEL 6. RISK AND RESILIENCE IN THE SUPPLY CHAIN

Tuesday, May 1, 2021	
12:15 p.m. – 1:30 p.m.	**Chair: Stephanie A. Douglas,** Executive Director, Industrial Operations, Naval Sea Systems Command (NAVSEA) ***Quantifying Systemic Risk and Fragility in the U.S. Defense Industrial Base*** John Ullrich, Raytheon Technologies John Kamp, The George Washington University ***Design and Development of a Risk Assessment Tool to Enhance Decision Making Capabilities of Federal Acquisition Data by Leveraging Web Information*** Ningning Wu, University of Arkansas Mihail Tudoreanu, University of Arkansas at Little Rock ***Adapting DoD Acquisition Processes to More Effective include Partners and Allies*** Jerry McGinn, George Mason University

Ms. Stephanie A. Douglas—currently serves as the Executive Director for Navy Regional Maintenance Center (NRMC) in Norfolk, Va. Her duties include oversight of the operations and management of the Regional Maintenance Centers in the execution of private-sector, depot- level maintenance and modernization, surface ship intermediate-level maintenance, and fleet technical and engineering support world-wide.

Ms. Douglas was selected for appointment to the Senior Executive Service in April 2016. She has six years of experience as a civil servant while serving as Deputy Director for Fleet Maintenance (N43B) at U.S. Fleet Forces Command. She was responsible for fleet maintenance policy and a $3.5B fleet maintenance budget supporting two Naval Shipyards, three Regional Maintenance Centers, and Trident Refit Facility (TRF), Kings Bay and Readiness Support Group, New London to deliver mission ready ships to combatant commanders and achieve expected service life.

Ms. Douglas previously served as a naval officer from 1981 to 2010, retiring in the grade of Captain with 29 years of service. During her active duty career as an Engineering Duty Officer, she served on USS Emory S Land (AS 39) as Electrical Repair Officer and at numerous maintenance activities. Her tours included Charleston Naval Shipyard (CNSY) as nuclear and non-nuclear ship superintendent for submarines; Portsmouth Naval Shipyard (PNSY) as Baseline Advanced Industrial Management (BAIM) Implementation Manager and Operations Planning Manager; Supervisor of Shipbuilding Conversion and Repair, Portsmouth (now a portion Of Mid-Atlantic Regional Maintenance Center), as Planning Officer; and Pearl Harbor Naval Shipyard and Intermediate Maintenance Facility (PHNSY& IMF) as Production Resources Officer. During this tour, she led the damage removal efforts in Guam for USS San Francisco (SSN 711) following the ship's collision. She served two tours at Ship Repair Facility and Japan Regional Maintenance Center, Japan, first as Production Officer and Executive Officer and, finally, as Commanding Officer, where she made all preparations for the arrival of USS George Washington (CVN 73), the first nuclear powered warship to be home ported overseas and successfully completed CVN 73's first Forward Deployed Naval Forces availability. Her staff assignments include Commander, Navy Surface Forces, Atlantic as the Surface Type Desk Coordinator for amphibious, auxiliary, mine warfare and command ships, Naval Sea Systems Command (NAVSEA 04) as the Deputy Director of the Shipyard Management Group, where she developed corporate policy and oversaw business operations for four naval shipyards. Her final active duty tour was as Surface Ship Program Director in the Fleet Maintenance Directorate (N43) on the staff of U.S. Fleet Forces Command.

Ms. Douglas graduated from Auburn University where she earned a Bachelor of Science Degree in Chemical Engineering and her commission from the NROTC program in 1981. She earned a Naval Engineers Degree, a Master of Science Degree in Mechanical Engineering and a Master of Science Degree in Ocean Systems Management from Massachusetts Institute of Technology (MIT) in 1992.

Ms. Douglas' military decorations include the Legion of Merit, Meritorious Service Medal with a Silver Star, Navy Commendation Medal and two Meritorious Unit Commendations. Ms. Douglas has been an active member of American Society of Naval Engineers (ASNE) since 1987, and has served as Secretary, Vice Chair and Chairperson of the Tidewater ASNE Section and as National Councilor.

Quantifying Systemic Risk and Fragility in the U.S. Defense Industrial Base

John Ullrich—is pursuing a Doctor of Engineering degree in engineering management from the George Washington University. He holds a Master of Systems Engineering degree from Johns Hopkins University and an undergraduate degree in packaging engineering from the University of Wisconsin–Stout. Ullrich is a Senior Program Manager in Global Supply Chain Management at Raytheon Technologies, supporting Land Warfare and Air Defense systems. Additionally, he teaches within the Systems and Industrial Engineering Department at the University of Arizona. His research interests include engineering management, C5ISR, and high-energy physics. Ullrich is a member of the American Physical Society and the International Council on Systems Engineering. [ullrich@email.arizona.edu]

John Kamp—received a Doctor of Engineering degree in engineering management from the George Washington University in 2019, a Master of Engineering degree in nuclear engineering from Iowa State University, and a Bachelor of Arts degree in mathematics and French from the University of Nebraska–Lincoln. Kamp joined the George Washington University research staff in 2019 and supports Dr. Etemadi's acquisition strategy research project. He is a retired naval submarine officer with extensive experience in research and development and program management. His research interests include engineering management, maritime systems, and acquisition system research. Kamp is a Fellow in the Royal Institution of Naval Architects and a member of several professional associations. [jckamp2018@gwu.edu]

Abstract

This research quantifies fragility within the U.S. Defense Industrial Base (DIB) and translates it into supplier risk. The proposed model identifies systemically critical suppliers, where critically is characterized in terms of the supplier either being highly coupled within the industrial base, operating in a limited competition space, or owning a disproportionately large market share within a specific commodity. Each of these properties is quantified using centrality and community detection methods. By correctly assessing critical suppliers in the defense base, it allows for a methodical approach to addressing standard failure modes that typically result in material disruptions in advance of realizing interruptions. Quantifying fragility in supply chains based on systemic centrality and communities is a novel effort. Direct application of this process within the DIB fundamentally approaches assessing and strengthening our supply base resiliency in a completely different manner.

***Keywords*:** Defense Industrial Base, Fragility, Centrality, Community Detection, Systemic Risk

Introduction

The U.S. Defense Industrial Base (DIB) comprises a massive network of suppliers who, in totality, offer the capabilities and capacities required to meet and sustain the demand for the raw material, components, subsystems, and end-item deliverable weapon systems. Government agencies have focused on quantifying multiple dimensions of supply base risk within the DIB. In general terms, risk in this context is the uncertainty of events that would disrupt the material flow or create system availability delays. The U.S. Government Accountability Office (GAO) has provided multiple reports to the House of Representatives Committee on Armed Services recommending quantification methodologies and practices to manage supply base risk within the DIB (GAO, 2017). There is a two-part challenge in undertaking the effective execution of supply chain risk management within this environment. First, the DIB is supporting a range of diverse products and technologies. Secondly, traditional supply-chain risk management approaches focus on programmatic impacts versus systemic impacts (Sinha et al., 2004). The Department of Defense (DoD)

Critical Asset Identification Process (CAIP) quantifies Task Critical Assets (TCA) annually (DoD, 2008). For material considered defense-critical or task-critical by the DoD, the rough cost and schedule impact to requalify an unknown supply source is as high as $10 million and 9 months.

To illustrate the challenges of criticality assessment and program-centric supplier risk focus, consider a sole-source of supply for precision machined parts. One of the critical risk characteristics the GAO identifies is sole source dependencies (only one qualified source). In this hypothetical scenario, suppose that our sole source of supply were to declare bankruptcy and immediately cease operations. Programmatic consequences manifest as unfavorable impacts to cost, schedule, or both. The first realized consequence manifests as a limitation of manufacturing at the system level to the material on-hand position. This degraded state persists until a qualified alternative source of supply can be established. The total impact will directly correlate with the material's complexity and relative criticality to system-level operational requirements. Suppose this supply source is limited to a single program. In that case, the impact is limited to the cost and schedule associated with requalification. In this example, identifying criticality and risk consequence is program-facing; there is no accounting for more extensive dependencies or consequences within the DoD.

In contrast, consider the same bankruptcy scenario with a tightly coupled supplier within the DIB, meaning multiple programs provide manufacturing demand to the supplier. Connecting demand amplifies the impact of supply loss across programs, prime contractors, defense agencies, and, ultimately, throughout an entire commodity (North American Industry Classification System [NAICS]). Government program offices have limited information on single sources of supply. This research provides a pragmatic approach to quantifying a supplier's criticality within the DIB relative to its potential negative impact within the defense acquisition spectrum and utilizes network analysis to visualize and quantify dependencies within the DIB, translating them into system-risk measures.

Literature Review

Existing principles, practices, and analytic tools support this research and allow open-source spending data to characterize dependencies and connectedness and quantify fragility in terms of systemic risk. Doing so supports two novel approaches to assessing and strengthening our industrial base. First, this approach quantifies the growing dependencies within the industrial base, supporting risk management of items like obsolescence, capacity, and availability. Second, it allows for a meta-level view of the DIB that supports dynamic modeling and simulation of supplier failure propagation through the network (Meyer et al., 2014).

Fragility and System Risk Within the Defense Industrial Base

We define fragility within the DIB's context as the impact of a failing supplier on other suppliers, where failing is any disruption in material flow (availability or capacity). The current DIB is a fragile network, less conducive to competition, and challenged to scale quickly, grow, or innovate (Aviles & Sleeper, 2016). In this sense, fragility manifests as a common-mode failure where causal effects leading to failures propagate through the supply chain network. These common-mode failures can come from a range of realized impacts stemming from a range of macro forces impacting the DIB: reliance on sole-sourcing, uncertainty in defense budgets, and "bull-whip" demand cycles. These forces are quantified in terms of risk and realized as network fragility before the systemic disruption. Additionally, these forces are fundamental to creating uncertainty that stifles industry investment and growth (DoD, 2018).

While not ideal from a resiliency standpoint, supplier fragility does not directly quantify the respective supplier's risk (Lambert & Cooper, 2000), which means being a critical supplier within the DIB is not necessarily a direct indication or a probability that the supplier may fail. However, tight coupling or high dependencies within the network are a way to characterize the consequence of a respective supplier failing, most notably in terms of material disruption. Disruptions in material flow represent systemic risk, where systemic risk is the uncertainty of DIB network disruption. Moreover, realized consequence is associated with the supply base being unable to perform at total capacity or efficiency. As stated, there is a shared failure mode associated with disruption; the consequences, however, map to multiple dimensions; fragility quantification must adequately discern these modes. In this view, the DIB is not dissimilar to a complex social network or a financial network. Understanding the potential influence of a supplier within a network is a viable strategy to both modernize the industrial base and ensure a continuous supply for Defense procurement.

Traditional Network Analysis Applicability

Quantifying critical firm financial network analysis provides a pragmatic and scalable approach for defining critical suppliers and dependencies in a supply chain. Jorge Chan-Lau (2018) offered a risk-dimension mapping framework to quantify systemic risk within a financial network. Chan-Lau suggested three dimensions: first, *"too-connected-to-fail,"* where a tightly coupled firm's failure represented a risk to its neighboring firms; second, *"too-important-to-fail,"* where failure represents a considerable impact, even if the system-wide impact is not significant; and finally, *"too-big-to-fail,"* where the firm has a disproportionally large share of the systemic activities (Chan-Lau, 2018). This proposed architecture aligns well to supply chain mapping, as within the DIB supply chain, there is a range of highly connected, niche capability or massive suppliers.

Characterization of Supply Chain Risk

Supply chain risk management (SCRM) relates to the strategic management approach of risks, issues, and opportunities impacting a supply chain based on an organizational approach to assessing a respective event's potential consequences (Hallikas, 2004). SCRM directly leverages risk management tools in collaboration with supply chain professionals, both internally and externally. They focus on translating uncertainties of logistic and material flow efforts, material availability, or resources into an actionable plan for execution. Supply chain networks are inherently complex and dynamic; therefore, SCRM frameworks focus on providing effective risk management over a broad operational environment. Fundamentally, SCRM is the principle that an enterprise needs to prevent material disruptions throughout its entire supply base or supply chain. Critical measures of effectiveness are a systematic means to identify potential disruption sources, an enterprise approach to be an assessment of internal supply chain risks as well as an assessment of supplier or sub-tier supplier risk, assigned cognizant supply chain professionals managing identified risk, and, finally, the systematic means for continuous monitoring of disruptions or disruption sources (Blackhurst et al., 2008).

By definition, a supply chain inherently relies on connected critical providers, knowledge points, or handoffs, where a failure within the chain disrupts its coupled partner. In each respective reliance, uncertainty manifests as vulnerability; SCRM reduces vulnerability throughout a supply chain's entire value stream (Hallikas, 2004). Supply chain risk exists in multiple dimensions: natural disasters, raw material shortages, market forces, distribution challenges, or product or part technical maturity. This broad range of risks translates into a considerable exposure position that scales with the enterprise's size and complexity (Finch, 2004). SCRM typically incorporates the following processes as part of the risk management framework: identification, assessment, mitigation, acceptance, and monitoring of supply chain risks (Chopra & Meindl, 2009). It is a relatively heterogeneous literature base for SCRM, and the majority rely on traditional risk measures to influence action that can improve the agility or robustness of a supply chain. Supply chain agility is the speed an enterprise can react with should disruption or threat emerge within the supply base.

Quantifying this measure is the manufacturing lead time for supplier material, where minimizing the make-span or procurement-span for the material is optimal. Also included in supply chain agility is the responsiveness to changing to market needs, where an organizational goal is to build a supply base capable of transition to a different or modified material solution without impacting delivery reliability. Supply chain robustness shares some similarities with supply chain agility, where the notable delta is not an organization's ability to adapt but rather the quantified incurred disruption of a singular event. In a robust supply chain, when a change occurs, the supply base inherently provides more time to plan a course of action. Additionally, a supply chain's robustness measures a supply chain's ability to carry out its functions in a degraded state. Using a major natural disaster as an example of a disruptor, if an impacted supply base can maintain deliveries without a strategic shift in execution, it would be fair to say that the base was robust and not impacted by a singular event.

Quantification of supply chain risk is the product of a consequence in terms of an event's cost impact or material disruption incurring schedule increase compared against the event's likelihood. Like traditional risk management approaches, this product approach prioritizes and characterizes the risk and opportunity spectrum (Hubbard, 2009). This approach is the most popular methodology for quantifying risk, both within a supply chain and in the broader sense of risk management. Regardless of this approach's debatable effectiveness, it is, as stated, widely accepted within supply chain professional organizations (Manuj & Mentzer, 2008).

Factors likely to disrupt a supplier, product, program, or service establish risk archetypes; these archetypes help suggest the likelihood of impact realization (Outdot, 2010). Traditional supply chain measures supporting this quantification are a supplier's financial viability, operational capacity or expertise, or a quantifiable supplier resiliency score. Additionally, certifications provided by compliance organizations such as ISO or the National Aerospace and Defense Contractors Accreditation Program (NADCAP) indicate a low likelihood of an adverse event occurring via a supplier's successful acquisition retainment of certification. Finally, quantification of supply chain risk in customer value is germane in industry practice, where the primary measure is on-time delivery and order correctness. Supply chains with risk in a customer value dimension manage threats associated with procuring the wrong or defective products within their demand portfolio (Nishat, 2006). This risk dimension aligns with traditional measures of quality management systems: defects per unit, the accuracy of an order, or rework cycles (Rao & Goldsby, 2009).

Centrality and Community Measures of Criticality

Provided a sufficiently complex network exists, there will inherently be relationships of either highly connected nodes or tightly coupled nodes within a localized area (Newman, 2008). The well-defined principles and power laws that support these concepts stem from social network analysis and are both long-standing and proven (Bonacich, 1987). There is a nearly endless amount of research available where the application of centrality measures supports critical nodes or vertices identification within a network for a range of practical purposes, most notably the continued evolution of the use case of modeling influence in a social network (Wang & Street, 2015). Beyond social networks, these methods are in use in biology research to identify critical species in pollination communities (González et al., 2010), in health research to assess associations between measures of network centrality and health in a retirement community (Schafer, 2011), and within the financial industry to identify and assess the risk of financial firms (Chan-Lau, 2018). The common link in each application's approach is a need to understand the network's relationships that support the characterization of node importance or insignificance.

Centrality and community indices directly answer what is fundamentally important to a node, vertex, or network. The output is a tangible function providing real-values for node and flow importance concerning the analyzed network. As stated previously, the word "importance" can relate to a range of actual definitions based on the analysis's intent. Two general categories of "importance" have been proposed (Vivas et al., 2019). First, centrality indices reflecting network flow are critical nodes predicated on the classification of centrality based on the flow considered vital to a network (Opsahl et al., 2010). As an example, in financial network analysis, this is the amount of money flowing from firm to firm, where the out-strength of a node reflects direct spend or transfer of funds, and the in-strength represents receipt or acceptance of funds (Chan-Lau, 2018). This example results in the quantification of node importance in a minimum of two dimensions, dependency on money distribution (out) and the total holdings or receipts (in). Second, "importance" can be measured in terms of the coupling of nodes within a network. For example, in the modeling of pollination generalist species of plants, a tightly coupled sub-network of nodes increases the probability of cross-pollination among the subsets (Alvarez-Socorro et al., n.d.).

Leveraging Centrality and Community to Quantify Systemic Risk

As a novel approach to quantifying risk, vulnerabilities, and imbalances within the DIB, this research proposes that centrality and community measures provide critical insight into two macro forces threatening a supply chain. First, connectedness-based risk rankings quantify systemic risk. Second, community measures quantify fragility. A supplier can be both systemically risky and fragile. In this paper, the following arguments establish systemic risk, fragility, and imbalance: systemic risk directly relates to a supplier's criticality within a supply chain network. A supplier with more influence carries a more significant negative impact on the overall network in the event of a disruption; it is, therefore, more systemically risky than a weakly-connected supplier. Fragility indicates vulnerability or the lack of supply chain network robustness (Perera et al., 2018). Larger communities with more outstanding overall systemic dependencies illustrate vulnerability within the supply chain network. Finally, imbalance represents disproportional levels of both risk and fragility for both commodities and suppliers.

In the remaining sections of the paper, Methodology details the network creation and structure and the applicability of specific centrality measures and community, thereby providing acquisition agencies with lower sub-tier visibility regardless of program or procurement authority. Results uses Aircraft NAICS as a use-case to apply network analysis; this analysis supports a key research objective of detecting, evaluating, and characterizing supply base threats capable of disrupting material availability. Lastly, Conclusions presents the conclusions of this research, with the intent that through further modeling and via a coupled methodical supplier development approach, a more resilient and responsive DIB can be developed.

Methodology

This section briefly describes the methods utilized to calculate centrality measures and assess modularity to support community identification.[20] Systemically critical suppliers exist as highly linked nodes throughout the network (central nodes), tightly coupled links within neighboring nodes (community nodes), or a state where the supplier is both central and tightly bound within a community.

Data Aggregation and Network Structure

This research is limited to unclassified, open-source acquisition data; no prime generated or propriety data is within the analysis. Therefore, the analysis is subject to contractor reporting accuracy for material spend disclosed per the Federal Funding Accountability and Transparency Act of 2006 (FFATA). The FFATA requires that any federal contract, grant, loan, and other financial assistance awards of more than $25,000 are on a publicly accessible and searchable website. Data reporting is limited to first-tier suppliers; subcontract award information contains awardee, DUNS information, parent company information, award date, program usage, and material type. The provided illustrations show the type of data and views available from open-source government data for Army Missile Procurement (U.S. Department of the Treasury, Bureau of the Fiscal Service, 2021).

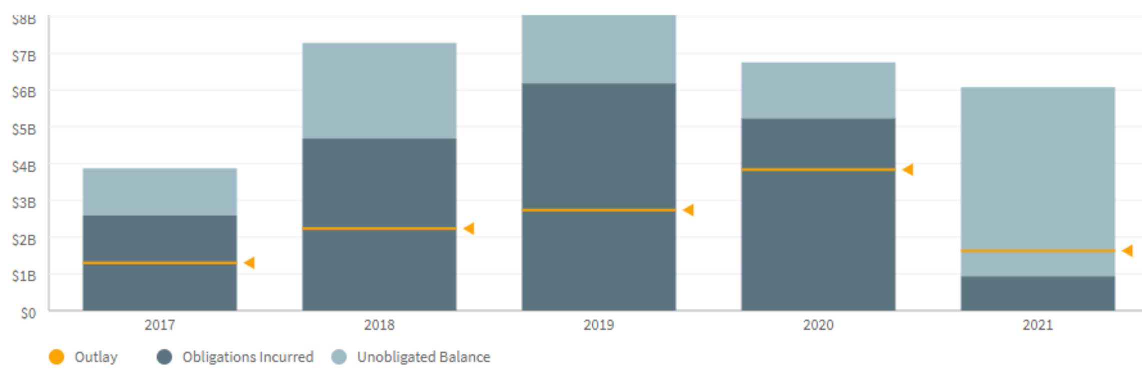


Figure 22. Spending Over Time (Fiscal Year [FY] 2017+ Army Missiles)

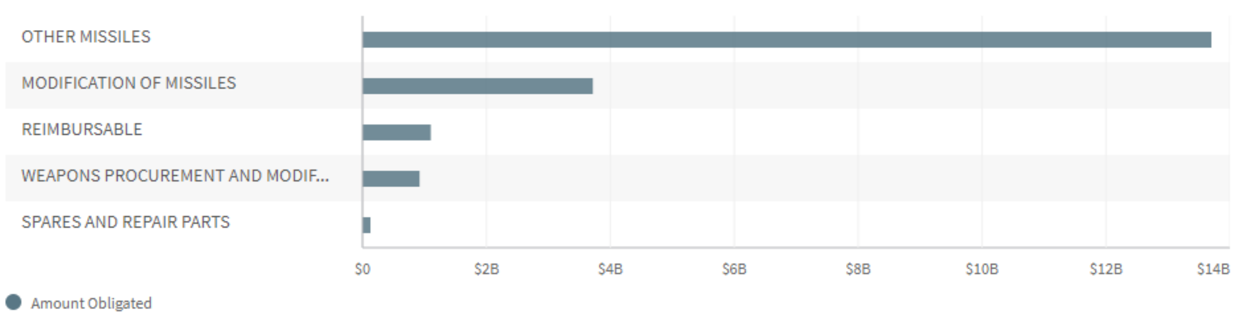


Figure 23. Spending by Category (FY2017+ Army Missiles)

[20] Underlying math foundations are provided as references.

Defense programs or NAICS commodities facilitate the analysis of relations between objects. Our vertices or nodes will represent the following organizations procuring agencies, prime contractors, and subcontractors (reference Figure 3). Edges will communicate both the existence of a relationship and a directed path or flow of acquisition dollars. Reference Figure 4 for an example of the visualization output.

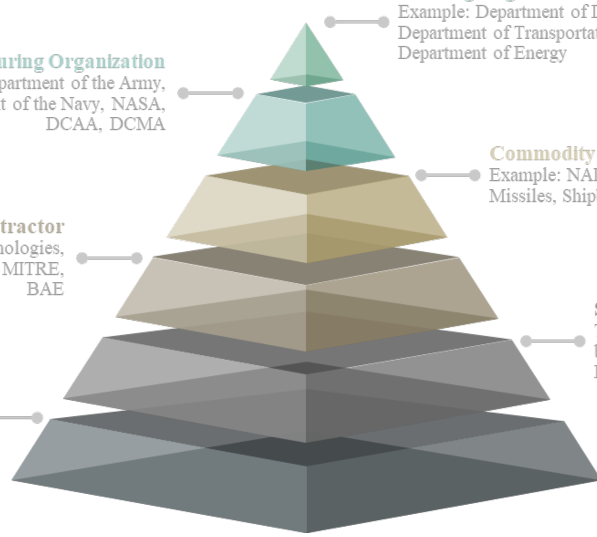


Figure 24. Network Structure

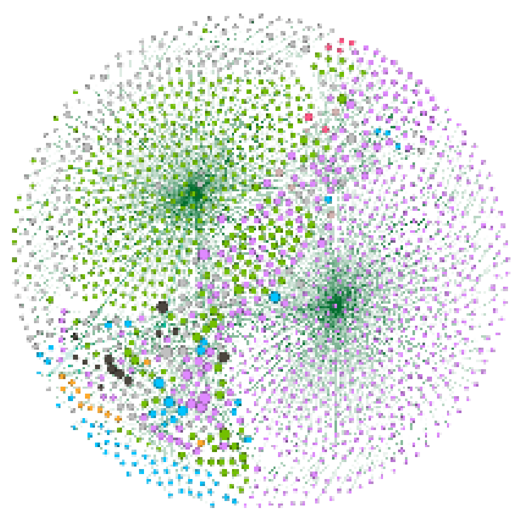


Figure 25. FY2019 Army Missile Procurement Visualization

Centrality (Node Level)

Centrality measures allow for identifying systemically critical suppliers in the supplier base; nodes reflect specific contractors and sub-contractors, node size reflects the centrality score, and color of the node reflects segregated communities' subsystems. Table 1 summarizes measures of centrality. Degree in this context is a local measure; the DIB financial network requires a global view of the supplier's connections. Alternative centrality measures are required to characterize systemically critical suppliers within the defense network correctly.

Table 6. Measures of Centrality

Item	Basis	Measure	DIB Applicability	Source
Degree	Importance score based on the number of links held by each node	Direct connections	In-degree and out-degree measures to better understand the flow of material	Perera et al., 2018
Betweenness	The number of times a node lies on the shortest path between other nodes	Network efficiency of flow	High betweenness indicates critical suppliers that are highly active within the network	Estrada et al., 2009
Closeness	Time required to spread information from a node to the other nodes in the network	Shortest paths between all nodes	Suppliers with high closeness centrality levels support mitigation of the impacts arising from bullwhip effect (Xu, et al., 2016)	Buechel & Buskens, 2013
EigenCentrality	Represents the relative strength or influence over other nodes in the network	Node influence	Quantifying the propagation of failure tied to disruption of a supplier	Ruhnau, 2000
PageRank	Similar to EigenCentrality, the assigned score reflects influence within the network, but PageRank also considers link direction and weight	Node Influence	The extent of failure propagated through a community of suppliers or across a commodity	Page, 1999

Communities (Network Level)

While centrality measures provide insight on systemically critical suppliers, the complexity and size of a macro-view of defense procurement requires an approach capable of accurately decomposing highly interconnected nodes into communities. Doing so supports the quantification of fragility in the multiple dimensions in which it can exist. The usage of community detection allows for analysis of tightly coupled suppliers, further facilitating quantification of likely common failure-mode points within the network. As the applicability of centrality measures, multiple methodologies of community detection are germane in network science. Table 2 shows some of these community measures.

Table 7. Measures of Community

Item	Basis	Measure	DIB Applicability	Source
Network Diameter	Edge count of the shortest path across the network	Complexity	Supports quantification of local community authority or the lack of authority across a commodity	Abd-El-Barr, 2009
Network Density	The level of interconnectivity between nodes	Connectivity	Higher density indicates a more robust supply chain	Bendle & Patterson, 2008
Clustering Coefficient	The level of coupling nodes demonstrated	Subsystem or neighborhoods	Assessing program, agency, or prime contractor supply chain dependencies	Brintrup et al., 2016
Modularity	The strength of the allocation of subsystems within a network	Subsystem or neighborhoods	Detecting community structure within an NAICS group	Fortunato & Barthelemy, 2007

Risk Association

Reference Figure 5 for an overview of the applied risk framework. This graph depicts relative community strength on the x-axis, where a higher assigned score represents a more substantial connected supplier. EigenCentrality scores compose the y-axis, indicating a supplier's strength or influence over other nodes in the network. Finally, the supplier node size represents a function of its relative community ranking and its overall authority within the network combined with the supplier's weighted indegree. Leveraging Centrality and Community to Quantify Systemic Risk proposed systemic risk, which is a risk to the overall supply chain network's efficiency or effectiveness, which could be determined using total supply-base influence measures. Two forms of risk are present: (1) the local criticality of a supplier, where subsequent supplier risk can be further defined using traditional defense industrial risk measures (reference Table 3) and (2) the systemic risk a node presents within its overall network or community.

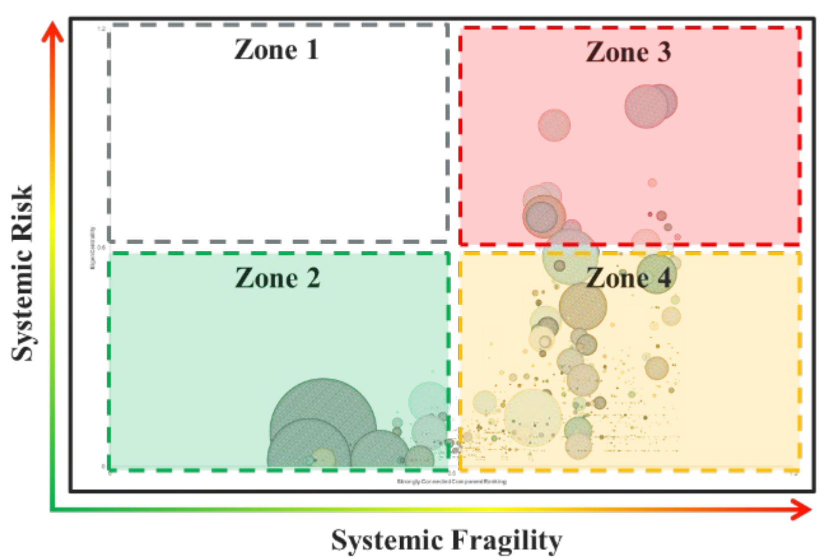


Figure 26. Mapping Risk to Centrality and Community

Four distinct zones form using this analysis technique. First is High-Systemic Risk and Low-Systemic Fragility; suppliers in this zone would carry considerable network influence but would have a lower strength of connection or community impact; this zone should be free of qualifying nodes, as a weakly connected supplier should not be supported with a high influence ranking. Second is Low-System Risk and Low Systemic-Fragility; suppliers in this zone would reflect both lower influence and community coupling, meaning they pose a low systemic threat; however, they carry considerable quantified local risk; as an example, prime contractors tend to appear here; they carry a disproportionally large total of spend with low in-strength. While these nodes are generally more central, nodes identified with more significant risk are more systemic. Third is High-Systemic Risk and High-Systemic Fragility; these suppliers are considered imbalanced; they are critical to the network from a community perspective; they also carry significant systemic risk. Additionally, their influence and in-degree can establish local risk. Fourth is Low-Systemic Risk and High-Systemic Fragility; these suppliers represent the absence of network robustness.

These measures are relative to the scope of the network analysis completed. For example, the Department of the Army spending analysis will result in a different set of identified risk, fragility, and imbalance than the same analysis focused on Department of the Navy spending. Moreover, the combination of both agencies will again shift quantification and output. Furthermore, analyzing modules or communities within an analysis will provide a different set of focus suppliers. This dynamic nature is critical for correctly identifying the specific threats for a cognizant program office or prime contractor and understanding the overlapping or shared risk.

Mapping Risk to Traditional Supply Chain Risk Areas

It is reasonable to leverage centrality and community measures against traditional risk areas. Table 3 provides GAO-identified risk areas threatening the DIB (GAO, 2018). By selecting node level measures with known supply chain network implications, further evaluation of systemically risky or fragile suppliers is achievable in terms of their local risk factors.

Table 8. Mapping Network Measure to Traditional Risk Areas

Traditional Risk Area (GAO)	Traditional Approaches	Concern	Pf Measures (Likelihood)	Cf Measures (Severity)
Financial Viability of Suppliers	Monitor – Monitor DUNS data as available	Shrinking DIB, inconsistent demand forecasting	DUNS Trend (6-month, 12-month) – Couple with community measures, the financial viability of the community	Highest betweenness levels within a community
Sole Source	Monitor – Quantitative at the program level	Single points of failure	Closeness centrality, ability to share demand	Highest Eigenvector measure within a network
Limited Production Capacity	Avoid - Qualitative, supplier RFPs	Inability to ramp quickly	Trend analysis supplier CAGR (increasing) Highest Eigenvector measure within a network; within a commodity	Highest Eigenvector measure within a network; within a commodity
Facility Damage by Disaster	Monitor - Quantitative concerning risk areas, qualitative regarding the impact	The failure mode of sole-source	Natural disaster probabilities/distributions	Supplier Geolocation – Number of programs/primes impacted. Highest Eigenvector measure within a network; within a commodity
Loss of Skill or Equipment	Accept – Difficult to quantify. Highly variable by program	Lack of manufacturing expertise and DIB investment funding	Trend analysis supplier CAGR (decreasing)	Highest Eigenvector measure within a network; within a commodity
Foreign Dependence	Mitigate - Quantitively at the prime level, qualitative at the subcontract level	Component dependencies external to the United States	DUNS Trend (6-month, 12-month) – Couple with community measures, the financial viability of the community, commercial market share	Parent DUNS, Highest Eigenvector measure within a network; within a commodity

As an example, reference Figure 6. This subset view of suppliers shows suppliers with the least betweenness centrality while still holding system risk. Closeness centrality is critical to the effectiveness of the supply chain in the presence of a degraded state or inaccurate demand planning (Perera et al., 2018); these detractors contribute to the "bull-whip" effect in supply chains (Xu et al., 2014). In our provided view, these are essentially critical suppliers within the network with limited or nonexistent closeness measures. They cannot share total demand and are therefore risk considerations for traditional concerns like sole sourcing, limited capacity, or loss of skill or equipment.

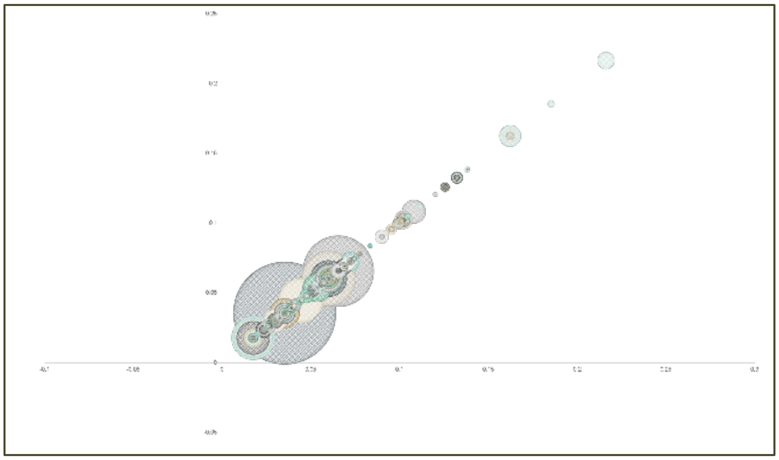


Figure 27. Network Measures Translated Into Sole-Source Risk

The following section provides the application of these processes. They show the use of the methods outlined as they relate to Aircraft manufacturing in FY2020. The output of this analysis will be the identification of systemic risk, fragility, and imbalance within the supply base.

Results

Application: FY2020 NAICS – Aircraft Manufacturing

This analysis evaluated roughly $25 billion in disclosed spend. Key prime contractors were BAE, Lockheed Martin, and Raytheon Technologies. NAICS analysis was limited to the following codes and their respective titles: Aircraft Manufacturing (336411), Aircraft Engine and Engine Part Manufacturing (336412), and Aircraft Parts and Auxiliary Equipment Manufacturing (336413). As a commodity, this represents deliverable items such as air vehicles, gas turbines, engine components, avionic subsystems, and engineering services. The primary procuring agencies are of the DoD, provided as follows in order of out-degree: Department of the Air Force, Department of the Navy, Defense Logistics Agency, Department of the Army, U.S. Special Operations Command, and the Defense Contract Management Agency.

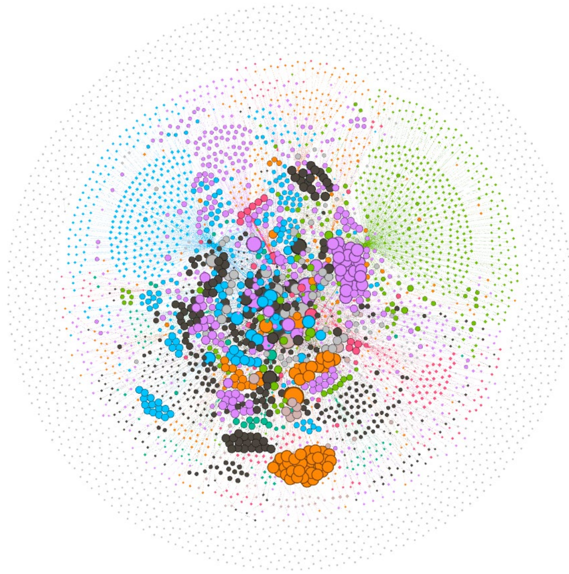


Figure 28. FY2020 NAICS Aircraft Supplier Network

Per Figure 7, the node's size reflects the network EigenCentrality score, conveying suppliers with network influence. The assigned node color indicates a subsystem within the network, and these are an output of the analysis of network modularity. Assigned modularity aligns with either a principal prime contractor, a specialized commodity provider with limited direct competitors (notable examples: sand castings, energetic materials, solid-state rocket motors), or a family of parent-company–owned sub-contract suppliers. As an example of wholly-owned subsidiaries driving communities, the suppliers United Technologies, Parker Hannifin, Aerojet, and L3 essentially build independent sub-tier mapping communities.

Aircraft Centrality Measures

Centrality for each supplier is quantified in four different measures: weighted degree, closeness, eigenvector centrality, and PageRank. These outputs provide the basis for fragility assessment; they inherently communicate the network's criticality based on a critical node's impact. The consequence in these terms is relative to the overall network versus a specific program or contractor; impact by a prime contractor can, however, directly map to an individual program. Table 4 summarizes the overlap of centrality-based network measures representing the systemic risk of sub-tier suppliers relative to the overall NAIC Aircraft supply base.

Table 9. Fragility Assessment Overlap (Centrality)

Overlapping Suppliers by Centrality Measure

	Weighted Degree	Betweenness	Closeness	PageRank	Eigen Centrality
	Top 10 Suppliers				
Weighted Degree		8	8	2	6
Betweenness	27		10	4	10
Closeness	26	45		5	10
PageRank	6	6	5		3
Eigen Centrality	25	28	28	6	
	Top 50 Suppliers				

	Weighted Degree	Betweenness	Closeness	PageRank	Eigen Centrality
	Top 25 Suppliers				
Weighted Degree		17	16	5	13
Betweenness	41		25	4	20
Closeness	45	68		5	23
PageRank	9	7	6		6
Eigen Centrality	31	28	28	7	
	Top 100 Suppliers				

	Weighted Degree	Betweenness	Closeness	PageRank	Eigen Centrality
	Top 10 Suppliers				
Weighted Degree		80%	80%	20%	60%
Betweenness	54%		100%	40%	100%
Closeness	52%	90%		50%	100%
PageRank	12%	12%	10%		30%
Eigen Centrality	50%	56%	56%	12%	
	Top 50 Suppliers				

	Weighted Degree	Betweenness	Closeness	PageRank	Eigen Centrality
	Top 25 Suppliers				
Weighted Degree		68%	64%	20%	52%
Betweenness	41%		100%	16%	80%
Closeness	45%	68%		20%	92%
PageRank	9%	7%	6%		24%
Eigen Centrality	31%	28%	28%	7%	
	Top 100 Suppliers				

When looking at the overlap of measures, suppliers' composition should draw attention to prime contractors' dependencies within the network. Table 5 shows the top 10 overlapping suppliers for this network.

Table 10. Top 10 Overlapping Suppliers (Aircraft NAIC FY2020)

SUPPLIER NAME	MODULARITY CLASS
UNITED TECHNOLOGIES CORPORATION	92
LEONARDO SPA	92
EATON CORPORATION PUBLIC LIMITED COMPANY	92
CURTISS-WRIGHT CORPORATION	92
BOEING COMPANY, THE	92
AMETEK INC.	92
L3HARRIS TECHNOLOGIES, INC.	93
Transdigm Group Incorporated	93
HEICO CORPORATION	93
NORTHROP GRUMMAN CORPORATION	122

The presence of crucial prime contractors results from their weighted degree per the analysis provided in Table 4. In this example, suppliers like United Technology, Boeing, and Northrop Grumman are sub-contracts to prime spending. Their complete failure concerning the network (e.g., bankruptcy) is improbable. Overlapping firms carry the highest overall fragility or concern. Their respective modularity classifications convey the interdependencies that exist. These suppliers are not only critical to the performance of the supply chain network but are also highly dependent on each other. Traditional monitoring methods and assignment of fragility, criticality, or risk based on total monetary spend, whether it be by program or supplier, are insufficient to characterize total industry fragility.

Expanding on the measurement intent outlay provided within the Methodology section, consider the following: The discernible differences in identified suppliers indicate that different centrality measures indicate that dimensions of fragility exist for supply chain networks by lumping measures together or looking myopically at total spending hides suppliers with considerable network influence. Table 6 provides a conceptual approach to matching dimensional fragility measures with traditional supply base risk measures.

Table 11. Centrality Based Fragility Mapped to Systemic Risk Drivers

Measure	Fragility Dimension	Systemic Risk Drivers
Weighted Degree	Primarily parent companies, or direct subcontract award to major prime contractors. The network is dependent on forecasted demand	• Demand Uncertainty • Budget Uncertainty • Natural Disaster or Malicious Attack
Betweenness	Composed of "bridge suppliers," this model moves to the first tier of the prime contractor supplier spend. As an effect, these are primarily parent suppliers or familiar sources of supply for generic material (electronic components, fasteners)	• Foreign Dependence • Single Sources of Supply
Closeness	Relatively high overlap of closeness and weighted degree indicates that the network's agility or speed depends on large tier suppliers. Respective capabilities and capacities should facilitate shorter paths through the network.	• Limited production capacity • Foreign Dependence • Natural Disaster or Malicious Attack
PageRank	The PageRank algorithm consistently highlights influential suppliers outside of the top spend.	• Obsolete Items • Financial Viability of Suppliers • Sole sourcing • Loss of skill or equipment
EigenCentrality	They are highly coupled or connected suppliers within the network; their dependencies cross over programs, procuring agencies, and even commodities.	• Limited production capacity • Foreign Dependence • Loss of skill or equipment • Financial viability • Sole source • Natural Disaster or Malicious Attack

Aircraft Supply Chain Network – Systemic Risk

The suppliers listed in Table 7 carry the highest systemic risk within this commodity code. The EigenCentrality measure dictates the supplier (displayed in Figure 8). There is a range of technical capability provisions listed, and this suggests critical suppliers across a broad spectrum of provided solutions. The influence of these suppliers propagates through the supply chain network. Their disruption impacts parent companies, prime contractors, and coupled procurement agencies. It is important to note that these suppliers share community measures; Aircraft Supply Chair Network – Systemic Module Risk describes the impacts to systemic risk at a community level.

Table 12. Top 10 Systemic Risk Suppliers

SUPPLIER NAME	PROVISION
GOODRICH CORPORATION	Lighting Systems, Actuation, and Control
HAMILTON CORPORATION	Propulsion Systems, Flight Control Systems
COBHAM INC.	Antenna, Electronic Subsystem, RAD-Hard
AMI INDUSTRIES, INC.	Emergency evacuation systems, Seating systems, Life rafts
B/E AEROSPACE, INC.	Structures
INTERTRADE LIMITED	Recertified airframe and engine parts
EXOTIC METALS FORMING LLC	Engine ducting and exhausts
L3HARRIS TECHNOLOGIES, INC.	R.F. equipment, Data Link Communication
WESCAM INC	Air Surveillance and Reconnaissance
CHELTON AVIONICS, INC.	Antenna systems, avionics systems, electronics systems

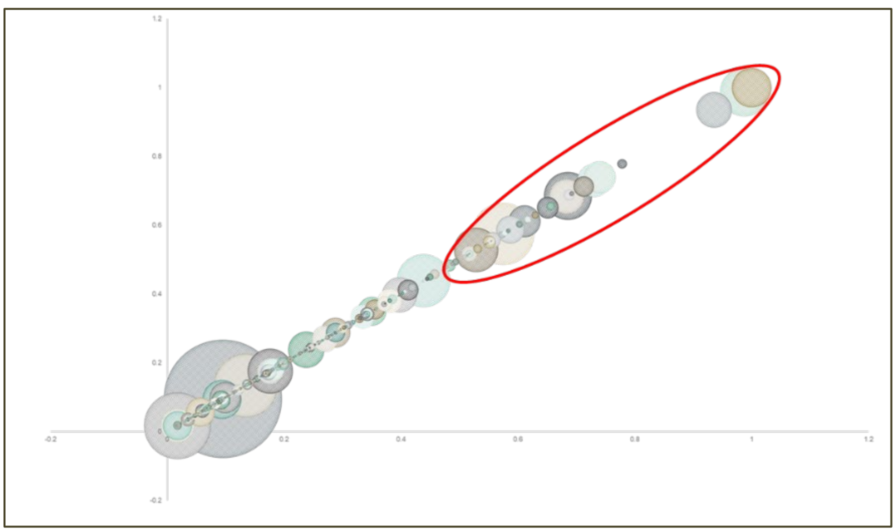


Figure 29. Aircraft Systemic Risk Visualization

- ***Aircraft Supply Chain Network – Systemic Module Risk***

A crucial module or communities that formed within the network centered around United Technologies Corporation (UTC). UTC is a parent company within this analysis; the basis for this community is derived from the material acquisition across NAICS either directly to UTC or one of their wholly-owned subsidiaries. While the complete list of suppliers will not result in a complete module composed of UTC subsidiaries, Table 8 provides a list of the top 10 systemically risky suppliers within the module. This analysis provides a supply chain manager insight into critical dependencies within a community. More notably, this analysis supports further risk characterization based on the supplier’s authority and valuation (size of the node); reference Figure 9.

Table 13. Systemic Risk - Module Analysis

SUPPLIER NAME	PARENT COMPANY
HAMILTON SUNDSTRAND CORPORATION	UNITED TECHNOLOGIES CORPORATION
GOODRICH CORPORATION	UNITED TECHNOLOGIES CORPORATION
Rockwell Collins, Inc.	UNITED TECHNOLOGIES CORPORATION
AMI INDUSTRIES, INC.	UNITED TECHNOLOGIES CORPORATION
B/E AEROSPACE, INC.	UNITED TECHNOLOGIES CORPORATION
INTERTRADE LIMITED	UNITED TECHNOLOGIES CORPORATION
GOODRICH ACTUATION SYSTEMS SAS	UNITED TECHNOLOGIES CORPORATION
ROSEMOUNT AEROSPACE INC.	UNITED TECHNOLOGIES CORPORATION
J. A. REINHARDT & CO., INC.	UNITED TECHNOLOGIES CORPORATION
GOODRICH LIGHTING SYSTEMS, INC.	UNITED TECHNOLOGIES CORPORATION
GOODRICH ACTUATION SYSTEMS LTD	UNITED TECHNOLOGIES CORPORATION

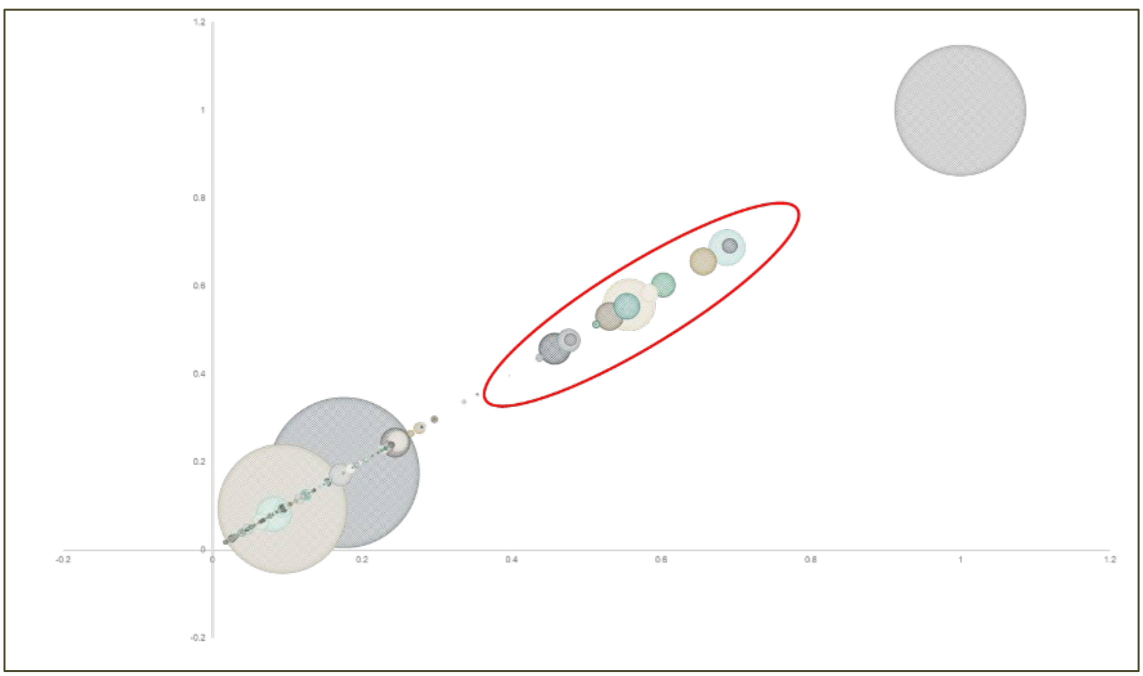


Figure 30. Systemic Risk - Module Analysis

Aircraft Supply Chain Network – Systemic Fragility

Table 9 lists top identified suppliers as a function of their authority and relative component strength measure. These represent weak points in the supply chain network.

They are network vulnerabilities with the general implication that there is no ready-made set of alternative sourcing options.

Table 14 – Top 10 Systemic Fragility Suppliers

SUPPLIER NAME	PROVISION
ACME EMBEDDED SOLUTIONS	Ruggedized Computing Systems
SIERRA ALLOYS COMPANY	Titanium Manufacturing
PERILLO INDUSTRIES, INC.	Power Subsystems
FIBREFORM ELECTRONICS, INC.	Precision Machining
TORAY ADVANCED COMPOSITES ADS, LLC	Composite Materials
S&L AEROSPACE METALS, LLC	Structural Machining
RIVERSIDE MACHINE & ENGINEERING, INC.	Precision Machining
MICROWAVE DEVELOPMENT LABORATORIES, INC.	Waveguide Components
ADVANCED CONVERSION TECHNOLOGY, INC.	Power Subsystems
BOEDEKER PLASTICS, INC.	Molded Plastics

Aircraft Supply Chain Network – Imbalance

Given the massive nature of this supply chain network, narrowing systemic risk and fragility to each category's top 10 drivers is less than ideal for taking a pragmatic approach to improving the base's robustness. The concept of imbalance introduced in the Literature Review can narrow systemic risk and fragility into network-specific threats, as shown in Figure 10.

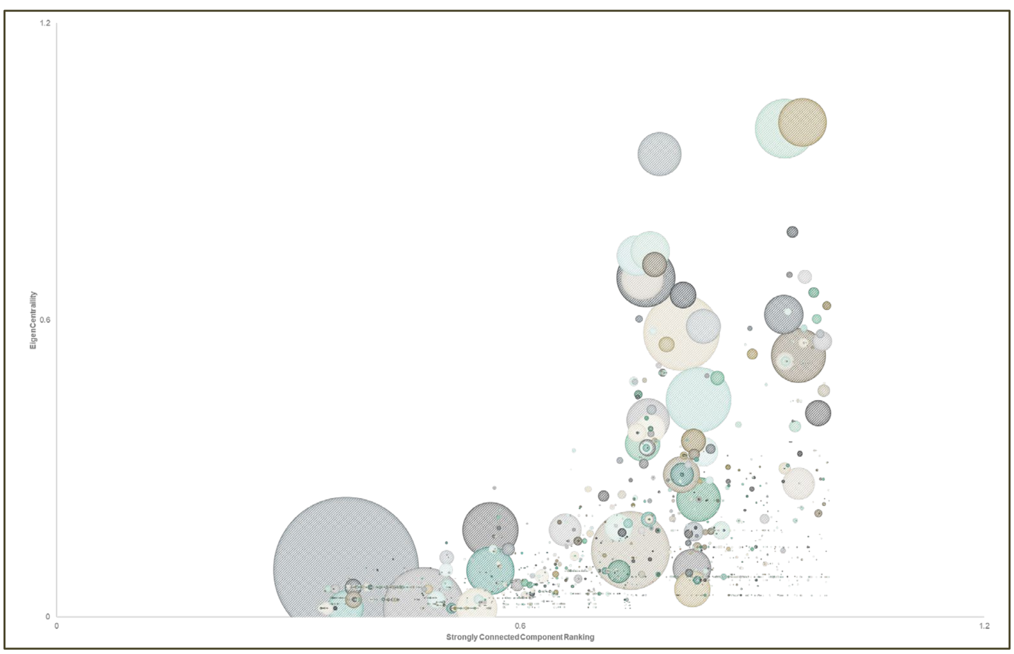


Figure 31. Aircraft NAICS - Imbalance Assessment

In Aircraft manufacturing, the following sub commodities are systemically risky and fragile: titanium manufacturing and forging, engine component manufacturing, structures, and precision machining. The suppliers providing this material share similar network influence measures and are primarily shared sources of supply regardless of the prime contractor, consequently resulting in a disproportional (imbalanced) amount of total spend distribution and, consequently, network criticality assignment. This analysis augmented with targeted supplier development efforts would both highlight and make mitigation activities actionable. These are critical points within the supply base that could be augmented with direct investment in capabilities, training, and long-term demand stabilization or additional suppliers developed to build redundancy in the overall network.

Conclusions

We have presented a detailed approach for leveraging centrality and community measures to quantify systemic risk, vulnerability, and imbalance in defense supply chain networks. This approach evaluates defense procurement supply-base resiliency by commodity, program execution office, or overarching defense procurement agency. As a result, the following three objectives and their conclusions are as follows: First, systemic risk is quantified using centrality measures to identify the most critical nodes within the network. A supplier with more influence carries a more significant negative impact on the overall network in the event of disruption and is, therefore, more systemically risky. Second, DIB fragility is quantified using community measures; facilitating identification of communities with more significant overall systemic dependencies illustrates vulnerability within the supply chain network. Third, imbalance represents disproportional levels of both risk and fragility in both specific commodities and suppliers.

To illustrate the application of these concepts, an FY2020 view of Aircraft manufacturing was provided. This analysis addressed 80,000+ records of subcontract procurement for material ranging from fasteners to avionic subsystems. This visualization facilitated the identification of suppliers in terms of systemic risk and fragility in the following technical areas: power subsystems, structures, forgings, microwave components, and electronic components. Furthermore, a novel approach to quantifying traditional risk measures using centrality and community detection was proposed, highlighting sole source risks within a network.

By leveraging network analysis principles and practices, we have demonstrated how application within the DIB can differentiate supplier criticality. Future work will refine supplier risk measures and integrate trend analysis to quantify industry contraction or expansion by commodity. Additionally, a dynamic version of this modeling application is in work, supporting modeling and simulation of the DIB to quantify the consequences of systemic failures further.

References

Abd-El-Barr, M. (2009). Review: Topological network design: A survey. *Journal of Network and Computer Applications*, *32*(3), 501–509. https://sciencedirect.com/science/article/pii/s108480450800101x

Alvarez-Socorro, A. J., Herrera-Almarza, G. C., & González-Díaz, L. A. (n.d.). Eigencentrality based on dissimilarity measures reveals central nodes in complex networks. *Scientific Reports, 5*, 17095. http://www.nature.com/articles/srep17095

Asbjørnslett, B. (2008). Assessing the vulnerability of supply chains. In G. A. Zsidisin and B. Ritchie (Eds.), *Supply chain risk: A handbook of assessment, management & performance* (pp. 15–33). Springer.

Aviles, L., & Sleeper, S. (2016). *Identifying and mitigating the impact of the Budget Control Act on high-risk sectors and tiers of the defense industrial base: Assessment approach to industrial base risks*. https://apps.dtic.mil/sti/citations/AD1016751

Bendle, L. J., & Patterson, I. (2008). Network density, centrality, and communication in a serious leisure social world. *Annals of Leisure Research, 11*, 1–19. http://tandfonline.com/doi/pdf/10.1080/11745398.2008.9686783

Blackhurst, J. V., Scheibe, K. P., & Johnson, D. J. (2008). Supplier risk assessment and monitoring. *International Journal of Physical Distribution & Logistics Management*, 143–165.

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, *92*(5), 1170–1182.

Brintrup, A., Ledwoch, A., & Barros, J. (2016). Topological robustness of the global automotive industry. *Logistics Research*, *9*(1), 1–17. https://link.springer.com/article/10.1007/s12159-015-0128-1

Buechel, B., & Buskens, V. (2013). The dynamics of closeness and betweenness. *Journal of Mathematical Sociology*, *37*(3), 159–191. https://ideas.repec.org/p/bie/wpaper/398.html

Chan-Lau, J. A. (2018). Systemic centrality and systemic communities in financial networks. *Quantitative Finance and Economics*, 468–496.

Chopra, S., & Meindl, P. (2009). *Supply chain management* (4th ed.). Prentice-Hall.

DoD. (2008). *Defense Critical Infrastructure Program (DCIP): DoD mission-based Critical Asset Identification Process (CAIP)* (DoDM 3020.45-V1). Office of the Under Secretary of Defense for Policy.

DoD. (2018). *Assessing and strengthening the manufacturing and defense industrial base and supply chain resiliency of the United States.* Office of the Under Secretary of Defense.

Estrada, E., Higham, D. J., & Hatano, N. (2009). Communicability betweenness in complex networks. *Physica A: Statistical Mechanics and Its Applications*, *388*(5), 764–774. https://www.sciencedirect.com/science/article/abs/pii/S0378437108009473

Finch, P. (2004). Supply chain risk management. *Supply Chain Management*, *9*(2), 183–196. https://emerald.com/insight/content/doi/10.1108/13598540410527079/full/html

Fortunato, S., & Barthelemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(1), 36–41. http://www.pnas.org/content/104/1/36.abstract

GAO. (2017). *Defense supply chain: DOD needs complete information on single sources of supply to proactively manage the risks* (GAO-17-768).

GAO. (2018). *Integrating existing supplier data and addressing workforce challenges could improve risk analysis* (GAO-18-435).

González, A. M., Dalsgaard, B., & Olesen, J. M. (2010). Centrality measures and the importance of generalist species in pollination networks. *Ecological Complexity*, *7*(1), 36–43. https://sciencedirect.com/science/article/pii/s1476945x09000294

Hallikas, J. (2004). Risk management processes in supplier networks. *International Journal of Production Economics*, *90*(1), 47–58.

Hubbard, D. (2009). *The failure of risk management: Why it's broken and how to fix it.* John Wiley & Sons.

Lambert, D. M., & Cooper, M. C. (2000). Issues in supply chain management. *Industrial Marketing Management*, *29*(1), 65–83. https://sciencedirect.com/science/article/abs/pii/s0019850199001133

Manuj, I., & Mentzer, J. T. (2008). Global supply chain risk management strategies. *International Journal of Physical Distribution & Logistics Management*, *38*(3), 192–223.

Meyer, M., Brintrup, A., & Windt, K. (2014). Linking product and machine network structure using nested pattern analysis. *Procedia CIRP*, *17*, 278–283. https://sciencedirect.com/science/article/pii/s2212827114003412

Newman, M. E. (2008). The mathematics of networks. In *The new Palgrave dictionary of economics* (2nd ed.), 3-4. http://www-personal.umich.edu/~mejn/papers/palgrave.pdf

Nishat, F. M., Banwet, D. K., & Shankar, R. (2006). Mapping supply chains on risk and customer sensitivity dimensions. *Industrial Management and Data Systems*, *106*(6), 878–895.

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, *32*(3), 245–251. http://toreopsahl.com/2010/04/21/article-node-centrality-in-weighted-networks-generalizing-degree-and-shortest-paths/

Outdot, J. M. (2010). *Journal of Public Policy*, *30*(2), 201–218.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the web*. Stanford University InfoLab Publication Server. http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf

Perera, S., Bell, M. G., & Bliemer, M. C. (2018). *Network science approach to modelling emergence and topological robustness of supply networks: A review and perspective*. arXiv: Physics and Society. https://arxiv.org/abs/1803.09913

Rao, S., & Goldsby, T. J. (2009). Supply chain risks: A review and typology. *The International Journal of Logistics Management*, *20*(1), 97–112.

Ruhnau, B. (2000). Eigenvector-centrality—A node-centrality? *Social Networks*, *22*(4), 357–365. https://sciencedirect.com/science/article/pii/s0378873300000319

Schafer, M. H. (2011). Health and network centrality in a continuing care retirement community. *Journals of Gerontology: Series B, Psychological Sciences and Social Sciences*, *66B*(6), 795–803. https://academic.oup.com/psychsocgerontology/article/66b/6/795/593397

Sinha, P. R., Whitman, L. E., & Malzahn, D. E. (2004). Methodology to mitigate supplier risk in an aerospace supply chain. *Supply Chain Management*, *9*(2), 154–168. https://emerald.com/insight/content/doi/10.1108/13598540410527051/full/html

U.S. Department of the Treasury, Bureau of the Fiscal Service. (2021). *USAspending.gov*. https://www.usaspending.gov/

Vivas, R. d., Sant'Anna, A. M., Esquerre, K., & Freires, F. G. (2019). Integrated method combining analytical and mathematical models for the evaluation and optimization of sustainable supply chains: A Brazilian case study. *Computers & Industrial Engineering*, 105670. https://sciencedirect.com/science/article/pii/s0360835219300476

Wang, W., & Street, W. N. (2015). Modeling influence diffusion to uncover influence centrality and community structure in social networks. *Social Network Analysis and Mining*, *5*(1), 15. https://link.springer.com/article/10.1007/s13278-015-0254-4

Xu, M., Wang, X., & Zhao, L. (2014). Predicted supply chain resilience based on structural evolution against random supply disruptions. *International Journal of Systems Science: Operations & Logistics*, *1*(2), 105–117.

Xu, N-R., Liu, J-B., Li, D-X., & Wang, J. (2016). Research on evolutionary mechanism of agile supply chain network via complex network theory. *Mathematical Problems in Engineering*, 1, 1–9.

Design and Development of Data-Driven Risk Assessment Through the Integration of Federal Acquisition Data with Open, Internet Sources

Ningning Wu—is Professor of Information Science at the University of Arkansas at Little Rock. She received a BS and an MS degree in electrical engineering from the University of Science and Technology of China and a PhD in information technology from George Mason University. Wu's research interests are data mining, network and information security, and information quality. She holds certificates of the IAIDQ Information Quality Certified Professional (IQCP) and the SANS GIAC Security Essentials Certified Professional. [nxwu@ualr.edu]

M. Eduard Tudoreanu—is Professor of Information Science at the University of Arkansas at Little Rock. Tudoreanu has expertise in human–computer interaction, information quality, advanced visualization of complex data, and virtual reality. He worked on visual data analysis and has extensive experience in software development and user interface design. Professor Tudoreanu was the founding Technical Director of the Emerging Analytics Center. He has been the keynote speaker at ABSEL 2010 and served as a panelist for the National Science Foundation and Missouri EPSCoR. He earned his Doctor of Science degree in computer science in 2002 from Washington University in St. Louis. [metudoreanu@ualr.edu]

Sowjanya Manasa Sistla—is an enthusiastic graduate student who always has a zeal for learning. She earned a bachelor's degree in computer science. She worked at Accenture, where she started learning to analyze data and learned how deeply it is having an impact in the current world of technology, where almost everything is becoming dependent on data. Her work on researching data, finding credible sources for the data, and analyzing data to give estimates on the given input has been an amazing experience and a great opportunity to learn data analytical skills. [ssistla@ualr.edu]

Abstract

In acquisition planning and management, all information needs to be considered to ensure the successful execution of a project. Key pieces of information, with high data quality, must be used for accurate risk management that requires a good understanding and quick identification of the potential factors that might affect a program. This paper investigates effective data science techniques for a quick identification of high risks posed by natural disasters to a supply/production chain. The paper proposes a tool for helping a project manager analyze the integrity of a supply chain and find alternative potential contractors in lower risk areas. We focus on assessing the risks related to natural disasters at a contractor's location, business type, and business relationships and networks. The system described here seeks to first identify critical contractors based on the natural disaster risk of the places they perform their main activity for a contract and the uniqueness of their business types. Second, we design a risk assessment framework and a visualization framework that enable an acquisition expert to assess a project's risk by identifying contractors that may become critical and require additional redundancy. The visualization tool also can be used to determine additional potential subcontractors that are located in lower risks areas of the country.

Key words: visualization dashboard, data analytics, risk assessment, natural disasters

Introduction

The goal of this work is to assess and assign the risk level of a contractor based on its location and business properties. We built a tool to help identify alternative contractors for the high-risk ones. In addition to designing a risk assessment framework, our work relies on visualization tools to enable an acquisition expert to assess a program's risk by identifying the potential of failure for various contractors. Our approach also allows an expert to compare multiple acquisition programs to identify whether the same contractors or the same high-risk geographic areas occur in more than one project, thus becoming a single potential point of failure or delay for multiple programs.

Assessing the natural disaster risk of an area is a challenging task because multiple metrics and dimensions must be considered for a comprehensive picture. National Centers for Environmental Information (NCEI) has been collecting natural disaster data for each U.S. county since 1950. The data cover a wide range of natural disasters, including flood, tornado, hurricane, blizzard, high wind, flash flood, hail, dust storm, and so on. However, there is no standard way to categorize a disaster by its intensity and damage level. Even though some disaster types, such as tornadoes and hurricanes, do have a categorization system for their intensity, it is often difficult to assess an incident's impact to the local communities without other supporting information. Furthermore, it is hard to compare the impacts of different disaster types. One challenge of this study is to identify appropriate techniques to categorize the severity levels of natural disasters. New data sources will be researched to complement the NECI data for a better risk categorization system. One possible source is the Federal Emergency Management Agency (FEMA), which can provide information about the recovery efforts after a natural disaster strikes.

Critical contractors are those who provide unique products and services or have some special characteristics so that if they failed, it would be hard to find alternatives to fill their roles. Critical contractors could be a weak link in a supply chain, and their early identification will help the project manager control the risks effectively. The study will leverage the North American Industry Classification System (NAICS) code along with other business information for identifying the critical contractors.

NAICS is the standard used by federal agencies to describe the business specialization of a company (Office of Management and Budget, 2017). A NAICS code can be attached to many products and many companies, and so it can be used to identify potential alternative contractors for a program or a collection of programs. Because there might be many companies with the same NAICS code, to effectively identify critical contractors, acquisition experts may need to find more business-related information using the Internet or a data science–based tool such as the business risk indicators by Gill et al. (2019) or the Internet text processing framework by Wu et al. (2018).

We considered two case studies to demonstrate our framework. The first one focused on https://www.fpds.gov/ and extracted places of performance for various types of business used by the Navy over a period of 10 years. The second case study focused on awards with a high number of contractors in high-risk areas. These data were extracted from historical records from a data set we built based on https://www.usaspending.gov/. Based on the natural disaster risk assessment framework we developed, each U.S. county is assigned a risk level of high, medium, or low, depending on the historical NCEI weather data of the area as well as the FEMA assistance programs that had been utilized in the past. Then, the database was searched for contracts with a high percentage of contractors located in the high-risk areas. Contracts with a high concentration of risky contractors are prone to fail and thus require an effective risk management plan. One approach to reducing the overall risk of a contract is to replace high-risk contractors with low-risk ones.

The technique we present here allows an expert to find projects with a high number of contractors in high-risk areas. Once that has been accomplished, we provide two mechanisms for an expert to find possible alternative contractors. One is to find geographical areas with low risk of natural disasters and where there is a concentration of industries that could provide products and services for a given project. The second is to find actual contractors that have been involved in federal awards and are in areas with low disaster risks, regardless of the concentration of related industries.

Related Work

Previously, policy-makers and researchers have recognized the need to employ data as a multifaceted means of increasing the agility of the acquisition process (Krzysko & Baney, 2012). To this end, research has looked at automatic means of dealing with the heterogeneous acquisition data sources from text processing (Zhao et al., 2015), systems engineering (Cilli et al., 2015), and business (Gaither, 2014) perspectives. Our paper is different both in content and in approach—in content in that we rely on big data to identify hidden risk factors, and in the approach in that our expertise in information visualization, data quality, data governance and policy (chief data officers), and in data science provides a value-based perspective.

Gill et al. (2019) took a step further and employed data science techniques to determine a model that would assist an expert in source-selection decisions. Using a set of decision forests approaches, they distilled a number of risk indicators to predict which contractors are most likely to succeed in their federal contract obligations. Their method also uses publicly available data, but it focuses primarily on the financial and business aspects of the contractors. Our approach seeks to include natural disaster risk into the risk analysis model.

Tudoreanu et al. (2018) investigated employment data in an attempt to correlate changes in employment with negative modifications to contracts. Such correlations can be exploited to infer hidden and undisclosed contractors that are part of the defense acquisition network. Hidden contractors may pose the risk of becoming a weak stress point of a project and would affect the overall outcome of the project.

Wu et al. (2018) proposed a framework based on data science approach that aims to utilize the online information to assess and improve acquisition database quality as well as to find the hidden patterns to further acquisition research. The main component of the framework is a web-search and text mining module, whose main function is to search the internet and identify the most credible and accurate information online.

Apte et al. (2016) explored the use of big data analytics techniques to explore and analyze large data sets that are used to capture information about Department of Defense (DoD) services acquisitions. The paper described how big data analytics could potentially be used in acquisition research. As the proof of concept, the paper tested the application of big data analytic techniques by applying them to a data set of Contractor Performance Assessment Report System (CPARS) ratings of 715 acquired services. It also created predictive models to explore the causes of failed services contracts. Since the data set used in the research was rather small and far from the scope of big data, the techniques explored by the paper mainly focus on traditional data mining techniques without taking into account big data properties.

Black et al. (2014) studied the quality of narratives in CPARS and their value to the acquisition process. The research used statistical analysis to examine 715 Army service contractor performance reports in CPARS in order to understand three major questions: (1) To what degree are government contracting professionals submitting to CPARS contractor performance narratives in accordance with the guidelines provided in the CPARS user's manual? (2) What is the added value of the contractor performance narratives beyond the value of the objective scores for performance? (3) What is the statistical relationship between the sentiment contained in the narratives and the objective scores for contractor evaluations?

Previous Research Results

This paper builds upon the natural disaster risk model by Wu et al. (2020), and we review the main points here. A natural disaster risk was calculated by integrating two sources to data, NCEI and FEMA. Weather data of all U.S. counties between the years 1950 and 2018 from https://www.ncei.noaa.gov/ and natural disaster assistance data from https://www.fema.gov/ covering 1953 to 2020 was used to create a weighted disaster score (WDS). That score uses the declared assistance programs to assess a natural disaster's intensity and damage level. The score itself is then analyzed and simplified to a three-level scale—namely high, medium, and low—based on the distribution of the WDS. Figure 1 shows the distribution of risk levels both geographically and by location of federal contractor.

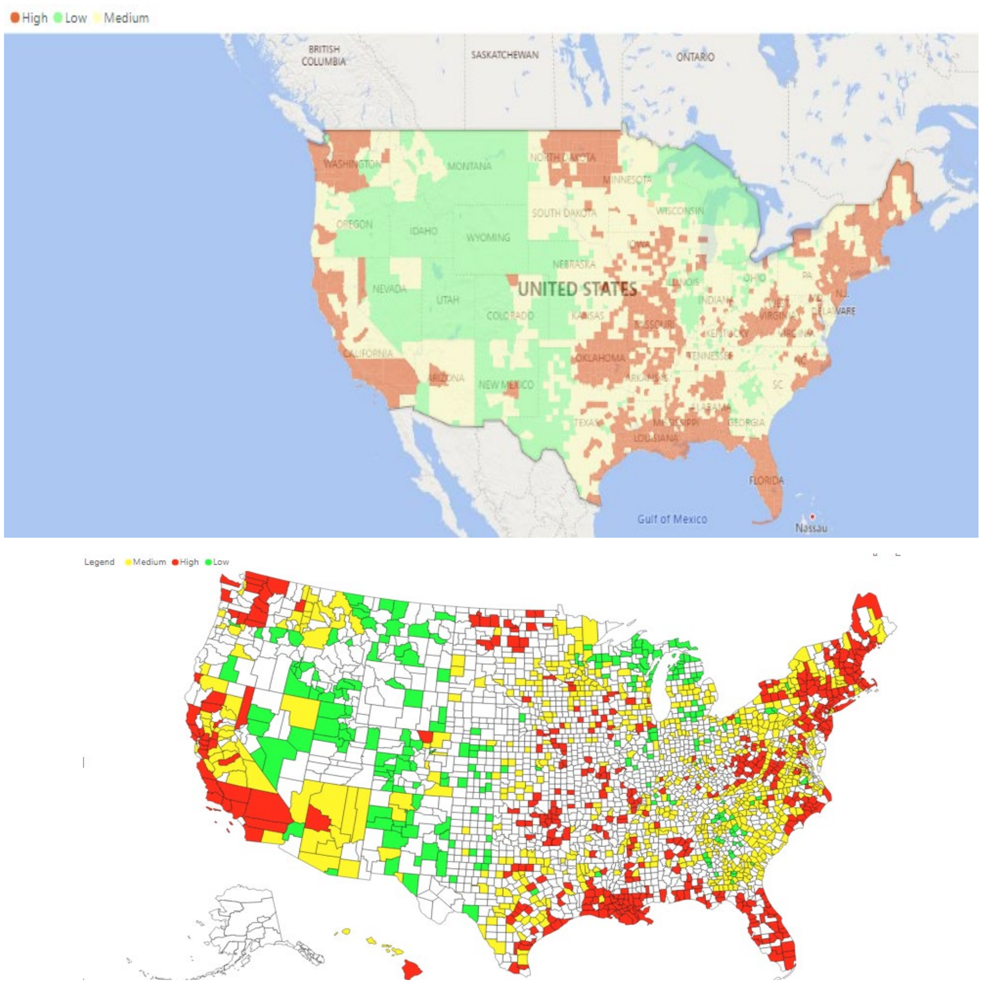


Figure 1. Natural Disaster Risk Level (Low, Medium, High), Displayed Geographically (Top) and by Location of Federal Contractor (Bottom) for the United States[21]

[21] The index takes into count both the number of occurrences of disasters and their magnitude.

Research Results

The work presented in this paper can be broken down into three main thrusts:

- Find awards with a large number of contractors in high-risk areas;
- Show alternative areas that have the desired industrial concentration, industry type, and lower natural disaster risk level;
- Show alternative individual contractors that are of the desired industry type and satisfy other geographical requirements a project might have.

All of these are meant to provide an acquisition expert the ability to assess the risk posed by natural disasters to a project or a set of projects. Furthermore, our goal is to also assist an expert in selecting alternative subcontractors in order to mitigate delays and failures posed by natural disasters.

Determine Projects with a Large Number of Contractors in High-Risk Area

The framework we propose is to maintain a database with the information from NCEI and FEMA. That database is periodically updated, and a new WDS is computed for every county in the United States. Specifically, we focus on FEMA disaster declarations and determine the number of disasters in an area during a period and the types of assistance programs declared. The data are cross-referenced with the NCEI to only consider natural disasters and eliminate man-made ones. WDS is calculated as follows:

$$s = \sum_{i=1}^{4} w_i \times n_i$$

where n_i is the number of a specific type of the assistance programs, and w_i is the corresponding weight for the type. The weight for each assistance program is defined as follows:

- Disaster mitigation: 0.25
- Public assistance: 0.50
- Housing assistance: 0.75
- Individual assistance: 1.0

The next step is to search for projects that have more contractors located in high-risk areas of natural disasters than in other types of areas. For https://www.usaspending.gov/, the process involves locating sub-awardees for each project, and it is relatively straightforward given the organization of the publicly available data. For https://www.fpds.gov/, the archives are provided in a different format, and we use a combination of data fields to match project to subcontractors. A different processing pipeline is used for each of the two data sources. For https://www.usaspending.gov/, database queries and Tableau are employed to obtain the dashboard, while for https://www.fpds.gov/, a Python and Jupyter Notebook pipeline was developed.

Alternative Geographical Areas for an Industry Type

We developed a technique to allow an acquisition expert to better inform their source-selection decision with regard to natural disaster risk. This thrust of our approach aims at providing the means to discover regions, in particular counties, where there exists some concentration of industry with a given profile. The level of concentration is to be set by an acquisition professional, and it involves the existence of a certain number of historical federal contracts that have been awarded in that region. The industry type can be specified through NAICS codes.

A program manager can examine the historical distribution of federal contractors for a given industry type (i.e., NAICS code). Figure 2 shows the distribution for a sample code that includes ventilation, heating, air-conditioning, and commercial refrigeration equipment manufacturing. In this scenario, the manager would then look at the areas that have medium or low risk of natural disaster and that also have been used in the past in many federal contracts. One possible view may be to show only those areas that have had an above-average number of federal awards (see Figure 3, top). If the selection is not sufficient, the analyst could lower the requirements on the total number of past federal contracts, which shows more potential regions with a lower natural-disaster risk (see Figure 3, bottom).

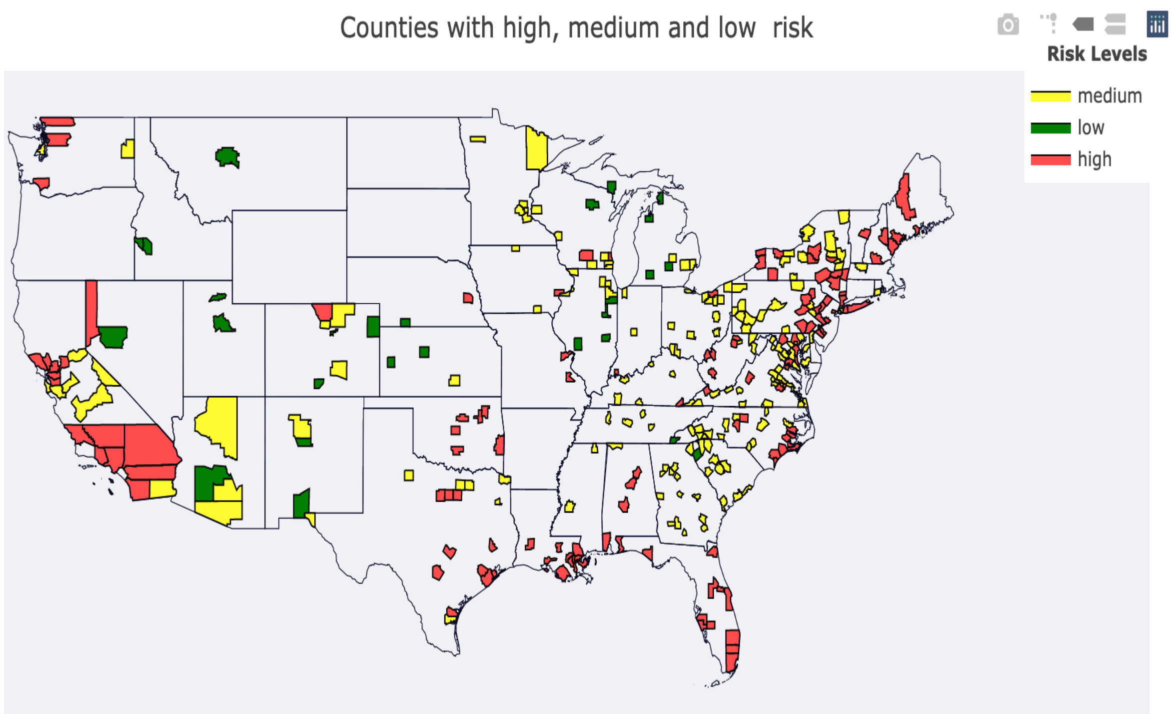


Figure 2. Distribution of Contractors from Historical FPDS Records for NAICS Code 33341, Department of the Navy[22]

[22] The color represents the natural disaster risk level for the place of performance.

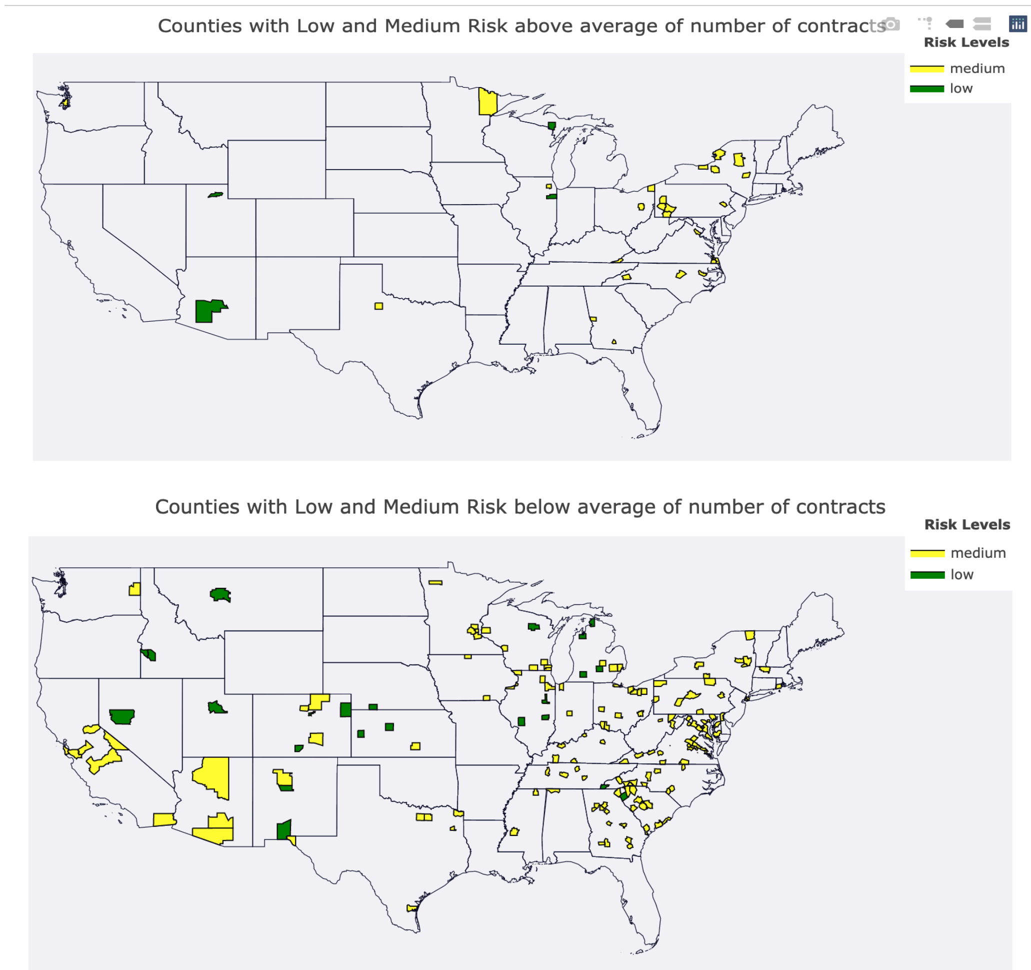


Figure 3. Regions in the Lower Risk of Natural Disasters That Have an Above Average of Past Federal Contracts (Top) and Below Average of Past Federal Contracts (Bottom)[23]

Alternative Contractors and Multiple Award Comparison

The data in https://www.usaspending.gov/ are categorized based on the types of spending, which include contracts, grants, loans, and other financial assistance. The information on contracts is organized into two tables: one for prime contracts and the other for subcontracts. The subcontract table was used in this study as it contained the detailed information of each award, including the award ID and funding agency's information, as well as the business name, address, DUNS number, and NAICS code of both primary contractor and each subcontractor. In addition, the address where each subcontractor physically performed the work is also listed in the table. Combining with the information from our natural disaster risk assessment framework, we are able to identify the awards with a high percentage of subcontractors located in the high-risk areas. Furthermore, we are able to

[23] Data is for NAICS code 33341, Department of the Navy contracts.

identify the subcontractors of a same business type based on their NAICS codes and partition them by their locations and the natural disaster risks of their locations.

For illustration purposes, two contracts with a higher-than-average concentration of contractors in high-risk areas were randomly selected and are apparent in Figure 4. One of the awards has some of the largest numbers of contractors; the other has a relatively small number of contractors. For each business type identified by a unique NAICS code, we partitioned the corresponding contractors into two groups: one located in a high-risk area, the other in an area with either a medium or low risk level. The contractors in the second group can be used to find alternatives to high-risk contractors when needed and are the subject of Figure 5.

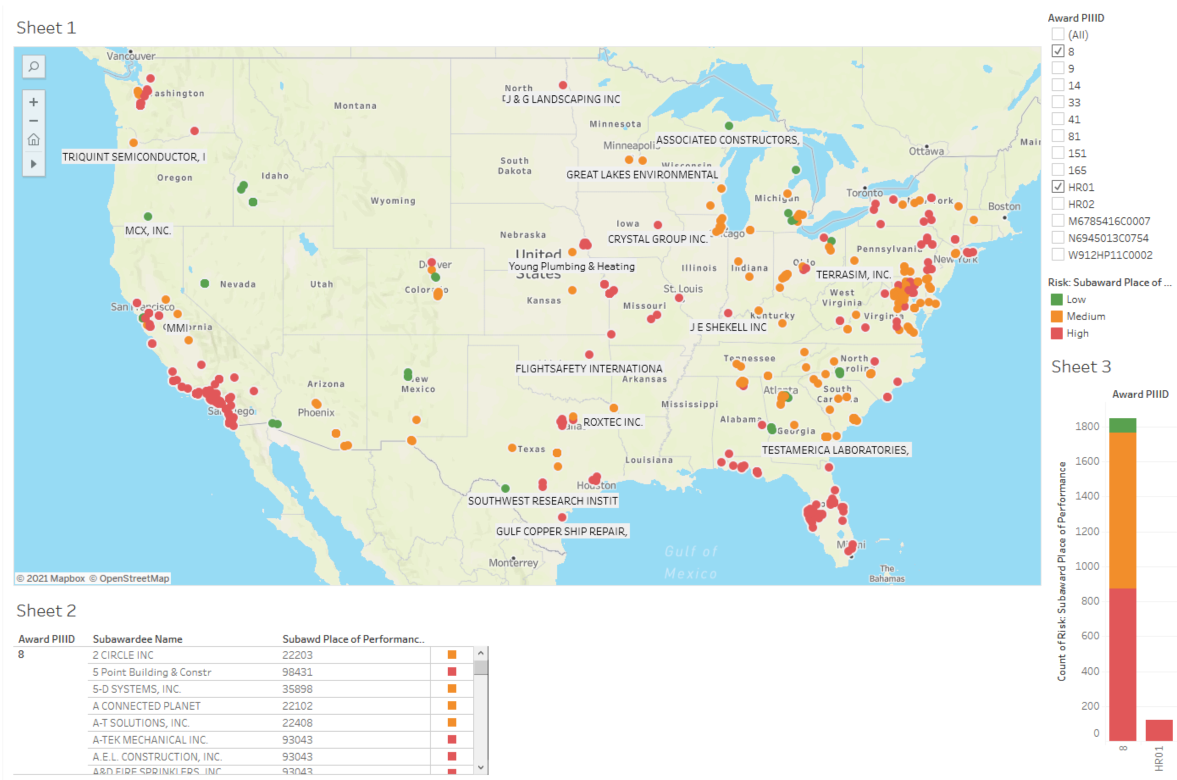


Figure 4. Dashboard Showing Sub-Awardees for Two Awards as Seen in the Checkboxes in the Top Right[24]

The first dashboard, Figure 4, is designed to provide a view of the distribution of subcontractors for a project. It consists of a map in which the places of performance are shown on a map and color-coded to reflect the natural disaster risk level. A list of awards of interest is provided on the top-right side. The bottom-right shows a bar chart, which gives an idea of the size of an award in terms of number of subcontractors. Stacked bars show how many contractors operate in areas with high, medium, or low risk. The final view, bottom-left, provides a list of contractors, places of performance of their services, and a color-coded risk

[24] The four panels are map, award selector, bar chart showing number and risk type for awardees, and individual contractor list. Panels are linked, and selection in any one of them determines what is displayed in the other three.

level. Note that the same contractor may be producing goods or performing services in multiple places.

The interaction with this dashboard allows an expert both to analyze a single award and to compare multiple ones. The panels are linked and selecting or clicking on one element in one panel results in changes in all the others. The map allows the selection of geographical regions and the analysis of the subcontractors in that area. The bar chart can be used to focus on subcontractors from any selected project and allows the user to focus further on those subcontractors who operate in the desired risk category: high, medium, or low. Finally, the individual list of subcontractors provides fine-grained details and allows the user to examine one contractor at a time in terms of location, map position, and natural disaster risk level.

In addition to analysis and comparison, a program manager may need to be able to find alternatives for some subcontractors in order to mitigate the risk posed by natural disasters. To this end, our framework can search existing and past federal projects to find such alternatives. The results are provided to a user visually through the dashboard shown in Figure 5. The data are automatically filtered to eliminate high-risk areas of performance, and the sheer number of potential solutions can be further narrowed down using NAICS codes. For this paper, we defaulted to a five-digit NAICS code, which can provide relatively closely related businesses. Just as the previous panel, the map is the main feature of the dashboard, and it is augmented by a NAICS code selection panel and a list of subcontractors.

The user of this panel would start by focusing on one or more likely related NAICS codes. Codes can be added into the analysis if a suitable subcontractor replacement is not found, or they can be removed if there are too many candidates. The map offers a way of finding subcontractors who are close to a geographical area, to waterways, or to highways if transportation is an issue to be considered by the project. Finally, the list of subcontractors provides both the location of the subcontractor and the place(s) of performance of the service or work.

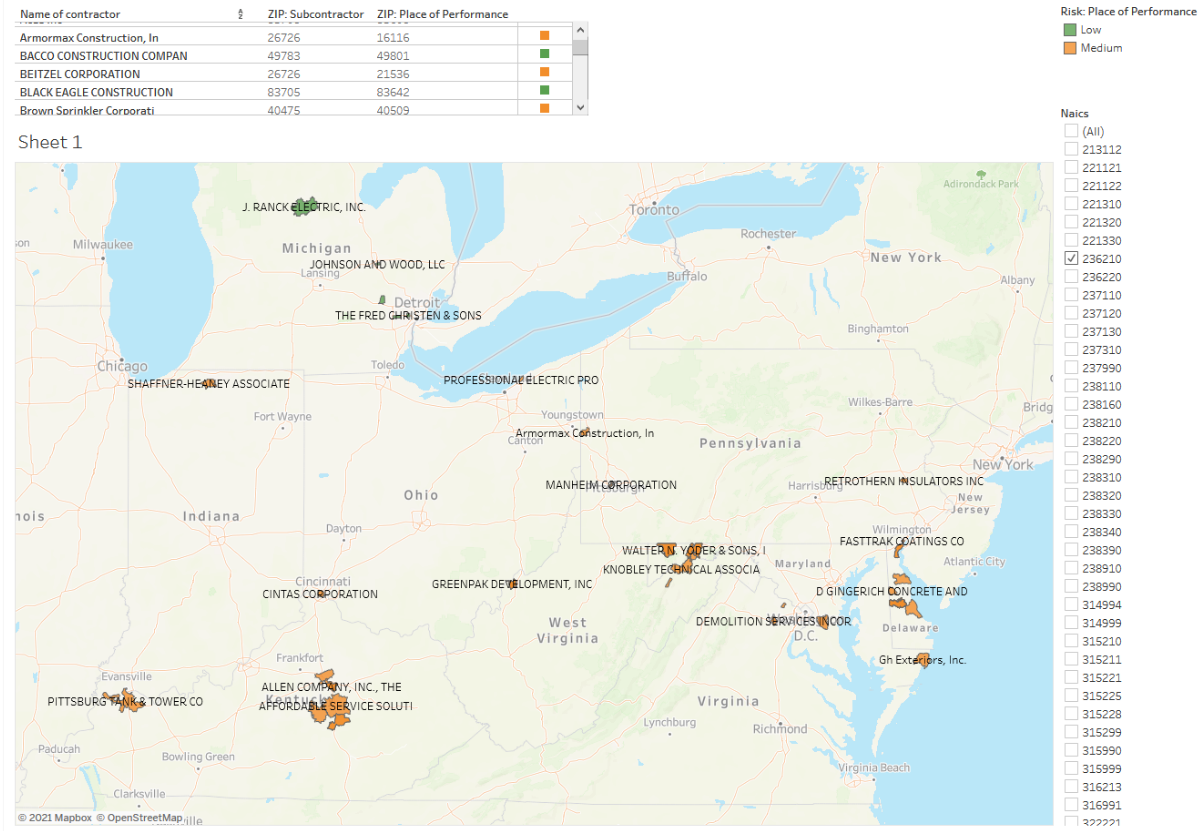


Figure 5. Dashboard for Finding Alternative Contractors by Name (Top Panel), NAICS Codes (Left Panel), or Geographical Location (Center Map)

Conclusion and Future Work

The paper presented a framework for allowing a user to manage the risk posed by natural disasters to a project or program. We developed visual tools to allow a program manager or planner to identify potential alternative subcontractors that operate in areas with a lower risk of natural disasters. Visual tools can also be used to analyze the existing risk of natural disasters for an existing project. The tool can provide various levels of details from an overview of the where and how many high-risk subcontractors are operating to individual listing of each contractor. Furthermore, a domain expert can compare multiple awards and look for common potential risk among those awards. The expert can also analyze concentrations of various industry types in locations that are both prone to natural disaster and relatively safe from them. Finally, all of these visual, user-friendly tools are supported by data science approaches to organize, manage, integrate, and transform the underlying data. The research results would be helpful for the acquisition management and planning to control the risks of natural disasters and their impacts to a project.

Our future work will focus on two directions. First, investigate other risk factors in order to develop a comprehensive, data-driven risk assessment framework that can be applied quickly to both large and small acquisition and purchasing projects. Second, explore the use of deep-learning and website crawling to provide more focused and real-time information about contractors to a domain expert.

References

Apte, U., Rendon, R., & Dixon, M. (2016). Big data analysis of contractor performance information for service acquisition in DoD: A proof of concept. *Proceedings of the 13th Annual Acquisition Research Symposium*, *2*, 398–418. https://calhoun.nps.edu/handle/10945/53512

Black, S., Henley, J., & Clute, M. (2014). *Determining the value of Contractor Performance Assessment Reporting System (CPARS) narratives for the acquisition process* (NPS-CM-14-022) [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun. https://calhoun.nps.edu/handle/10945/42583

Cilli, M., Parnell, G. S., Cloutier, R., & Zigh, T. (2015). A systems engineering perspective on the revised Defense Acquisition System. *Systems Engineering*, *18*(6), 584–603.

Gaither, C. C. (2014). Incorporating market based decision making processes in defense acquisitions. *International Journal of Defense Acquisition Management*, *6*, 38–50.

Gill, D., Muir, W. A., & Rendon, R. G. (2019). Predicting federal contractor performance issues using data analytics. *Proceedings of the 16th Annual Acquisition Research Symposium, 2*, 16–27. https://calhoun.nps.edu/handle/10945/62901

Krzysko, M., & Baney, B. (2012). The need for "acquisition visibility." *Journal of Software Technology*, *15*(1), 4–9.

Office of Management and Budget. (2017). *North American Industry Classification System*. Executive Office of the President. https://www.census.gov/eos/www/naics/2017NAICS/2017_NAICS_Manual.pdf

Tudoreanu, M. E., Franklin, K., Rego, A., Wu, N., & Wang, R. (2018). Searching hidden links: Inferring undisclosed subcontractors from public contract records and employment data. *Proceedings of the 15th Annual Acquisition Research Symposium*, *2*, 491–499. https://calhoun.nps.edu/handle/10945/58788

Wu, N., Tudoreanu, M. E., Siruvolu, S., Wang, R., & Talburt, J. R. (2020). Analysis of contractor data in federal acquisition databases. *Proceedings of the 17th Annual Acquisition Research Symposium*, 1–11. https://calhoun.nps.edu/handle/10945/64767

Wu, N., Tudoreanu, M. E., & Wang, R. (2018). Leveraging public data for quality improvement and pattern discovery of federal acquisition data*. Proceedings of the 15th Annual Acquisition Research Symposium*.

Zhao, D., MacKinnon, D., Douglas, J., Gallup, D., & Shelley, P. (2015). *Lexical Link Analysis (LLA) application: Improving web service to Defense Acquisition Visibility Environment* (NPS-AM-15-129). Naval Postgraduate School. https://dair.nps.edu/handle/123456789/2653

Building Industrial Resilience with a Little Help from Our Friends: Adapting DoD Acquisition Processes to Facilitate Allied and Partner Engagement

John G. (Jerry) McGinn—is the Executive Director of the Center for Government Contracting in the School of Business at George Mason University (GMU). In this role, he has established and is leading the first-of-its-kind university center for research, education and training, and collaboration on issues facing the more than $500 billion government contracting industry. McGinn is also a trusted strategic advisor sought after for his expertise in U.S. industrial policy, supply chain, the Defense Production Act (DPA), industrial security, export control, foreign military sales, and industrial base policies. He currently serves on the Economic Security Subcommittee of the Homeland Security Advisory Council. [jmcginn5@gmu.edu]

Abstract

This paper focuses on numerous existing and recent initiatives and programs involving allied and partner support to the U.S. defense industrial base and explores how they are currently instantiated in DoD acquisition processes. Are DoD acquisition processes able to effectively utilize efforts such as the National Technology Industrial Base (NTIB), Reciprocal Defense Procurement Agreements, Defense Production Act (DPA) Title III, and others to foster programs with partners that build industrial resilience in the defense industrial base? Research and interviews found that most of these efforts have promising foundations, but DoD acquisition processes need adaption to effectively increase allied and partner involvement. The author makes a series of recommendations to address these findings about the programs analyzed and the acquisition system itself. The author concludes that reframing acquisition processes and programs to better include partners and allies is ultimately a win-win proposition for all parties involved. Pursuing this approach will help to provide a concrete foundation for the future of international industrial collaboration and will build the industrial resilience needed to face the national security challenges of today and tomorrow. The paper also points to areas for further research to further refine the recommendations and explore specific areas in greater depth.

Research Question

This paper focuses on the numerous existing and recent programs and initiatives focused on building resilience in the U.S. defense industrial base and examines how current DoD acquisition processes support or hinder the involvement of allied and partner governments and companies. Specifically, there are a large number of existing and recent programs or initiatives that are designed, at least in part, to enable the involvement of companies based outside of the United States in U.S. defense acquisition programs. These efforts include Reciprocal Defense Procurement Agreements, the National Technology Industrial Base (NTIB), Trusted Capital Marketplace (TCM), and Defense Production Act (DPA) Title III. How are these programs impacting DoD acquisition processes, and are close allies and partners able to effectively contribute to building resilience in the U.S. defense industrial base?

Need for Increased U.S. Defense Industrial Base Capacity

The annual Department of Defense (DoD) industrial capabilities report to Congress lays out key trends in the defense industrial base. To quote from the Fiscal Year (FY) 2020 report, “The defense industrial base is the key to preserving and extending U.S. competitive military dominance in the coming century and, with it, deterrence that will keep Americans safe and keep the peace” (DoD, 2021). This report and previous annual reports showed fragility at lower levels of the defense supply chain, workforce challenges, troubling trends in the sole sourcing of materials such as rare earth elements and chemicals from China, and chronic weaknesses in areas such as microelectronics.

China is the principal challenge as outlined in the current National Security Strategy (White House, 2017) and National Defense Strategy (DoD, 2019). These legacy documents from the Trump administration will change in priorities and focus under the Biden administration, but the new administration has made clear in its Interim National Security Strategic Guidance (White House, 2021) and other venues that a focus on China as the "pacing" national security challenge will remain (Shelbourne, 2021).

China has co-opted the free market system for its purposes. Today, China is no longer focused on being the locus for low-cost manufacturing and is looking to move up the value chain and gain intellectual property. The Made in China 2025 strategic plan clearly demarcated the areas where Chinese companies and state-owned enterprises would invest (robotics, AI, autonomy, etc.) and the investment has followed (Sutter, 2020). The approach for this is multifaceted: from licit business transactions to illicit intellectual property theft. Moreover, the truth of the matter is that any Chinese company, under the 1993 state law, is compelled to provide information deemed important for national security to the Chinese government (Brown & Singh, 2018).

These industrial base concerns have led to a greater focus on reshoring industrial capabilities previously conducted overseas to domestic locations. This prioritization began in earnest after the whole of government review of the U.S. manufacturing and defense industrial base launched in July 2017 by Executive Order No. 13,806 (2017). This effort underscored similar themes found in earlier reports but with greater fidelity and detail. The recommendations coming out of the final report, released in September 2018, focused first and foremost on addressing industrial base weaknesses where the United States was reliant on Chinese single and sole sources in areas such as rare earth elements, specialty chemicals, and small unmanned aerial systems (Executive Order No. 13,806, 2017).

DoD industrial base investment has concentrated on these and other priorities, but the concern about China has also led to an increased focus on Buy America efforts in some quarters. During 2020, for example, a congressional effort to expand Buy America in the FY2021 National Defense Authorization Act (NDAA) would have required 75% of Pentagon major defense acquisition programs to be procured solely from U.S. sources by 2021 and 100% by 2026. This provision would have obviated the industrial contributions by companies headquartered in allied countries, many of whom have significant U.S. subsidiaries (Greenwalt, 2020). This Buy America approach undermines those close allies that support thousands of American jobs through the development and purchase of U.S. defense systems. Because of existing Buy America regulations and national security priorities, defense systems are already one of the strongest domestic manufacturing sectors. Eliminating key international suppliers, many of whom have significant U.S. physical and economic presence, from our defense industrial base is counterproductive. Moreover, these countries—mainly our NATO allies and close partners like Japan, Australia, and Israel—buy billions of dollars of U.S. defense systems each year (McGinn, 2020).

Rather than excluding them, it is important to look at how partners and allies can contribute to efforts to building resilience in the U.S. defense industrial base.

International Involvement in DoD Acquisition

The principal objective of DoD acquisition system is to get the absolute best capabilities to meet U.S. warfighter needs. DoD programs do this primarily through contracting with U.S. companies large and small. However, given the global nature of technology in general and the aerospace and defense business in particular, allied and partner governments and companies have been participating in DoD programs for decades. Participation occurs in three principal ways: (1) as a contributor to a program, (2) as a customer, and (3) as part of an international cooperative program. I will briefly discuss each in turn.

International Contributions to DoD Programs

The rationale for involving non-U.S. companies in DoD programs derives directly from the objective of the defense acquisition system. In some cases, the best technology or system for a program can be obtained from a non-U.S. source. The same goes for technology, subsystems, and components procured by U.S. prime integrators. Working as prime contractors, subcontractors, and suppliers, foreign sources have made significant contributions to the U.S. defense industrial base.

In addition, many foreign companies have U.S.-based subsidiaries that manufacture products or conduct services for unclassified and classified DoD programs. For those conducting classified work, these subsidiaries operate under foreign ownership, control, or influence (FOCI) regulations governed by the Defense Counterintelligence and Security Agency, which limits communications and sharing of information between the parent company and the U.S. subsidiary (Defense Counterintelligence and Security Agency, n.d.-a). Regardless, U.S.-based subsidiaries of foreign-owned companies are considered U.S. companies for purposes of DoD acquisition.

International Purchases of U.S. Systems

The purchase of DoD systems by international customers through foreign military sales (FMS) or direct commercial sales (DCS) is another major avenue for international participation in DoD acquisition. FMS are conducted via government-to-government agreements overseen by the Department of State (State) and executed by the DoD through the Defense Security Cooperation Agency (DSCA), while DCS are conducted via government-to-industry competitions, and sales are licensed by State or the Department of Commerce under their respective export control regimes governing military or commercial dual-use items.

U.S. international defense sales are significant and number in the tens of billions of dollars each year, but it is very difficult to track actual dollar expenditures associated with specific FMS and DCS transactions. This is because FMS are reported by DSCA in proposed letters of offer and acceptance that cite an “estimated” price for a “possible” sale (Defense Counterintelligence and Security Agency, n.d.-b). This is required under the U.S. Arms Export Control Act. DCS transactions, meanwhile, are reported by the purchasing governments and/or the U.S. company producing the system. The U.S. government reports the granting of an export license to authorize the sale of items via DCS, but the authorized value reported does not equate to a sale or the actual sale dollar amount (Department of State, 2021). The Stockholm International Peace Research Institute (SIPRI) provides the most comprehensive estimate of international transfer in its Arms Transfers Database (Stockholm International Peace Research Institute, n.d.).

International Cooperative Programs

DoD International Cooperative Programs (ICPs) are acquisition partnership arrangements between the United States and foreign countries that are established via memoranda of understanding (MoU) or project agreements (PAs) that often combine both foreign government and foreign commercial source participation. ICPs began in the 1960s, most prominently with the NATO Sea Sparrow missile, and continue today (Kenlon, 2018).

The largest ICP by far is the F-35 Joint Strike Fighter, with eight countries initially participating in the program. Each country contributed various amounts to the research and development of the program, receiving in return various levels of participation (Gertler, 2020). Numerous other major programs have ICPs, including the Guided Multiple Launch Rocket System (GMLRS), P-8 Maritime Patrol Aircraft, NATO Alliance Ground Surveillance, and SM-3 Block IIA. There have only been a few new major system ICPs established in the past decade, however (Kenlon, 2018).

International Impacts on the U.S. Defense Industrial Base

These international contributions in DoD programs have both direct and indirect impacts on the U.S. defense industrial base:

- **Direct impacts.** Many of these contributions have direct impacts on the health of the U.S. defense industrial base through employment and purchases. Specifically,
  - U.S. subsidiaries of foreign-headquartered firms employ tens of thousands of American workers through the production of systems and the performance of services in direct support of DoD programs. The United Kingdom (UK), for example, has estimated that U.S. subsidiaries of UK companies employ more than 56,000 U.S. personnel (Pierce, 2020).
  - The purchase of DoD systems by foreign government customers through FMS and DCS support the development and production of U.S. programs. Some programs, most notably the F-16 fighter, continue in production today solely because of these continued international sales.
  - ICPs such as the F-35, GMLRS, and P-8 support the continued production and product improvement of these U.S.-based programs.
- **Indirect impacts.** Other international participation in DoD acquisition programs also have a more indirect impact on the U.S. defense industrial base. For example,
  - The purchase of foreign-produced subsystems and parts by DoD programs contribute to the production and sustainment of programs across the DoD enterprise. These contributions do not have a direct impact on U.S. jobs or facilities, but they do enable the successful performance of a program and therefore support the economic impact of the program over the acquisition life cycle.
  - Partnering relationships established by the DoD and U.S. industry with their foreign counterparts can lead to additional teaming in technology development and product improvement efforts beyond the scope of the original program, leading to expanded U.S. industrial base opportunities in the global defense marketplace.[25]

Successes and Challenges

There have been recent notable successes in the involvement of international-headquartered firms in DoD programs. For example, Fincantieri Marinette Marine, the U.S. subsidiary of the Italian shipbuilding Fincantieri, won the Navy's Future Frigate competition in April 2020. The company invested nearly $180 million in its Wisconsin facility over several years, which helped the company position for its win of this almost $800 million contract (Fincantieri Marinette Marine, 2020). In addition, in this year's submission of proposals for

25. Thanks to Frank Kenlon for suggesting this impact.

the preliminary design phase of the U.S. Army's Optionally Manned Fighting Vehicle (OMFV), three of the four announced bidders are led non-U.S. headquartered companies: BAE Systems (teamed with Elbit Systems of America), Rheinmetall (teamed with Raytheon Technologies and L3Harris), and Hanwha (teamed with Oshkosh; Callan, 2021).

In the regulatory space, the FY2019 NDAA created an exemption for NTIB entities in the United States operating under a special security agreement to not be required to obtain a national interest determination for access to proscribed information (John S. McCain National Defense Authorization Act [NDAA], 2018b). This streamlined the involvement of U.S. subsidiaries of NTIB countries performing highly classified programs for U.S. government agencies and was fully implemented in December 2020 (National Industrial Security Program Operating Manual [NISPOM], 2020).

However, significant challenges remain. Most of these challenges are either cultural within the DoD acquisition system or are an artifact of existing acquisition practices and processes. For example, in the development of system requirements in many DoD programs, requirements documents are frequently classified as SECRET NOFORN (or higher). This requires non-U.S. companies to obtain DoD technology security and foreign disclosure authorizations to even engage in classified discussions with DoD program offices or formally respond to DoD contracting officer requests for information that have classified aspects. This hinders the ability of these companies to better understand program office needs at early stages in program requirements formulation. It can even prevent them from responding to or making them less competitive for eventual DoD solicitations.

Even after winning a program, these types of challenges can continue. For instance, the intellectual property for a DoD program is designated U.S.-only. That is appropriate in most cases but can be problematic when a U.S. subsidiary is the prime. In the Navy's Future Frigate program, Fincantieri Marinette Marine was not permitted to speak with the parent company engineers to discuss engineering challenges during the design stage. This created difficulties that could have been more easily addressed if they could have obtained a U.S. government export license that would have enabled the U.S. subsidiary to conduct authorized technical discussions with parent company engineers (R. Hunt, personal communication, April 2, 2021).

Recommendations

There are several ways that the DoD can educate acquisition professionals and adapt existing processes to facilitate greater involvement of partners and allies in DoD programs:

- The DoD should promote current DoD 5000 series processes for incorporating allied and partner companies' technologies and systems into DoD programs.
- The DoD should improve the ability of acquisition professionals to develop requirements documents that facilitate the early involvement of allied and partner companies in DoD programs (e.g., avoid citing classified, U.S.-only documents in either informal or formal requests for information or solicitations where possible).
- The DoD should facilitate communications between U.S. subsidiary firms performing on DoD contracts and their parent foreign firms in areas of engineering with direct impact on the conduct of the system being developed or fielded (e.g., working with U.S. government export control organizations to authorize appropriate engineer-to-engineer engagement throughout the program life cycle).
- The DoD should reexamine its approach to ICPs to identify different types of cooperative opportunities leading to a new generation of programs in the coming years.

Existing and Emerging Initiatives and Programs

Reciprocal Defense Procurement and Acquisition Policy Memoranda of Understanding (RDP MOUs)

- ***Current Use***

There are currently 26 countries that have RDP MOUs with the United States (Defense Pricing and Contracting, n.d.). They are

- Australia
- Belgium
- Canada
- Czech Republic
- Denmark
- Egypt
- Estonia
- Federal Republic of Germany
- Finland
- France
- Greece
- Israel
- Italy
- Japan
- Latvia
- Luxembourg
- Netherlands
- Norway
- Poland
- Portugal
- Slovenia
- Spain
- Sweden
- Switzerland
- Turkey
- United Kingdom of Great Britain and Northern Ireland

These MOUs establish agreed upon procurement principles that foster transparency and openness to competition in each country's respective defense marketplace.

The largest tangible benefit for the non-U.S. signatory countries is that they are waived from Buy America provisions when competing for DoD programs (DFARS 225.872-1, 2021). The existence of this exemption is often not well recognized in some program offices or on Capitol Hill, and others are opposed to these exemptions in the first place, as described above. Moreover, the Buy America focus of the current administration, which is directly at odds with its Interim National Security Strategy Guidance emphasis on strengthening relationships with allies and partners, sends conflicting signals to allies and partners (White House, 2021).

In addition to its RDP MOU, Canada also has a Defence Production Sharing Agreement (DPSA), which guides defense trade procedures between the two allies (Canadian Commercial Corporation, n.d.). The DPSA is further codified within the U.S. Defense Federal Acquisition Regulation Supplement 225.870 (DFARS 225.870, 2021) and permits the Canadian Commercial Corporation to help Canadian firms compete for DoD opportunities.

- ***Potential Future Uses***

There is a clear lack of awareness of the RDP MOUs and what they mean for DoD acquisition and for strengthening the defense industrial base. More detailed analysis of the impact of the presence *and* spending of RDP MOU countries, for example, would help to explicate the benefits of allied participation in DoD acquisition. This analysis would assist in educating acquisition professionals, DoD officials, and congressional staff and members about these tangible benefits. This could also spur greater involvement in DoD programs, leading to more investment as well as economic and national security benefits. In addition, written statements by the Biden administration on their support for RDP MOUs would go a long to reassure signatory countries.

- ***Recommendations***
    - The administration would be well served to formally articulate their support for RDP MOUs as part of their emphasis to strengthen relationships with allies and partners.
    - The DoD or Congress should request an analysis of the impact of RDP MOU countries' contributions to the U.S. defense industrial base through participation in DoD programs *and* the purchase of U.S. defense systems through foreign military or direct commercial sales.
    - The DoD should work to increase awareness of and educate acquisition professionals across the services about the Buy America exemption for RDP MOU countries as well as the Canadian DSPA DFARS clause to help spur additional competition and innovative solutions in the U.S. defense industrial base.

Security of Supply Arrangements

- ***Current Use***

There are currently 9 bilateral Security of Supply Arrangements (SoSAs) between the United States and partner countries. Specifically, the following countries have SoSAs with the United States:

- Australia
- Canada
- Finland
- Italy
- Netherlands
- Norway
- Spain
- Sweden
- United Kingdom

Not surprisingly, all of the SoSAs are with RDP MOU countries. These arrangements implement part of the Declaration of Principles in the RDP MOUs and recognize the "mutual interdependence of supplies needed for national security" as well as calling for the signatories to "explore solutions for achieving assurance of supply" (Industrial Policy, n.d.-b). Some of the signatory nations have established industry codes of conduct as a measure of reliance of their respective industry partners to support defense priorities.

The most telling part of these efforts, however, is the fact that they are *arrangements*, not agreements. That underscores the relatively informal and voluntary nature of these bilateral initiatives. There arrangements are confidence-building measures, and there is value in that, but they are not formal commitments by the respective government signatories. Thus, it is not surprising that these arrangements have not been invoked directly in any specific case to date (Hasik, 2021).

- ***Potential Future Uses***

It is time to reexamine SoSAs and their use for today's national security challenges. They can continue to be utilized as confidence building measures, but the United States and signatory nations should also consider methods for strengthening them. For example, the DoD could explore ways to partner with countries on mutually beneficial efforts through SoSAs. An arrangement with Japan for microelectronics and arrangements with Brazil and India for chemicals, for instance, could strengthen industrial capacities in those critical areas.

- ***Recommendation***
  - The DoD should conduct a review of SoSAs with partner countries to determine the future of SoSAs, specifically how these (and any future) arrangements could be adapted to make them more relevant for today's global industrial security challenges.

- **National Technology Industrial Base**

- ***Current Use***

The NTIB has deep roots. Initially born out of the North American Technology and Industrial Base Organization (NATIBO; Government of Canada, n.d.), the NTIB was first codified in law in 1992. At that time, Congress required the DoD to report annually on "steps to foster and safeguard" the NTIB. The NTIB was defined at the outset to include the United States and Canada (Hunter et al., 2017). NATIBO conducted periodic bilateral studies on industrial base issues, but neither NATIBO nor NTIB gained significant visibility (or impact) until the NTIB was expanded to include Australia and the United Kingdom in the 2017 National Defense Authorization Act (NDAA, 2016).

The expansion of the NTIB brought greater attention to the need to increase industrial cooperation between these key allies (Greenwalt, 2019). The NTIB countries established an initial governance structure and identified areas for initial focus. The increased dialogue has been favorably viewed by the NTIB governments, but the NTIB made an immediate impact in area of foreign direct investment (FDI).

Prior to 2010, for example, the vast majority of FDI in the United States came from allied and partner countries. In less than a decade, those ratios shifted dramatically. From 2016 to 2018, transactions originating from China were the largest proportion of cases filed: 26.5%. Moreover, the nature of the Chinese transactions drew increased scrutiny because the vast majority of these proposed acquisitions (84%) were focused on the manufacturing, finance, information, and services sectors (Department of the Treasury, 2021).

This shift drew significant bipartisan attention on the Committee on Foreign Investment in the United States (CFIUS), which reviews foreign transactions for national security concerns, and led directly to a significant strengthening of CFIUS authorities through the Foreign Investment Risk Review Modernization Act of 2018 (FIRRMA; NDAA, 2018a). As the United States has strengthened its position on FDI, Chinese investment started to focus on other countries with advanced technology companies. Thus, NTIB governments have undertaken significant efforts to share best practices among NTIB countries to counter potential national security impacts.

NTIB was also featured in the Executive Order 13,806 final report of the U.S. manufacturing and defense industrial base. The final report, published in October 2018, recognized the global nature of the defense industrial base and underscored the importance of allies and partners. For example, the broadening of the NTIB is cited favorably: “These types of agreements [i.e., NTIB] with partners and allies provide economies of scale and scope, help facilitate cost-effective defense production, and increase Warfighter interoperability” (DoD, 2018).

- ***Potential Future Uses***

The governments are focusing on NTIB governance and on sharing best practices in FDI and will continue to do that. What else can be done? Analysts such as William Greenwalt (2019) argue that the NTIB should be used to foster export control reform among member countries. While that goal is laudable, there has been little progress on this front since the NTIB expansion. One potential related area that could be promising to pursue is releasability. For example, when companies from NTIB countries are working with program executive offices and program offices, those offices sometimes establish releasability provisos on the program’s technology. These provisos require companies not based in the United States to file for an export license. That delays the time before a foreign company can examine technical data and speak at a technical level with a DoD customer for a potential or actual solicitation, thereby making them less competitive. Establishing releasability criteria at the outset to include appropriate NTIB-based companies, for example, could create opportunities for greater competition for DoD customers.

The recently passed FY2021 NDAA had a provision that directed the DoD to establish criteria for expanding the NTIB to include additional countries (NDAA, 2021). This clearly indicates congressional intent to increase NTIB countries, and there have been numerous countries seeking to be part of the NTIB. As part of establishing any criteria for expansion, the DoD should look at creating specific opportunities for strengthening the ability of NTIB-based companies to contribute to the U.S. defense industrial base.

An immediate opportunity for success is to create acquisition pathways for DoD projects and programs to employ NTIB-based companies that are able to provide leading edge technology and affordable solutions to emerging DoD requirements. A concrete way for the DoD to incentivize NTIB contributions to U.S. efforts to strengthen the defense industrial base, for example, would be to create opportunities for companies based in NTIB countries to compete for projects and programs by simply changing acquisition rules to incentivize DoD acquisition personnel to consider facilitating NTIB offerings at prime, subcontractor, and supplier levels. The DoD has started to do this to a modest degree. In a March 2019 memorandum to DoD acquisition officials, for example, then Acting Principal Director of Defense Pricing and Contracting Kim Herrington specifically recommended the “inclusion of” NTIB members in innovation-focused Other Transactions (OT) consortia (Herrington, 2019). Memos like this are useful to set the conditions for change, but the NTIB needs to be formalized through rule changes and DFARS clauses that can be deployed in solicitations and contracts. Once the rules are changed, education and training can help to expand opportunities for NTIB companies (McGinn, 2021).

This education need also extends to industry. One of the biggest failures of the previous U.S.–UK and U.S.–Australian defense trade cooperation treaties, for example, is the fact that the governments did not get industry engaged early enough to incentivize companies to use the treaties. These treaties, approved by the U.S. Senate in late 2010, were designed to create a “trusted community” of companies that could share technology and compete for opportunities in this trusted community (Directorate of Defense Trade Controls, 2010). Unfortunately, they never realized their potential, and while they have been used for government-to-government efforts, they have almost never been used by industry.

The NTIB provides an opportunity to do better, but the governments need to help create pathways where industry can see the potential business benefit. DoD and NTIB governments can and should make the value proposition of NTIB clear to industry and then let the resulting business relationships grow and flourish. Industry will not always pursue these incentives, but there is 100% certainty that they will not pursue them in the absence of a clearly defined pathway to success.

- ***Recommendations***
  - Conduct rule-making and establish DFARS clauses focused on facilitating NTIB participation and membership in opportunities such as OT consortia, DPA Title III, IBAS, and other appropriate programs.
  - Educate the acquisition workforce on the use of NTIB clauses for use in programs across the DoD.
  - Advertise these NTIB-inclusive opportunities to NTIB countries and trade associations to facilitate additional solutions to U.S. industrial base challenges.

- **DPA Title III**

- ***Current Use***

The Defense Production Act (DPA) is a long-standing authority that derives from the Korean War. Passed in 1950 and drawing on the War Powers Acts of World War II, the DPA is a broad set of authorities to help the U.S. government strengthen the defense industrial base to respond to national emergencies. Title III of the DPA is focused specifically on the expansion of productive capacity and supply and utilizes grants, loans, loan guarantees, purchases, and purchase commitments to build industrial capacity (Cecire & Peters, 2020). The DoD has been delegated authority to execute Title III projects and has used this authority for decades to expand industrial capacity in areas such as the creation of a domestic beryllium production facility to complex forgings for naval propulsion shafts (Air Force Research Library, 2013; *Earmark Declaration*, 2009).

The COVID-19 pandemic response has put DPA Title III into overdrive. Where DPA annual appropriations fluctuated between $40 million and $100 million during the decade preceding the pandemic, $1 billion was appropriated for DPA Title III in the March 2020 CARES Act. The recently passed American Recovery Act dramatically upped the ante, appropriating $10 billion for current and future pandemic response that almost certainly will be allocated via DPA Title III (American Rescue Plan Act, 2021). The focus of these Title III projects is to increase domestic production capacity to reduce reliance on non-U.S. sources for items such as vaccine production, personal protective equipment (PPE), testing equipment, and so on. The primary focus is on reducing dependencies on Chinese sources, as COVID had exposed the dominant positions of China in areas such as PPE and antibiotics production (Bradsher & Alderman, 2020; Swanson, 2020).

- ***Potential Future Uses***

In efforts to reshore or onshore manufacturing capacity, one relatively unknown provision of the DPA defines the term *domestic source* as a business concern

> that performs in the United States or Canada substantially all of the research and development, engineering, manufacturing, and production activities required of such business concern under a contract with the United States relating to a critical component or a critical technology item. (Defense Production Act, 1950)

This permits Canadian firms to apply for DPA Title III grants and for Title III projects to be conducted in Canada as well. This is a legacy of long-standing U.S.–Canadian industrial base collaboration in the Cold War North American Aerospace Defense Command (NORAD) and NATIBO (Hunter et al., 2017). Given the geographic proximity and the expertise of Canadian industry in areas such as mining and chemicals, this could significantly benefit the U.S. defense industrial base. If the DPA were amended to include other NTIB countries in the definition of domestic source, this would substantially expand the opportunities for firms in Australia and the UK to contribute to strengthening the U.S. industrial base.

- ***Recommendations***
  - The DoD and Canada should promote the ability for Canadian-based firms to contribute to DPA Title III projects.
  - The DoD should submit a legislative proposal to amend the DPA to include all NTIB countries to reflect the fact that they are already part of the U.S. industrial base and make their respective industrial capabilities available to strengthen that base.

Industrial Base Analysis and Sustainment Program

- ***Current Use***

The Industrial Base Analysis and Sustainment (IBAS) program was established in 2014 in the DoD to fund the mitigation of defense industrial base issues (Nelson, 2016). Defined in 10 U.S.C. § 2508 (Industrial Base Fund, 2011), IBAS has four principal functions:

> (1) to support the monitoring and assessment of the industrial base … ;
> (2) to address critical issues in the industrial base relating to urgent operational needs;
> (3) to support efforts to expand the industrial base; and
> (4) to address supply chain vulnerabilities. (Industrial Base Fund, 2011)

With a similar mandate to DPA Title III, IBAS conducts projects to build industrial capabilities to support DoD priorities. IBAS has both an open Broad Agency Announcement (BAA) and the Cornerstone OT Authority to support a broad range of industrial base requirements (Industrial Policy, n.d.-a). IBAS has ranged from $10 to $100 million in appropriations annually, depending on congressional adds, and has conducted a wide array of projects ranging from updating naval propulsion foundry and electron beam welding to munitions and missile improvements (Defense and Aerospace Competitive Intelligence Service, 2019).

- ***Potential Future Uses***

IBAS is an active program that will continue to receive funding for industrial base projects. IBAS does not have a clause in its BAA or OT like DPA Title III explicitly including Canadian or other non–U.S.-headquartered firms, but there is no explicit restriction either. Firms based or headquartered in allied or partner countries are currently eligible to join the Cornerstone OTA, where the majority of IBAS contract opportunities are posted, on a case-by-case basis. Designating NTIB countries as eligible for IBAS projects through a DFARS clause or by changing the DPA definition of domestic source would create a number of new eligible firms to help strengthen the U.S. defense industrial base.

- ***Recommendations***
  - The DoD should consider creating and adding a DFARS clause to the IBAS BAA and Cornerstone OT making companies based in NTIB countries eligible to compete for IBAS opportunities to reflect the fact that they are already part of the U.S. industrial base.

- The DoD should advertise and encourage NTIB-based companies to join the Cornerstone OTA and to develop solutions to meet IBAS solicitations in the coming months.

Trusted Capital Management

- ***Current Use***

Trusted Capital Management (TCM) is one of the newest DoD efforts to strengthen the defense industrial base. After several previous aborted efforts, TCM formally launched in late 2020 with the creation of its digital marketplace. The overall objective of TCM is to reduce the vulnerability of high-technology start-ups funded by venture capital or private equity to funding from sources of adversarial capital, principally from China. Chinese technology priority areas in recent years have included many high-tech areas such as robotics, autonomy, and artificial intelligence. In addition to funding domestic sources of innovation in these areas, Chinese-based private equity and venture funds have invested in U.S.-based start-ups. Some of these investments have been shielded from easy discovery by start-ups. In response to some publicly revealed instances of adversarial investment, the DoD created TCM (Trusted Capital, n.d.).

TCM is intended to create a trusted clearinghouse for companies and investors to conduct business free of potential adversarial investment. This effort is still in its early stages, but approximately 50 venture capital firms and companies have been vetted, and additional firms are in the pipeline (McLeary, 2021).

- ***Potential Future Uses***

TCM was created first and foremost for U.S. companies and investors, but there has always been an appreciation for non-U.S. companies and investors in allied and partner countries given the global nature of investment and technology. To that end, TCM is open to venture capital (VC) funds and companies based in other countries as long as they go through the same vetting process of other firms in the trusted capital marketplace and are currently on a DoD contract.[26]

In addition, there is strong DoD interest in the development of TCM-like regimes in partner and allied countries or even in organizations such as the North Atlantic Treaty Organization (NATO). These could develop over time because U.S. allies are facing the same adversarial challenge of Chinese investments in their markets.

- ***Recommendations***
  - The DoD should promote TCM to allied- or partner-based VC firms and companies focused on DoD business.
  - The DoD and State should work with allied governments and NATO to establish TCM-like organizations in their respective countries or jurisdictions.

---

2. The latter requirement can be challenging for many start-ups, however, because current Small Business Administration rules do not permit foreign firms to compete for early-stage opportunities such as Small Business Innovation Research funding.

Conclusions

Allies and partners have long played a productive role in our defense acquisition system and have contributed to the U.S. defense industrial base. Given the common threat that we face in our respective supply chains, it is imperative to eliminate exposure to Chinese suppliers in critical national security areas. It makes little sense, however, to reshore all industrial capacity in a Buy America "Only" approach. Instead, we should focus first and foremost on those manufacturing areas where we are most vulnerable to China. In that effort, there are numerous manufacturing and resource areas where we can work with our close allies and partners to help achieve that common goal. From mining and chemicals to microelectronics and hypersonics, government and industry partnerships are synergistic and mutually beneficial for parties involved.

The focus, therefore, should be on actions to foster true international industrial collaboration. That is done through actual participation in defense programs. Interestingly, there is a surprising amount of that happening right under our noses. Leonardo partnered with Boeing to win the Huey helicopter replacement program; Saab partnered with Boeing to win the Air Force T-X deal; SAIC partnered with Singapore's ST Engineering and Belgium's CMI Defence to prototype a light tank for the Army; Fincantieri Marinette Marine's won the Navy's Future Frigate program and the Army OMFV partnerships described above; and the list goes on and on (Judson, 2018). Not all of these partnerships have been or will be successful, but they create more competition for the DoD customer and lead to more industrial base resilience. And that, ultimately, is the goal.

We need to create opportunity spaces for companies to operate within groups of "trusted communities," to borrow a phrase from the treaties. The preceding has detailed how we have started to build these communities, but there is a long way to go. The NTIB is best postured to become one of those trusted communities. Whether NTIB companies are small or medium-sized enterprises operating exclusively in one of these countries or if they are subsidiaries of U.S.-headquartered primes, these companies are now part of *one* industrial base. The RDP MOU countries are a different trusted community, and we can build on these over time.

In sum, reframing acquisition processes and program to more effectively include partner and allied government and industry participation is ultimately a win-win proposition for all parties involved. Pursuing programs, initiatives, and recommendations like those described in this paper will help to provide a concrete foundation for the future of international industrial collaboration and will build industrial resilience we need to face the national security challenges of today and tomorrow.

Summary Table of Recommendations

Area	Recommendation
General	• The DoD should promote current DoD 5000 series processes for incorporating allied and partner companies' technologies and systems into DoD programs.
	• The DoD should improve the ability of acquisition professionals to develop requirements documents that facilitate the early involvement of allied and partner companies in DoD programs (e.g., avoid citing classified, U.S.-only documents in either informal or formal requests for information or solicitations where possible).
	• The DoD should facilitate communications between U.S. subsidiary firms performing on DoD contracts and their parent foreign firms in areas of engineering with direct impact on the conduct of the system being developed or fielded (e.g., working with U.S. government export control

Area	Recommendation
	organizations to authorize appropriate engineer-to-engineer engagement throughout the program life cycle).
	• The DoD should reexamine its approach to ICPs to identify different types of cooperative opportunities leading to a new generation of programs in the coming years.
RDP MOUs	• The administration would be well served to formally articulate their support for RDP MOUs as part of their emphasis to strengthen relationships with allies and partners.
	• The DoD or Congress should request an analysis of the impact of RDP MOU countries' contributions to the U.S. defense industrial base through participation in DoD programs *and* the purchase of U.S. defense systems through foreign military or direct commercial sales.
	• The DoD should work to increase awareness of and educate acquisition professionals across the services about the Buy America exemption for RDP MOU countries and the Canadian DPSA DFARS clause to help spur additional competition and innovative solutions in the U.S. defense industrial base.
SoSAs	• The DoD should conduct a review of SoSAs with partner countries to determine the future of SoSAs, specifically how these (and any future) arrangements could be adapted to make them more relevant for today's global industrial security challenges.
NTIB	• Conduct rule making and establish DFARS clauses focused on facilitating NTIB participation in solicitations for industrial base opportunities such as DPA Title III, IBAS, and other appropriate programs.
	• Educate the acquisition workforce on the use of NTIB clauses for use in programs across the DoD.
	• Advertise these NTIB-inclusive opportunities to NTIB countries and trade associations to facilitate additional solutions to U.S. industrial base challenges.
DPA Title III	• The DoD and Canada should promote the ability for Canadian-based firms to contribute to DPA Title III projects.
	• The DoD should submit a legislative proposal to amend the DPA to include all NTIB countries to reflect the fact that they are already part of the U.S. industrial base and make their respective industrial capabilities available to strengthen that base.
IBAS	• Conduct rule-making and establish DFARS clauses focused on facilitating NTIB participation and membership in opportunities such as OT consortia, DPA Title III, IBAS, and other appropriate programs.
	• The DoD should advertise and encourage NTIB-based companies to join the Cornerstone OTA and to develop solutions to meet IBAS solicitations in the coming months.
TCM	• The DoD should promote TCM to allied- or partner-based VC firms and companies focused on DoD business.
	• The DoD and the Department of State should work with allied governments and NATO to establish TCM-like organizations in their respective jurisdictions.

References

Air Force Research Laboratory. (2013, September 17). *Defense Production Act Title III project establishes domestic source for beryllium*. Wright-Patterson Air Force Base. https://www.wpafb.af.mil/News/Article-Display/Article/819343/defense-production-act-title-iii-project-establishes-domestic-source-for-beryll/

American Rescue Plan Act of 2021, H.R. 1319, 117th Cong. (2021). https://www.congress.gov/bill/117th-congress/house-bill/1319/text

Bradsher, K., & Alderman, L. (2020, April 2). The world needs masks. China makes them, but has been hoarding them. *The New York Times.* https://www.nytimes.com/2020/03/13/business/masks-china-coronavirus.html

Brown, M., & Singh, P. (2018, January). *China's technology transfer strategy: How Chinese investments in emerging technology enable a strategic competitor to access the crown jewels of U.S. innovation*. Defense Innovation Unit Experimental. https://admin.govexec.com/media/diux_chinatechnologytransferstudy_jan_2018_(1).pdf

Callan, B. (2021, April 20). *OMFV bidders—Potential for eventual seismic shifts in U.S. armor*. Capital Alpha Partners.

Canadian Commercial Corporation. (n.d.). *Defence production sharing agreement between Canada and the United States*. Retrieved April 23, 2021, from https://www.ccc.ca/wp-content/uploads/2019/05/defence-production-sharing-agreement-en.pdf

Cecire, M. H., & Peters, H. M. (2020, March 2). *The Defense Production Act of 1950: History, authorities, and considerations for Congress* (CRS Report No. R43767). Congressional Research Service. https://crsreports.congress.gov/product/pdf/R/R43767

Defense and Aerospace Competitive Intelligence Service. (2019, February). *Exhibit R-2, RDT&E budget item justification: PB 2020 Office of the Secretary of Defense*. https://www.dacis.com/budget/budget_pdf/FY20/RDTE/D/0607210D8Z_199.pdf

Defense Counterintelligence and Security Agency. (n.d.-a). *Foreign ownership, control or influence*. Retrieved April 20, 2021, from https://www.dcsa.mil/mc/ctp/foci/

Defense Counterintelligence and Security Agency. (n.d.-b) *Major arms sales*. Retrieved April 20, 2021, from https://www.dsca.mil/press-media/major-arms-sales

Defense Pricing and Contracting. (n.d.). *Reciprocal defense procurement and acquisition policy memoranda of understanding*. Retrieved April 13, 2021, from https://www.acq.osd.mil/dpap/cpic/ic/reciprocal_procurement_memoranda_of_understanding.html

Defense Production Act of 1950, 50 U.S.C. § 55 (1950). https://uscode.house.gov/view.xhtml?path=/prelim@title50/chapter55&edition=prelim

Department of the Treasury. (2021). *CFIUS reports and tables*. https://home.treasury.gov/policy-issues/international/the-committee-on-foreign-investment-in-the-united-states-cfius/cfius-reports-and-tables

DoD. (2018, September). *Assessing and strengthening the manufacturing and defense industrial base and supply chain resiliency of the United States: Report to President Donald J. Trump by the Interagency Task Force in fulfillment of Executive Order 13806*. https://media.defense.gov/2018/oct/05/2002048904/-1/-1/1/assessing-and-strengthening-the-manufacturing-and%20defense-industrial-base-and-supply-chain-resiliency.pdf

DoD. (2019, July). *Summary of the 2018 national defense strategy of the United States of America: Sharpening the American military's competitive edge*. https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf

DoD. (2021, January). *Fiscal Year 2020 annual industrial capabilities report to Congress*. https://www.businessdefense.gov/resources/

DFARS 225.870, Contracting with Canadian Contractors (2021). https://www.acq.osd.mil/dpap/dars/dfars/html/current/225_8.htm

DFARS 225.872-1, Contracting with Qualifying Country Sources (2021). https://www.acq.osd.mil/dpap/dars/dfars/html/current/225_8.htm#225.872-1

Directorate of Defense Trade Controls. (n.d.). *2019 Section 665 Report reports over $55 billion in "authorized value" and over $2.5 billion in "shipped value," but the actual value of 2019 DCS sales is not tracked or reported*. Department of State. Retrieved April 20, 2021, from https://www.pmddtc.state.gov/sys_attachment.do?sysparm_referring_url=tear_off&view=true&sys_id=3b39303bdb1414105c3070808c9619e6

Directorate of Defense Trade Controls. (2010). *United Kingdom & Australia defense trade cooperation treaties*. Department of State.

https://www.pmddtc.state.gov/ddtc_public?id=ddtc_kb_article_page&sys_id=2d21e8b3dbb8d300d0a370131f96190b
Earmark declaration, 111th Cong. (2009) (testimony of Charles W. Dent). https://www.govinfo.gov/content/pkg/CREC-2009-07-29/html/CREC-2009-07-29-pt1-PgE2057-3.htm
Exec. Order No. 13,806, 3 C.F.R. 34597 (2017). https://www.govinfo.gov/content/pkg/FR-2017-07-26/pdf/2017-15860.pdf
Fincantieri Marinette Marine. (2020, April 30). *Fincantieri Marinette Marine lands future frigate contract*. https://fincantierimarinettemarine.com/fincantieri-marinette-marine-lands-future-frigate-contract/
Gertler, J. (2020, May 27). *F-35 joint strike fighter (JSF) program* (CRS Report No. RL30563). Congressional Research Service. https://fas.org/sgp/crs/weapons/RL30563.pdf
Government of Canada. (n.d.). *Doing business with the United States military*. Retrieved April 15, 2021, from https://www.canada.ca/en/department-national-defence/services/doing-business-with-foreign-defence-markets/doing-business-with-united-states-military.html
Greenwalt, W. (2019, April). *Leveraging the national technology industrial base to address great-power competition: The imperative to integrate industrial capabilities of close allies*. Atlantic Council. https://www.atlanticcouncil.org/in-depth-research-reports/report/leveraging-the-national-technology-industrial-base-to-address-great-power-competition/
Greenwalt, W. (2020, September 10). House declares defense trade war on best U.S. allies: F35 sales in jeopardy? *Breaking Defense.* https://breakingdefense.com/2020/09/house-declares-defense-trade-war-on-best-us-allies-f-35-sales-in-jeopardy/
Hasik, J. (2021). *Atlantic Council report on SoSAs* [Unpublished manuscript].
Herrington, K. (2019, March 5). *Clarification of Other Transactions Guide, November 2018, and nontraditional defense contractors supplies and services* [Memorandum]. DoD.
Hunter, A. P., Obecny, K., Sanders, G., Rudelinger, J., & Ellman, J. (2017). *U.S.–Canadian defense industrial cooperation.* Center for Strategic and International Studies. https://csis-website-prod.s3.amazonaws.com/s3fs-public/event/170628_Obecny_USCanadianDefenseCoop_Web.pdf
Industrial Base Fund, 10 U.S.C. § 2508 (2011). https://www.govinfo.gov/content/pkg/USCODE-2010-title10/pdf/USCODE-2010-title10-subtitleA-partIV-chap148-subchapII-sec2508.pdf
Industrial Policy. (n.d.-a). *Industrial base analysis and sustainment (IBAS)*. DoD. Retrieved April 9, 2021, from https://www.businessdefense.gov/IBAS/
Industrial Policy. (n.d.-b). *Security of supply*. DoD. Retrieved April 9, 2021, from https://www.businessdefense.gov/security-of-supply/
John S. McCain National Defense Authorization Act of 2019, Pub. L. No. 115–232 § 253, 132 Stat. 1638 (2018a). https://www.congress.gov/115/plaws/publ232/PLAW-115publ232.pdf
John S. McCain National Defense Authorization Act of 2019, Pub. L. No. 115–232 § 842, 132 Stat. 1644 (2018b). https://www.congress.gov/115/plaws/publ232/PLAW-115publ232.pdf
Judson, J. (2018, October 30). U.S. Army's modernization efforts foster boom in US-foreign partnerships. *Defense News.* https://www.defensenews.com/land/2018/10/30/army-modernization-fosters-boom-in-us-foreign-partnerships/
Kenlon, F. (2018, January 5). *Making the case for international cooperative programs.* Defense Acquisition University, International Acquisition Management Blog. https://www.dau.edu/training/career-development/intl-acq-mgmt/blog/Making-the-Case-for-International-Cooperative-Programs
McGinn, J. (2020, September 15). Re-shoring does not mean "Buy America" only. *Defense News*. https://www.defensenews.com/opinion/commentary/2020/09/15/reshoring-does-not-mean-buy-america-only/
McGinn, J. (2021, February 4). Why "Buy American" doesn't do enough for the economy and national security. *MarketWatch*. https://www.marketwatch.com/story/why-buy-american-doesnt-do-enough-for-the-economy-and-national-security-11612405671
McLeary, P. (2021, April 12). DoD's new pushback against Chinese money in U.S. defense industry. *Breaking Defense*. https://breakingdefense.com/2021/04/dod-pushes-back-against-prc-money-in-us-supply-base/

National Defense Authorization Act for Fiscal Year 2017, Pub. L. No. 114-328 § 881 (2016). https://www.congress.gov/114/plaws/publ328/PLAW-114publ328.pdf

National Defense Authorization Act for Fiscal Year 2021, H.R. 6395, 116th Cong. (2021). https://www.congress.gov/bill/116th-congress/house-bill/6395/text

National Industrial Security Program Operating Manual (NISPOM); Final Rule, 32 C.F.R. 117 (2020). https://www.govinfo.gov/content/pkg/FR-2020-12-21/html/2020-27698.htm

Nelson, B. K. (2016, October). Keeping track of horseshoe nails: Industrial base analysis and sustainment. *Defense AT & L Magazine*. https://www.dau.edu/library/defense-atl/DATLFiles/Sep-Oct2016/Nelson.pdf

Pierce, K. (2020). *Letter to Senator Jack Reed*.

Shelbourne, M. (2021, January 19). SECDEF nominee Austin affirms threat from China, will "update" national defense strategy. *USNI News*. https://news.usni.org/2021/01/19/secdef-nominee-austin-affirms-threat-from-china-will-update-national-defense-strategy

Stockholm International Peace Research Institute. (n.d.). *SIPRI arms transfer database*. Retrieved April 20, 2021, from https://www.sipri.org/databases/armstransfers

Sutter, K. (2020, August 11). *"Made in China 2025" industrial policies: Issues for Congress* (CRS Report No. IF10964). Congressional Research Service. https://fas.org/sgp/crs/row/IF10964.pdf

Swanson, A. (2020, March 11). Coronavirus spurs U.S. efforts to end China's chokehold on drugs. *The New York Times*. https://www.nytimes.com/2020/03/11/business/economy/coronavirus-china-trump-drugs.html

Trusted Capital. (n.d.). *Program objective.* DoD. Retrieved April 10, 2021, from https://www.acq.osd.mil/tc/

White House. (2017, December 18). *National security strategy of the United States of America*. https://trumpwhitehouse.archives.gov/wp-content/uploads/2017/12/NSS-Final-12-18-2017-0905.pdf

White House. (2021, January 25). *Executive order on ensuring the future is made in all of American by all of America's workers*. https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/25/executive-order-on-ensuring-the-future-is-made-in-all-of-america-by-all-of-americas-workers/

White House. (2021, March). *Interim national security strategic guidance*. https://www.whitehouse.gov/wp-content/uploads/2021/03/NSC-1v2.pdf

PANEL 7. ACCOUNTING FOR UNCERTAINTY IN COST ESTIMATING

Tuesday, May 11, 2021	
12:15 p.m. – 1:30 p.m.	**Chair: Richard Burke,** Deputy Director for Cost Assessment and Program Evaluation ***Cost Recovery in Commercial Item Contracts*** Gregory Tharp ***Risk-Based Modeling of Lifecycle and Total Ownership Cost*** Johnathan Mun, Naval Postgraduate School ***A New Learning Curve for Department of Defense Acquisition Programs: How to Account for the "Flattening Effect"*** John Elshaw, Air Force Institute of Technology Jonathan Ritschel, Air Force Institute of Technology Clay Koschnick, Air Force Institute of Technology Dakotah Hogan, Air Force Cost Analysis Agency (SAF/FMCA) Adedeji Badiru, Air Force Institute of Technology

Dr. Richard Burke—has served as the Deputy Director, Cost Assessment, in the Office of the Secretary of Defense, Cost Assessment and Program Evaluation (OSD/CAPE) since June, 2009. Prior to this he served as the Deputy Director for Resource Analysis, Office of the Secretary of Defense, Program Analysis and Evaluation, and as the Chairman of the Cost Analysis Improvement Group (CAIG) beginning in November 2002. He joined the Office of the Secretary of Defense in April 1988 prior to his service in DoD. Dr. Burke served in several program management positions at Sandia National Laboratories in Albuquerque, New Mexico. He is an International Affairs Fellow of the Council on Foreign Relations in New York, and served as a visiting scholar at Stanford University during 1992‐93. Educated at the Massachusetts Institute of Technology, he received a doctorate in nuclear engineering and decision analysis in 1984. His published work includes studies of the economic and international aspects of commercial nuclear power reactors, the economic risks of power reactor accidents, and export controls on high‐technology industries.

Cost Recovery in Commercial Item Contracts

Greg Tharp—has more than 15 years of experience as a librarian and researcher. He previously served as a member of the State of Connecticut Library Advisory Council on Library Planning and Development. Tharp holds an Master's Certificate in Acquisitions and Contract Management from American Graduate University, an Advanced Certificate in Archives Management from Simmons University, a Master of Library Science from Southern Connecticut State University, and a Bachelor of Science from Sacred Heart University. Tharp holds a Certificate of Professional Librarianship from the Commonwealth of Massachusetts and is a Second-Degree Black Belt. [g.tharp@mindspring.com]

Abstract

Purpose: For uniformity, present value methodologies and clauses are needed.

Design: Based on a review of literature, FASB Concept Statement No. 7, and the FAR; a way to calculate present value of commercial item contracts is identified as well as present value clauses.

Findings: Commercial items involving an uncertain degree of risk and dynamic costs use a fuzzy net cash flow methodology, forming basis of net present value, to calculate cash flow in order account for risk and changing costs. Clauses should be inserted into commercial item contracts to allow for greater clarity as to how present value is calculated with certainty.

Practical Implications: No method to calculate the present value of recovery of cost of work performed prior to termination in commercial contracts exists. Present value clauses and methods are consistent with the intention of framers of the FAR.

Originality/Value: Due to lack of a current methodology to calculate the present value of commercial item contracts, FASB Concept Statement No. 7, present value, and fuzzy net cash flow are used to calculate present value of commercial items.

Background

The "Section 809 Panel" (n.d.) was tasked with improving efficiency of acquisition regulations and recommended improving acquisition of commercial items; but did not address what is fair and reasonable profit on work performed prior to termination. Thus, the following research question is addressed: How is present value identified in commercial item contracts and what corresponding contract clause should be inserted into the contract?

While the termination for convenience clause has "been around since the end of the Civil War"(Acquisitions.gov, n.d.) when the government needed to terminate contracts entered into during wartime when the need doesn't exist during peacetime, it has not been defined what is fair and reasonable.

The Federal Acquisition Regulation (FAR) might define what is fair and reasonable since the FAR regulates acquisitions with most executive agencies.

One consults FAR 2.101 on Acquisition.gov since it contains the definitions used in the FAR. However, FAR 2.101 does not define what fair and reasonable profit on work performed prior to termination is.

One looks to the FAR conventions as fair and reasonable profit on work performed prior to termination could be a "permissible exercise of a contracting officer's authority" (Acquisitions.gov, n.d.). However, a contracting officer's interpretation of what is fair and reasonable based on prior experience, regulations, etc. may be different from a contractor's definition of fair and reasonable based on the usage of estimated costs in its accounting system.

For instance, a computer is estimated to cost $100 by the government while contractor actually incurs $150 to make the computer, which is what a reasonable business person would expect to pay for a computer given reasonable efficiencies and market conditions; however they yield different results when the contract is terminated.

FAR 31.201-1's five part-time for allowability defines reasonable costs. Reasonable costs are not uniformly applied as FAR 31.201-3(b) states "What is reasonable depends upon a variety of considerations and circumstances, including generally accepted sound business practices."

A reasonable cost for a fixed-price contract to paint a house differs from reasonable cost for a fixed-price contract with economic price adjustment because price of paint went up or other factors. Reasonable cost fails to address valuation of the unfinished portion of contracts terminated for convenience varies from contract to contract even if it performed with the same contractor. Inconsistent valuation of the unfinished portion of the contract could leave either contractor or government wondering if they were compensated fairly for their efforts.

Reviewing DFARS (Defense Federal Acquisition Regulation Supplement, n.d.), PGI, and Uniform Commercial Code (UCC) fails to address the question of fairness of valuation of the unfinished portion of contracts. DFARS 212.102 (Defense Federal Acquisition Regulation Supplement, n.d.) applicability is limited to acquisitions over $1,000,000, and DFARS 249.5 (Defense Federal Acquisition Regulation Supplement, n.d.) applicability is limited to major acquisitions over $25,000,000 for RDT&E and more than $100,000,000 for production inventory. Since many commercial item contracts are below the DFARS threshold, the UCC is consulted.

UCC Section 1-204 defines "value for rights if a person acquires them in return for any consideration sufficient to support a contract" (Cornell Law School, n.d.). A contractor could give a dollar in exchange for a ballpoint pen. It may not be fair to both parties; however, it creates a legally sufficient contract or that it adequately compensates the contractor for their effort they put in making the ballpoint pen.

How does the lack of a uniform definition of fair and reasonable compensation impact acquisitions practice?

Perhaps the answer is in FAR 12.403(d)(i)(B)(ii) since fair compensation includes "any charges the contractor can demonstrate directly resulted from the termination."

Contractors may demonstrate such charges using its standard record keeping system (including Generally Accepted Accounting Principles or GAAP). While the FAR does not preclude the usage of analytical tools or methods to determine the amount of settlement of commercial item contracts terminated for convivence, the word *may* implies that the contractor has the discretion of using GAAP to prove that its settlement costs represent fair compensation (see Kingdomware Technologies, Inc. v. United States) which makes FAR 12.403(d)(i)(B)(ii) a discretionary rule which all contractors must follow. Contractors not using GAAP to demonstrate termination claims could use an alternative method to meet the requirement to prove fair and reasonable compensation for cost recovery in commercial item contracts. FAR 49.201(a) mandates through the usage of the word *should* that the contractor needs to be fairly compensated for work and preparations associated with terminating contracts for convenience.

Recouping cost for a computer valued using estimated costs to be $100 by the government and valued using actual costs by the contractor as being $150, the contractor could use GAAP to value the computer at $150. It may seem to be fair since the contractor used GAAP; however the government can argue that the estimated costs of $100 is fair and reasonable valuation given that the $100 is "consideration sufficient to support a simple contract" under section 1-204 of the Uniform Commercial Code and is consistent with the government's estimating methodologies.

The difference on what is fair valuation for a commercial item, such as a computer, makes it difficult to determine if the contractor should be compensated $70 (70% of $100) using estimated costs or if the contractor should be compensated $105 (70% of $150) using actual costs. Negotiations of termination settlements may not reflect what each party perceives as the effort that they put into the contract since FAR 12.403(d)(i)(B)(ii) does not specifically identify a standard recordkeeping system. Record keeping can be non-GAAP or GAAP. Current practice in commercial environments for contractors is to choose valuation methodologies depending on their industry norms, which results in varying valuations of commercial items (e.g., commodities) due to market conditions thus leading to the perception that termination settlements are not being valued fairly.

It is proposed that net present value be used as a standard by which to value commercial item contracts terminated for convenience to allow both the government and contractors to be fairly compensated for the effort, they put into contracts terminated for convenience.

Terminating for Convenience on Commercial Contracts vs. Non-Commercial Contract Terminations

Termination for Cause termination may be used for cost recovery and negotiation. FAR 12.403(c)(2) states "the government's rights after a termination for cause shall include all the remedies available to any buyer in the marketplace." Termination for Cause centers on acquiring similar items, such as commercial items, for instance, exchanging a lawnmower with the make and model because it contained a defective blade. The lawnmower would still cost $100.

It is not addressed how a partially completed item is fairly valued by the government when a contract is terminated for convenience is not addressed.

For instance, a $100,000 widget is only 30% completed. While contractor may use GAAP to value the 70% of the widget contract that is incomplete, there is not a guarantee that the contractor will do so since FAR does not make the usage of GAAP mandatory. Since GAAP is not applied uniformly to value terminations, widgets purchased under the same conditions by the same contractor may be valued using both GAAP and non-GAAP measures which leads contractors to question if they are getting a fair valuation for their efforts.

So, attention must be turned to alternative uniform ways to calculate cost recovery.

- **Introduction to Cost Recovery in Commercial Item Contracts**

To provide context for a discussion of a potential uniform way to calculate cost recovery in commercial item contracts, GAAP, Generally Accepted Accounting Principles, are discussed. GAAP principles include:

- ***Recognition***: What items should be recognized in the financial statements (for example as assets, liabilities, revenues, and expenses).
- ***Measurement***: What amounts should be reported for each of the elements included in the financial statements.
- ***Presentation***: What line items, subtotals, and totals should be displayed in the financial statements and how line items might be aggregated within the financial statements.
- ***Disclosure***: What specific information is most important to the users of the financial statements. Disclosures both supplement and explain amounts in the statements.

GAAP is followed by most organizations and was developed and established by FASB (Financial Accounting Standards Board).

Uniformity is an issue when calculating amount “contractors are entitled to recover the amount of cost of the contract work performed prior to termination” in contracts terminated for convenience. To calculate cost recovery using the present value of actual costs completed with certainty as required by Lisbon Contractors, Inc. v. United States, 828 F. 2d 759, 765 (Fed. Cir. 1987), FASB Concepts Statement No. 7 (FASB, 2000) is used to calculate the present value of actual costs completed. To address the issue of uniformity as well as to ensure the cost recovery is calculated with certainty, this paper seeks to address the following research question: How should the present value of commercial item contracts should be calculated and what corresponding contract clause should be inserted into the contract? This research question is relevant because there is no empirical evidence or support or current literature which already answers this question.

Review of Existing Literature

A literature review was conducted to determine if present value is currently being used for commercial item contracts. This not the case, as present value is currently being used for other contracting functions. For instance, the General Services Administration uses a “Present Value Analysis Model” (General Services Administration, 2019) for lease proposals which does not pertain to commercial items or to termination of contracts. The General Services Administration’s model only takes into account variables (e.g., utilities) which go into rental agreements. Chapter 9 of the Contract Pricing Guide discusses net present value in relation to cost price and analysis, which does not address net present value for commercial item contract cost recovery. Thus, there is no existing literature that addresses net present value usage in commercial item contract cost recovery.

What Are Commercial Items?

To provide context for potential commercial item termination reform, commercial items are defined. For non-government commercial acquisitions, the Federal Register identifies commercial items as including installation services, maintenance services, and other services procured to support a commercial item as well as products that were created by integrating commercial subsystems and components into a unique system (National Archives, n.d.).

The Federal Acquisition Reform Act (FARA) of 1996 (United States Department of Labor, 1995) and Federal Acquisition Streamlining Act (United States Congress, 1994) identifies that the government prefers to purchase of commercial items for a myriad of reasons including minimizing acquisition lead time. FAR Council implementation of FARA with FAR Part 12, Acquisition of Commercial Items, streamlined commercial acquisition procedures and made them contingent upon a commercial item determination.

However, commercial item reform has not established a uniform way to value the terminated portion of a commercial item contract in a manner that is perceived to be both fair and reasonable by the contractor and the government.

FASB: What is the FASB and Why Use Concept Statement No. 7

Discussion of what the FASB (Financial Accounting Standards Board) is and why Concept Statement No. 7 is used helps to give context to the calculation of present value and fuzzy present value. The FASB (Financial Accounting Standards Board) is an independent, non-for-profit organization that establishes financial accounting and reporting standards for public and private companies and non-profit organizations that follow GAAP. FASB is recognized as authoritative guidance by the Securities and Exchange Commission (SEC), state Boards of Accountancy, and the American Institute of Certified Public Accountants (AICPA) among other organizations.

Widespread acceptance of FASB pronouncements as being authoritative by the SEC, AICPA, and state boards of accountancy makes it makes likely that usage of a FASB Concept Statement to calculate the present value of actual costs completed will be unchallenged by expert witness testimony before the GAO, Boards of Contracting Appeals, or United States Court of Federal Claims during the appeal of a contracting officer's decision. Recent decisions by the GAO, Armed Services Board of Contracting Appeals, Civilian Board of Contracting Appeals, and the Court of Federal Claims do not identify any instances where the FASB Concept Statement was used by either the government or the contractor to calculate the present value of actual completed. While usage of the FASB Concept Statement goes into uncharted territory in terms of whether it will be blessed by the judiciary, the authoritative nature of FASB and the acceptance of FASB pronouncements by the SEC, AICPA, and state boards of accountancy makes it prudent for the present value of actual costs completed to be calculated using a FASB Concept Statement. Since usage of GAAP is a widely accepted accounting practice recognized by the SEC, AICPA, and state board of accountancy, it is sound business judgment to award a contract using GAAP using FASB Concepts Statement No. 7.

Why Use FASB Concept Statement No. 7

FASB Concepts Statement No. 7 (FASB, 2000) is used to calculate present value of actual costs completed to determine amount of cost to be recovered due to the lack of a mechanism in the marketplace to readily observe the present value of actual costs completed to be recovered. While commercial item contracts may be "distinguished from one another in timing and uncertainty (FASB, 2000, p. 7), present value measurement helps to establish an economic difference between the commercial item contracts. Noted in FASB Concepts Statement No. 7 (FASB, 2000, p. 7–9), elements of present value include: (a) estimate of future cash flow, or in more complex cases, series of future cash flows at different times; (b) expectations about possible variations in the amount or timing of these cash flows, (c) time value of money, represented by risk-free rate of interest; (d) price for bearing the uncertainty inherent in the asset; (e) other, sometimes unidentifiable, factors including illiquidity and market imperfections. The contracting officer shall consider these elements of present value in determining amount of cost to be recovered because some commercial items as identified in the Commercial Item Handbook may be: (a) "noncommercial modification" (Commercial Item Handbook, n.d.), (b) minor modifications of a type not customarily available in the commercial marketplace made to meet the government's requirement, (c) evolved items, or (d) a type not identical to those in the commercial marketplace. Except modifications of commercial items available in the marketplace and the usage of a commercial item already on the marketplace meeting the government's requirement, commercial items identified in the Commercial Item Handbook do meet the elements of present value set forth in FASB Concept Statement No. 7 since they are illiquid in that they do not have a marketplace outside of the government. Illiquid commercial items may be commodities, such as gold, being purchased specifically for use by the government. These illiquid items would most likely be valued by consulting with experts on the particular item or by other estimating techniques permissible in the FAR and other procurement regulations.

Calculation of the value of cash flow of cost to be recovered, either using best estimate or expected present value, applies to commercial items because they meet the elements of present value. Best estimate is used to determine amount of cost to be recovered for commercial items available in the marketplace and modifications of commercial items in the marketplace since there is a commercial item already in the marketplace to compare it to.

How to Identify Cash Flow for Present Value

The following steps, based on FASB Concept Statement No.7, are recommended to identify cash flow for commercial items using the best estimate method:

1. Comparison of commercial item in the contract to another commercial item existing in the marketplace that has an observed interest rate (this is commonly referred to as "the rate commensurate with the risk" [FASB, 2000]).
2. Identification of set of discounted cash flows and comparison of cash flow sets between two commercial items.
3. Evaluation of characteristics or elements in commercial items which are different from each other.
4. Evaluate if changing economic conditions will cause the two commercial item cash flows to behave differently.

Changing economic conditions may include labor strike increasing or decreasing the price of one commercial item. Different characteristics or elements of commercial items may have different capabilities (such as the speed of a machine lathe). Economic conditions may be outside of government or contractor control; however, it is considered when calculating present value.

Best estimate method cannot be used for commercial items for which no market for the item or comparable item exists. For instance, actual cost to complete a supercomputer may be $100 million, comparable supercomputers on the market are valued at $200 million and $300 million based on market research. Best estimate or most likely cash flow necessary to complete the supercomputer is $200 million. Therefore, contractor can expect to recover $200 million on the contract to build a supercomputer.

Estimated Cash Flow Example

An expected cash flow estimate method example is when actual cost to complete building a commercial (e.g., supercomputer computer) item may be $100 million, comparable supercomputers on the market are valued at $200 million and $300 million with probabilities of completion of actually completing the supercomputer at 10%, 60%, and 30%, respectively. The expected cash flow is $220 million using the following formula: ($100 million x .1) + ($200 million x .6) + ($300 million x .30) =$220 million. Therefore, contractor can expect to recover $220 million on the contract to build a supercomputer.

Although the estimated cash flow estimate method results in a higher cost of recovery for the contractor ($220 million versus $200 million using the best estimate method), the estimated cash flow method is preferred because it accounts for the uncertainty in the timing of the cash flow. Assigning risk in commercial item contracts to calculate present value of cost recovery is not an exact science. Economic or other conditions may contribute to the uncertainty of the contractor recovering these actual costs. Also, risk needs to be assessed on the commercial items under contract in relationship to the extent the commercial item adds to or diminishes total risk in the total commercial items under contract. Returns on one commercial item vary with the market for all commercial items or recent experience and framing of decisions by contracting officers influence the price of commercial items in the market. Alternatively, behavioral economics might also influence price of commercial items in the marketplace. Economic factors and different contracting officers may assign different degrees of risk to a commercial item necessitate market research to be conducted to justify calculations of actual costs to be recovered on a commercial item contract.

Examples of economic factors impacting price of commercial items would be inflation or deflation.

- ***Why Use Estimated Cash Flow Method?***

One of the objectives of the FAR is to "minimize administrative costs" per FAR 1.102. Usage of estimated cash flow method to calculate estimated cost to complete a contract is a consist method to determine how best to calculate the percentage of contract to be completed.

The cash flow estimate method results in a higher cost of recovery for the contractor ($220 million versus $200 million using the best estimate method), which results in cost recovery being calculated using a generally accepted method. Thus, contracting officer oversight and legal scrutiny of contracts for commercial items using estimated cash flow method is reduced.

Different interpretations trigger contradictory results since some contractors elect to use different net present value analysis techniques and some not electing to use any techniques at all. Inevitably, court decisions will differ on the same sets of facts if contractors use different present value analysis techniques or no present value techniques, promoting doubt as to amount of cost to be recovered. A contractor using estimated cash flow method does not risk a court not finding fair compensation in terminations for convenience because of usage of generally accepted method to calculate net present value.

Valuation Based on Past Performance

Actual cost to complete may be valued solely using value of the commercial item multiplied by probability of completion based on past performance in instances when prices of comparable commercial items cannot be obtained. For example, a supercomputer contract costing $100 million and a probability of completion of 10% would result in actual cost recovery to complete of $10 million.

This approach is used in contingency operations; defense or recovery from cyber, nuclear, biological, chemical, or radiological attack; major disasters or emergency assistance required under the Stafford Act as declared by the president; and Humanitarian or Peacekeeping Operations. A contingency operation might be a hurricane because the president issues emergency assistance declarations for those natural disasters. During contingencies, it may be impractical to gather other commercial sources for the purposes of valuing the commercial item or it must be so urgent that the contract be terminated, and another contractor selected that it would be impractical to gather other commercial sources for comparison. Other sole source acquisitions scenarios may also apply other than contingency operations and the same valuation procedure would apply.

Unique Systems: Present Value

A unique system could range from a weapons system to a supercomputer. Thus, “fuzzy net cash flow” (Maravas & Pantouvakis, 2012) is useful in determining present value of the commercial item. The following formula based on Maravas and Pantouvakis (2012) is proposed in order to calculate cash flow: Cash Flow = Time Savings + Operating Cost Savings + Accident Savings + Environmental Savings – Investment Cost – Operation and Maintenance Cost. Accident Savings includes manpower and facilities damage and is determined by the contractor subject to verification by government auditors.

Investment cost is the cost of acquiring the commercial item, including research and development expenses. Time savings include manpower as well as facilities overhead and other incurred direct and indirect costs.

Due to lack of empirical data and usage of fuzzy cash flow, administrative burden to government and industry may or may not be increased or decreased. Data is collected for each of the year(s) of operation and may be reduced or increased by a certain percentage and variables of time savings may be reduced or increased by a certain percentage in order to determine the valuation of cash flow. This quantifies risk inherent in the commercial item system. In the initial years of a system, the investment costs of the project, such the cost of component commercial items, may be present while there are no benefits or maintenance costs present. Time savings, operational cost savings, environmental savings, and operation and maintenance costs are realized in later years of the project based on contractor provided data.

The system becomes unsustainable when investment cost (e.g., cost of replacement parts) outweighs the benefits and savings of the project. Since it is unknown how risky a project will be, cash flow calculated is from the previous formula will be multiplied by the probability of the unique system being successful in order to derive the present value of a system over time. In FASB Concept Statement No. 7 page 21, the expected present value of a system is the sum of the present value of the present values.

For example, present value of a system being 10% successful is 95.24 in year 1, 541.64 as the probability increases to 60% in year 2, and 255.48 as the probability decreases to 30% in year 3, the total expected value is 892.36. Contract was 30% complete at time of termination; therefore contractor is entitled to recover 30% of 892.36 as the amount of contract work performed prior to termination.

Watson as an Example of a Unique System and Present Value

Watson, the IBM supercomputer used in the popular television series Jeopardy, is a unique system. In year 1, Watson would be unsustainable because of tweaks needed for its algorithm, thus a 10% success rate is assigned to Watson and for the purposes of this example it is valued at $95.24 million in year 1, and the next year the probability increases to 60% because the kinks in algorithm were worked out thus probability increases to 60%, which results in a value of $541.64 million. Valuation is increased to $892.36 million in year 3. If the contract to build Watson was terminated when it was 30% completed, the contractor is entitled to recover 30% of $892.36 million as the amount of contract work performed prior to termination.

- ***Why Use Fuzzy Net Cash Flow?***

An objective of the FAR is to "minimize administrative costs" as documented in FAR 1.102. The estimated cash flow method to calculate estimated cost to complete a contract is a consistent method. Thus, it reduces the need for contracting officer oversight and legal scrutiny of commercial item contracts using fuzzy net cash flow method. Currently, no contractors use fuzzy net cash flow since it is not mentioned in the FAR.

Inevitably, court decisions will differ on the same sets of facts if contractors use different present value analysis techniques or no present value techniques, promoting doubt as to amount of cost to be recovered. Contractors using the estimated cash flow method may have courts find fair compensation in terminations for convenience because of the usage of a generally accepted method to calculate net present value.

Are These Changes Permissible in the FAR?

FAR 12.403(d)(1)(i)(a) ("Welcome to FARSite", n.d.) states the percentage of the contract price reflecting percentage of work performed is the amount that the contractor should be paid when terminating a contract for convenience. FAR 12.403(d)(1)(i)(a) states in part: (d) *Termination for the Government's convenience.* (1) When contracting officer terminates a contract for commercial items for the Government's convenience, the contractor shall be paid --

(i) (A) The percentage of the contract price reflecting the percentage of the work performed prior to the notice of the termination for fixed-price or fixed price with economic price adjustment contracts.

- **Net Present Value Clause**

Lack of existing literature on an NPV clause for commercial item contracts terminated for convenience warrants rationale for this clause based on policy goals of the FAR. A policy goal of the FAR is uniformity. Contractors can't elect not to use NPV, a type of analytical method, just because FAR 12.403(d)(i)(B)(ii) only includes the word *may*, rather than *shall*, which the Supreme Court has held to be discretionary.

The drafters of the FAR required "coordination, simplicity and uniformity in the Federal acquisition process" (Mason, 2000, p. 724). Requiring net present value in all cases of government contracting furthers this goal. Conversely, different interpretations would occur with some contractors electing to use present value analysis and some electing not to do so.

Inevitably, court decisions will differ on the same sets of facts if contractors used GAAP as a means of fair compensation merely because of the usage of the term *may*, promoting doubt as to if the contractor should use net present value analysis as promogulated by GAAP. A contractor using NPV clause is fair compensating in terminations for convenience because of the lack of the word *shall* in the clause.

Alternative 1: The term *net present value* (NPV) shall mean present value of cash payments generated by a commercial item(s), calculated using a discount rate determined by an actuary selected by the government and determined in accordance with GAAP. Probability shall be determined by conducting an analysis of alternatives analysis. Contractors shall calculate cost recovery or charges from commercial item terminations based on NPV or fuzzy net cash flow.

Alternative 2: The term *net present value* (NPV) shall mean present value of cash payments generated by a commercial item(s), due in the future reduced by a discount rate equal to 100% of the Applicable Federal Rate (as defined in Code Section 1274(d). Contractors shall calculate cost recovery or charges from commercial item terminations based on NPV. Probability shall be determined by conducting an analysis of alternatives analysis.

Internal Revenue Code Section (IRC) 1274(d) applies to debt instruments which are publicly traded or issued for publicly traded property, such as government property. As noted in IRC 1275(a)(1), a debt instrument includes:

> Daily portion of original issue discount for any day shall be determined under section 1272(a) (without regard to paragraph (7) thereof and without regard to section 1273(a)(3)). In the case of an obligor of a short-term obligation (as defined in section 1283(a)(1)(A)) who uses the cash receipts and disbursements method of accounting, the original issue discount (and any other interest payable) on such obligation shall be deductible only when paid.

An example of a debt instrument may be a loan from a bank to a contractor.

So, alternative 2 applies to commercial items purchased using debt financing. Conversely, a contractor who does not use net present value clause risks a court not finding fair compensation in terminations for convenience because of the usage of word *may* in FAR 12.403(d)(i)(B)(ii).

A final policy benefit is consistent enforcement of the usage of present value in termination of commercial item contracts. Definition of NPV, either in alternative 1 or alternative 2, and uniform application gives contractors a clear path to take prior to court intervention. This "minimizes administrative operating costs" as required in FAR 1.102, by having government spend less time and money in oversight functions (such as audit) ensuring that contracting officers fairly compensate contractors in terminations since it is mandatory not optional to do so.

- ***Example of Net Present Value Clause***

Using Alternative 1, an acquisition would be valued at $61,446 today assuming a return of $10,000 per year over 10 years having a discount rate of 10%. Without using net present value, the present value of the acquisition at year 10 would be $3,855. If the contract is terminated prior to completion without using net present when it is 70% completed, the contractor would be entitled to recovery $2698.50 compared with $43012.20 using net present value due to the time value of money. It is not practical to provide an example of Alternative 2 because the AFR changes from month to month within a given year.

Conclusion

Calculation of cost recovery in commercial item contracts terminated for convenience is a challenge due to a lack of current literature and a lack of mandatory techniques to be used.

Usage of FASB Concepts Statement No. 7 after calculating a fuzzy net cash flow, present value of commercial item contracts is determined for complex commercial acquisitions. Usage of fuzzy net cash flow reduces the overall administrative burden for contract administration since it is a consistent technique being used. Present value is calculated using FASB Concepts Statement No. 7 for other commercial items.

Usage of present value reduces the overall administrative burden for contract administration since it is a consistent technique being used. The difference between present value and fuzzy net cash flow is that fuzzy net cash takes into account other variables impacting valuation, such as operating time, which may provide a more accurate projection of present value for complex commercial items. Two present value clauses are proposed in order to assist contracting officers in determining the present value of commercial item contracts and to serve a variety of policy benefits as to the construction of the FAR. Mandatory usage of NPV clauses provides for uniformity while providing a framework to base negotiations upon.

It is unknown if the present value, fuzzy value, and present value clauses will hold up to scrutiny by the Armed Services Board of Contracting Appeals and other courts due to the lack of cases which address these issues. There is a lack of empirical evidence or existing literature regarding the usage of present value and fuzzy value in commercial item contracts since these methods have not been tried in the field yet.

References

Acquisitions.gov. (n.d.). *FAR.* Retrieved December 20, 2019, from https://www.acquisitions.gov.

Contract Pricing and Finance Guide. (n.d.). *9.0: Chapter introduction.* Retrieved December 20, 2019, from https://www.acq.osd.mil/dpap/cpf/docs/contract_pricing_finance_guide/vol2_ch9.pdf

Cornell Law School. (n.d.). *Uniform commercial code.* Retrieved March 25, 2020, from https://www.law.cornell.edu/ucc/1/part_2

Defense Federal Acquisition Regulation Supplement. (n.d.). Retrieved March 25, 2020, from https://www.acquisition.gov

Federal Acquisition Regulation: Definition of “Commercial Item.” (n.d.). Retrieved July 2, 2019, from www.federalregister.gov

Financial Accounting Standards Board. (n.d.). *About the FASB.* Retrieved August 19, 2019, from https://www.fasb.org/jsp/FASB/Page/SectionPage&cid=1176154526495

Financial Accounting Standards Board of the Financial Accounting Foundation. (2000). *Statement of financial accounting concepts no. 7: Using cash flow information and present value in accounting measurements.* http://www.fasb.org/jsp/FASB/Page/PreCodSectionPage&cid=1176156317989

General Services Administration. (2019). *Policies and tools present value analysis model*. https://www.gsa.gov/real-estate/real-estateservices/leasing-policy-procedures/policy-and-tools/tools/present-value-analysis-model

Legal Information Institute. (n.d.).

https://www.law.cornell.edu/uscode/text/26/1274

Loving, J. T. (2004, October). *Pricing commercial items convivence terminations.* https://www.ncmahq.org

Malone, L. K. (2018, March). *Recovery when the government terminates for convenience.* https://www.oles.com/blog/recovery-when-thegovernment-terminates-for-convenience/

Maravas, A., & Pantouvakis, J.-P. (2012). Project cash flow analysis in the presence of uncertainty in activity duration and cost. *International Journal of Project Management*, *30*(2012), 374–384. https://doi.org/10.1016/j.ijproman.2011.08.005

Mason, A. B. (2000). Implied-in-fact contracts under the Federal Acquisition Regulation: Why Pacord got it wrong. *William & Mary Law Review*, *41*(2) 709–739. https://scholarship.law.wm.edu/cgi/viewcontent.cgi?article=1521&context=wmlr

National Archives. (n.d.). *Federal register: The daily journal of the United States government*. Retrieved September 26, 2019, from https://www.federalregister.gov

Office of the Secretary of Defense. (n.d.) *Commercial Item Handbook (Version 2.0)*. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=2ahUKEwiZvfi4-I7kAhULqlkKHROsDe4QFjABegQIDBAE&url=https%3A%2F%2Fwww.acq.osd.mil%2Fdpap%2Fcpic%2Fdraftcihandbook08012011.docx&usg=AOvVaw3didT12tN2eE7IblabIjAb

Oles Morrison. (n.d.). *Recovery when the government terminates for convenience.* Oles Morrison. https://www.Oles.com

Section 809 Panel. (n.d.). Retrieved March 26, 2020, from https://section809panel.org

Supreme Court of the United States. (2015, October). *Kingdomware Technologies, Inc. v. United States*. https://www.supremecourt.gov/opinions/15pdf/14-916_6j37.pdf

United States Congress. (1994). *Federal acquisition streamlining act of 1994.* https://www.govinfo.gov/content/pkg/BILLS-103s1587enr/pdf/BILLS-103s1587enr.pdf

United States Department of Labor. (1995). *Federal acquisition reform act of 1995.* https://www.thecre.com/fedlaw/legal25/fara.htm

Welcome to FARSite. (n.d.). Retrieved July 2, 2019, from https://www.farsite.hill.af.mil

Risk-Based Modeling of Life-Cycle and Total Ownership Cost

Dr. Johnathan Mun is a specialist in advanced decision analytics, quantitative risk modeling, strategic flexibility real options, predictive modeling, and portfolio optimization. He holds 22 patents and patents pending and has authored 24 books; has created over a dozen software applications in advanced decision analytics; and has written over 100 technical notes, journal articles, and white papers. He is currently the CEO of Real Options Valuation Inc., and his prior positions include Vice President of Analytics at Oracle/Crystal Ball and a Senior Manager at KPMG Consulting. Mun holds a PhD in finance and economics from Lehigh University, an MBA and an MS from Southeastern University, and a BS in physics and biology from the University of Miami. [jcmun@nps.edu]

LT Eliah Ledbetter, MBA Student

LT Katelyn George, MBA Student

Abstract

In this research, we answer the following primary question: Would an advanced analytical model be a more effective metric to estimate total ownership cost (TOC) with life-cycle cost under uncertainty and risk than the current method of life-cycle cost estimates for surface electro-optical infrared (EO/IR) sensors? To accomplish this, we developed and analyzed a computational model for Total Ownership with Life-Cycle Cost Model Under Uncertainty for Surface Electro-Optical Infrared Sensors. During the development of the model, we identified the required data and examined the current Department of Defense (DoD) method for determining system life-cycle costs for defense systems and determined that the proposed model is a useful alternative to the current method of determining the life-cycle costs for EO/IR sensors on surface ships. Finally, we concluded that the developed model can be applied to cost estimating in other sectors of DoD cost projections.

Introduction

Research Purpose

The purpose of this research is to develop a model to estimate total ownership with life-cycle costs under uncertainty associated with surface electro-optical infrared (EO/IR) sensors. We examine the basics of total ownership cost (TOC) modeling over the life cycle of the EO/IR sensors, including the inception phase of acquisition costs, followed by annual operations and maintenance (O&M) expenses, along with a final set of disposition costs at the end of life of the sensor. This model will allow managers to have better decision analytics of the costs of said sensors for use in subsequent cost comparisons across sensor platforms, return on investment analysis, portfolio allocation of resources, and analysis of alternatives.

Research Focus

In this research, we answer the following primary question: Would an advanced analytical model be a more effective metric to estimate TOC with life-cycle cost under uncertainty and risk than the current method of life-cycle cost estimates for surface EO/IR sensors? To accomplish this, we develop and analyze a Total Ownership with Life-Cycle Cost Model Under Uncertainty for surface EO/IR sensors. In the development of the model, we determine what data are required to implement our proposed model for surface ship EO/IR sensors. We also examine the current Department of Defense (DoD) method for determining system life-cycle costs for defense systems and consider whether the proposed model is a useful alternative to the current method of determining the life-cycle costs for EO/IR sensors on surface ships. Finally, we consider whether the developed model can be applied to cost estimating in other sectors of DoD cost projections.

Research Summary

While executing a standard life cycle–based TOC analysis, we assume that, before the system is operational, there are substantial acquisition costs. These costs are usually referred to as Year 0, followed by the operational years where operation and maintenance costs will apply. The final price analyzed is the salvage cost, or the cost to properly dispose of, sell, or render the system inoperable. The sum of these three expenses is called the life-cycle cost. Unfortunately, the accurate calculation of these costs is not as straightforward as their descriptions. To accurately incorporate these three factors, it is essential to consider economic theory. The elements of time valuation of money are critical in the analysis of alternatives. The economic growth, annual discount rate, inflation, and opportunity cost of investing in a specific system are essential to our study. Other factors include budgetary cutbacks and changes in technology. The model will allow the user to input these changes to manually adjust for each of these. Utilizing this model will serve as a proof of concept to understand how this approach could be used to reduce cost overflow and prevent budget overruns. It will provide greater insight into the true nature of the cost of cash outflow and the life cycle of the product and its associated costs. These results would give leaders a more effective metric to analyze TOC under uncertainty, therefore allowing leadership to make more informed decisions in the DoD acquisition process.

Literature Review

Introduction

This background and literature review provide a comprehensive overview of the topics pertinent to our project. We first examine the concepts and best practices in the field of cost and cost estimation and their application inside of the DoD. We then investigate the DoD's acquisition process as a whole to analyze how the DoD can utilize cost estimation to influence decision-making. After covering basic cost estimation and the acquisition system, we then discuss TOC and life-cycle cost estimations and how these factors play a role in calculating the overall cost of a system. The review also covers the topics of risk and uncertainty to explain the relationship and the differences between the two as well as to highlight the importance of properly accounting for both factors. We conclude with an overview of our model's subject, the EO/IR sensor. We give a brief rundown of the capabilities as well as the applications that these sensors have on Navy surface vessels, along with their rapidly changing technology, and state why it is imperative that the Navy continues to buy these sensors while ensuring the cost stays at a rational price point.

Cost Estimation

The DoD receives a limited amount of funds every fiscal year and must decide how those funds are used in support of U.S. national strategies and goals. Specifically, those decisions fall into one of three categories: long-term planning, budgeting, or choosing among alternatives (Mislick & Nussbaum, 2015). The government is tasked with spending taxpayers' dollars effectively and efficiently. This means that the DoD decision-makers must ensure they make strategic investments, including the acquisition of new programs and systems. Before a program is implemented or a system is purchased, decision-makers must understand the full cost that will be incurred and its effect on the DoD's limited budget.

The projected costs of major acquisitions are produced through a process known as *cost estimation*. Cost estimation is defined as "the process of collecting and analyzing historical data and applying quantitative models, techniques, tools, and databases in order to predict an estimate of the future cost of an item, product, or task" (Mislick & Nussbaum, 2015, p. 11). In basic terms, cost estimation is performed by running relevant data from the past through a model or database to predict what an item will cost in the future. It is important to note that reliable historical data are fundamental to this process.

In order to produce cost estimates, we must first gather available historical data. Collecting data is often the most time-consuming and costly step of the entire cost estimation process (Mislick & Nussbaum, 2015). Only after the historical data have been obtained can the cost analyst start the "organization, normalization, and management of that historical data" (Mislick & Nussbaum, 2015, p. 11). *Normalization* refers to taking the historical data and "applying adjustments to that data to gain consistent, comparable data to be used in your estimates" (Mislick & Nussbaum, 2015, p. 78). Normalizing the data set allows the analyst to compare data across different periods of time by adjusting for different factors. The data set must be normalized three different ways: for content, for quantity, and for inflation (Mislick & Nussbaum, 2015). Normalizing for content ensures comparison across the same category or type of data (Mislick & Nussbaum, 2015). Normalizing for quantity ensures comparison of data at the same point on the learning curve of production and of equal quantities (Mislick & Nussbaum, 2015). Finally, the data are adjusted to account for inflation when comparing data from different years (Mislick & Nussbaum, 2015).

The second component of cost estimation is the quantitative model that is used to turn normalized historical data into a future cost estimate. Mislick and Nussbaum (2015) explain that the "profession of cost estimating is scientifically grounded by using transparent, rationally defensible and reviewable quantitative methods" (p. 12). The development of a high-quality quantitative model is key in cost estimation. If a poor quantitative model is used, then the quality and reliability of the cost estimate will also be poor. This highlights the importance of the quality cost models for EO/IR sensors.

The third part of Mislick and Nussbaum's (2015) definition of cost estimation is to predict. The ultimate goal of cost estimation is to predict a future cost. The prediction is based on the information available at the time. We can only "estimate the conditions that will pertain later when the project is executed" and must rely on the information available in the present (Mislick & Nussbaum, 2015, p. 12). While no one can forecast the future with 100% accuracy, through historical data and quantitative models, we are able to provide a more accurate prediction that, while not perfect, is still a useful tool for decision-makers in the acquisition process.

One of the most important characteristics of a quality cost estimate is that it must be understandable to the user or decision-maker in order to be an efficient decision-making tool (Mislick & Nussbaum, 2015). To this end, a complex approach to cost estimation should be avoided and a simpler approach should be used (Mislick & Nussbaum, 2015). An understandable estimate also clearly lays out the assumptions and ground rules that were used in the process (Mislick & Nussbaum, 2015). With the diversity among people's background and experiences, there can be differing underlying assumptions in the cost estimation process. Therefore, the assumptions used must be clearly stated, and a sensitivity analysis should be performed to accommodate additional variations of assumptions (Mislick & Nussbaum, 2015).

Cost Overview

Before comprehending cost estimation methods, it is important to become familiar with the terms associated with cost estimation. To begin with, an understanding of "cost" provides a solid foundation in the cost estimation process. If we do not understand what we are trying to predict, then we will not produce a quality or credible estimation. The term *cost* is often used interchangeably with the term *price*; however, they do not have the same meaning. There is an important distinction between the two terms. Mislick and Nussbaum (2015) define cost as the total amount of money needed to produce a certain item, or a quantitative measurement that accounts for all resources needed to produce an item. However, they refer to price as the amount of money that a person must pay for an item. When we go into a store, we normally ask the salesperson "What does this item cost?" Answering the literal question of what an item costs would encompass every resource that went into the development and production of that item. Instead, the accurate question is, "What's the item's price?" or "How much money must I exchange to receive that item?"

Because the term *cost* can refer to a number of different types or categories, the type of cost is important to understand during the cost estimation process. One of the first distinctions is between recurring and nonrecurring costs. A recurring cost is “repetitive and occurs each time a company produces a unit” (Mislick & Nussbaum, 2015, p. 26). When a bottling company produces a bottled beverage, each bottle cap has an associated cost. The cost of each bottle cap is recurring. In contrast, a nonrecurring cost is “not repetitive and cannot be tied to the quantity of the items being produced” (Mislick & Nussbaum, 2015, p. 26). The cost associated with purchase of the bottling machine would be considered nonrecurring. Closely related to recurring and nonrecurring costs are fixed and variable costs. Variable costs are associated and vary with the level of production (Mislick & Nussbaum, 2015). The more units produced, the more the total variable cost. However, fixed costs are unaffected by the level of production and are “generally associated with nonrecurring costs” (Mislick & Nussbaum, 2015, p. 27). No matter how many units are produced, the fixed cost will remain unchanged.

Another distinction between types of cost is direct and indirect costs. A direct cost can be “reasonably measured and allocated to a specific output, product, or work activity” (Mislick & Nussbaum, 2015, p. 26). The material used to produce an item is a direct cost. An indirect cost “cannot be attributed or allocated to a specific output, product, or work activity” (Mislick & Nussbaum, 2015, p. 27). The maintenance required for the upkeep of a machine used in production is indirect. Operating costs that are not direct labor or material, such as electricity and property taxes, are classified as overhead costs (Mislick & Nussbaum, 2015).

The Theory of Predictive Modeling in Cost

Generally, forecasting can be divided into quantitative and qualitative approaches. Qualitative forecasting is used when little to no reliable historical, contemporaneous, or comparable data exist. Several qualitative methods exist, such as the Delphi or expert opinion approach (a consensus-building forecast by field experts, marketing experts, or internal staff members), management assumptions (target growth rates set by senior management), and market research or external data or polling and surveys (data obtained through third-party sources, industry and sector indexes, or active market research). These estimates can be either single-point estimates (an average consensus) or a set of prediction values (a distribution of predictions). The latter can be entered into Risk Simulator as a custom distribution, and the resulting predictions can be simulated (i.e., running a nonparametric simulation using the prediction data points as the custom distribution).

For quantitative forecasting, the available data or data that need to be forecasted can be divided into time series (values that have a time element to them, such as revenues at different years, inflation rates, interest rates, market share, failure rates, and so forth), cross-sectional (values that are time independent, such as the grade point average of sophomore students across the nation in a particular year, given each student’s levels of SAT scores, IQ, and number of alcoholic beverages consumed per week), or mixed panel (mixture between time-series and panel data; e.g., predicting sales over the next 10 years given budgeted marketing expenses and market share projections, which means that the sales data are time series, but exogenous variables such as marketing expenses and market share exist to help to model the forecast predictions). Here is a quick review of each of the most commonly used forecasting methodologies.

Life-Cycle Cost

In developing a cost estimate, we first must understand a program’s or project’s life cycle. A life cycle follows the project or program from its inception to its disposal, or “cradle to grave.” It includes “the various stages of activity or phases through which the project progresses on its way from beginning to completion” (Rendon & Snider, 2008, p. 3). The life cycle starts at a program’s development; flows through its production, operation, and maintenance; and finally concludes after proper disposal. The costs associated with this process are classified as the program’s life-cycle cost.

The Defense Acquisition University (DAU) defines *life-cycle cost* as the direct cost of the acquisition program as well as the indirect cost that can be logically attributed to the program over the entire life cycle (Defense Acquisition University [DAU], n.d.-b). It includes the cost to the government to "acquire, operate, support (to include manpower), and where applicable, dispose" of a system or program (DAU, n.d.-b). There are multiple stakeholders in the DoD—such as Congress, the program manager and office, and contractors—that view a program's life-cycle cost from different perspectives. These multiple perspectives have led to three different methods of breaking down and displaying life-cycle cost.

The first method is breaking down program life-cycle costs by five different appropriation categories (DAU, n.d.-b): research, development, test, and evaluation (RDT&E); procurement; operations and maintenance (O&M); military construction (MILCON); and military personnel (MILPERS). This method is used to develop and submit budget requests to Congress (DAU, n.d.-b).

However, program managers and program offices would not find the first method as useful as Congress does. Instead, they utilize program life-cycle costs that are broken down by Work Breakdown Structure (WBS; DAU, n.d.-b). The DAU describes a *Work Breakdown Structure* as a framework that displays "the total system as a product-oriented family tree composed of hardware, software, services, data, and facilities" (DAU, n.d.-b). The WBS relates all of the work elements to each other and eventually to the final product (DAU, n.d.-b). A WBS encompasses all of the work necessary to produce a product (Huynh & Snider, 2008). This breakdown shows the relationship between costs and different elements of a system, which is a useful tool for program managers and contractors.

The Office of the Secretary of Defense for Cost Assessment and Program Evaluation (OSD CAPE) outlined the third life-cycle cost display method in its *Operating and Support Cost-Estimating Guide* (Office of the Secretary of Defense for Cost Assessment and Program Evaluation [OSD CAPE], 2014). The OSD CAPE defines a program's life-cycle cost as the summation of four different cost categories or phases: research and development (R&D), investment, operating and support (O&S), and disposal. Figure 1 provides a graphical representation of the four cost categories over a program's life cycle.

R&D is the initial cost category or phase in a program's life cycle. These costs are the first incurred in the research, design, and development of a new system or program. They can also include the "system design and integration; development, fabrication, assembly, and test of hardware and software for protypes and/or engineering development models" (OSD CAPE, 2014, pp. 2–3).

Following R&D is the investment cost category. These costs are incurred from "procurement and related activities from the beginning of low rate initial production (LRIP) through completion of deployment" (OSD CAPE, 2014, pp. 2–3). *Low rate initial production* refers to the production of the minimal number of a product or system that is required for initial operational test and evaluation (IOT&E; DAU, n.d.-c). Investment costs can include program management, initial spares, technical publications, and equipment training (OSD CAPE, 2014).

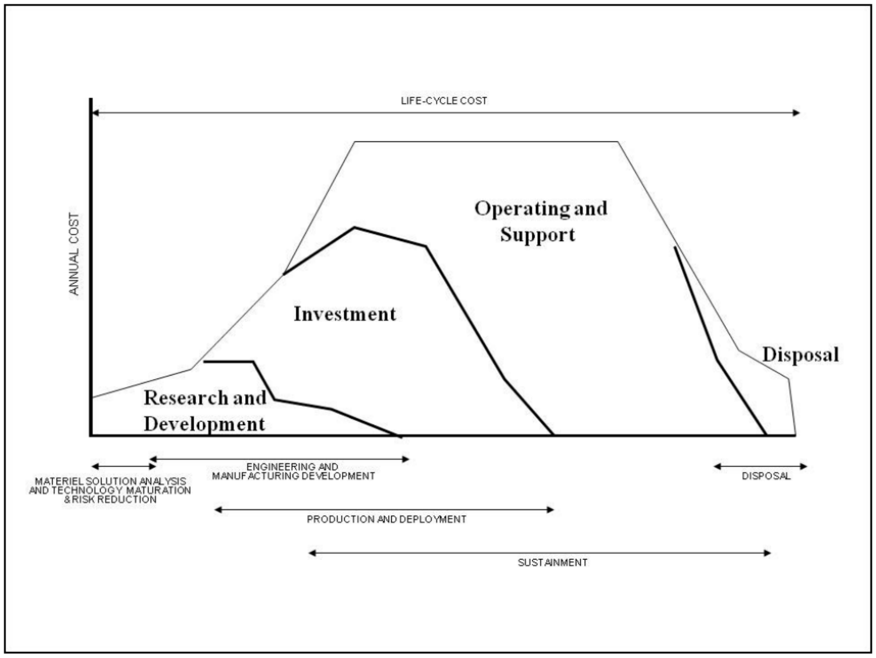


Figure 1. Notional Profile of Annual Program Expenditures by Major Cost Category over the System Life Cycle. Source: OSD CAPE (2014).

The O&S phase is the third phase in the OSD CAPE definition of life-cycle cost. The O&S phase normally accounts for a majority of a project's life-cycle costs (OSD CAPE, 2014). O&S consists of all of a system's operation and sustainment cost from initial deployment to the end of its operational life. This includes all the costs associated with "operating, maintaining, and supporting a fielded system" (OSD CAPE, 2014, pp. 2–3). Specifically, costs can include "personnel, equipment, supplies, software, and services associated with operating, modifying, maintaining, supplying, and otherwise supporting a system" (OSD CAPE, 2014, pp. 2–3).

The fourth and final OSD CAPE cost category is disposal. Disposal costs are those associated with the proper disposal or demilitarization at the end of a system's operational life (OSD CAPE, 2014). These costs can include "disassembly, materials processing, decontamination, collection/storage/disposal of hazardous materials and/or waste, safety precautions, and transportation of the system to and from the disposal site" (OSD CAPE, 2014, pp. 2–5). However, disposal costs can also be incurred during the sustainment phase due to unplanned system losses (OSD CAPE, 2014). We revisit this method of life-cycle costing in our discussion of total ownership costing.

Cost Estimation in the Department of Defense

Cost estimation is an important and required tool used by decision-makers in defense acquisitions. The requirement for a cost estimation is outlined in Department of Defense Instruction 5000.02, *Operation of the Defense Acquisition System*. Specifically, the instruction mandates that the

> DoD Component will develop a DoD Component Cost Estimate that covers the entire life cycle of the program for all Major Defense Acquisition Programs (MDAPs) prior to Milestone A, B, and C reviews and the Full-Rate Production Decision; and for all Major Automated Information System (MAIS) programs at any time an Economic Analysis is due. (DoD, 2017, p. 135)

This means that before the acquisition process can move beyond the MSA, TMRR, and EMD phases and ultimately continue on to full production, a cost estimate encompassing the entire program life cycle must be produced. In addition to the DoD's Component cost estimate, a separate, independent cost estimate is also required. *DODI 5000.02* requires the Milestone Decision Authority to consider an "independent estimate of the full life-cycle cost of a program, prepared or approved by the Director of Cost Analysis and Program Evaluation (DCAPE)" (DoD, 2017, p. 135). The DoD Component and DCAPE cost estimates are typically classified as Life-Cycle Cost Estimations (LCCEs). Mislick and Nussbaum (2015) describe an LCCE as a "a cost estimate for the totality of the resources that will be necessary throughout the product's life cycle" (p. 18).

There are four main cost estimating techniques used in the DoD to develop an LCCE, and they can be used in different phases of a program's life cycle (Ambrose, 2017). The first method is parametric cost estimating and involves the use of statistical inferences to generate an estimate based on system performance and design (Ambrose, 2017). Using historical data from similar systems, cost estimation relationships (CERs) and patterns are identified. Those patterns are assumed to hold true in the future and are used to predict cost (Mislick & Nussbaum, 2015). The second method is analogy cost estimating, whereby a new system is compared to a similar existing system. The analogy method is a relatively quick and inexpensive method; however, it may not be as precise as other methods (Ambrose, 2017). The parametric and analogy methods are normally used early on in the acquisition process during the materiel solution analysis (MSA), technology maturation and risk reduction (TMMR), and engineering and manufacturing development (EMD) phases (Ambrose, 2017). The third and most time-consuming method is engineering cost estimation. In this method, the system is broken down into its WBS elements in which individual detailed estimates are conducted. These estimates are then summed together to create the overall estimate (Mislick & Nussbaum, 2015). The engineering method is used during the TMMR phase and through the remaining acquisition process (Ambrose, 2017). The last main method used by the DoD is actual costing. This method uses the actual costs from a system that were incurred in the past to predict the cost of producing that system in the future (Ambrose, 2017). This method can be used after a program has entered the production and deployment (P&D) phase.

Total Ownership Cost

While LCCEs are a useful tool for decision-makers, they present a narrower scope when a broader perspective may be more beneficial (Kobren, 2014). Thus, we introduce the concept of total ownership cost (TOC). The DAU defines *total ownership cost* as including the "elements of life-cycle cost as well as other infrastructure or business process costs not normally attributed to the program" (Kobren, 2014). Infrastructure refers to "all military department and defense agency activities that sustain the military forces assigned to the combatant and component commanders" (Kobren, 2014). The major infrastructure categories are support to equipment, support to military personnel, and support to military bases (Kobren, 2014). Not normally included in a traditional LCCE, other support activities to consider in a cost estimate are recruiting, environmental and safety compliance, management headquarters functions, and logistics infrastructure activities (Kobren, 2014).

DoD Directive 5000.01 states that

> DoD Components shall plan programs based on realistic projections of the dollars and manpower likely to be available in future years. To the greatest extent possible, the MDAs shall identify the total costs of ownership, and at a minimum, the major drivers of total ownership costs. (DoD, 2003)

This requires the DoD to expand beyond the basic life-cycle cost estimation and include the support activities and infrastructure costs. To support the DoD directive, the Department of the Navy (DoN) issued its *Total Ownership Cost (TOC) Guidebook* in which it describes "new departmental and naval processes" that support the DoD policy of the identification of total costs of ownership (DoN, 2014, p.

6). Specifically, the guidebook assists the DoN and its organizations in developing, understanding, and applying the TOC requirements of the DoD.

The DoN outlines the importance of TOC: “As the DoD (and Navy) funding remains constant or declines, and as Navy’s purchasing power declines as a result, increasing the decision weight priority for alternatives that can mitigate and reduce TOC becomes our clearest path to a capable an optimally affordable Fleet” (DoN, 2014, p. 8). For this reason, we focus our model on TOC instead of a standard life-cycle cost.

Risk and Uncertainty

A key point that we need to understand in cost estimating is that the future is uncertain. Therefore, an essential pillar in developing a defensible and credible cost estimate is ensuring that risk and uncertainty are incorporated. A cost estimate can be severely affected by factors such as technological maturity, schedule slips, software requirements, or any other unforeseen event (Mislick & Nussbaum, 2015). Unknown factors make any “point estimate” or any exact answer extraordinarily unlikely (Mislick & Nussbaum, 2015). A more accurate estimate uses a central tendency centered on the original point estimate and a range both higher and lower to define the bounds of the estimate.

Though similar and related, risk and uncertainty are not synonymous. In the simplest terms, *risk* is the “probability” of the occurrence of a negative or unfavorable event, while *uncertainty* is the lack of certainty, or the realization that definitively knowing the outcome of any future event is completely impossible (Mislick & Nussbaum, 2015). Unlike with risk, with uncertainty we are not able to predict the possibility of any future outcome. In Johnathan Mun’s book, *Readings in Certified Quantitative Risk Management (CQRM),* he states,

> The concepts of risk and uncertainty are related but different. Uncertainty involves variables that are unknown and changing, but uncertainty will become known and resolved through the passage of time, events, and action. Risk is something one bears and is the outcome of uncertainty. Sometimes risk may remain constant while uncertainty increases over time. (Mun, 2015, p. 28)

A good way to think about risk and uncertainty is to imagine going on a sky diving trip with a friend. As the plane takes off, you and your friend realize that there is only one parachute and that the parachute is looking like it is somewhat past its service life. Your friend, being slightly more adventurous than you, decides to grab the parachute and take the jump. Both you and your friend share the same level of uncertainty about whether the parachute will open and whether your friend will live to tell the story. However, only your friend will assume the risk of jumping out of the plane and falling to his death.

Electro-Optical Infrared Sensors

Electro-optics (EO) are the field systems that convert electrons into photons (Driggers & Nichols, 2012). These systems are designed to respond to wavelengths within the 0.4–0.07 micrometer wavelength (Driggers & Nichols, 2012). They deliver images that are analogous to human vision; some EO systems are even capable of processing the near or short infrared spectral region (Driggers & Nichols, 2012). Figure 2 shows the basic components of an EO/IR sensor system.

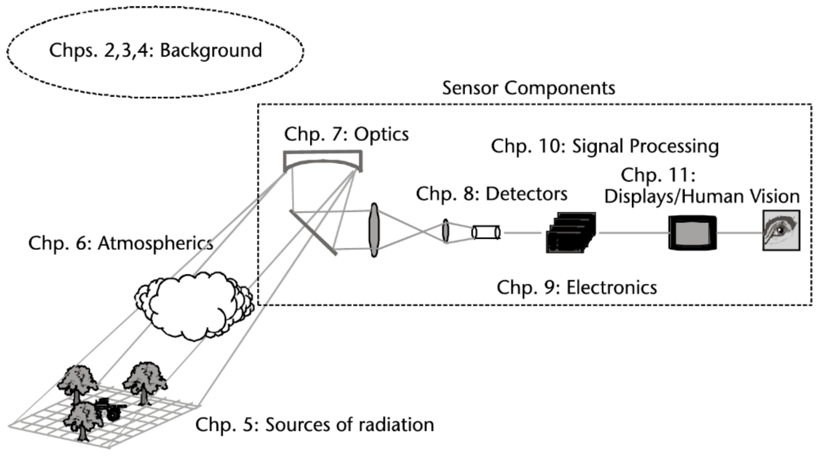


Figure 2. EO and IR Sensors (Driggers & Nichols, 2012)

The term *target* is used to describe the desired image that we are looking for with an EO sensor. The signal from a target usually has a large reflective component typically in the EO wavelength band. The target is provided this reflective component by moonlight, starlight, sunlight, or any artificial light source (Driggers & Nichols, 2012). The light sources reflecting off of the background and the target are known as external radiation. Radiation reflected by targets and background does not go directly to the EO sensor. The reflected radiation must first transition through the atmosphere, where it experiences scattering, before being processed by the EO sensor (Driggers & Nichols, 2012). Scattering is a phenomenon where particles in the atmosphere such as smoke, smog, or mist interfere with the reflection. Once the reflected radiation meets the EO sensor, it is passed through the sensing element, which could be detectors, tubes, or image intensifiers (low light situations; Driggers & Nichols, 2012). Next, the output of the sensor element is digested by the electronics and sent to a human interface for the operator (human) to gather some information from the process. This information could take a myriad of shapes such as detection, recognition, or identification of targets such as a warship. In short, EO sensors are essentially products of the light reflected from the scene (Driggers & Nichols, 2012). Figure 3 represents a typical EO sensor scenario.

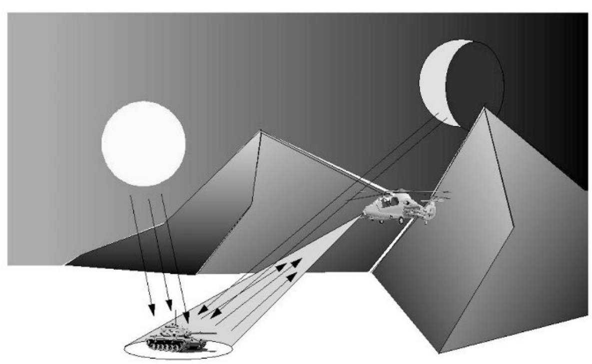


Figure 3. Typical EO Sensor Scenario (Driggers & Nichols, 2012)

Infrared is able to digest the spectral region from 0.7 to 14 micrometer wavelengths and is divided into four subregions:

> The near-infrared (NIR) region is from 0.7 to 1.1 mm, the short-wave infrared (SWIR) region is from 1.1 to 3 mm, the midwave infrared (MWIR) region is from 3 to 5 mm, and the long-wave infrared (LWIR) region is from 8 to 14 mm. Infrared is primarily used in night operations. (Driggers & Nichols, 2012)

The science of infrared is based on the science supporting Planck's law, which states that all bodies above the temperature of absolute zero emit electromagnetic radiation. The electromagnetic radiation is exploited to uncover the electromagnetic signatures given off that do not correlate to the wavelengths visible by the human eye or EO sensors.

> As the temperature of the object gets hotter, the peak wavelength moves to shorter wavelengths so that at very hot temperatures the radiation is perceived by the eye as light. The emissive surface characteristics of the hot object determine the spectral emission weighting of the radiation. The radiation emitted travels through the atmosphere, where it will then meet the aperture of the sensor. (Driggers & Nichols, 2012, p. 7)

EO/IR Sensors on Surface Ships

Before the advent of EO, direct optics were a commander's main resource in support of tactical decision-making. Binoculars, stadimeters, and periscopes were the keys to situational awareness and obtaining fire control solutions for torpedoes and gun engagements (Davidson, 2015). With the invention of EO, warfighters are no longer restricted to the limitations of the human eye. The application of using television cameras and the discovery of light-sensitive semiconductor materials allow images to be converted into electrical signals that are fed into displays for humans to process information. EO sensors paired with the ability of infrared detection allow warfighters to discern a target in the most vast and unlit environments (Davidson, 2015).

In Stefan Nitschke's (2007) article, "New Generation Naval Electro Optics," he states, "Electro Optical/Infrared technology is an invaluable aid for the 21st century battlespace arena. It provides surface warships, submarines, and maritime aviation operating in the varying naval environment with extensive image gathering, navigational, and targeting capabilities" (p. 87). The constant advances in EO/IR systems have led to the development of sensors with integral lasers that are used to measure distances with extreme accuracy and are a fraction of the size of the range finders of legacy ships (Davidson, 2015). In the report given by the Institute of Defense Analyses entitled *A Tutorial on Electro-Optical/Infrared (EO/IR) Theory and Systems*, it is stated that "the performance of an EO/IR sensor depends on the optics, detector, display, target-background contrast, and the intensity of the illumination source" (Koretsky et al., 2013, p. 5).

Technological advances have emphasized the importance of the opportunity and the necessity to reinvest in the newest technologies and systems. These advances in technology will drive future EO/IR systems purchases by the DoD. These system acquisitions will require credible and reliable cost estimations to ensure that the DoD manages its budget effectively. With the complexity and uniqueness of EO/IR systems, an efficient cost estimation model is needed to account for all life-cycle costs. The additional aspect of uncertainty should also be considered in the estimation. The cost estimation model we are proposing considers TOCs and uncertainty for the acquisition of EO/IR systems for U.S. Navy surface ships. This model will serve as a proof of concept to help future DoD decision-makers understand the cost associated with EO/IR systems so they can make strategic investments.

Model Application and Results

The inputs for this model were sourced from the program components lists provided by the research sponsor, NAVSEA, for the (generic or specific) EO/IR sensor. The cost estimates for this model were sourced using rough of order magnitude (ROM) values. The values fluctuate slightly between the five different systems to illustrate the differing systems' costs between contract estimates. These values were explicitly created to further the proof of concept of the model and, therefore, do not necessarily reflect the accurate value for component, part, or salary of support team members. However, these values do show how the simulation can provide an estimate of an entire system and demonstrate how much impact each variable will have on the overall life-cycle cost estimate. In this example, we simulate a cost estimate of an EO/IR system being implemented on 55 platforms with a service life of 20 years.

Model Inputs and Data

The Total Ownership Cost is calculated by summing the initial Acquisition Cost, Operation Cost, Maintenance Cost, and Disposal Cost. The model accounts for these four phases, beginning with the Acquisition Cost. In a real-world scenario, a cost analyst would utilize the technical specifications given by the program office to enter the required values. From the technical specifications, the analyst would insert two crucial metrics. The first is the number of platforms that will receive the system, and the second is the number of components required in each system. Since real-world data are not available for this notional model, this research uses the ROM system to fill in the blanks. In Systems A–E, the model uses 55 as the number of platforms.

The Acquisition Unit Cost accounts for all of the planning, design, and construction costs to make each component possible. The model also considers the estimated cost for a replacement component. The estimated cost for replacement parts should be considerably lower than the initial Acquisition Cost because developed technology will only need to be reproduced instead of being redeveloped. The Operational Cost per year is an estimate of the amount required to run the component for a year. The Operation Cost includes equipment depreciation, costs of the energy source used to power the component, cost of damage due to use, and so on. Similarly, the Maintenance cost is an estimate based on the amount required to maintain the equipment every year. Figure 4 shows the categories for Acquisition and Operation and Maintenance Costs.

Once the cost analyst has entered the acquisition cost for the hardware and software required for the system, the analyst must remember to account for the human element. The analyst will need to ensure that the cost required to pay for those responsible for the design, logistics, management, and technology are represented in the model. This model uses the Acquisition Cost column to record the initial salary of each job. The Number of Platforms column describes the number of teams required for each system. The Number of Units per System column describes the number of people required on each team. The Operation Cost column is used to annotate the continuing salary for the human element for the remainder of the program's life. Essentially, this is how an analyst would annotate a recurring salary payment. Throughout the five systems, the number of people per team and the amount requested per salary will vary. Figure 5 shows an example of where salaries are inputted into the model.

All of the costs mentioned previously are recurring costs, costs that will be multiplied by the number of years of the program and summed to get the total cost. Analysts must be sure not to forget to account for all of the one-time costs associated with the origins of any project. Figure 6 shows the list of nonrecurring costs accounted for in the model.

Finally, we account for all of the disposal and end-of-life-cycle costs that will also be one-time costs. Figure 7 shows the nonrecurring end-of-life-cycle costs.

Categories	Number of Units per System	Number of Platforms	Acquisition Cost (Unit)	%	Operational Costs (Unit) Per Year	%	Maintenance (Unit) Per Year	%	Replacement (Unit) Per Year	%	Total Acquisition Cost	%	Total Annual O&M	%
Grand Total			##############		$217.00		$1,164.00		$42,391.00		$45,863,500.00		$15,308,443.00	
Narrow-Medium Field of View (NFOV) Sensors	**43**	**935**	**$7,110.00**	**0.2%**	**$73.00**	**33.6%**	**$594.00**	**51.0%**	**$4,880.00**	**11.5%**	**$960,850.00**	**2.1%**	**$753,720.00**	**4.9%**
NF-DIR (NFOV Director)	2	55	$400.00	0.0%	$5.00	2.3%	$30.00	2.6%	$300.00	0.7%	$44,000.00	0.1%	$36,850.00	0.2%
NF-TIS (Thermal Imaging Sensor) - TIS #1	3	55	$350.00	0.0%	$6.00	2.8%	$23.00	2.0%	$150.00	0.4%	$57,750.00	0.1%	$29,535.00	0.2%
NF-TIS (Thermal Imaging Sensor) - TIS #2	2	55	$460.00	0.0%	$7.00	3.2%	$25.00	2.1%	$300.00	0.7%	$50,600.00	0.1%	$36,520.00	0.2%
NF-EOS (Electro-Optic Sensor) - EOS #1	3	55	$230.00	0.0%	$5.00	2.3%	$34.00	2.9%	$200.00	0.5%	$37,950.00	0.1%	$39,435.00	0.3%
NF-EOS (Electro-Optic Sensor) - EOS #2	3	55	$340.00	0.0%	$3.00	1.4%	$45.00	3.9%	$220.00	0.5%	$56,100.00	0.1%	$44,220.00	0.3%
NF-EOS (Electro-Optic Sensor) - EOS #3	3	55	$450.00	0.0%	$2.00	0.9%	$56.00	4.8%	$250.00	0.6%	$74,250.00	0.2%	$50,820.00	0.3%
NF-LRF (Laser Rangefinder)	2	55	$560.00	0.0%	$3.00	1.4%	$45.00	3.9%	$560.00	1.3%	$61,600.00	0.1%	$66,880.00	0.4%
NF-LDR (Laser Designator/Rangefinder)	2	55	$430.00	0.0%	$4.00	1.8%	$34.00	2.9%	$220.00	0.5%	$47,300.00	0.1%	$28,380.00	0.2%
NF-LDRFI (Laser Designator/Rangefinder/Illuminator)	2	55	$460.00	0.0%	$5.00	2.3%	$23.00	2.0%	$140.00	0.3%	$50,600.00	0.1%	$18,480.00	0.1%
NF-LP (Laser Pointer)	5	55	$450.00	0.0%	$6.00	2.8%	$45.00	3.9%	$270.00	0.6%	$123,750.00	0.3%	$88,275.00	0.6%
NF-LOI (Laser Optical/Ocular Interrupter)	1	55	$560.00	0.0%	$6.00	2.8%	$65.00	5.6%	$320.00	0.8%	$30,800.00	0.1%	$21,505.00	0.1%
NF-LI (Laser Illuminator)	3	55	$430.00	0.0%	$3.00	1.4%	$43.00	3.7%	$540.00	1.3%	$70,950.00	0.2%	$96,690.00	0.6%
NF-IRU (Inertial Reference Unit)	2	55	$430.00	0.0%	$3.00	1.4%	$34.00	2.9%	$450.00	1.1%	$47,300.00	0.1%	$53,570.00	0.3%
NF-BSM (Boresight Module)	1	55	$230.00	0.0%	$3.00	1.4%	$23.00	2.0%	$220.00	0.5%	$12,650.00	0.0%	$13,530.00	0.1%
NF-EU (Electronics Unit)	2	55	$670.00	0.0%	$3.00	1.4%	$23.00	2.0%	$330.00	0.8%	$73,700.00	0.2%	$39,160.00	0.3%
Ancillary Material (cabling, mounting hardware, etc.)	3	55	$430.00	0.0%	$3.00	1.4%	$23.00	2.0%	$200.00	0.5%	$70,950.00	0.2%	$37,290.00	0.2%
Other: __________	4	55	$230.00	0.0%	$6.00	2.8%	$23.00	2.0%	$210.00	0.5%	$50,600.00	0.1%	$52,580.00	0.3%
Wide Field of View (WFOV) Sensors	**23**	**385**	**$25,600.00**	**0.6%**	**$24.00**	**11.1%**	**$245.00**	**21.0%**	**$14,200.00**	**33.5%**	**$4,240,500.00**	**9.2%**	**$2,487,100.00**	**16.2%**
WF-DIR (Director)	2	55	$4,500.00	0.1%	$4.00	1.8%	$35.00	3.0%	$2,000.00	4.7%	$495,000.00	1.1%	$224,290.00	1.5%
WF-TIS (Thermal Imaging Sensor)	3	55	$3,500.00	0.1%	$3.00	1.4%	$43.00	3.7%	$1,200.00	2.8%	$577,500.00	1.3%	$205,590.00	1.3%
WF-EOS (Electro-Optic Sensor)	1	55	$4,500.00	0.1%	$2.00	0.9%	$23.00	2.0%	$3,200.00	7.5%	$247,500.00	0.5%	$177,375.00	1.2%
WF-IRU (Inertial Reference Unit)	2	55	$5,300.00	0.1%	$6.00	2.8%	$22.00	1.9%	$2,300.00	5.4%	$583,000.00	1.3%	$256,080.00	1.7%
WF-EU (Electronics Unit)	4	55	$1,000.00	0.0%	$3.00	1.4%	$55.00	4.7%	$1,000.00	2.4%	$220,000.00	0.5%	$232,760.00	1.5%
Ancillary Material (cabling, mounting hardware, etc.)	5	55	$2,300.00	0.1%	$2.00	0.9%	$45.00	3.9%	$2,100.00	5.0%	$632,500.00	1.4%	$590,425.00	3.9%
Other: __________	6	55	$4,500.00	0.1%	$4.00	1.8%	$22.00	1.9%	$2,400.00	5.7%	$1,485,000.00	3.2%	$800,580.00	5.2%
EO/IR Sensor Manager (ESM)	**17**	**330**	**$1,910.00**	**0.0%**	**$45.00**	**20.7%**	**$124.00**	**10.7%**	**$1,400.00**	**3.3%**	**$282,150.00**	**0.6%**	**$275,990.00**	**1.8%**
Processing Equipment	3	55	$340.00	0.0%	$4.00	1.8%	$15.00	1.3%	$150.00	0.4%	$56,100.00	0.1%	$27,885.00	0.2%
Processing Software	4	55	$230.00	0.0%	$6.00	2.8%	$23.00	2.0%	$230.00	0.5%	$50,600.00	0.1%	$56,980.00	0.4%
Recording Equipment	5	55	$240.00	0.0%	$5.00	2.3%	$40.00	3.4%	$430.00	1.0%	$66,000.00	0.1%	$130,625.00	0.9%
Docking Station Equipment	2	55	$350.00	0.0%	$5.00	2.3%	$21.00	1.8%	$230.00	0.5%	$38,500.00	0.1%	$28,160.00	0.2%
Ancillary Material (video converters, encoders, ethernet switches, racks, cabling, etc.)	1	55	$210.00	0.0%	$12.00	5.5%	$10.00	0.9%	$210.00	0.5%	$11,550.00	0.0%	$12,760.00	0.1%
Other: __________	2	55	$540.00	0.0%	$13.00	6.0%	$15.00	1.3%	$150.00	0.4%	$59,400.00	0.1%	$19,580.00	0.1%

Figure 4. Categories for Acquisition and Operation and Maintenance Costs

Categories	Number of Units per System	Number of Platforms	Acquisition Cost (Unit)	%	Operational Costs (Unit) Per Year	%
Grand Total			$4,202,920.00		$1,907,716.00	
Manpower and Personnel	30	6	$321,000.00	7.6%	$240,000.00	12.6%
Program Management Office Team	8	1	$80,000.00	1.9%	$80,000.00	4.2%
Manning and military occupational series training	6	1	$40,000.00	1.0%	$40,000.00	2.1%
Depot Activation	5	1	$60,000.00	1.4%	$55,000.00	2.9%
Software Sustainment	4	1	$40,000.00	1.0%	$35,000.00	1.8%
Initial Fielding Support	4	1	$56,000.00	1.3%	$30,000.00	1.6%
Other: __________	3	1	$45,000.00	1.1%	$0.00	0.0%

Figure 5. Manpower and Personnel Salary Input Section

Nonrecurring Acquisition and End of Lifecycle Costs	Total	%
Acquisition and Procurement	$467,800.00	
Bid Specifications Development	$10,000.00	2.1%
Proposal Evaluation	$2,000.00	0.4%
Data Collection	$40,000.00	8.6%
Data Analysis	$12,000.00	2.6%
Contracts Development	$3,000.00	0.6%
Program Planning	$4,000.00	0.9%
Hardware Purchases	$10,000.00	2.1%
Personal Computers	$10,000.00	2.1%
Peripherals	$15,000.00	3.2%
Storage	$60,000.00	12.8%
Networking	$23,000.00	4.9%
Related Equipment	$35,000.00	7.5%
Other costs	$10,000.00	2.1%
Administrative Cost	$34,000.00	7.3%
Asset Management	$15,000.00	3.2%
Overseeing Contractor Services	$4,000.00	0.9%
In-House Training for Staff	$5,000.00	1.1%
Product Maintenance	$2,000.00	0.4%
Help Desk Support	$10,000.00	2.1%
IT Support for Database Management	$20,000.00	4.3%
Network Management Support	$42,000.00	9.0%
Software Upgrades	$12,000.00	2.6%
Hardware Upgrades	$2,100.00	0.4%
Internet and Network Access Cost	$14,000.00	3.0%
Furniture and Equipment	$10,000.00	2.1%
Energy Costs	$3,400.00	0.7%
Informal Training	$4,300.00	0.9%
Downtime Support and Outsource	$24,000.00	5.1%
Other costs	$32,000.00	6.8%

Figure 6. Nonrecurring Acquisition and Procurement Costs

Nonrecurring End of Lifecycle Costs	Total	%
End of Lifecycle	$109,000.00	
Administrative Cost	$40,000.00	36.7%
Asset Management	$20,000.00	18.3%
Vendor Contract Procurement	$4,000.00	3.7%
Staging, Sanitizing, Testing	$10,000.00	9.2%
Follow-Up Support	$10,000.00	9.2%
Recycling and Disposal Fees	$5,000.00	4.6%
Value of Sold Products and Materials	$20,000.00	18.3%

Figure 7. Nonrecurring End-of-Life-Cycle Costs

Results and Analysis

Once the data have been manually inputted into the model, the cost analyst can utilize the multitude of charts, graphs, and tools to analyze the TOC of the systems. These graphs, charts, and tools allow the analyst to compare multiple cost estimates over the entire life of the system at the same time. This research analyzed the following tables and charts to highlight the functionality of the model: Total Net Life-Cycle Cost, Present Value of Discounted Total Net Life-Cycle Cost, Cash Total Net Cost at 5-Year Increments, Total Ownership Cost Forecast Statistic Table, Simulation Probability Charts, and the Tornado Analysis.

Total Net Life-Cycle Costs and Cash Total Net Cost at 5-Year Increments

Figure 8 shows the Total Net Life-Cycle Cost for all five systems over a span of 30 years. The table and graph show the cost for the systems broken down into 5-year estimates. The model projects the life span of the system past the 20-year expected service life. This extension allows the cost analyst to consider cost out to the 30-year point, as many DoD systems tend to exceed their expected service lives. However, the 5-year increments also allow a decision-maker to understand the total net cost of disposing of a system before its 20-year service life. The side-by-side comparison enables a decision-maker to graphically perceive the potential differences between the cost estimates of the multiple systems. When choosing between alternatives, Figure 8 can be a beneficial decision aid.

In the analysis table in Figure 8, the 20-Year Cash Total Net Cost ranges from $554 million (System C) to $771 million (System D). If cost were the determining factor, a decision-maker could quickly determine that System C should be selected. To make the comparison even easier to analyze, Figure 9 provides a side-by-side comparison of all five systems at each of the 5-year increments. Looking at the 20-Year Total Net Cost Graph, it can be clearly seen that System C has the lowest Total Net Cost.

Cost analysis should only be one part of the picture when it comes to making the correct strategic decision. For example, each system’s specifications and capabilities—its military benefits or returns—should also be computed, such that each system will have its own return on investment (ROI). Nonetheless, the major component of any ROI analysis is cost. The focus of this research is to determine this cost computation. Another aspect of TOC analysis is its use in cost mitigation, cost savings, and cost deferred, which constitute another point of view of cost-based decision analytics.

Analysis Period/Type	System A	System B	System C	System D	System E
5 Year Cash Total Net Cost	$192,078,759.39	$199,950,888.59	$158,074,801.16	$206,656,401.49	$161,982,378.15
10 Year Cash Total Net Cost	$348,972,742.05	$366,909,094.54	$280,489,136.93	$381,039,284.24	$291,014,854.14
15 Year Cash Total Net Cost	$517,992,119.86	$546,770,499.13	$412,364,142.71	$568,899,174.37	$430,019,476.51
20 Year Cash Total Net Cost	$700,073,991.93	$740,532,313.22	$554,430,976.94	$771,277,628.99	$579,766,932.65
25 Year Cash Total Net Cost	$896,227,880.11	$949,268,816.09	$707,477,304.94	$989,296,700.88	$741,087,471.77
30 Year Cash Total Net Cost	$1,107,541,326.15	$1,174,137,311.67	$872,351,665.95	$1,224,165,159.57	$914,875,508.07

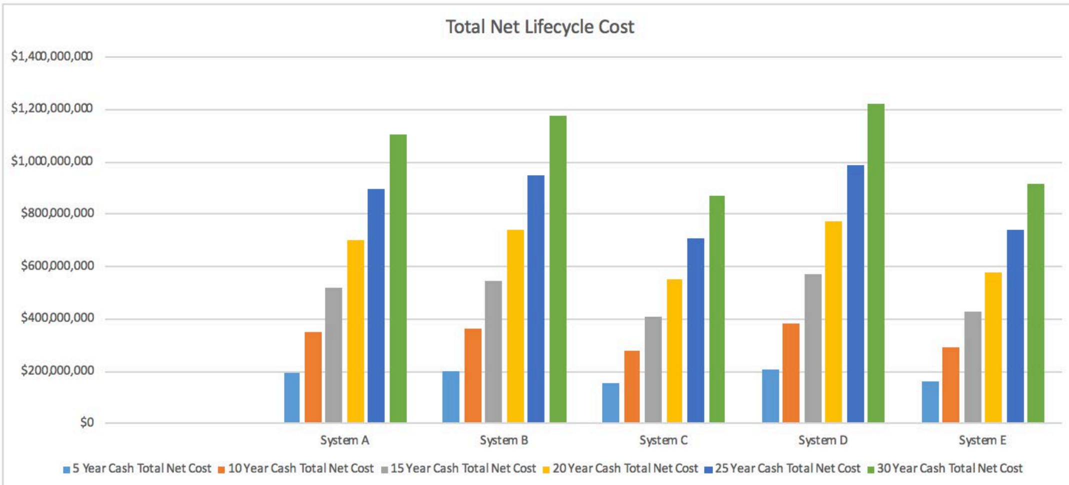


Figure 8. Total Net Life-Cycle Cost

Present Value of Discounted Total Net Life-Cycle Cost

While Figure 10 shows the Total Net Life-Cycle Cost, it does not include consideration of economic factors such as the time value of money and uncertainty risk. To mitigate these factors in the model, Figure 10 incorporates a Net Present Value Life-Cycle Cost estimate using a discount rate of 3% (i.e., the government's cost of money, where we can use 20-year and 30-year Treasury bond yields as proxies). In the analysis table in Figure 10, the 20-Year Total Net Cost ranges from $554 million (System C) to $771 million (System D), but when looking at the more realistic Present Value Discounted Net Life-Cycle Cost, the range between Systems C and D decreases to $418 million and $577 million. Not only do the estimates for the minimum and maximum values decrease when the discount factor is applied, but the delta of the range between the values also shrinks by $57.8 million. Incorporating the discount rate into the model gives the decision-maker a complete analysis of the costs. Specifically, it shows the value of the lifetime cost of a system in today's money, thereby putting all systems with different life cycles and life spans on an equal footing with each other for a better cost comparison.

Analysis Period/Type	System A	System B	System C	System D	System E
5 Year Cash Cost in Present Values	$179,704,285.34	$186,783,594.12	$148,416,499.73	$192,904,219.83	$151,802,725.35
10 Year Cash Cost in Present Values	$303,544,126.37	$318,568,174.68	$245,037,964.84	$330,549,870.97	$253,648,577.73
15 Year Cash Cost in Present Values	$418,626,211.89	$441,033,105.19	$334,826,703.72	$458,461,345.77	$348,292,207.24
20 Year Cash Cost in Present Values	$525,569,818.93	$554,837,402.90	$418,265,845.06	$577,326,978.99	$436,242,875.34
25 Year Cash Cost in Present Values	$624,950,442.22	$660,593,492.07	$495,804,364.31	$687,786,438.92	$517,973,842.03
30 Year Cash Cost in Present Values	$717,302,888.41	$758,870,497.06	$567,859,497.04	$790,434,167.10	$593,924,909.86

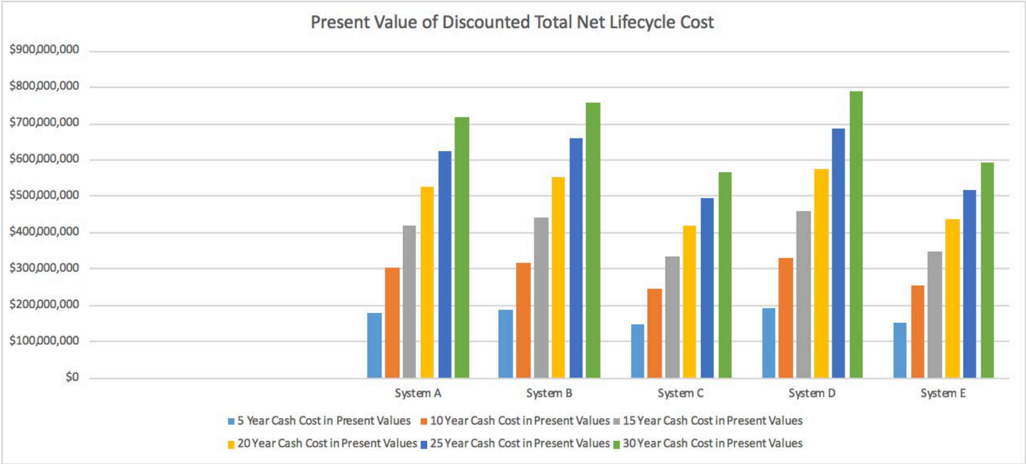


Figure 9. Present Value of Discounted Net Life-Cycle Cost

Forecast Statistics Table - TOC Model

Cell	Total Lifetime Cost for System A (20 Years)	Total Lifetime Cost for System B (25 Years)	Total Lifetime Cost for System C (20 Years)	Total Lifetime Cost for System D (10 Years)	Total Lifetime Cost for System E (15 Years)	Total PV Lifetime Cost for System A (20 Years)	Total PV Lifetime Cost for System B (25 Years)	Total PV Lifetime Cost for System C (20 Years)	Total PV Lifetime Cost for System D (10 Years)	Total PV Lifetime Cost for System E (15 Years)
Name	S42	S42	S42	S42	S42	S43	S43	S43	S43	S43
Number of Datapoints	10,000	10,000	10,000	10,000	10,000	10,000	10,000	10,000	10,000	10,000
Mean	$700,128,499.57	$740,532,963.43	$554,392,425.60	$771,338,025.77	$579,775,293.62	$525,611,374.46	$554,838,179.48	$418,238,633.17	$577,371,510.62	$436,248,097.08
Median	$700,141,006.85	$740,535,580.46	$554,384,371.67	$771,364,875.28	$579,759,448.54	$525,610,342.27	$554,835,058.62	$418,245,556.21	$577,382,472.47	$436,233,772.78
Standard Deviation	$6,257,100.80	$5,845,829.50	$4,467,888.27	$7,611,267.45	$4,630,641.37	$4,594,559.43	$4,292,302.35	$3,283,147.51	$5,583,858.78	$3,400,790.39
Coefficient of Variation	0.89%	0.79%	0.81%	0.99%	0.80%	0.87%	0.77%	0.78%	0.97%	0.78%
Maximum	$721,408,465.72	$760,794,257.46	$568,380,140.14	$792,265,719.52	$595,514,481.66	$541,361,082.89	$569,528,468.16	$428,545,963.69	$592,883,001.86	$447,659,644.97
Minimum	$679,850,811.09	$719,523,385.66	$540,339,232.67	$748,640,296.95	$564,025,304.82	$510,709,204.65	$539,393,598.53	$407,787,597.45	$560,676,945.01	$424,678,356.35
Range	$41,557,654.63	$41,270,871.80	$28,040,907.47	$43,625,422.57	$31,489,176.84	$30,651,878.24	$30,134,869.63	$20,758,366.24	$32,206,056.85	$22,981,288.62
Skewness	-0.0032	-0.0042	-0.0148	-0.0026	0.0087	-0.0023	-0.0028	-0.0144	-0.0026	0.0091
Kurtosis	-0.4009	-0.2633	-0.3281	-0.5003	-0.3088	-0.3995	-0.2616	-0.3249	-0.4998	-0.3093
25% Percentile	$695,719,253.67	$736,446,455.02	$551,284,478.55	$765,796,106.33	$576,509,002.84	$522,378,543.35	$551,839,791.47	$415,958,053.49	$573,311,007.82	$433,851,693.64
75% Percentile	$704,532,335.37	$744,600,466.38	$557,526,620.08	$776,618,455.92	$582,946,797.53	$528,851,703.11	$557,822,109.37	$420,545,180.57	$581,253,957.83	$438,577,466.99
Error Precision at 95%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%
5% Percentile	$689,682,929.25	$730,959,587.63	$546,946,259.66	$758,769,396.90	$572,144,154.21	$517,951,178.33	$547,793,197.77	$412,759,859.00	$568,154,901.40	$430,640,138.12
10% Percentile	$691,747,640.51	$732,794,898.81	$548,533,001.64	$761,225,549.84	$573,689,454.88	$519,468,239.36	$549,181,125.39	$413,946,175.50	$569,990,913.70	$431,776,478.59
20% Percentile	$694,602,871.32	$735,367,935.12	$550,510,692.26	$764,515,420.38	$575,778,821.69	$521,563,700.07	$551,045,883.57	$415,386,396.82	$572,378,313.33	$433,315,521.79
30% Percentile	$696,704,619.98	$737,452,394.84	$551,979,547.59	$767,095,731.78	$577,226,617.51	$523,096,629.77	$552,578,355.90	$416,460,293.34	$574,253,331.14	$434,382,267.57
40% Percentile	$698,493,546.06	$739,011,582.52	$553,233,344.80	$769,309,931.15	$578,543,844.11	$524,394,394.43	$553,713,041.68	$417,379,609.28	$575,880,312.72	$435,341,189.38
50% Percentile	$700,141,006.85	$740,535,580.46	$554,384,371.67	$771,364,875.28	$579,759,448.54	$525,610,342.27	$554,835,058.62	$418,245,556.21	$577,382,472.47	$436,233,772.78
60% Percentile	$701,843,734.05	$742,099,383.62	$555,579,609.61	$773,403,235.14	$580,989,590.40	$526,887,691.99	$555,979,070.58	$419,103,492.37	$578,891,015.17	$437,139,948.28
70% Percentile	$703,601,587.76	$743,714,003.76	$556,847,632.09	$775,491,755.81	$582,222,603.44	$528,161,141.23	$557,172,586.27	$420,044,672.96	$580,410,041.35	$438,057,186.64
80% Percentile	$705,593,527.64	$745,597,406.41	$558,265,766.20	$778,104,442.05	$583,805,528.28	$529,626,241.70	$558,543,630.43	$421,090,371.98	$582,332,648.63	$439,205,172.05
90% Percentile	$708,345,917.64	$748,088,346.78	$560,256,120.90	$781,498,101.24	$585,875,474.23	$531,647,003.55	$560,385,422.61	$422,546,755.03	$584,823,963.64	$440,711,757.31
95% Percentile	$710,307,855.95	$750,205,085.99	$561,774,937.44	$784,001,472.92	$587,517,826.61	$533,116,251.14	$561,929,004.53	$423,658,279.36	$586,692,336.33	$441,943,293.84
99% Percentile	$713,991,274.06	$753,529,283.26	$564,370,570.06	$787,642,529.91	$589,971,521.84	$535,763,348.04	$564,406,101.50	$425,570,623.73	$589,320,737.18	$443,747,515.40

Figure 10. Total Ownership Cost Forecast Statistics Table

Stochastic Total Ownership Cost Forecast Statistics Table

The Forecast Statistics Table, shown in Figure 10, summarizes the distribution of the Total Life-Cycle Cost and the Total Present Value (PV) Life-Cycle Cost for the five systems at different points in the life cycle of the system based on risk-based simulation and stochastic TOC models used to value the alternative cost paths. Figure 10 highlights the outcomes of running 10,000 trials using the Monte Carlo Risk Simulator. The takeaways from this figure are the mean, standard deviation, maximum, minimum, and range data points. These metrics provide a decision-maker with a better understanding of how uncertainty can affect the Total Life-Cycle Cost and Total PV Life-Cycle Cost of a system.

System C looks at the cost over a 20-year life span. Using the Monte Carlo Risk Simulator, the maximum Total Life-Cycle Cost of the system is $568 million, while the minimum is $540 million. These values represent the worst- and best-case scenarios, respectively. The simulations produced a Total Life-Cycle Cost range of $28 million and a mean value of $554 million. The standard deviation of Total Life-Cycle Cost simulations for System C is $4.5 million, meaning that 68.2% of the estimates will fall within ±$4.5 million of the mean if the distribution is somewhat normally distributed. Figure 10 also shows the same metrics for the PV of the Total Life-Cycle Cost for all systems.

Simulation Probability Charts

A simulation probability chart is a histogram or frequency distribution of all of the total life-cycle costs of a system based on 10,000 simulation runs or trials. The probability chart produces a graphic representation of the information contained in the forecast statistics table. Figure 11 shows the frequency distribution of the total life-cycle cost for System A over a 20-year life. In the figure, System A's frequency distribution is shaped as a roughly symmetrical bell curve centered on a mean of $700 million. Using this chart, an analyst could confidently conclude that the total life-cycle cost for this system will fall between $679 million and $721 million. The figure also shows the 90% confidence interval of the TOC to be between $690 million and $710 million. This means that there is a 90% chance that given all uncertainties that exist in each of the input assumptions, the 20-year total lifetime cost for System A will be between these two values. In addition, there is only a 5% chance that the cost can be below $690 million and a 5% chance it can exceed $710 million. Figure 12 uses the same frequency distribution over the same 20-year system life as in Figure 11; however, Figure 12 takes into account the discount rate to better illustrate the economic factor of inflation over time. Similarly, the 90% confidence interval in present values is between $518 million and $533 million.

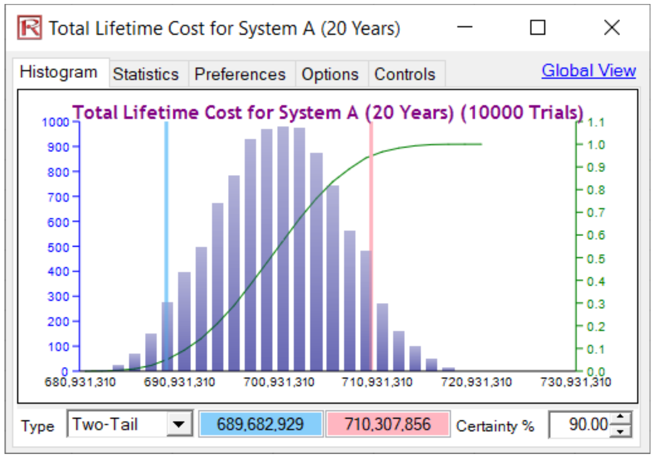


Figure 11. Total Life-Cycle Cost for System A (20 Years)

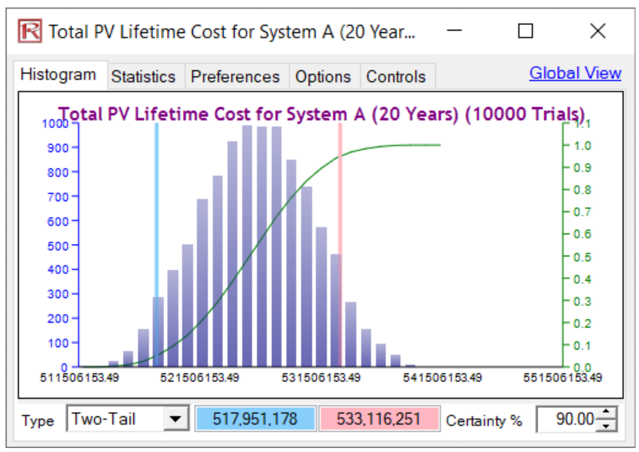


Figure 12. Total Present Value Life-Cycle Cost for System A (20 Years)

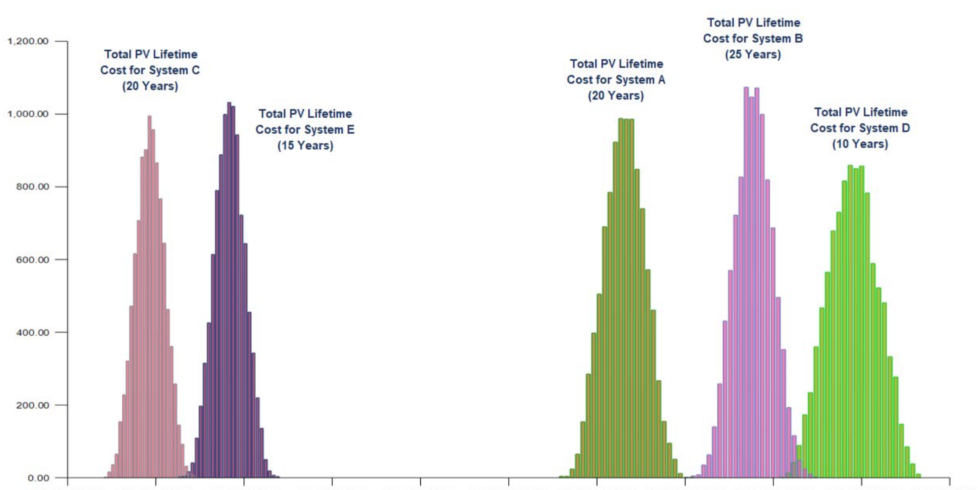


Figure 13. Probability Distribution Cost Overlay of the Five Systems

Tornado Analysis

The tornado analysis chart gives decision-makers the ability to break down which variables have the most significant impact on the overall outcome of the simulation. By focusing on the top critical factors, decision-makers can focus on cost reduction techniques in places that will have the most effect. The tornado analysis allows the decision-makers to adjust how many critical variables to display. Figure 14 shows the tornado analysis chart detailing the 20 most impactful variables on the TOC model. Based on the notional cost values inputted into the model, the number of platforms containing that ancillary material is the most critical factor.

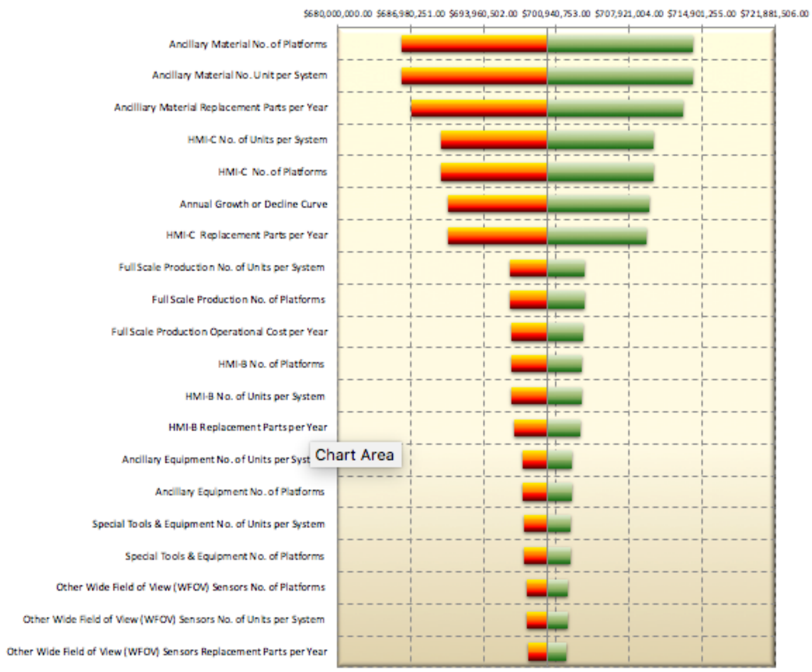


Figure 14. Tornado Analysis

Conclusion

Key Conclusions

The purpose of this report was to develop a total ownership with life-cycle cost model while considering uncertainty for EO/IR sensors on U.S. Navy surface ships. Through the examination of TOC modeling over the life cycle of EO/IR sensors, including the inception phase of acquisition costs, followed by annual O&M expenses, along with a final set of Disposition Costs, we were able to develop a useful model for TOC estimations. Using Monte Carlo risk simulation, our model accounts for risk and uncertainty when producing cost estimates. The model also provides analysts with a more realistic estimate by factoring in economic theory, such as economic growth, annual discount rate, and inflation.

As discussed, the cost analysis models presented should be only one part of a larger picture when it comes to making the correct strategic investment decisions. For example, each system's specifications, capabilities, military benefits, or financial and noneconomic returns should also be computed, such that each system will have its own return on investment (ROI). Nonetheless, the major component of any ROI analysis is cost. The focus of this current research is to determine a suitable method to compute critical life-cycle cost. Another use of TOC modeling is in determining cost mitigation, cost savings, and cost deferred, that is, what the cost differential might be or an Analysis of Alternatives, which constitutes another point of view of cost-based decision analytics. The model allows decision-makers to have better decision analytics of the costs of surface EO/IR sensors. These analytics can be used in subsequent cost comparisons between different sensor platforms, Analysis of Alternatives, and portfolio allocation of resources. Specifically, Program Executive Office Integrated Warfare Systems (PEO IWS) and NAVSEA can utilize this model in future program cost estimation development. Since the model is tailorable to different sensor configurations, it can provide clarity in analyzing different and complex alternative sensor systems to develop and outfit the fleet. The results of this model give decision-makers a more effective metric to analyze TOC under uncertainty; this can reduce cost overflow and prevent budget overruns. Ultimately, the model allows leadership to make more informed decisions in the DoD acquisition process and maximize the use of its limited resources.

Current Research Limitations and Follow-on Research

The main limitation of the current study is that notional cost data were used to provide a proof of concept that the model functions as designed. However, this presents an opportunity for future research whereby additional follow-on research with empirical data should be conducted. This model can analyze cost data in past, present, and future EO/IR models.

Beginning with historical data, a cost analyst could compile a list of program components associated with a system that is either retired or currently in use. Once the list of components is obtained, the analyst can then associate the estimated historical cost assigned to each component during the program's initial cost estimate (e.g., a program cost estimate developed in 1992). Using the original cost data and component list, the analyst could then run the new total ownership with life-cycle cost model under uncertainty. This would produce a new cost estimate for the program, which could then be compared to the original estimate and the actual life-cycle cost of the program. Executing this study would determine whether the TOC model developed in this thesis is a superior method of cost estimation for the DoD.

Another follow-on study could be done using the data from a program that is currently undergoing its initial cost estimation. The cost estimate could be done in conjunction with the DoD's current methods of cost estimation. Another researcher could partner with PEO IWS and the new system's program office to complete a cost estimate using the TOC model developed in this thesis. This process would allow for real-time cost comparisons at different stages in the acquisition process. The comparison between the two estimates would provide decision-makers with another method of verifying assumptions and validating that their cost estimates are reasonable and credible. Concurrently conducting the cost estimates allows researchers and cost estimators to compare their estimates to actual cost data at the different increments throughout the program's life cycle. This comparison would determine which method of cost estimation was more accurate at different points in the system's life cycle.

These follow-on studies require real-world cost data from historical or current EO/IR programs. While data collection may prove difficult and time-consuming, this research would be beneficial to the DoD and well worth the investment. Working with PEO IWS and the program office's cost estimation teams could result in model improvements and provide an even more robust total ownership with life-cycle cost model under uncertainty.

Other Applications and Conclusions

This research focuses specifically on the application of this TOC model with regard to EO/IR sensors on surface ships; it barely scratches the surface of the model's potential. This model could be applied to any one of the thousands of acquisition projects in the DoD. The model's use is not confined to EO/IR sensors on surface ships but can be adjusted and developed for various programs. The process and the strength of the results that the model would provide would be the same; the only necessary change a cost analyst would need to make is to alter the list of components to reflect whichever system or program is being analyzed. In the same fashion, this model could also provide contractors and non-DoD organizations with an additional method of cost estimation.

Cost estimation is not an exact science; however, this model provides a coherent method of estimating the total ownership with life-cycle costs under uncertainty for EO/IR sensors on surface ships. It gives a decision-maker another tool when evaluating alternative programs and courses of action. The ultimate goal of this model is to provide a more effective tool in determining how the DoD spends its limited resources on competing priorities. While follow-on research needs to be conducted to validate the efficacy of the model, this thesis offers a proof of concept and takes a step toward DoD portfolio optimization.

References

Ambrose, M. (2017). *Cost estimating methods*. Defense Acquisition University. https://media.dau.mil/media/Cost+Estimating+Methods/0_943h37ey

Davidson, P. (2015). Naval electro optics. *Naval Forces*, *36*, 15–21.

Defense Acquisition University. (n.d.-a). *Joint Capabilities Integration and Development System (JCIDS)*. Retrieved March 24, 2021, from https://www.dau.mil/acquipedia/pages/articledetails.aspx#!371

Defense Acquisition University. (n.d.-b). *Life cycle cost*. Retrieved March 24, 2021, from https://www.dau.mil/acquipedia/pages/articledetails.aspx#!419

Defense Acquisition University. (n.d.-c). *Low Rate Initial Production (LRIP) of production and deployment phase*. Retrieved March 24, 2021, from https://www.dau.mil/acquipedia/pages/articledetails.aspx#!473

DoD. (2003, May 12). *The defense acquisition system* (DODI 5000.01). http://www.acqnotes.com/Attachments/DoD%20Directive%205000.01.pdf

DoD. (2017, August 10). *Operation of the defense acquisition system* (DODI 5000.02; incorporating Change 3). https://acqnotes.com/wp-content/uploads/2014/09/DoD-Instruction-5000.02-The-Defense-Acquisition-System-10-Aug-17-Change-3.pdf

DoN. (2014). *Department of the Navy total ownership cost (TOC) guidebook.* http://www.acqnotes.com/Attachments/Navy%20Total%20Ownership%20Cost%20Guidebook%203%20June%202014.pdf

Driggers, R., & Nichols, J. (2012). *Introduction to infrared and electro-optical systems* (2nd ed.). Artech House.

Huynh, T., & Snider, K. F. (2008). Systems engineering. In R. G. Rendon and K. F. Snider (eds.), *Management of defense acquisition projects* (pp. 45–62). American Institute of Aeronautics and Astronautics Inc.

Kobren, B. (2014, March). *Life cycle costs (LCC) and total ownership costs (TOC): A study in contrasts*. Defense Acquisition University. https://www.dau.edu/training/career-development/logistics/blog/Life-Cycle-Costs-(LCC)-and-Total-Ownership-Costs-(TOC)---A-Study-in-Contrasts

Koretsky, G, Nicoll, J., & Taylor, M. (2013, January). *A tutorial on electro-optical/infrared (EO/IR) theory and systems*. Institute for Defense Analysis.

Mislick, G., & Nussbaum, D. (2015). *Cost estimation methods and tools*. John Wiley & Sons.

Mun, J. (2015). *Readings in certified quantitative risk management (CQRM): Applying Monte Carlo risk simulation, strategic real options, stochastic forecasting, portfolio optimization, data analytics, business intelligence, and decision modeling* (3rd ed.). Thomson-Shore, ROV Press, and IIPER Press.

Nitschke, S. (2007). New generation naval electro optics. *Military Technology*, 87–95.

Office of the Secretary of Defense for Cost Assessment and Program Evaluation. (2014, March). *Operating and support cost-estimating guide*. DoD. https://www.cape.osd.mil/files/OS_Guide_v9_March_2014.pdf

Rendon, R., & Snider, K. F. (2008). Project management concepts. In *Management of defense acquisition projects* (pp. 1–16). American Institute of Aeronautics and Astronautics Inc.

A New Learning Curve for Department of Defense Acquisition Programs: How to Account for the "Flattening Effect"

John J. Elshaw—is an Assistant Professor of systems engineering in the Department of Systems Engineering and Management at the Air Force Institute of Technology. He holds a BS in accounting from the University of Akron, an MBA from Regis University, and a PhD in management with a specialization in Organizational Behavior and Human Resource Management from Purdue University. He is a graduate of the United States Air Force Squadron Officer School and Air Command and Staff College. His areas of interest include technology and the impact on individual and group behavior, learning curve analysis and the relationship to manufacturing and production, organizational causes of high consequence errors, organizational climate and culture, research methods with a focus on hierarchical linear modeling, structural equation modeling, and multivariate analysis. He is a member of several professional associations such as the Academy of Management (AoM) and American Psychological Association (APA), and he has won several awards for his teaching, leadership, and writing. Elshaw has recently advised several students in research specifically dealing with learning curve analysis as it relates to cost estimation. Elshaw has also authored several peer-reviewed journal publications in the area of cost estimation and learning curves, and most recently co-authored a paper which won the 2016 IIE Industrial and Systems Engineering Research Conference Best Track Paper—Safety, Human Factors and Ergonomics.

Lt Col Clay M. Koschnick—is an Assistant Professor of systems engineering in the Department of Systems Engineering and Management at the Air Force Institute of Technology. He holds a BS in operations research from the United States Air Force Academy, an MS in operations research from The Georgia Institute of Technology, and a PhD in systems engineering with a specialization in operations research and econometrics from the University of Florida. He is a graduate of the United States Air Force Squadron Officer School, Air Command and Staff College, and Air War College. His areas of interest include financial engineering, engineering economics, decision analysis, equipment replacement analysis in the presence of emerging technologies, and research methods with a focus on econometric analysis. Koschnick has advised several students in research specifically dealing with using statistical models to estimate acquisition program costs. One of these students has papers published in the *Defense Acquisition Research Journal* and the *Defense Analytics and Logistics Journal,* and another presented their research at the 2017 Military Operations Research Symposium (MORS).

Capt Dakotah W. Hogan—is a Cost Analyst at the U.S. Air Force Cost Analysis Agency. His professional experience to date includes aircraft production and sustainment cost estimating in the KC-46 program office. Hogan holds an MS in Cost Analysis from the Air Force Institute of Technology and a BS in economics from the U.S. Air Force Academy.

Dr. Jonathan D. Ritschel—is an Assistant Professor of cost analysis in the Department of Systems Engineering and Management at the Air Force Institute of Technology (AFIT). He holds a BBA in Accountancy from the University of Notre Dame, an MS from AFIT in cost analysis, and a PhD in economics from George Mason University. He is a graduate of the United States Air Force Squadron Officer School, Air Command and Staff College, and Air War College. His research interests include cost analysis, factor analysis, defense acquisitions, acquisition reforms, defense economics, public choice, and economic institutional analysis. Ritschel has recently advised over 10 students in research specifically related to the cost analysis field. In addition, Ritschel has authored over 20 peer-reviewed publications related to cost analysis and defense cost economics with recent presentations at the International Cost Estimating and Analysis Association conference.

Dr. Adedeji B. Badiru—is Dean and Senior Academic Officer for the Graduate School of Engineering and Management at the Air Force Institute of Technology (AFIT). He was previously Professor and Head of Systems Engineering and Management at the AFIT, Professor and Department Head of Industrial & Information Engineering at the University of Tennessee in Knoxville, and Professor of Industrial Engineering and Dean of University College at the University of

Oklahoma, Norman. He is a registered professional engineer (PE), a certified Project Management Professional (PMP), a Fellow of the Institute of Industrial Engineers, and a Fellow of the Nigerian Academy of Engineering. He holds a BS in industrial engineering, an MS in mathematics, and an MS in industrial engineering from Tennessee Technological University. He also holds a PhD in industrial engineering from the University of Central Florida. His areas of interest include mathematical modeling, project modeling and analysis, economic analysis, systems engineering, and efficiency/productivity analysis and improvement. He is the author of over 25 books, 34 book chapters, 70 technical journal articles, and 110 conference proceedings and presentations. He also has published 25 magazine articles and 20 editorials and periodicals.

Abstract

Traditional learning curve theory assumes a constant learning rate regardless of the number of units produced; however, a collection of theoretical and empirical evidence indicates that learning rates decrease as more units are produced in some cases. These diminishing learning rates cause traditional learning curves to underestimate required resources, potentially resulting in cost overruns. A diminishing learning rate model, Boone's Learning Curve (2018), was recently developed to model this phenomenon. This research confirmed that Boone's Learning Curve is more accurate in modeling observed learning curves using production data of 169 Department of Defense (DoD) end-items. However, further empirical analysis revealed deficiencies in the theoretical justifications of why and under what conditions Boone's Learning Curve more accurately models observations. This research also discovered that diminishing learning rates are present but not pervasive in the sampled observations. Additionally, this research explored the theoretical and empirical evidence that may cause learning curves to exhibit diminishing learning rates and be more accurately modeled by Boone's Learning Curve. Only a limited number of theory-based variables were useful in explaining these phenomena. This research further justifies the necessity of a diminishing learning rate model and proposes a framework to investigate learning curves that exhibit diminishing learning rates.

Introduction

The Budget Control Act of 2011 subjected the Department of Defense (DoD) to a more fiscally constrained and financially conscious environment than ever before, contrasted with a demand for new acquisition programs of almost every type. As an increasing number of programs are terminated, managers at every level in the DoD are expected to ensure the DoD's shrinking budget is being used in the most cost-effective way. The increased scrutiny adds greater emphasis on the accuracy of program office cost estimates given that an approved program cost estimate supports every major acquisition program funded by the DoD.

In order to obtain reliable cost estimates, cost estimating models and tools within the DoD must be evaluated for their relevance and accuracy. The current learning curve methods used within the DoD's cost estimating procedures are from the 1930s (Wright, 1936). As automation and robotics increasingly replace human touch-labor in the production process, the current 80-year-old model may no longer be appropriate for accurate learning curve (and cost) estimates. Robotics and automation machines do not learn; however, they are inevitably a part of future production. New learning curve methods that incorporate automated production should be examined as a possible tool for cost estimators in the acquisition process. Additionally, we examine the flattening effect that occurs toward the end of the acquisition process and the impact this has on different learning curve formulas. Originally published by Badiru (2012), the half-life of a learning curve is the incremental production level required to reduce cumulative average cost per unit to half its initial level. Conversely, a half-life forgetting curve is the amount of time it takes for performance to decline to half of its initial level. A model that more accurately reflects true cost is critical for planning, especially over the long-term when there are hundreds of millions of dollars involved. The results of this research will inform the acquisition community as well as provide a tool that has the potential to significantly outperform the cost estimation and learning curve models currently in use.

The purpose of this research ultimately is to investigate new learning curve methods, develop learning curve theory within the DoD, and pursue a more accurate cost estimation model.

Research Approach (Methodology)

The premise of this research is to apply a new learning curve in such a way that the estimated learning rate is modeled as a decreasing function over time as opposed to the constant learning rate that is currently in use with Wright's Learning Curve. The current model in use today mathematically states that for every doubling of units there will be a constant gain in efficiency. For example, if the manufacturer observed a 10% reduction in man-hours in the time to produce unit 10 from the time to produce unit 5, then they should expect to see the same 10% reduction in man-hours in the time to produce unit 10 to the time to produce unit 20. Unfortunately, in real world situations this constant rate of decay is rarely the case. We propose that more accurate cost estimates could be made if a decay factor was taken into consideration. The proposed modification may take this form:

$$\mathrm{Cos}\,t(X) = aX^{f(x)}$$

Where:

Cost(X) = the cumulative average time (or cost) per unit

X = the cumulative number of units produced

A = time (or cost) required to produce the first unit

F(x) = the learning curve slope represented as a function of units produced

The specific function used for the slope is what this research will attempt to understand. Figure 1 shows the phenomena this research will attempt to model, where the black (flatter) line depicts the traditional curve used to model learning, the red (steeper) line represents the hypothesized learning structure, and the blue line represents actual data.

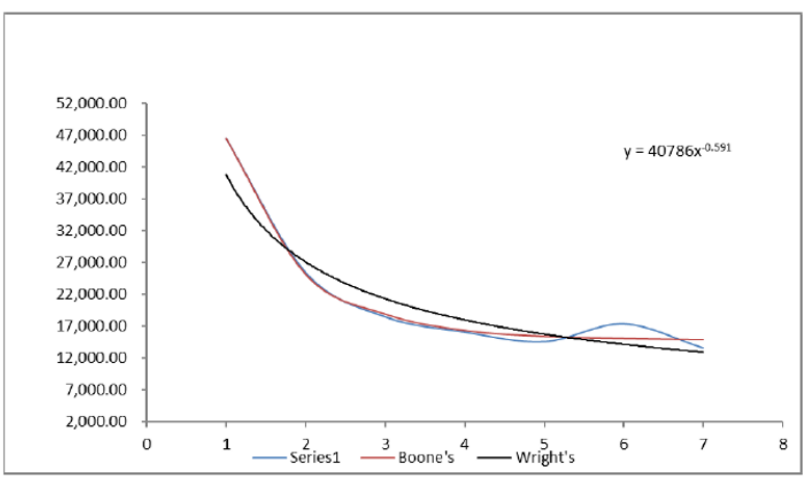


Figure 32. Learning Curve Comparison

Learning Curve Background

The concepts of learning and learning curves are intuitive: as a worker repetitively performs tasks to assemble a product, the worker will gain efficiencies. These efficiencies should decrease the time the worker spends on each unit as more units are produced. These efficiencies translate to a continuous reduction in labor hours and cost savings over time. The GAO Cost Estimating and Assessment Guide (2009) lists four reasons for these gains in efficiency that broaden the scope of learning. First, as more units are produced, workers tend to become “more physically and mentally adept” at performing tasks, and supervisors become more efficient at utilizing workers. This aspect of learning is considered to be learning-by-doing at the laborer level and is termed *autonomous learning* (Levy, 1965). Second, the work environment is improved to include the “climate, lighting, and general working conditions” (Leonard, 2009). Third, the production process is changed to “optimize the placement of tools and material and simplify tasks.” Lastly, market forces in the competitive business environment will require suppliers to improve efficiency to survive. These last three aspects of learning are considered to be organizational learning by continuous improvement efforts termed *induced learning* (Levy, 1965; Dutton & Thomas, 1984).

Several terms are used to describe learning curves that include cost improvement curve, cost/quantity relationship, manufacturing process function, experience curve, and product improvement function (International Cost Estimating and Analysis Association [ICEAA], 2014, p. 7). The original and most generalized term *learning curve* will be used in this research. Although learning curves are used most popularly in aircraft manufacturing, the concept can be applied to “relatively large and complex products that require various types of fabrication and assembly skills” (Asher, 1956, p. 5). The Air Force Cost Analysis Handbook includes several specific situations in which learning curves apply that include “the manufacture of a complex end-item, limited changes to product characteristics or technology, continuous manufacturing process, constant management pressure to improve, and consistent production rates” (Department of the Air Force, 2007, pp. 8-1–8-2). The Handbook also includes other criteria, including “a high proportion of manual labor, labor efficiency/job familiarization, standardization, specialization, and methods improvements, improved materiel flow and reduced scrap, improved production procedures, tools, and equipment, improved workflow and engineering support, and product redesign improvements” (Department of the Air Force, 2007, pp. 8-1–8-2). These situations describe aspects of the manufacturing process that enable organizational learning and allow for labor efficiencies, although learning can occur without all criteria being present.

Cumulative Average Learning Curve Theory

The concept of a learning curve was first formally recorded by Theodore Paul Wright in 1936 in his work “Factors Affecting the Cost of Airplanes.” Wright identified the learning curve concept in a pre-World War II production environment of a small two-seater aircraft (ICEAA, 2014, p. 16). He observed that as a worker repeatedly performs the same task, the time required to complete that task will decrease at a constant rate. More specifically, Wright formulized that as the number of aircraft produced doubles, the cumulative average labor cost would decrease at a constant rate (Wright, 1936). This relationship is described in Equation 1 and is the Cumulative Average Theory widely in use. When learning curves utilize this Cumulative Average Theory, they are frequently called “Wright Curves” (ICEAA, 2014, p. 16).

Equation 1

$$\bar{Y} = Ax^b \quad (1)$$

where Y¯is the cumulative average cost of the first x units, A is the theoretical cost to produce the first unit, x is the cumulative number of units produced, and b is the learning curve slope (LCS) divided by the natural logarithm of 2.

The learning curve slope in the learning curve exponent "b" of Equation 1 defines how each doubling of produced units reduces cumulative average costs. For example, Wright used his empirical data to calculate a learning curve slope of 80%. Therefore, as the cumulative number of units doubles, the cumulative labor cost of the doubled units would be 80% of the original undoubled amount, resulting in a 20% cumulative average reduction in labor cost (Wright, 1936). This 20% cumulative average reduction can also be called the rate of learning. Higher rates of learning lead to greater reductions in labor costs. Although Wright's model cited 80% as a universal learning curve slope, learning curve theory evolved to realize that other slopes are possible based on an end-item's unique manufacturing characteristics (Jaber, 2006).

Wright's Cumulative Average Learning Curve is cumbersome to use because cumulative average costs are calculated in place of the unit cost. Figure 2 depicts Wright's Learning Curve with an 80% learning curve slope based on a first unit cost of 100. Figure 2 shows that as the cumulative number of units produced doubles, the cumulative average cost decreases by 20%. The corresponding unit cost is also displayed to highlight how the cumulative average cost differs from the unit cost. The unit cost is calculated by summing the cumulative costs up to but not including the unit number and subtracting it from the total cost. In Figure 2, the first unit cost for both cumulative average cost and unit cost is 100. For unit two, the cumulative average cost of the first two units is 80; this is a 20% reduction due to an 80% learning curve slope. The total cost of the first two units is 160. Because the first unit cost is known to be 100, the second unit cost can be calculated from the difference to be 60. These same calculations can be used to obtain the unit costs for the remaining units.

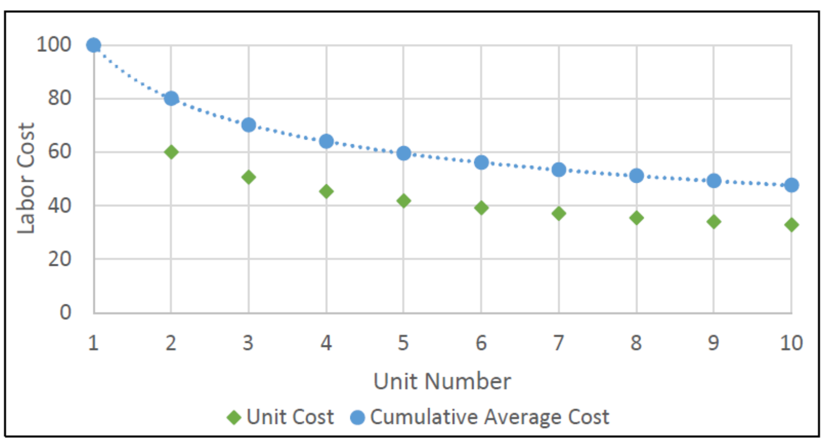


Figure 33. Wright's Cumulative Average Theory at an 80% Learning Curve Slope

Wright illustrated Equation 1 using a graph with vertical and horizontal axes displayed in logarithmic rather than linear scale. Wright illustrated his equation on this logarithmic graph in order to highlight the straight line representing a constant rate of learning (Wright, 1936). The same function can be graphed in linear scale by transforming

Equation 1 into a log-linear form by taking the natural logarithm of both sides. This log-linear transformed equation is shown in Equation 2. The parameter definitions for Equations 1 and 2 are the same.

Equation 2

$$ln\,\bar{Y} = lnA + b\,ln\,x \qquad (2)$$

Equation 2 also allows analysts to apply linear regression analysis in order to estimate the parameters A and b from a set of cumulative average cost data (Mislick & Nussbaum, 2015, p. 185). Figure 3 displays Wright's Cumulative Average Theory at an 80% learning curve slope transformed into a log-linear form. The parameters are identical to the parameters of Figure 2 with a first unit cost of 100. The constant learning curve slope indicated by the linear slope in Figure 3 is a crucial concept of this traditional learning curve theory.

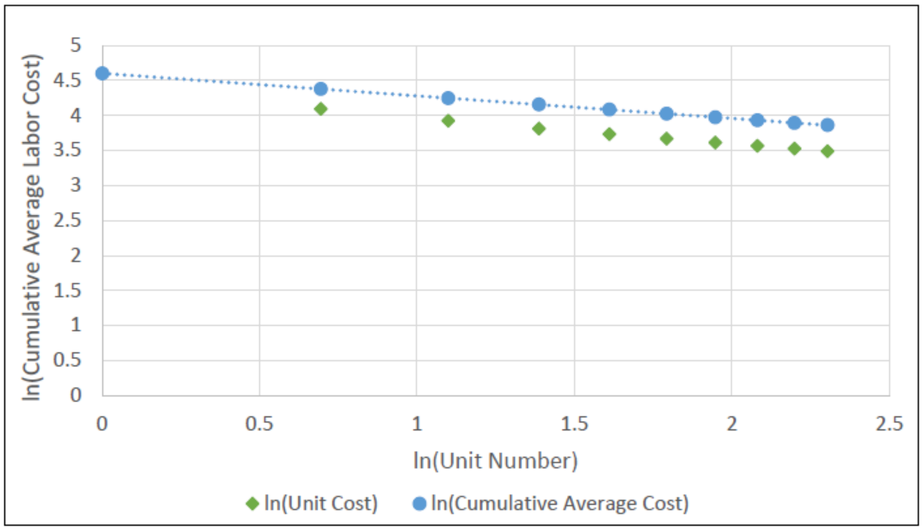


Figure 34. Wright's Cumulative Average Theory at an 80% Learning Curve Slope in Log-Linear Form

Unit Learning Curve Theory

Several years following Wright's Cumulative Average learning curve theory, J. R. Crawford formulated the Unit Learning Curve Theory, formally written in 1944. Together, these theories form the basis of the traditional learning curve theory. Crawford proposed his Unit Theory first in an undated manual prepared for Lockheed Aircraft Company personnel after realizing the difficulty of calculating unit costs from Cumulative Average Learning Curve Theory equations (Asher, 1956, pp. 21–22). As shown in Equation 3, Crawford's Learning Curve yields an estimated unit cost given the unit's sequential unit number within the production line, a learning curve slope, and a theoretical first unit cost. Crawford's Unit Theory is the same as Wright's aside from these differences in variable interpretation. Learning curves are often called "Crawford Curves" when they utilize Crawford's Unit Theory (ICEAA, 2014, p. 31).

Equation 3

$$Y = Ax^{b} \qquad (3)$$

where Y is the individual cost of unit x, A is the theoretical cost of the first unit, x is the unit number of the unit cost being forecasted, and b is the LCS divided by the natural logarithm of 2.

Using Crawford's Unit Theory with a learning curve slope of 80% and a first unit cost of 100 labor hours, the cost of the second unit is 80 labor hours. This 20% reduction in labor hours or rate of learning is due to the 80% learning curve slope. Figure 4 illustrates Crawford's Unit Theory using the same parameters from Wright's Cumulative Average Theory shown in Figures 2 and 3. The cumulative average costs are not shown in Figure 4 because these are not germane to Unit Theory.

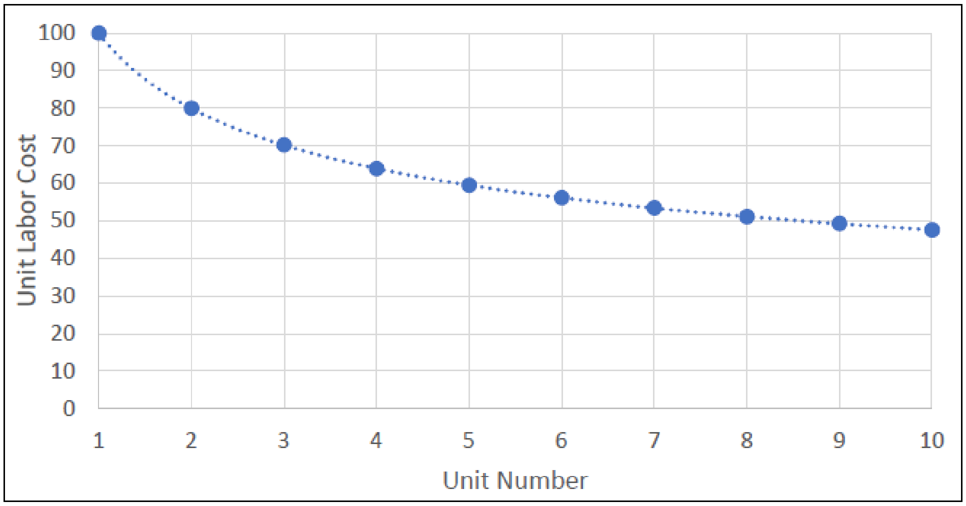


Figure 35. Crawford's Unit Theory at an 80% Learning Curve Slope

Crawford's Unit Theory Learning Curve is transformed into a log-linear form using the same methodology used to derive Equation 2. Crawford's Learning Curve in log-linear form is shown in Equation 4 using the same parameter definitions from Equation 3.

Equation 4

$$ln\, Y = lnA + b\, ln\, x \quad (4)$$

Crawford's Unit Theory Learning Curve is shown in logarithmic scale in Figure 5. Similar to Cumulative Average Theory, the constant learning indicated by the linear slope in Figure 5 is a vital concept of this theory: the rate of learning remains constant as the units double regardless of the number of units produced.

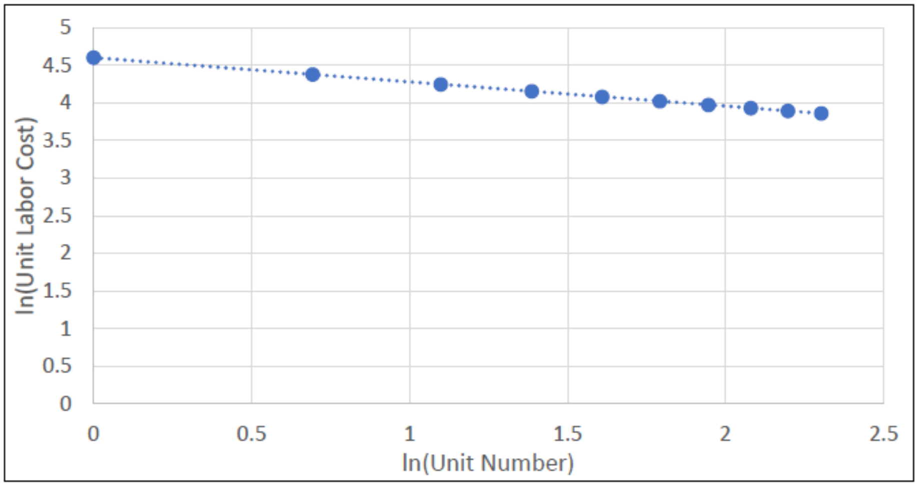


Figure 36. Crawford's Unit Theory at an 80% Learning Curve Slope in Log-Linear Form

Crawford's Unit Theory parameters are straightforward to estimate when data are available by each unit (also called unitary data); however, manufacturers generally report cost data in the form of production lots that include the total lot cost and the number of units in that lot (Mislick & Nussbaum, 2015, p. 191). Unlike Wright's Cumulative Average Theory, Crawford's Unit Theory must utilize lot midpoints to estimate parameters when unitary cost data is unavailable. The algebraic lot midpoint is defined as "the theoretical unit whose cost is equal to the average unit cost for that lot on the learning curve" (Mislick & Nussbaum, 2015, p. 192). In other words, the lot midpoint is the unit that will divide the area under the learning curve evenly within the lot (Mislick & Nussbaum, 2015, p. 192). This lot midpoint is used in the Unit Theory learning curve formula as the sequential unit number or independent variable "x" in Equations 3 and 4. The lot midpoint supplants using sequential unit numbers because sequential unit numbers are unavailable when using lot cost data. When the lot midpoint is the independent variable in Equation 3, the dependent variable will yield the average lot cost. The average lot cost results because this x-coordinate is the most representative point for the lot (ICEAA, 2014, p. 40).

Lot midpoints are calculated in a two-step approach due to the lack of a closed-form solution. A closed-form solution does not exist because the lot cost is a function of the learning curve exponent *b* from Equations 3 and 4 used to estimate the lot midpoint. However, the lot midpoint is also used to estimate the learning curve exponent "b." The first step in calculating a lot midpoint utilizes a parameter-free approximation formula to estimate the lot midpoint. These lot midpoint estimates are then used to estimate the learning curve exponent "b." The second step is to use a lot midpoint formula that includes an estimate of the learning curve exponent "b" and iterate until successive values of the estimated lot midpoints and "b" are sufficiently small (Mislick & Nussbaum, 2015, pp. 200–201). The parameter-free lot midpoint approximation is shown in Equations 5 (Mislick & Nussbaum, 2015, p. 193).

Equation 5

$$\text{Lot Midpoint (LMP)} = \frac{F+L+2\sqrt{FL}}{4} \qquad (5)$$

where F is the first unit number in a lot and L is the last unit number in a lot. These lot midpoint estimates are then used to estimate the learning curve parameters for Crawford's model (Equation 3) using the GRG non-linear optimization algorithm.

Several parameter lot midpoint approximations exist, but a simple and popular lot midpoint approximation is Asher's Approximation shown in Equation 6. The same parameter definitions presented in Equation 5 also apply to Equation 6, and the learning curve exponent "b" is the same as shown previously in Equations 1–4 (Mislick & Nussbaum, 2015, p. 201).

Equation 6

$$\text{Lot Midpoint} \approx \left[\frac{\left(L+\frac{1}{2}\right)^{b+1}-\left(F-\frac{1}{2}\right)^{b+1}}{(L-F+1)(b+1)}\right]^{\left(\frac{1}{b}\right)} \quad (6)$$

where F is the first unit number in a lot, L is the last unit number in a lot, and b is the estimated value from Equation 1.

Comparison of Cumulative Average and Unit Theory

Cumulative Average Theory and Unit Theory will produce different predicted costs provided the same set of data despite all predicted costs being normalized to unit costs. Figure 6 demonstrates this point where Unit Theory was used to generate data using a first unit cost of 100 and a learning curve slope of 90%. The original Unit Theory data was converted to cumulative averages in order to estimate Cumulative Average Theory Learning Curve parameters. Cumulative Average Theory estimated a learning curve slope of 93% and a first unit cost of 101.24. These Cumulative Average Theory parameters were then used to predict cumulative average costs. These predicted costs were then converted to unit costs. This conversion allows for the Cumulative Average predictions to be directly compared to the original Unit Theory generated data.

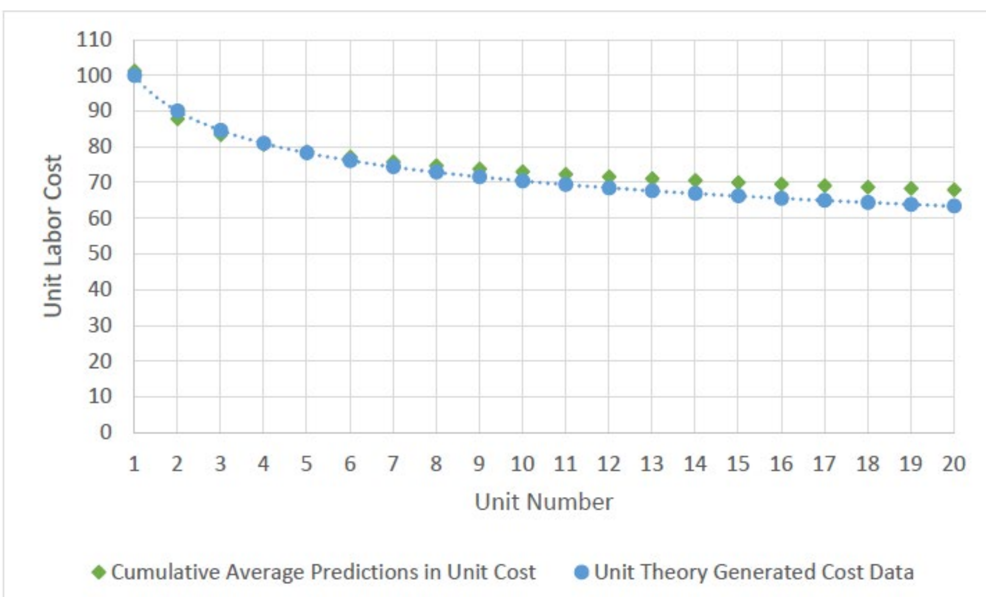


Figure 37. Cumulative Average and Unit Learning Curve Theory Comparison

As shown in Figure 6, the Cumulative Average Learning Curve predictions first overestimate, then underestimate, and ultimately overestimate the generated Unit Theory data for all remaining units. A similar case would occur if Cumulative Average Theory were used to generate data and Unit Theory learning curve parameters were estimated from this data. Figure 6 highlights that these two theories are inherently different due to differences that occur when estimating learning curves parameters using unit costs or cumulative average costs. Several factors can assist an analyst in deciding which theory to apply; however, solely relying on goodness-of-fit statistics will likely bias the decision toward Cumulative Average Theory (Mislick & Nussbaum, 2015, p. 215; Cullis et al., 2008). Frequently goodness-of-fit statistics to include the coefficient of determination (R2) and standard error are used to determine which model best explains variation in a dataset. The coefficient of determination is the total variation of the dependent variable explained by the independent variable (Hilmer, 2014, p. 90). The standard error is the measure of how far on average the data tend to fall from the predicted learning curve (Hilmer, 2014, p. 91). These goodness-of-fit statistics can be used when comparing models of the same units, which is not the case when comparing Cumulative Average and Unit Theory learning curves.

Researchers investigated this Cumulative Average Theory goodness-of-fit statistic bias and presented at a Society of Cost Estimating and Analysis (SCEA) conference (Cullis et al., 2008). The researchers used a methodology similar to that used to produce the example illustrated in Figure 6. Cumulative Average data were first generated, and a Unit Theory learning curve was fit to this data. Unit Theory data were then generated, and a Cumulative Average learning curve was fit to this data. The goodness-of-fit statistics reliably indicated the correct learning curve theory to model these perfect data. Next, artificial variation was injected into the generated data, and the researchers repeated the process. When the researchers injected variation or error in the data, the goodness-of-fit statistics overwhelmingly favored selecting Cumulative Average Theory over Unit Theory even with small amounts of variation in the data (Cullis et al., 2008). In other words, Cumulative Average Theory Learning Curves will tend to have a higher coefficient of determination and lower standard error when compared to Unit Theory learning curves. This bias in the goodness-of-fit statistics is because Cumulative Average Theory is a cumulative running average, so the curves are generally smoother and closer to the data points (Mislick & Nussbaum, 2015, p. 215). Therefore, bias exists in favor of Cumulative Average Theory, so more subjective judgments are warranted to determine which learning curve theory to utilize.

The GAO Cost Estimating and Assessment Guide (2009) provides factors to consider when choosing which learning curve theory to model data. Analysts should review which theories were applied to analogous systems that are similar in form, fit, or function to the current system being considered (GAO, 2009, p. 369). Next, some industries have standards that prefer one theory over the other (GAO, 2009, p. 369). Experience should also be considered by reviewing which theory has been applied to the contractor in the past (GAO, 2009, p. 369). Lastly, some aspects of the production environment can indicate which theory is best to apply (GAO, 2009, p. 369). For example, Cumulative Average theory is best when “the contractor is starting production with prototype tooling, has an inadequate supplier base, expects early design changes, is subject to short lead times,” or where there is a risk of concurrency between development and production phases (GAO, 2009, p. 369). In contrast, Unit Theory is more suited for contractors that are well-prepared to begin production (GAO, 2009, p. 369).

Other factors must be considered when deciding which learning curve theory to use. Cumulative Average learning curve theory will provide more conservative estimates and is less responsive to trends than the Unit Learning Curve theory (Mislick & Nussbaum, 2015, p. 215). For these reasons, Unit Learning Curve theory is frequently favored by government negotiators when negotiating contracts (Mislick & Nussbaum, 2015, p. 215). Cumulative

Average Learning Curve theory also relies on continuous data and is unable to be calculated with missing prior data using traditional estimation techniques.

Finally, when ordinary least squares (OLS) regression is used to estimate Cumulative Average learning curve parameters, the cumulative averaging technique violates the OLS regression assumption of independence. For OLS regression to provide an unbiased estimator, the data must be obtained through independent random sampling (Hilmer, 2014, pp. 111–112). In other words, the unit labor cost and its associated unit number cannot be statistically related to other observations of unit labor cost and their associated unit numbers. This assumption is violated due to the costs of earlier observations being a function of the costs of later observations from cumulative averaging calculations. This violation biases the learning curve parameters to produce expected values that are not equal to the population parameter being estimated (Hilmer, 2014, p. 109). Despite this violation, Cumulative Average Learning Curves estimated using OLS regression are widely used and remain a valid method for estimating learning curves.

Cost Accounting for Learning Curves

The fundamental aspects of traditional learning curve theory apply only to a subset of total program costs. Hence appropriate costs must be considered when applying the theory to yield viable parameter estimates and predictions. In a complex program, costs can be presented in units of hours or dollars and organized as recurring and non-recurring for various cost elements of the end-item or the program as a whole. For each cost element, labor costs are also categorized into further groups. The analyst must select the applicable subset of costs and consider their units when utilizing learning curve theory.

Costs are generally categorized as recurring and non-recurring costs. Non-recurring costs are one-time costs that are not directly attributable to the number of end-items being produced (Mislick & Nussbaum, 2015, p. 26). Recurring costs are costs that are incurred repeatedly for each unit produced (Mislick & Nussbaum, 2015, p. 26). At the basis of learning curve theory is the idea that costs vary as more units of an end-item are produced. Therefore, non-recurring costs are excluded from learning curve analysis due to the inability to relate these costs with the number of units produced (Mislick & Nussbaum, 2015, p. 180). Traditional research also has limited insight into how learning applies to nonrecurring costs. For these reasons, learning curve analysis focuses solely on recurring costs in estimating learning curve parameters and predicting recurring costs (Mislick & Nussbaum, 2015, p. 180).

Manufacturing program costs are also organized broadly as labor, material, and overhead. T. P. Wright initially claimed that all these categories vary with the number of units produced, although he specifically focused on labor hour costs when forming his seminal theory (Wright, 1936). Due to this focus and the intuitive idea of learning at the laborer level, researchers have since focused solely on labor costs, including J. R. Crawford. Crawford exclusively studied labor learning and elaborated at length on how learning occurs from the laborer’s perspective (Asher, 1956, p. 24). Both fundamental theorists also focused on the laborers who manufactured the aircraft by considering touch labor (Asher, 1956, pp. 16, 21). Additionally, cost estimating standard practice and guidance concerning learning curves also provides the basis for considering touch labor costs only (ICEAA, 2014, p. 7; Department of the Air Force, 2007, p. 8-1).

Defense contractors are often contractually required to submit costs incurred when producing large, complex end-items for the U.S. government. These costs have historically been submitted on the Defense Department (DD) Form 1921 report series to include the Functional Cost-Hours Report DD Form 1921-1 and Progress Curve Report DD Form 1921-

2. The DoD transitioned to using the Cost and Hour Report "FlexFile" and Quantity Data reports on May 15, 2019 (Burke, 2019). However, historical program data will likely remain in the legacy 1921 series forms, and "FlexFile" and Quantity Data reports can easily be manipulated to legacy 1921 forms. The 1921-1 form is organized by work breakdown structure (WBS) elements that include various functional cost categories both in units of hours and dollars. Three broad functional cost categories: labor, material, and other costs are included in both forms of recurring and non-recurring costs. This form also has four functional labor categories that include manufacturing, tooling, engineering, and quality control labor. These four labor category costs, when summed with the material costs and other costs, comprise the total cost for each WBS element for recurring and non-recurring costs. This total cost is provided in units of dollars due to the underlying units of material and other costs. A document accompanies the 1921-1 to describe the elements of the form called a 1921-1 Data Item Description (DID). The 1921-1 DID defines these various functional cost categories to include the four labor categories whose definitions are useful in determining which categories pertain to learning curve analysis.

The definition for the manufacturing labor cost category most clearly aligns with the extant literature to be the focus as the pertinent labor cost category for learning curve research. According to the 1921-1 DID, the manufacturing labor category "includes the effort and costs expended in the fabrication, assembly, integration, and functional testing of a product or end-item. It involves all the processes necessary to convert raw materials into finished items" (1921-1 Data Item Description, p. 12). This manufacturing labor category aligns with the categories examined by Wright (1936), which he called "assembly operations" (p. 124), along with those cost categories Crawford studied, which he called "airframe-manufacturing processes" (Asher, 1956, p. 21). A RAND learning curve study also defined the direct labor used in the study as "those expended to manufacture the airframe and install the equipment required to transform the airframe into a complete, flyable airplane" (Asher, 1956, p. 49). Therefore, the manufacturing labor cost category as defined by the 1921-1 DID is associated with the types of labor costs studied by traditional learning curve theorists and succeeding research. However, data availability can prompt analysts and researchers to use total costs instead. Although these curves remain valid according to Wright (1936), they are composite learning curves with caveats to be discussed later.

The 1921-1 is organized into WBS elements defined for each program. A WBS element is a method to display, define, and organize the overall end-item into sub-products while maintaining their relationship with the end-item and other sub-products (DoD, 2018, p. 4). For example, WBS elements for an aircraft program could include the airframe, wings, and engines, among other elements. These WBS elements are comprised of lower-level elements as well. For example, the airframe element may include elements such as the forward, middle, and aft airframe. WBS elements can also comprise activities instead of physical components such as testing the aircraft. Although some of these activities may experience efficiencies over time, traditional learning curve theory focuses exclusively on the production of physical components. WBS elements are frequently organized into various cost categories that can comprise elements suitable and unsuitable for learning curve analysis.

WBS elements are organized into various cost categories to include procurement costs, weapons system costs, and flyaway/rollaway/sail-away costs among others. Not all WBS elements and their respective cost categories are pertinent for learning curve analysis. The group of WBS elements in which learning is relevant is prime mission equipment and its sub-elements. Prime mission equipment is all hardware and software WBS elements installed on the weapon system such as "propulsion equipment, electronics, armament, etc."

(Flyaway Costs, n.d.). The prime mission equipment WBS aligns with those elements that experience learning according to the traditional learning curve theorists. The prime mission equipment WBS group excludes elements such as systems engineering and program management (SE/PM) and system test and evaluation (STE); these costs are instead included in flyaway costs (Flyaway Costs, n.d.). These latter elements are activities tangentially related costs that may not experience learning as theorized by the traditional learning curve literature.

Recent learning curves research has considered equivalent WBS elements. Moore et al. (2015) scoped their research to consider airframe costs, which is a sub-element of prime mission equipment due to the homogeneity of the programs they analyzed. Honious et al. (2016) also used airframe costs. Boone used prime mission equipment WBS elements to perform analysis due to the wide variety of programs researched (Boone, 2018, pp. 22–23).

Another cost accounting item to consider is whether to use hours or dollars as the units for labor cost. The total cost for each WBS element is provided in dollars due to material and other costs not having associated labor hours. Therefore, if the total WBS cost is used for analysis, units of dollars will be used. In contrast, the four labor categories to include manufacturing labor has both dollars and hours associated with each. Ideally, labor hours would be analyzed for a variety of reasons as discussed by the Air Force Cost Analysis Handbook (2007). First, labor dollars must be normalized to remove the effects of escalation (Department of the Air Force, 2007, p. 8-65). Escalation effects comprise economy-wide price changes as well as industry-specific price changes. Normalization removes these price variations that allow for labor costs to be compared across different fiscal years. The escalation indices used to normalize data are estimates of the escalation that the industry experienced that year. Therefore, the use of escalation indices can inject error into the data. In contrast, if labor hour data were used, labor costs between years could easily be compared without normalization. Furthermore, changes in labor rates can also bias the labor cost learning curve. Senior personnel are brought on to a program initially due to the initial complexity (Department of the Air Force, 2007, p. 8-65). Once the program stabilizes and production increases, the program usually transitions to more junior labor (Department of the Air Force, 2007, p. 8-65). This labor rate effect, when combined with the effect of normal learning, artificially steepens the learning curve (Department of the Air Force, 2007, p. 8-65). Therefore, the slopes of learning curves utilizing labor dollars will likely be steeper due to the influence of declining average labor rates as the workforce builds towards full-rate production (Department of the Air Force, 2007, p. 8-65). Although using labor cost data in dollars does not invalidate analysis and is frequently utilized due to data availability, labor hour data would ideally be used for these reasons.

In summary, the literature indicates using direct, recurring, manufacturing labor costs in the form of hours. These costs should be considered only for the WBS elements that include prime mission equipment and its lower-level elements. Using these specific WBS costs in the form of hours ensures alignment with the original costs and elements considered to be affected by learning in the traditional models. Although this review exclusively examined the DD 1921-1 form, the 1921-2 form reports the same cost data, albeit in a different format for specific use in learning curve analysis. The methodology on which WBS elements and costs to consider is translatable between the two legacy forms along with the current Cost and Hour Report “FlexFile” and Quantity Data reports.

Variations to Traditional Learning Curve Theory

The traditional learning curve models assume a constant learning environment comprised of a stable production line and invariable end-item design. Due to the realities of changing production environments and modifications to the end-item configuration, many researchers have investigated non-stable learning environments. These areas include production rate changes, changes to the end-item design during production, and breaks in production, among other topics.

Production rates of end-items can vary as the program proceeds through the production life cycle. Researchers in a 1974 RAND report first formally proposed that production rate effects can alter unit costs (Large et al., 1974). The researchers hypothesized that as more units are produced, fewer costs would be allocated to each unit due to fixed costs within the manufacturing process (Large et al., 1974). When fixed costs are allocated over more units, each unit will be less expensive. The researchers also hypothesized that as more units are produced, the contractor may be able to take advantage of volume discounts, resulting in lower material costs per unit (Large et al., 1974). This modified learning curve equation, termed Unit Theory with Rate Adjustment, is shown in Equation 7.

Equation 7

$$Y = Ax^b R^c \tag{7}$$

Where:

Y = cost of the xth unit

A = theoretical cost to produce the first unit (T1)

x = sequential unit number of the unit being calculated

b = $\frac{\ln Learning\ Curve\ Slope}{ln}$

R = Annual production rate

c = $\frac{\ln Rate\ Slope}{\ln 2}$

Despite these logical hypotheses, the researchers rarely found the rate term to be statistically significant. Several factors can cause the rate term to be not statistically significant, such as how the contractor responds to the production rate changes. There are also statistical challenges when investigating rate effects due to the likely presence of multicollinearity. The independent variables “x” and “R” in Equation 7 do not make independent contributions to describe the dependent variable because there is no means to hold “x” constant while changing “R.” For these reasons, statistical analysis is unable to discern the effects that either variable has on the dependent variable. The presence of multicollinearity tends to cause one or both independent variables to be not statistically significant when using linear regression analysis (Department of the Air Force, 2007, pp. 8-31–8-32). Also, researchers have studied how production breaks alter the learning curve. Production breaks occur when the manufacturer of the end-item stops production for a period. During the time lapse between the completion of a unit and the start of another unit, a loss of learning can occur. This learning loss results in an increased cost for the first unit following a production break and all subsequent units (Honious et al., 2016).

One popular method to assess the loss of learning and subsequent unit costs is the Anderlohr Method (1969). This method identifies five factors that when weighted appropriately account for the amount of learning lost during the production break. The amount of lost learning is used to regress up the original learning curve before production resumes. Once production resumes, the manufacturing process resets to the cost of a previously produced unit and progresses back down the revised learning curve at the original learning rate for all units produced after the production break (Anderlohr, 1969).

The final area in which research has focused is changes to the end-item during production. These changes include additions, deletions, and substitutions of components of the end-item. These modifications can occur due to conventional engineering change orders to the configuration of the end-item or due to concurrency between development testing and production that reveals necessary design changes. The addition, deletion, and substitutions of components of an end-item during production can cause the following units' costs to differ significantly from what is predicted using traditional learning curve theory (Department of the Air Force, 2007, pp. 8-50–8-56). Additionally, configuration changes can also affect the rate at which the manufacturing process learns, which alters and often steepens the original learning curve slope (Honious et al., 2016). The learning curve slope is especially affected during concurrency between development testing and production due to the continual flow of design changes, which tend to flatten the learning curve slope (Department of the Air Force, 2007, p. 8-50–8-56).

This research into production rates, production breaks, and changes to end-items demonstrates the importance of a stable production environment with a constant end-item configuration. Without these tenets, traditional learning curve analysis becomes challenging due to confounding variables. The influence of confounding variables obscures how unit costs are related to the number of units produced.

Forgetting & Plateauing Phenomena

An implicit assumption in the traditional learning curve theories is that knowledge obtained through learning does not depreciate (Epple et al., 1991). However, empirical evidence demonstrates that knowledge depreciates in organizations (Argote, 1993; Argote et al., 1990). Argote et al. (1990) have shown that knowledge depreciation occurs at both the laborer level and the organizational level. Many variations of the traditional models make use of a concept of performance decay and forgetting to model non-constant rates of learning. Forgetting and its effects on lost learning can take many forms and is essential to consider in contemporary learning curve analysis. Forgetting is the concept that laborers and the organization as a whole will experience a decline in performance over time resulting in non-constant rates of learning. Badiru (2012) theorizes that forgetting and resulting performance decay is a result of factors "including lack of training, reduced retention of skills, lapse in performance, extended breaks in practice, and natural forgetting" (p. 287). According to Badiru (2012), these factors may be caused by internal processes, including training policy or external factors such as breaks in production.

Badiru (2012) lists three cases in which forgetting arises. First, forgetting may occur continuously as a worker or organization progresses down the learning curve due in part to natural forgetting (Badiru, 2012). In other words, the impact of forgetting may not wholly eclipse the impact of learning but will hamper the rate of learning while performance continues to increase at a slower rate. Second, forgetting may occur at distinct and bounded intervals such as during a scheduled production break (Badiru, 2012). Third, forgetting may intermittently occur at random times and for stochastic intervals such as during times of employee turnover (Badiru, 2012). Figure 7 illustrates this third case where intermittent periods of forgetting degrade the regular learning curve path to result in a degraded performance curve. In this illustration, the learning curve is shown as an increase in performance rather than a decrease in time or cost (Badiru, 2012). Others have expanded on the causes of forgetting and have drawn similar conclusions to Badiru (2012), such as Glock et al., 2019; Jaber, 2006; and Jaber & Bonney, 1997.

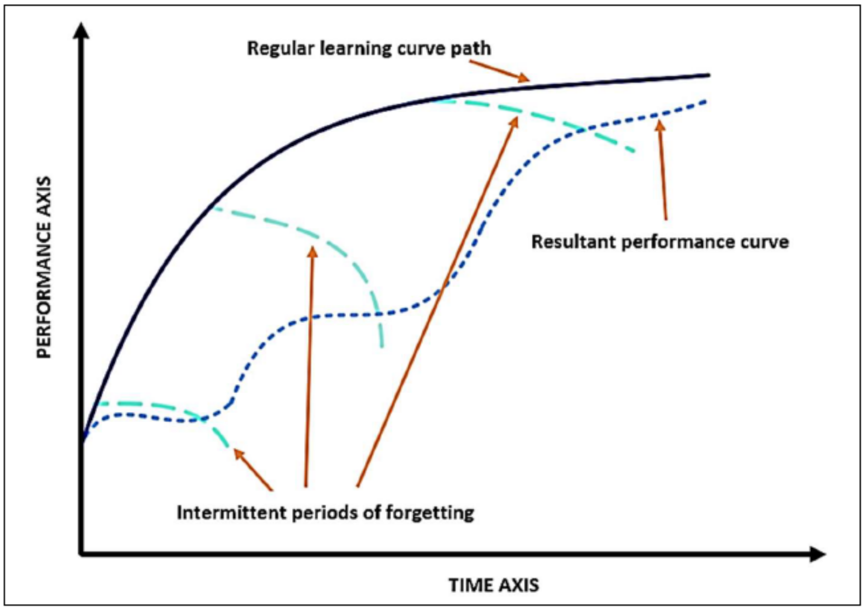


Figure 38. Effects of Forgetting on the Traditional Learning Curve

This decline in performance decays the rate of learning that causes longer manufacturing times and higher costs than would be forecasted using traditional learning curve theory. Although forgetting is common when production breaks occur as previously discussed, forgetting can occur without interruptions to the production line as discussed by Argote et al. (1990) and Badiru (2012). Many contemporary learning curve models attempt to incorporate the concept of forgetting. Models that incorporate variations to traditional learning curves—including rates of production, breaks in production, and configuration changes to end-items—attempt to model Badiru's (2012) second case of forgetting. Badiru's (2012) third case is challenging to model due to the stochastic nature of when and for how long forgetting will occur.

The concept of forgetting and its impact on decaying and non-constant rates of learning has proven relevant in contemporary learning curve research. Several forgetting models have been developed to include the learn-forget curve model (LFCM; Jaber & Bonney, 1996), the recency model (RCM; Nembhard & Uzumeri, 2000), the power integration and diffusion (PID) model (Sikström & Jaber, 2002), and the Depletion-Power-Integration-Latency (DPIL) model (Sikström & Jaber, 2012) among others (Glock et al., 2019). However, these forgetting models focus solely on the phenomenon of forgetting due to interruptions of the production process and most directly model Badiru's (2012) second case of forgetting (Glock et al., 2019; Anzanello & Fogliatto, 2011; Jaber, 2006). Jaber (2006) states that "there has been no model developed for industrial settings that considers forgetting as a result of factors other than production breaks" (p. 30-13) and mentions this as a potential area of future research. Although forgetting models have emerged after Jaber (2006), a review of the popular forgetting models cited confirms Jaber's statement.

Therefore, Badiru's (2012) first case of forgetting along the learning curve and its effect on the curve should be investigated.

A related concept to the forgetting phenomenon is the plateauing phenomenon. According to Jaber (2006), plateauing occurs when the learning process ceases. This ceasing of learning results in a flattening or partial flattening of the learning curve corresponding to rates of learning at or near zero. These near-zero rates of learning are in contrast to forgetting curves where rates of learning may become negative resulting in inverted learning curves. There remains debate as to when plateauing occurs in the production process or if learning ever ceases completely (Asher, 1956; Crossman, 1959; Honious et al., 2016; Moore et al., 2015). Jaber (2006) provides several explanations to explain the plateauing phenomenon that include concepts related to forgetting. Baloff (1966, 1970) recognized that plateauing is more likely to occur when capital is used in the production process as opposed to labor. According to some researchers, plateauing can be explained by either having to process the efficiencies learned before making additional improvements along the learning curve or to forgetting altogether (Corlett & Morcombe, 1970). According to other researchers, plateauing can be caused by labor ceasing to learn or management's unwillingness to invest in capital to foster induced learning (Yelle, 1980). Related to this underinvestment to foster induced learning, management's doubt as to whether learning efficiencies related to learning can occur is cited as another hindrance to constant rates of learning (Hirschmann, 1964).

Li and Rajagopalan (1998) investigated these explanations and concluded that no empirical evidence supports or contradicts them while ascribing plateauing to depreciation in knowledge or forgetting. Jaber (2006) concludes that "there is no tangible consensus among researchers as to what causes learning curves to plateau" and alludes that this is a topic for future research (p. 30-9).

Boone's Learning Curve: Accounting for the Flattening Effect

In an attempt to address the previous issues highlighted in constant learning rates, Boone (2018) developed a learning curve model with a rate of learning that diminishes as more units are produced. The traditional learning curve theories diminish the rate of cost reductions as more units are produced because costs will decrease at a constant rate only when the number of units produced doubles. Because the rate at which units double decreases as more units are produced, the rate of cost reductions will also decrease as more units are produced. However, the literature review cited various theoretical and empirical evidence indicating that the cost reductions that occur with each doubling of units may not be constant as the number of units produced increases. Therefore, Boone (2018) sought to attenuate the cost reductions that occur with each doubling of produced units by reducing the amount that each doubled unit's cost decreases as the number of units increases. This attenuation of cost reductions was accomplished by decreasing the rate of learning as the number of units increases.

Boone (2018) began by formulating a model where the exponent of the traditional learning curve equation is a function of the number of units produced. This amendment was intended to vary the learning curve exponent with the independent variable "x" in order to alter the degree of cost efficiencies experienced as the number of units produced changes. He then devised a series of specific models that decreased the learning curve exponent "b" as the number of units produced "x" increased. Boone (2018) first created a model without an additional parameter that aimed to reduce the learning curve exponent "b" directly by the unit number. However, he claimed that it resulted in too drastic of changes to the exponent value and did not model data appropriately (Boone, 2018, p. 20). To temper the effect each additional unit has on the parameter “b,” a qualifier was added. This qualifier “c” was named Boone’s Decay Value with an initially studied value ranging from zero to 5,000 (Boone, 2018, p. 21). The resulting Boone’s Learning Curve is shown in Equation 8.

Equation 8

$$\bar{Y} = Ax^{b/(1+\frac{x}{c})} \tag{8}$$

Where:

$\bar{Y}$ = the cumulative average time (or cost) per unit

X = the cumulative number of units produced

a = time (or cost) required to produce the first unit

b = slope of the function when plotted on log-log paper = log of the learning rate/log of 2

c = Boone’s Decay Value (range can be anything except 0)

Boone found this curve was not only flatter near the end of production but was also steeper in the early stages in comparison to the traditional theory learning curve (Boone, 2018, p. 21). Holding the cumulative number of units produced “x” constant, as Boone’s Decay Value approaches zero, the parameter “b” approaches zero representing a learning curve slope approaching 100%. As Boone’s Decay Values approaches infinity, the parameter “b” remains unchanged, and Boone’s Learning Curve simplifies to the traditional learning curve (Boone, 2018, p. 23).

Boone (2018) proceeded to test his model using Cumulative Average Theory against Wright’s model. Boone (2018) utilized prime mission equipment cost data in units of total dollars for fighter, bomber, and cargo aircraft programs along with missile and munition programs (p. 21). He constrained his data to require at least five lots per program to prevent overfitting the data (Boone, 2018, p. 22). In total, 46 weapon system platforms were tested (Boone, 2018, p. 23). The OLS regression method is unable to estimate the parameters for Boone’s Learning Curve because of the non-constant rate of learning; Boone’s Learning Curve is convex in logarithmic scale. Instead, Boone utilized Microsoft Excel’s Solver package to minimize the sum of squares errors (SSE) by iteratively adjusting the theoretical first unit cost “T1,” the learning curve slope “b,” and Boone’s Decay Value “c” (Boone, 2018, p. 24). The conventional OLS methodology was maintained to estimate the parameters for Wright’s Cumulative Average learning curve because this model remains linear in logarithmic scale (Boone, 2018, p. 24). Microsoft Excel’s Solver requires bounds for each parameter when solving for the combination of parameter values that minimize the SSE. Wright’s Learning Curve parameters informed these bounds to assist in estimating Boone’s Learning Curve parameters (Boone, 2018, pp. 24–25).

Boone’s theoretical first unit cost parameter minimum bound was equal to half of Wright’s theoretical first unit cost and twice the value of Wright’s first unit cost for the maximum bound (Boone, 2018, pp. 24–25). Boone’s “b” parameter bounds were set between -3 and 3 times Wright’s “b” specific to each estimated learning curve (Boone, 2018, pp. 24–25). These two bounds’ values varied for each learning curve estimated due to their dependence on Wright’s Learning Curve parameters. Lastly, Boone’s Decay Value was bound from zero to 5,000 for all estimated learning curves (Boone, 2018, pp. 24–25). The only limits that were found to be binding when solving for optimal values were the upper limit of Boone’s Decay Value (Boone, 2018, p. 25).

Boone then estimated the parameters for his curve and Wright's curve for each of the 46 programs using Cumulative Average Learning Curve theory. He obtained goodness-of-fit statistics in the form of the SSE and MAPE for each estimate in order to compare the accuracy of both curves (Boone, 2018, pp. 25–26). Boone then performed a paired difference t-test for both SSE and MAPE statistics (Boone, 2018, pp. 25–26). Both the SSE and MAPE paired difference t-tests rejected the null hypothesis that the means were equal to zero (Boone, 2018, pp. 29–30). These tests indicate that Boone's Learning Curve more accurately explains the cost data in comparison to Wright's Learning Curve at a significance level (α) of 0.05. Therefore, Boone demonstrated his learning curve to be more accurate to a statistically significant degree in comparison to Wright's Learning Curve (Boone, 2018, pp. 30–31).

Boone did not assess the predictive capacity of his model in comparison to Wright's by estimating parameters for a subset of early units and then extrapolating to future lots using the same estimated parameters. This approach departs from previous research including Moore et al. (2015) and Honious et al. (2016); however, this may be due to data availability as several lots are required for predictive analysis. Additionally, Boone did not hold constant the traditional learning curve parameters and estimate his new parameter, Boone's Decay Value, with these values fixed. Instead, Boone allowed all three of his parameters to change from Wright's estimated parameter values. Boone's methodology is also a departure from the previous methodology where Moore et al. (2015) and Honious et al. (2016) estimated the traditional learning curve, held the estimated parameters constant, and then added additional parameters. Boone's methodology may also be justified because Boone's Decay Value is a result of estimating learning curves and requires a different learning curve slope and first unit cost be considered. This Decay Value is unlike the parameters Stanford-B parameter and incompressibility parameter "M," which are measurable values that describe the manufacturing process itself.

Methodology Population and Sample

In order to test Boone's Learning Curve against the traditional learning curve theories, quantitative data from a diverse set of DoD programs was gathered. The population studied is DoD programs that have produced several complex end-items over time. These complex units can include aircraft, land vehicles, and missiles along with their complex sub-systems and sub-components. The data sample consists of programs with the necessary information required for learning curve analysis. This required program data included direct recurring labor costs in units of labor hours or total dollars per production lot along with the number of units per lot.

The direct recurring manufacturing labor cost category for each applicable WBS element was used to obtain labor hour data for each program. If this labor hour data were unavailable, the total recurring cost for each applicable WBS element was utilized instead. The total recurring dollar cost comprises the costs of all functional categories of labor along with materials costs and other costs for each WBS element. Unlike labor hours, costs in units of dollars must be normalized to be compared over time; therefore, all costs in units of dollars were normalized using escalation rates based on Producer Price Index (PPI) 3364 Aerospace Products and Parts. Removing the effects of escalation using PPI 3364 is common practice when normalizing costs in the aerospace industry. Additionally, total costs in units of dollars were provided and maintained in units of thousands of dollars. These labor cost data included costs at the prime mission equipment. WBS level prime mission equipment costs are directly related to touch labor and experience learning. Depending on data availability, additional elements below the prime mission equipment WBS elements were also analyzed to include engines and wings among other sub-systems and sub-components. Therefore, one program may contribute several unique components for learning curve analysis.

This final sample included direct recurring cost data from bomber, cargo, and fighter aircraft along with missiles and munitions. The programs in this dataset are both historical and contemporary spanning 1957–2018 and include a variety of defense contractors. This diverse dataset tested the generalizability of Boone's Learning Curve due to the varying levels of analysis along with the multitude of platforms, contractors, and production periods that foster various learning environments. In total, data from 123 weapon system programs were gathered with 258 unique components. Learning curve analysis will be performed on these unique components.

Data Collection

This dataset was created using DD Form 1921-1 "Functional Cost-Hour Report" and DD Form 1921-2 "Progress Curve Report" data obtained from the CADE Defense Automated Cost Information Management System (DACIMS). CADE DACIMS is a repository of DoD weapons system program cost data available to DoD analysts. Some historical data was also extracted from the AFLCMC Cost Research Library.

Business rules were created to avoid overfitting the data and to ensure learning curve analysis was appropriate to model each component's cost. The first business rule omitted programs with production lots of four or fewer. This business rule is consistent with Boone (2018, p. 22) and limited the sample from 258 to 169 unique components. A second business rule was also necessary when performing Cumulative Average Theory analysis. Cumulative Average Theory relies on continuous data because each lot's cumulative average cost and cumulative quantity is a function of all previous lots' costs and quantities. Therefore, if a program's production lot was missing cost or quantity data, all lots after that missing lot were removed for that program; however, all lots before the missing lot were retained. These lot removals decreased the number of lots in the total program. This reduced number of lots warranted the complete removal of some programs by applying the first business rule. This second business rule limited the dataset to 140 unique components for Cumulative Average Theory analysis. Despite these business rules, there was not a systematic elimination of any characteristic of the program labor cost data; a diverse dataset remained with the previously stated attributes.

Data Analysis

This analysis will examine whether Boone's Learning Curve more accurately explains variability in program labor cost data than the traditional theories. Both Cumulative Average Theory and Unit Theory will be used to make these comparisons. In order to test these hypotheses, learning curve parameters were estimated using each program's labor cost data for Boone's Learning Curve and the traditional learning curve models. Next, parameters from Boone's Learning Curve and the respective traditional theory were used to predict a learning curve. These predicted learning curves were then compared to the observed data. In order to utilize Unit Learning Curve Theory, Boone's Learning Curve was adapted from its original Cumulative Average Theory form to Unit Learning Curve Theory form. When Boone's Learning Curve utilized Cumulative Average Theory, Boone's Learning Curve was compared to Wright's Learning Curve. When Boone's Learning Curve utilized Unit Theory, Boone's Learning Curve was compared to Crawford's Learning Curve.

Parameters for each learning curve were estimated using non-linear optimization techniques in Microsoft Excel. The traditional learning curve theories could be estimated using OLS regression. However, non-linear optimization was utilized to estimate the traditional curves for equitable comparison with Boone's Learning Curve. In contrast, Boone's Learning Curve required the use of nonlinear optimization techniques. This requirement spawns from the fact that Boone's Learning Curve is not linear when logarithmically transformed due to the decaying learning curve slope. This non-linearity of Boone's Learning Curve precludes the parameters from being estimated using OLS regression. The learning curve parameters for Boone's Learning Curve (i.e., "A," "b," and "c") and the traditional theories (i.e., "A" and "b") were estimated by minimizing the SSE using Excel's Generalized Reduced Gradient (GRG) Nonlinear Solver and Excel's Evolutionary Solver engines. The SSE is calculated by squaring the vertical difference of the observed data and predicted data for each lot and summing these squared differences across all lots. The SSE is calculated separately for Boone's Learning Curve and the traditional learning curves. For each model, Excel Solver is set to minimize the objective cell that is set as the SSE. The changing variable cells are the learning curve parameters specific to each learning curve model. These parameters are iteratively solved for using optimization techniques specific to each engine for each learning curve model. When using Evolutionary Solver, it is also necessary to bound the changing variable cells and choose a set of values from which to begin the optimization process. Due to the inherent differences in both Cumulative Average Theory and Unit Learning Curve Theory, different specific processes were used to estimate parameters for each.

Cumulative Average Learning Curve Theory

The following process was implemented to estimate parameters for Wright's Learning Curve and Boone's Learning Curve using Cumulative Average Theory for each program.

1. Wright's Learning Curve parameters "A" and "b" were initially estimated using OLS regression. a. Cumulative Average Cost was the dependent variable, while Cumulative Number of Units Produced was the independent variable.
2. These initial learning curve parameter estimates were used as starting values to more precisely estimate Wright's Learning Curve parameters using GRG Non-Linear Solver. This process generated final estimates for Wright's Learning Curve parameters.
3. Boone's Learning Curve parameters "A," "b," and "c" were estimated using Excel's Evolutionary Solver. This process generated initial estimates for Boone's Learning Curve parameters.
    a. Final estimates for Wright's Learning Curve parameters were used to calculate the upper and lower bounds of Evolutionary Solver.
    b. The starting values were calculated from the upper and lower bounds.
4. The Evolutionary Solver learning curve parameter estimates for Boone's Learning Curve were used as starting values to more precisely estimate parameters using GRG Non-Linear Solver.

This process produced final estimates for Boone's Learning Curve parameters. When estimating Wright's Learning Curve parameters, the GRG Nonlinear Solver technique should produce SSE at least equal to the SSE using OLS regression. The GRG Nonlinear Solver technique was appropriate to estimate Wright's Learning Curve parameters because this technique is used to find locally optimal solutions of smooth and non-linear functions (Solver Technology–Smooth Nonlinear Optimization, 2012). Because OLS regression was used to provide starting values for GRG Nonlinear Solver, it was reasonable to assume that the global minimum is within this local region approximated using OLS regression. However, GRG Nonlinear Solver cannot guarantee that a global minimum is found. The GRG Nonlinear Solver Multistart method was also utilized to ensure this technique yielded a global minimum. However, the Multistart method failed to provide consistent, reliable parameter estimates, unlike GRG Nonlinear Solver.

When estimating Boone's Learning Curve, the GRG Nonlinear Solver optimized values for Wright's Learning Curve parameters were used to calculate bounds for Boone's Learning Curve. For the "A" parameter, the lower bound was half of Wright's "A" parameter, and the upper bound was twice that of Wright's "A" parameter. For bounds on Boone's Learning Curve "b" parameter, the absolute value of Wright's Learning Curve "b" parameter was multiplied by 3 to yield upper (positive) and lower (negative) bounds. Lastly, for Boone's Decay Value "c," bounds were set between 0 and 500,000 in contrast to Boone's original bounds of 0 and 5,000. This bound was increased in comparison to Boone (2018) due to the upper bound being binding in Boone's analyses. Except for Boone's Decay Value, these bounds are consistent with Boone (2018) and provide the optimization model with a restricted range to decrease the search time for an optimal solution but a broad enough range to not constrain the model. Similar to Boone (2018), none of these constraints were binding except for the Decay Value "c" upper bound despite relaxing this constraint in comparison to Boone (2018). Further relaxing this constraint would not have led to substantive changes to Boone's Learning Curve; as Boone's Decay Value "c" approaches infinity, Boone's Learning Curve transforms into Wright's Learning Curve.

In order to estimate Boone's Learning Curve, the starting values were set as the midpoint of the negative values of the slope parameter "b" because negative values are those that represent learning curve slopes below 100%. This starting point is reasonable because most programs in this analysis will experience cost efficiencies from learning. Thus, these programs will have learning curve slopes at or below 100% that translate to a negative learning curve exponent "b" parameter. The starting values for Boone's Learning Curve first unit cost "A" parameter were the midpoint of the upper and lower bounds. These two starting values and bounds depend on the parameters estimated by Wright's Learning Curve. The last starting value was set as the midpoint between the upper and lower bounds of the Boone's Decay Value "c," which was static at 250,000 due to the static bounds.

Once these starting values and bounds were set, Evolutionary Solver was used to estimate a globally optimal solution. The Evolutionary Solver technique was appropriate to estimate Boone's Learning Curve parameters because this technique is used to find a globally optimal solution of smooth and non-smooth functions (Solver Technology–Global Optimization, 2016). In contrast to estimating Wright's Learning Curve, a local region with the global minimum cannot be reliably approximated before using this optimization technique. This local region cannot be reliably approximated because OLS regression cannot be used to provide starting values for Boone's Learning Curve. Similar to GRG Nonlinear Solver, Evolutionary Solver cannot guarantee the parameter estimates produce a global minimum. However, using the Evolutionary Solver solution as starting values, the GRG Nonlinear Solver was then used to ensure the solution is locally optimal.

Unit Learning Curve Theory

Unit Learning Curve Theory parameter estimation maintained the Cumulative Average Learning Curve Theory parameter estimation methodology for calculating various bounds and starting values. Additionally, the justification for utilizing both Excel Solver engines also remains the same. However, the inclusion of lot midpoint calculations required different analysis techniques. The following process was implemented to estimate parameters for Crawford's Learning Curve and Boone's Learning Curve using Unit Theory for each program.

1. Parameter-free lot midpoint approximations (Equation 6) were calculated for each production lot.

2. Crawford's Learning Curve parameters "A" and "b" were initially estimated using OLS regression. a. Average Unit Cost was the dependent variable while Lot Midpoint, calculated in Step 1, was the independent variable.

3. These initial learning curve parameter estimates were used as starting values to more precisely estimate Crawford's Learning Curve parameters using GRG Non-Linear Solver. This process generated intermediate estimates of Crawford's Learning Curve parameters.

4. The intermediate estimate of Crawford's Learning Curve "b" parameter was used to calculate a more precise set of lot midpoints using Asher's Approximation (Equation 7).

5. Applying these more precise lot midpoint approximations, Crawford's Learning Curve parameters "A" and "b" were more accurately estimated using GRG Nonlinear Solver.

a. Steps 4 and 5 were repeated until the iterative process converged on a solution to produce final estimates of Crawford's Learning Curve parameters and lot midpoint approximations.

6. Parameter-free lot midpoint approximations (Equation 6) were used to estimate Boone's Learning Curve parameters "A," "b," and "c" using Excel's Evolutionary Solver. This process generated intermediate estimates of Boone's Learning Curve parameters.

a. The final estimates for Crawford's Learning Curve parameters were used to calculate upper and lower bounds.

b. The starting values were calculated from the upper and lower bounds.

7. The intermediate estimates of Boone's Learning Curve parameters "b" and "c" were used to approximate a more precise set of lot midpoints using Asher's Approximation adapted for Boone's Learning Curve (Equation 14).

8. Applying these more precise lot midpoint approximations, Boone's Learning Curve parameters were more accurately estimated using Evolutionary Solver. This process generated Evolutionary Solver parameter estimates of Boone's Learning Curve.

9. These Evolutionary Solver parameter estimates were used as starting values to improve further the accuracy of Boone's Learning Curve parameters estimates "A," "b," and "c" using GRG Non- Linear Solver.

a. Steps 7, 8, & 9 were repeated until the iterative process converged on a solution to produce a final estimate of Boone's Learning Curve parameters and lot midpoint approximations.

For both Crawford's and Boone's Learning Curves, an iterative process is used to calculate precise parameter estimates and lot midpoint approximations. This iterative process was repeated until a solution converged. A solution converged when small changes in the learning curve exponent "b" parameter were calculated between iterations. This process of iterative solving was adapted from Hu and Smith's "Accuracy Matters" (2013). For Boone's Learning Curve, a limit of 10 iterations was placed on the iterative process. This limit of 10 iterations was reached a limited number of times and still produced relatively small differences of Boone's Learning Curve exponent "b" between iterations. In order to estimate lot midpoints and Boone's Learning Curve parameters, Asher's Approximation from Equation 6 was adapted to incorporate Boone's decaying learning curve slope.

Statistical Significance Testing

The estimated parameters for Boone's Learning Curve and the traditional learning curves were used to create predicted learning curves. These predicted curves were then compared to observed data. Total model error was calculated by comparing the difference between observations and predicted values to determine which learning curve theory more accurately explained variability in the data. Two measures were used to determine the overall model error. The first error measure was RMSE. RMSE is calculated by dividing the total SSE by the number of observations; in our case, the number of lots in that program (McClave et al., 2014, p. 14-25). RMSE is a scale-dependent measure. Additionally, RMSE is not robust to outliers; therefore, the greater the magnitude of an outlier from the average error values, the more influence this outlier will have on RMSE. RMSE was used instead of total SSE because RMSE transforms units from squared units to original units. This transformation eases interpretation. RMSE can be interpreted as the average amount of error of the model in the model's original units.

The second measure used to determine the overall model error was MAPE. MAPE is calculated by subtracting the predicted value from the observed value, dividing this difference by the observed value, taking the absolute value, and multiplying by 100%; these absolute percent errors are then summed over all observations and divided by the total number of observations (McClave et al., 2014, p. 14-25). MAPE provides a unitless measure of accuracy and can be interpreted as the average percentage the model is inaccurate. MAPE is robust to outliers, so the effects of outliers do not unduly influence this measure.

After calculating these measures of overall model error, a series of paired difference t-tests were conducted to determine if reductions in error from Boone's Learning Curve are significantly different than zero or due to random chance. In order to conduct the first paired difference t-test, Boone's Learning Curve RMSE using Cumulative Average Theory will be subtracted from Wright's Learning Curve RMSE, and the difference will be divided by Wright's Learning Curve RMSE by observation. This calculation will yield a percentage difference rather than raw difference to compare programs of varying differences in magnitude equitably. Only programs with errors reported in total dollars will be examined first to examine the results of the different unit measures. The null hypothesis for this test is that the percentage difference in RMSE is equal to or less than zero. This null hypothesis represents that Boone's Learning Curve results in an equal amount of or more error in predicting observed values in comparison to Wright's Learning Curve. The alternative hypothesis is that the percentage difference is greater than zero. The alternative hypothesis represents that Boone's Learning Curve results in less error in predicting observed values than Wright's Learning Curve. This same methodology was repeated five more times to examine both learning curve theories, each with two measures of model error for programs examined using labor hours and total dollars.

Results

The diverse dataset utilized in Phase 1 was augmented with observed learning curves estimated at the flyaway cost level to form the basis of the dataset used in Phase 2. Phase 2 utilized Unit Theory learning curve predictions generated from analyses in Phase 1. Observed learning curve data, Crawford's Learning Curve data, and Boone's Learning Curve data were used to answer a variety of research questions. These analyses sought to determine 1. if the plateauing and forgetting phenomena occur using traditional learning curve theory, 2. how Boone's Learning Curve models observed learning curve data in comparison to traditional learning curve theory, and 3. which program attributes, if any, can explain the programs that are best modeled by Boone's Learning Curve as well as explain the prevalence of plateauing and forgetting phenomena.

In order to understand if the plateauing and forgetting phenomena occur using the traditional learning curve theory, the mean percentage error of Crawford's Learning Curve in the fourth quarter was isolated. These fourth quarter mean percentage errors indicated that Crawford's Learning Curve did not systematically underestimate observed learning curves. In fact, the statistical tests across all quarters indicated Crawford's Learning Curve systematically overestimated observed learning curves with a high amount of variability. These findings suggest that the plateauing and forgetting phenomena are not systematically present in observed data when modeled with Crawford's Learning Curve. This conclusion limits instances where Boone's Learning Curve can more accurately model observed learning curve data in comparison to Crawford's Learning Curve.

The mean percentage error of Crawford's Learning Curve in the fourth quarter were then analyzed using proportions. Crawford's Learning Curve underestimated 47.6% of the observed learning curves tested. Therefore, these hypothesis tests of proportions also failed to indicate that Crawford's Learning Curve systematically underestimates observed learning curve data in the fourth quarter. Lastly, for the proportion analysis, the proportion of learning curves that Crawford's Learning Curve underestimated in both the first and last quarter was investigated. These instances represent learning curves with high rates of learning decaying to low rates of learning; Boone's Learning Curve models these observed learning curves exceptionally well. The proportion of learning curves that experienced high rates of learning decaying to low rates of learning was 35% when modeled using Crawford's Learning Curve. This proportion highlights that if an observed learning curve is experiencing high rates of learning in the first quarter, it is also more likely to experience low rates of learning in the last quarter when modeled with the Crawford's Learning Curve. Although these results failed to show a systematic presence of plateauing, they emphasize opportunities for Boone's Learning Curve to improve upon Crawford's Learning Curve.

Next, Boone's Learning Curve was analyzed to determine how it models observed learning curve data in comparison to the traditional learning curve theory. Because Boone's Learning Curve contains an additional, empirically-estimated parameter, it is expected to improve upon Crawford's Learning Curve at an aggregate level. However, if Boone's Learning Curve more accurately models observed learning curves that plateau and remains approximately equal to Crawford's Learning Curve in terms of error for observed learning curves that do not plateau, then Boone's Learning Curve provides inherent value in modeling the plateauing phenomenon. To investigate this, a confusion matrix was used to determine how learning curves that plateau interact with learning curves that are more accurately explained by Boone's Learning Curve. The confusion matrix indicated that Boone's Learning Curve more accurately modeled observed learning curves based on if those observed learning curves plateaued for 77% of observations. These results provide mixed conclusions as to if Boone's Learning Curve improves upon Crawford's Learning Curve by more accurately modeling the plateauing phenomenon of observed learning curves or because of its an additional, empirically-estimated parameter.

To further investigate how Boone's Learning Curve models observed learning curves, the MAPE percentage differences between Boone's Learning Curve and Crawford's Learning Curve in the first, middle, and last quarters were investigated. Hypothesis tests were conducted to determine in which quarter Boone's Learning Curve improved upon Crawford's Learning Curve to a statistically significant degree. Outliers were excluded, and the dataset was limited to observations in which Boone's Learning Curve was a significant improvement in error. The hypothesis tests indicated that Boone's Learning Curve improved upon Crawford's Learning Curve to a statistically significant degree in the first quarter only. Furthermore, the tests indicated that the error improvements in the last quarter were not statistically different from zero. The middle quarters' hypothesis test indicated that Boone's Learning Curve was systematically worse at explaining observed data in comparison to Crawford's Learning Curve. Each of these tests had a high amount of variability. These results suggest that Boone's Learning Curve improves upon Crawford's Learning Curve more than it merely having an additional empirically estimated parameter. However, Boone's Learning Curve was not shown to model the plateau effect in the last quarter to a statistically significant degree. There is a possibility observed learning curves improved in different sections of the learning curves while performing worse in others in comparison to Crawford's Learning Curve. These error improvements in different sections of the learning curve would have caused substantial variations of error in each section. These substantial variations of error improvement within each section could significantly obscure the statistical results due to the variability included in each hypothesis test.

Regression analysis was also used to investigate which program attributes, if any, can explain the programs that are best modeled by Boone's Learning Curve as well as explain the plateauing and forgetting phenomena. For the first part of this regression analysis, the percentage difference between Boone's Learning Curve MAPE and Crawford's Learning Curve MAPE was used as a dependent variable. A variety of independent variables were created using the Literature Review and operationalized using available data. The independent variables representing the level of aggregation of the learning curve was significant in the model. The regression analysis indicated that Boone's Learning Curve MAPE would decrease 26% on average when modeling learning curves at the highest (flyaway cost) and second highest (air vehicle cost) levels of aggregation, respectively, in comparison to Crawford's Learning Curve. This finding is consistent with the research hypotheses. The independent variable that measured the amount of time that spanned between the first and last years of production was also statistically significant although the coefficient estimate's sign was opposite of the hypothesized sign. In a separate regression, the independent variable representing the learning curve units of measure in total dollars was also statistically significant. This independent variable indicated that Boone's Learning Curve is associated with a 14.9% decrease in error on average when modeling observed learning curves with units of measure in total dollars rather than hours in comparison to Crawford's Learning Curve. Despite these pool of significant variables, several hypothesized were not significant in the model. Additionally, the Adjusted R2 for these models was relatively low at 6.7% and 5.3%, respectively. These Adjusted R2 values highlight that a large amount of variability remains to be explained in the dependent variable. These theory-based independent variables were also tested using data-mining independent variables using a mixed stepwise regression. This methodology was employed to potentially expose statistically significant theory-based independent. Some groups of data-mining independent variables were statistically significant and able to explain more variability in the dependent variable; however, no additional statistically significant theory-based independent variables were revealed. Furthermore, the stepwise regression model results and statistical inferences may have been invalid due to not passing overall model tests.

For the second part of this regression analysis, the percentage error between Crawford's Learning Curve and observed data in the fourth quarter of the observed learning curve served as a dependent variable. This variable provided a measure of plateauing indifferent to how Boone's Learning Curve models observed data. After testing various independent variables and validating the model with statistical tests, no hypothesized independent variables were significant in the model. This finding suggests that either the operationalization of plateauing was inappropriate, there exists too much variability in the data to explain instances of plateauing, or the hypothesized independent variables are unsuitable or were not appropriately operationalized. These theory-based independent variables were also tested using a mixed stepwise regression in order to potentially expose other statistically significant theory-based independent. Some groups of data-mining independent variables were statistically significant and able to explain more variability in the dependent variable; however, no statistically significant theory-based independent variables were revealed. Furthermore, the second set of stepwise regression model results and statistical inferences may have also been invalid due to not passing overall model tests.

Summary

This research sought to determine 1. if the plateauing and forgetting phenomena occur using traditional learning curve theory, 2. how Boone's Learning Curve models observed learning curve data in comparison to traditional learning curve theory, and 3. which program attributes, if any, can explain the programs that are best modeled by Boone's Learning Curve as well as explain the plateauing and forgetting phenomena. The various statistical results indicate that the plateauing and forgetting phenomena do not systematically occur using the traditional Unit Learning Curve theory. Despite these results, there remain instances where plateauing and forgetting occur that provide opportunities for Boone's Learning Curve to improve upon Crawford's Learning Curve.

Additionally, results provide mixed conclusions as to if Boone's Learning Curve improves upon Crawford's Learning Curve by more accurately modeling the plateauing phenomenon of observed learning curves or because of its additional, empirically-estimated parameter. Tests also did not show that Boone's Learning Curve improved upon Crawford's Learning Curve to a statistically significant degree in the last quarter but instead in the first quarter. This finding brings into question if Boone's Learning Curve does in fact more accurately model the plateauing phenomenon.

Finally, few hypothesized independent variables were significant in explaining when Boone's Learning Curve more accurately models observed learning curves, and a significant amount of variability in the data remained. Despite these results, OLS regression analysis confirmed that the level of aggregation, the units of measure, and the time span of a program explain a small amount of variability of the degree to which Boone's Learning Curve more accurately models observed learning curves. The OLS regression results for explaining plateauing independent of Boone's Learning Curve indicated that no hypothesized independent variables were significant in the model even when non-theoretically based independent variables were included in the OLS regression model.

These results indicate that Boone's Learning Curve may provide value by more accurately explaining observed learning curves. However, the theoretical explanations as to where Boone's Learning Curve improves upon the traditional learning curve theory and under what circumstances Boone's Learning Curve is more appropriate to use than the traditional learning curve theory remains to be discovered.

References

Anderlohr, G. (1969). Determining the cost of production breaks. *Management Review*, *58*(12), 16–19.

Anzanello, M. J., & Fogliatto, F. S. (2011). Learning curve models and applications: Literature review and research directions. *International Journal of Industrial Ergonomics*, *41*(5), 573–583.

Argote, L. (1993). Group and organizational learning curves: Individual, system and environmental components. *British Journal of Social Psychology*, *32*(1), 31–51.

Argote, L., Beckman, S. L., & Epple, D. (1990). The persistence and transfer of learning in industrial settings. *Management Science*, *36*(2), 140–154.

Asher, H. (1956). *Cost–quantity relationships in the airframe industry* (Doctoral dissertation, The Ohio State University). https://etd.ohiolink.edu/!etd.send_file?accession=osu1485795295566357&disposition=inline

Badiru, A. B. (1992). Computational survey of univariate and multivariate learning curve models. *IEEE Transactions on Engineering Management*, *39*(2), 176–188. https://doi.org/10.1109/17.141275

Badiru, A. B. (2012). Half-life learning curves in the defense acquisition life cycle. *Defense Acquisition Research Journal*, *19*(3), 283–308. http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA582714

Baloff, N. (1966). Startups in machine-intensive production systems. *Journal of Industrial Engineering*, *17*(1), 25.

Baloff, N. (1970). Startup management. *IEEE Transactions on Engineering Management*, *EM-17*(4), 132–141.

Boone, E. R. (2018). *An analysis of learning curve theory and the flattening effect at the end of the production cycle* (No. AFIT-ENV-MS-18-M-181). Air Force Institute of Technology.

Burke, R. P. (2019). *Implementation of cost and hour report (FlexFile) and quantity data reports within the Cost and Software Reporting (CSDR) System*. DoD.

Corlett, E. N., & Morecombe, V. J. (1970). Straightening out learning curves. *Personnel Management*, *2*(6), 14–19.

Crawford, J. R. (1944). *Estimating, budgeting and scheduling*. Lockheed Aircraft Co.

Crossman, E. R. F. W. (1959). A theory of the acquisition of speed-skill. *Ergonomics*, *2*(2), 153–166.

Cullis, B., Coleman, R., Braxton, P., & McQueston, J. (2008). *CUMAV or unit? Is cum average vs. unit theory a fair fight?* SCEA Conference, 1–17. ICEAA. Published presentation by Lockheed Martin Corporation.

DoD. (2017). *Financial management regulation* (DoD 7000.14-R, Vol 2A). Under Secretary of Defense (Comptroller). https://comptroller.defense.gov/FMR/fmrvolumes.aspx

DoD. (2018). *Work breakdown structures for defense materiel items*. Department of the Air Force.

Department of the Air Force. (2007). *Air Force cost analysis handbook*.

Dutton, J. M., & Thomas, A. (1984). Treating progress functions as a managerial opportunity. *Academy of Management Review*, *9*(2), 235–247.

Epple, D., Argote, L., & Devadas, R. (1991). Organizational learning curves: A method for investigating intraplant transfer of knowledge acquired through learning by doing. *Organization Science*, *2*(1), 58–70.

Flyaway Costs. (n.d.). In *Defense Acquisition University glossary*. https://www.dau.edu/glossary/Pages/Glossary.aspx

Glock, C. H., Grosse, E. H., Jaber, M. Y., & Smunt, T. L. (2019). Applications of learning curves in production and operations management: A systematic literature review. *Computers & Industrial Engineering*, *131,* 422–441.

GAO. (2009). *GAO cost estimating and assessment guide: Best practices for developing and managing capital program costs* (GAO-09-3SP).

GAO. (2018). *Weapons system annual assessment: Knowledge gaps pose risks to sustaining recent positive trends* (GAO-18-360SP).

Hilmer, C. E., & Hilmer, M. J. (2014). *Practical econometrics.* McGraw-Hill Education.

Honious, C., Johnson, B., Elshaw, J., & Badiru, A. (2016). The impact of learning curve model selection and criteria for cost estimation accuracy in the DoD (SYM-AM-16-075). *Acquisition Research Symposium: Vol. 2* (pp. 421–486). Naval Postgraduate School.

Hu, S. P., & Smith, A. (2013). Accuracy matters: Selecting a lot-based cost improvement curve. *Journal of Cost Analysis and Parametrics*, *6*(1), 23–42.

International Cost Estimating and Analysis Association. (2013). Learning curve analysis: How to account for cost improvement. *Cost estimating body of knowledge training, Unit III, Module 7.*

Jaber, M. Y. (2006). Learning and forgetting models and their applications. *Handbook of Industrial and Systems Engineering*, *30*(1), 30–127.

Jaber, M. Y., & Bonney, M. (1996). Production breaks and the learning curve: The forgetting phenomenon. *Applied Mathematical Modelling*, *2*(20), 162–169.

Knecht, G. R. (1974). Costing, technological growth and generalized learning curves. *Journal of the Operational Research Society*, *25*(3), 487–491.

Large, J. P., Hoffmayer, K., & Kontrovich, F. (1974). *Production rate and production cost* (Vol. 1609, No. PA/E). RAND.

Levy, F. K. (1965). Adaptation in the production process. *Management Science*, *11*(6), B-136.

Li, G., & Rajagopalan, S. (1998). A learning curve model with knowledge depreciation. *European Journal of Operational Research*, *105*(1), 143–154.

McClave, J. T., Benson, P. G., & Sincich, T. (2014). *Statistics for business and economics.* Pearson.

Mislick, G. K., & Nussbaum, D. A. (2015). *Cost estimation: Methods and tools*. John Wiley & Sons.

Moore, J. R., Elshaw, J. J., Badiru, A. B., & Ritschel, J. D. (2015). Acquisition challenge: The importance of incompressibility in comparing learning curve models. *Defense Acquisition Research Journal*, *22*(4), 416–449.

Nembhard, D. A., & Uzumeri, M. V. (2000). Experiential learning and forgetting for manual and cognitive tasks. *International Journal of Industrial Ergonomics*, *25*(4), 315–326.

Sikström, S., & Jaber, M. Y. (2002). The power integration diffusion model for production breaks. *Journal of Experimental Psychology: Applied*, *8*(2), 118.

Sikström, S., & Jaber, M. Y. (2012). The Depletion–Power–Integration–Latency (DPIL) model of spaced and massed repetition. *Computers & Industrial Engineering*, *63*(1), 323–337.

Solver Technology–Global Optimization. (2016, September 12). https://www.solver.com/global-optimization-technology

Solver Technology–Smooth Nonlinear Optimization. (2012, March 12). https://www.solver.com/smooth-nonlinear-technology#Generalized Reduced Gradient Method

Wright, T. P. (1936). Factors affecting the cost of airplanes. *Journal of the Aeronautical Sciences*, *3*(4), 122–128.

Yelle, L. E. (1979). The learning curve: Historical review and comprehensive survey. *Decision Sciences*, *10*(2), 302–328.

Acquisition Research Program
Graduate School of Defense Management
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net