Software Engineering Institute

Quantifying Uncertainty in Early Lifecycle Cost Estimation (QUELCE)

Software Engineering Measurement and Analysis (SEMA)
Cost Estimation Research Group

Robert Ferguson
Dennis Goldenson
James McCurley
Robert Stoddard
David Zubrow
Debra Anderson

December 2011

TECHNICAL REPORT
CMU/SEI-2011-TR-026
ESC-TR-2011-026

http://www.sei.cmu.edu

Carnegie Mellon

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

This report was prepared for the

Contracting Officer
ESC/CAA
20 Shilling Circle
Building 1305, 3rd Floor
Hanscom AFB, MA 01731-2125

Table of Contents

List of Figures

List of Tables

Acknowledgments

We are indebted to many people who have contributed to our thinking about the work described in this report. Several encouraged us early on to further elaborate our approach to early cost estimation, most notably Joe Dean and Don Reifer. Others have been particularly helpful at various stages of our research, including Michael Cullen, Rob Flowe, Jim Judy and Keith Miller. The same is so for Adam Porter of the University of Maryland and our Software Engineering Institute (SEI) colleagues Anita Carleton, Erin Harper, John Robert, Dave Scherb, and Doug Schmidt.

Our colleagues Eduardo Miranda of Carnegie Mellon® University and Ricardo Valerdi of the University of Arizona have given us counsel about our approach from the beginning. At this writing, we are organizing a series of empirical trials of aspects of the method with them and their graduate students. We are especially grateful to Tektronix Communications for their participation and contributions. Their feedback provided valuable insights for implementing and refining our methods. David Hazard of Aerospace Corporation arranged access for us to the Transformational Satellite Communications System (TSAT) documents discussed in this report, and he has expressed interest in collaborating with us on future work. Our SEI colleague Julie Cohen was instrumental in introducing us to Dave. She also participated along with John Foreman, Tim Morrow, and Mike Phillips in the workshops we did with members of the SEI's Acquisition Support Program. Dan Galorath kindly discussed with us how best to couple the outcomes of our methods as inputs to traditional cost estimation models.

Abstract

Difficulties with accurately estimating the costs of developing new systems have been well documented, and cost overruns in new systems development are well known. The headline of a recent defense magazine article referred to the true cost of a weapon as "anyone's guess," reflecting this widely acknowledged fact. The difficulty of accurate cost estimation is compounded by the fact that estimates are now prepared much earlier in the acquisition lifecycle, well before there is concrete technical information available on the particular program to be developed. This report describes an innovative synthesis of analytical techniques into a cost estimation method that models and quantifies the uncertainties associated with early lifecycle cost estimation.

The method described in this report synthesizes scenario building, Bayesian Belief Network (BBN) modeling, and Monte Carlo simulation into an estimation method that quantifies uncertainties, allows subjective inputs, visually depicts influential relationships among program change drivers and outputs, and assists with the explicit description and documentation underlying an estimate. It uses scenario analysis and design structure matrix (DSM) techniques to limit the combinatorial effects of multiple interacting program change drivers to make modeling and analysis more tractable. Representing scenarios as BBNs enables sensitivity analysis, exploration of scenarios, and quantification of uncertainty. The methods link to existing cost estimation methods and tools to leverage their cost estimation relationships and calibration. As a result, cost estimates are embedded within clearly defined confidence intervals and explicitly associated with specific program scenarios of alternate futures. This report provides a step-by-step description of the method with examples and ideas for future research and development.

1 Introduction

The inaccuracy of early cost estimates for developing major Department of Defense (DoD) systems is well documented, and cost overruns have been a common problem that continues to worsen. The headline of a recent article, “As Pressure Grows to Cut Spending, the True Cost of a Weapon is Anyone’s Guess,” [Erwin 2011] reflects this widely acknowledged fact. Another author has referred to acquisition programs as being in a state of “perpetual scandal”[Cancian 2010].

The difficulty of accurate cost estimation is compounded by the fact that estimates are now prepared much earlier in the acquisition lifecycle, well before there is concrete technical information available on the particular program to be developed. Thus, the estimates are often based on a desired capability, or an abstract concept, rather than a concrete solution to achieve the desired capability.

As a result, early estimates rely heavily on expert judgments about cost factors. Many assumptions about the desired end product are made by experts in deriving the estimates, but these assumptions are often unstated and vary from one expert to the next. Little attention is paid to the way in which factors that influence cost may change over the lifecycle of program development and implementation. It is no surprise, then, that the resulting estimate is often far short of the actual cost of a new system.

The QUELCE (Quantifying Uncertainty in Early Cost Estimation) method overcomes many of these issues by bringing to bear the knowledge and experience of domain experts and estimators in new ways. QUELCE elicits information about program change driver uncertainties that are common to program execution in a DoD Major Defense Acquisition Program lifecycle. The information is transformed into a Bayesian Belief Network (BBN), which models the interdependencies and their impacts on cost via likely scenarios of program execution. Monte Carlo simulation is used to estimate the distribution of program cost through traditional cost estimation tools used within the DoD.

The QUELCE method thus

- makes use of available information not normally employed for program cost estimation
- provides an explicit, quantified consideration of the uncertainty of the program change drivers
- enables calculation (and re-calculation) of the cost impacts caused by changes that may occur during the program lifecycle
- enhances decision-making through the transparency of the assumptions going into the cost estimate

In this report, we explain the acquisition lifecycle, the scope of the problem, and our novel approach to achieving a more rigorous estimate of costs for DoD acquisition programs.

2 The Problem with Early Cost Estimation

2.1 Department of Defense Acquisition Lifecycle

The Defense Acquisition System is the management process for all DoD acquisition programs. The system is an event-based process: acquisition programs proceed through a series of milestone reviews and other decision points that may authorize entry into a significant new program phase. Acquisition categories are used as part of the process, and programs of increasing dollar value and management interest are subject to increasing levels of oversight. The most expensive programs are known as Major Defense Acquisition Programs (MDAPs) or Major Automated Information System (MAIS). These two program categories have the most extensive statutory and regulatory reporting requirements.

An overview of the DoD acquisition lifecycle is depicted in Figure 1 [DAU 2011].

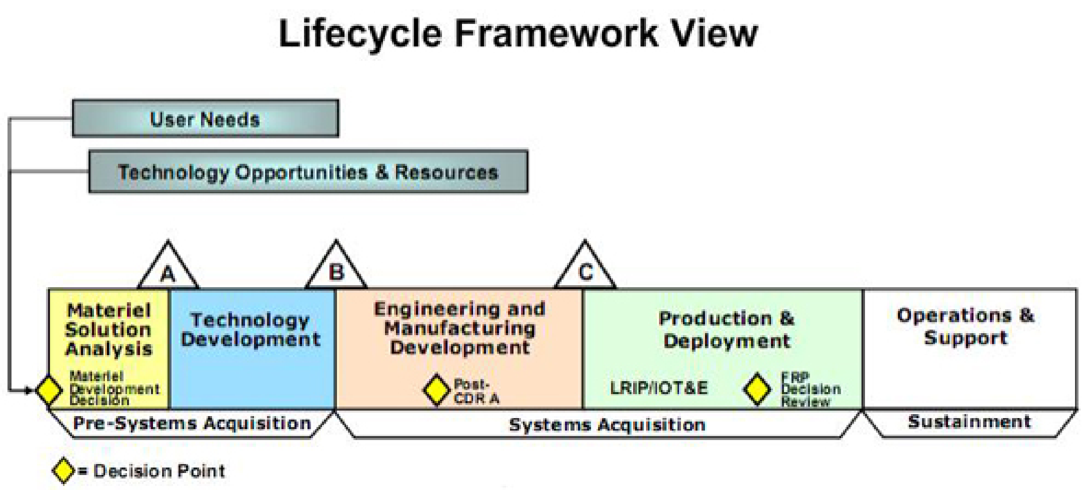


Figure 1 Acquisition Lifecycle

Significant program milestones are shown by triangles A, B, and C in the above diagram.

MDAP and MAIS acquisition programs start with an Initial Capabilities Document (ICD), which seeks to lay out desired capabilities related to specific mission-oriented needs and summarizes the Capabilities-Based Assessment (CBA), a process for assessing capabilities and user needs. The document also identifies gaps in existing capabilities and requires an analysis of doctrine, organization, training, materiel, leadership and education, personnel, and facilities.

If a materiel need is identified, the acquisition process continues with a Materiel Solution Analysis (MSA) phase. During this phase, an analysis of alternatives is undertaken to assess potential materiel solutions to the previously defined capability need. Key technologies are identified and lifecycle costs are estimated, considering commercial-off-the-shelf and custom solutions from both large and small businesses. At the end of the analysis, at Milestone A, a materiel solution to a capability need has been identified, and a Technology Development Strategy (TDS) has been completed.

The TDS assesses industrial and manufacturing capability for the desired materiel solution, and addresses the following:

- specific cost, schedule, and performance goals, including exit criteria, for the Technology Development Phase
- cost and production schedule estimates to support management reviews
- production feasibility and cost and schedule impact analyses to support tradeoffs among alternatives
- available manufacturing processes and techniques
- design producibility risks
- probability of meeting delivery dates
- availability of critical and long-lead time materials
- production equipment availability
- realistic production unit cost goal
- recommendations for planned production testing and demonstration efforts
- methods for conserving critical and strategic materials and mitigating supply disruption risks and program impacts
- a preliminary acquisition strategy, including overall cost, schedule, and performance goals for the total research and development program

The TDS also includes a discussion of key assumptions and variables, and sensitivity to changes in these.

The Milestone A decision is made prior to development of the requirements and design work that is undertaken during the Technology Development Phase, between Milestone A and B. Prior to a 2005 policy change, the cost analysis prepared for Milestone A was limited to the cost of activities between Milestone A and Milestone B. More recently, however, the focus has shifted to an early (pre-Milestone A) need for estimates regarding the entire program lifecycle, including operations and support. MDAP lifecycles usually last for decades.

2.2 The Size of the Problem

Uncertainty in DoD program development causes enormous cost overruns, significant schedule delays, and compromises technical proficiency that seriously affects the DoD's ability to plan for the future in a flexible, responsive, and cost-effective manner. Department of Defense studies and the Government Accountability Office (GAO) have frequently cited poor cost estimation as one of the reasons for cost overrun problems in acquisition programs. Software is often a major culprit. One study by the Naval Postgraduate School found a 34 percent median value increase in software cost over the estimate [Dixon 2007]. The DoD Performance Assessments and Root Cause Analyses (PARCA) office studied ten acquisition programs with serious cost/schedule overruns in 2009-2010 and found that six were caused by unrealistic cost/schedule estimates [Bliss 2011]. Cost overruns lead to onerous congressional scrutiny, and an overrun in one program often leads to depletion of funds from others. Better cost estimates cannot make programs less expensive, but can reduce the size of cost overruns where cost growth is a function of the estimate's accuracy. Table 1 illustrates the growing disparity between early MDAP estimates and actual program performance [GAO 2008a].

Table 1 Cost Overruns in MDAP Portfolios

Analysis of DOD Major Defense Acquisition Program Portfolios

Fiscal year 2008 dollars	Fiscal year		
	2000 portfolio	**2005 portfolio**	**2007 portfolio**
Portfolio size			
Number of programs	75	91	95
Total planned commitments	$790 Billion	$1.5 Trillion	$1.6 Trillion
Commitments outstanding	$380 Billion	$887 Billion	$858 Billion
Portfolio performance			
Change to total RDT&E costs from first estimate	27 percent	33 percent	40 percent
Change in total acquisition cost from first estimate	6 percent	18 percent	26 percent
Estimated total acquisition cost growth	$42 Billion	$202 Billion	$295 Billion
Share of programs with 25 percent or more increase in program acquisition unit cost	37 percent	44 percent	44 percent
Average schedule delay in delivering initial capabilities	16 months	17 months	21 months

Source: GAO analysis of DOD data.

2.3 The Source of the Problem

A cost estimate is always developed from a definition of the scope of work to be performed. The scope defines what, where, and how many products and services will be delivered and to whom. The estimate will be calculated by making some set of historical comparisons. Usually estimators attempt to judge some "size" and "type" relationship as proportional to the work effort. Thus, a home builder can provide a preliminary cost estimate based on usable area, number of rooms, and some basic quality factors such as frame or brick.

The earliest stage of product development work determines the potential value of solving a problem, with little understanding of the cost to develop the solution. In the business world, costing a proposed solution involves estimating marketability and profitability, but in the DoD the driving concept is capability. For major programs in the DoD, the Joint Requirements Oversight Council (JROC) issues a Materiel Development Decision for the conceptual development of a solution to achieve a capability. The Materiel Solution Analysis Phase of the acquisition lifecycle is initiated with the commencement of various studies (discussed in more detail below). A successful outcome is the authorization to issue a Request for Proposal (RFP) for specifying and prototyping the desired product solution. A detailed estimate for prototype development cost and total lifecycle cost of the product, along with an Independent Cost Estimate (ICE), is required for Milestone A Certification. It is the preparation of these estimates that drives our current research.

A chart of the DoD Acquisition lifecycle with fully interactive guidance is available at https://ilc.dau.mil/. The Materiel Solution Analysis Phase is the stage of work preceding Milestone A.

A wealth of information is generated during the MSA phase, otherwise known as pre-Milestone A. An Analysis of Alternatives (AOA) identifies potential technologies and compares costs. Capability-Based Assessments (CBAs) determine technical performance needs in operational contexts. A Technology Development Strategy (TDS) details a plan to proceed from research to production, deployment, and sustainment.

Encompassing all this information and more, the proposed Materiel Solution essentially lays out a plan and high level requirements for implementing an idea to achieve specified capabilities, along with the estimated costs. However, all estimates will contain numerous assumptions about growth and uncertainty. When submitted for approval, the Independent Cost Estimate (ICE) can differ greatly from the Program Office estimate due to differences in these assumptions. For MDAPs, the ICE is performed by the Cost Assessment and Program Evaluation Office (CAPE). We have seen differences as large as an order of magnitude. As shown in Figure 2 [Feickert 2011], this can lead to rework that may take up to a year.

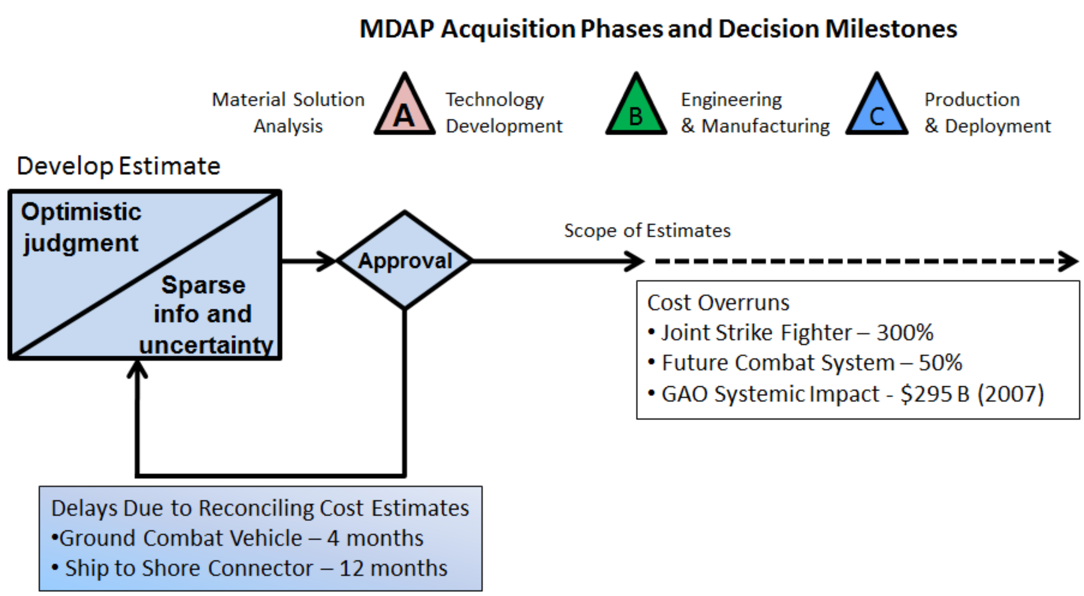
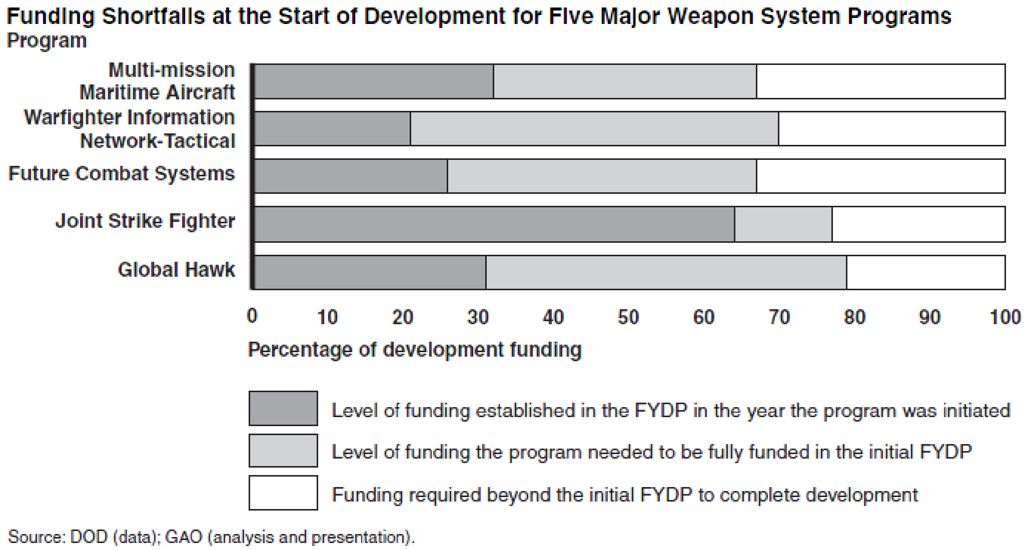


Figure 2 Major Phases and Decision Milestones for MDAPs

The early estimates made prior to product systems engineering and requirements work pose a number of problems for estimators, and poor estimates are known to be one of the main causes of cost growth and program breach [Hofbauer 2011]. In the last year, the Performance Assessments and Root Cause Analysis (PARCA) office investigated the reasons for six Nunn-McCurdy breaches and an additional four MDAPs with problems. They reported that six of these ten cases used unrealistic cost/schedule estimates [Bliss 2011]. The early estimate at Milestone A forms the basis of the plan that defines the program cost and schedule commitments moving forward. Inaccuracy in the estimate also affects the DoD funding process, which plans expenditures for years in advance. The resulting shortfalls in funding cause program instability in the form of reduced capabilities, schedule delays, and reduced procurement quantities as funds are shifted from other programs. The following table illustrates the severe underfunding of just five MDAPs as of 2008 [GAO-2008b].

Table 2 Cost Overruns in DoD Acquisitions



That same GAO report to Congress states: "DoD's flawed funding process is largely driven by decision makers' willingness to accept unrealistic cost estimates and DoD's commitment to more programs than it can support. DoD often underestimates development costs—due in part to a lack of knowledge and optimistic assumptions about requirements and critical technologies." Faced with investment decisions based on needed capabilities, problems encountered in creating estimates at this early stage are described below.

Limited Input Data: Very few requirements are documented. The required system performance, the maturity of the technology for the solution, and the capability of the vendors is not fully understood at the time of the estimate.

Uncertainty in Analogy-Based Estimates: Most early estimates are based on making analogies to existing products, and a properly documented analogy can provide useful data for the estimate. Many factors may be similar, particularly those relating to functionality and product scope. In addition to product description, measures of usage and physical size of the existing system may provide additional connection with development costs and schedule data. Technology, however, will be different: functionality will be added and new performance characteristics will be required. Software product size depends heavily on the implementation technology, and the technology heavily influences development productivity.

Expert Judgment Challenges: The DoD cost estimation community, and the domain experts who support them, leverage a vast array of knowledge and experience to produce and review cost estimates. The end results, of necessity, rely heavily on expert judgment. Given the uncertainties in predicting program performance years in advance, wide variation in judgment can exist between experts. Indeed, an individual expert's judgment can vary over time. Methods exist to sharpen the consistency and precision of such judgments, which we believe would prove very beneficial to the estimation process.

Methods Compound the Uncertainty: Methods for estimating require the repeated use of select data at multiple stages of the estimate. The uncertainty in the inputs then makes the estimate even less trustworthy. For example, the same information about product or project complexity may be used more than one time during the development of the estimate. As a result, any error in an input has a larger impact on the resulting estimate. Lack of transparency in the assumptions further compounds the problem.

Unknown Technology Readiness: Technology readiness may be over- or under-estimated. The contractor in charge of the product development work may not be familiar with the use of the selected technology. Thus, even if the technology has been demonstrated elsewhere, the contractor may require significant time to change internal processes and capabilities.

3 The QUELCE Method—A Proposed Solution

3.1 Overview

As explained in Section 2, DoD cost estimates do not make explicit all assumptions that may impact cost when forecasting several years into the future. They also do not account for the possibility and/or probability of change in numerous program-dependent variables that affect cost ("program change drivers") and the resulting magnitude of change that may be encountered. The QUELCE method thus

- makes use of available information not normally employed for program cost estimation
- provides an explicit, quantified consideration of the uncertainty of the program change drivers
- enables calculation (and re-calculation) of the cost impacts caused by changes that may occur during the program lifecycle
- enhances decision-making through the transparency of the assumptions going into the cost estimate

Figure 3 depicts the flow of information in the typical MDAP Acquisition process. Our approach provides a basis to identify, discuss, and assess the uncertainty of a diverse set of program change drivers that may be known prior to Milestone A. We require interaction with program domain experts due to the heavy reliance on their judgment during the Materiel Solution and Analysis phase, as depicted in Figure 3. The blue boxes represent the contributions from our approach.

A more detailed explanation of the specific steps in the QUELCE method is presented in the following sections.

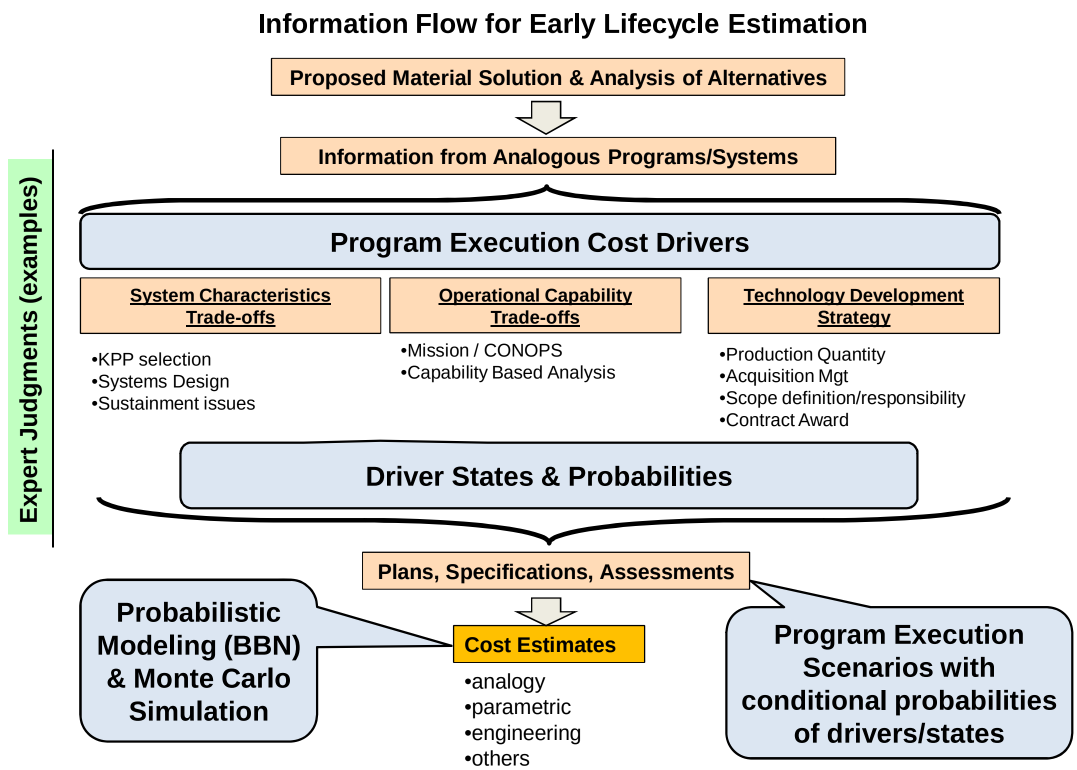


Figure 3 The Role of Expert Judgment in the MSA Phase

3.2 Steps in the QUELCE Method

The QUELCE method consists of the following steps in order:

1. Identify program change drivers: workshop and brainstorm by experts.
2. Identify states of program change drivers.
3. Identify cause-and-effect relationships between program change drivers, represented as a dependency matrix.
4. Reduce the dependency matrix to a feasible number of drivers for analysis, using the Design Structure Matrix method.
5. Construct a BBN using the reduced dependency matrix.
6. Populate BBN nodes with conditional probabilities.
7. Define scenarios representing nominal and alternative program execution futures by altering one or more program change driver probabilities.
8. Select a cost estimation tool and/or Cost Estimating Relationships (CERs) for generating the cost estimate.
9. Obtain program estimates of size and/or other cost inputs that will not be computed by the BBN.
10. For each selected scenario map BBN outputs to the input parameters for the cost estimation model and run a Monte Carlo simulation.

11. Report each scenario result independently for comparison to the program plan.

Steps 1 through 3 are conducted in a workshop setting. Program domain experts identify program change drivers, such as changes in mission, program stakeholders, or supplier relations. Each program change driver has an assumed, *nominal* state, which is identified. Experts then brainstorm about possible changes in the condition of each driver that may occur during the program lifecycle (see Table 3). Once these changed conditions, referred to as potential driver states, are fully identified, workshop participants then subjectively evaluate the cause and effect relationships among the drivers. Expert judgment is applied to rank the causal effects (see Figure 8).

Step 4 uses the Design Structure Matrix technique to reduce the number of drivers to those which comprise most of the potential impacts to cost. The technique is a well established method to reduce complicated dependency structures to a manageable size. In our case, this reduction facilitates the building of a Bayesian Belief Network.

Step 5 is the construction of a BBN using the program change drivers derived from Step 4 and their cause and effect relationships established in Step 3. The BBN is a probabilistic model that dynamically represents the drivers and their relationships as envisioned by the program domain experts. Figure 4 depicts an abbreviated visualization of a BBN, in which the circled nodes represent program change drivers and the arrows represent either cause and effect relationships or leading indicator relationships. In this example, one can see that a change in the Mission and CONOPS driver most likely will cause a change to the Capability Analysis driver, which in turn will likely effect a change in the Key Performance Parameter (KPP) driver and subsequently the Technical Challenge outcome factor. The three outcome factors (Product Challenge, Project Challenge, and Size Growth) are then used to predict some of the input values for traditional cost estimation models.

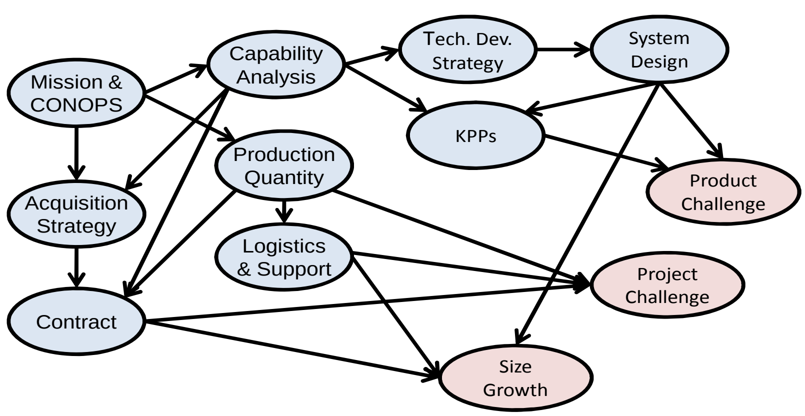


Figure 4 Example Bayesian Belief Network

In **Step 6** we assign conditional probabilities to the nodes (drivers) in the BBN (see Section 3.6.1, Populating Relationships Within a Bayesian Belief Network). Each node can assume a variety of states, each of which has an associated likelihood identified by the domain experts. This allows us to calculate outcome distributions on the variables: Technical Challenge, Project Challenge, and Size/Scope.

Step 7 requires the domain experts to use the BBN in defining scenarios. That is, we can specify the realization of a potential state in a particular node and recalculate the cascading impacts to other nodes and the resulting change in the outcome variables. Any change in one or more nodes (drivers) constitutes a scenario. Once the experts are satisfied that a sufficient number of scenarios are specified, we then solicit their judgment to rank them for likely impacts to cost.

In **Step 8** a decision is made as to which cost estimating tool(s), CERs, and/or other methods will be used to form the cost estimate. In our current research, we have developed the relationships between BBN-modeled program change drivers and COCOMO. We are exploring use of the commercial SEER cost estimating tool with its creator.

In **Step 9** we use the Program Office estimates of size and/or other cost inputs as the starting point, which we will adjust by applying the distributions calculated by the BBN. Often these values are estimated by analogy and aggregation.

In **Step 10** outcomes from each selected scenario (Step 7) are used to parameterize a Monte Carlo simulation. Using the information from Step 9, this provides probability distributions for adjusting the input factors to the cost estimating models. This also provides explicit confidence levels for the results.

We report the final cost estimates for each scenario in **Step 11**, including the nominal program plan. The explicit confidence levels and the visibility of all considered program change drivers allows for quick comparisons and future re-calculations. The transparency afforded by the consideration of alternative scenarios enables improved decision making and contingency planning.

3.3 Our Approach—A Focus on the Importance of Experts

The QUELCE approach originated in the context of current cost estimation practice and research. The DoD estimation process requires at least two independently prepared estimates. Typically, for an MDAP, one is prepared by the nascent program office, one is prepared by the Service's own cost experts,[1] and one is prepared by the CAPE. Since these estimates are prepared independently, their final cost totals may vary by a factor of 10 or more. Since such large discrepancies are very difficult to reconcile, the milestone approval may be delayed—sometimes by as much as several months.

Cost estimators for DoD MDAPs are expert, well-trained, and highly skilled. Provided with high-quality input data, they produce estimates that can reasonably be applied to program plans and budgets. As we mentioned, the data that is available at Milestone A is not similar to the data usu-

1 Naval Center for Cost Analysis (NCCA), Air Force Cost Analysis Agency (AFCAA), Office of the Deputy Assistant Secretary of the Army for Cost and Economics (ODASA-CE)

ally used for Milestone B estimates (at the beginning of Engineering and Manufacturing Development Phase), when better tools and better data about the technology are available. At Milestone A, however, the information is quite vague. It misses most of the technical specification; the technical performance measures and productivity data about the contractor must be assumed.

The objectives of our method include

- Make effective use of existing tools and estimation skills.
- Enhance the estimators' understanding of the potential for program change.
- Forecast the frequency and effects of program change.
- Document assumptions and change possibilities as clearly as possible.

Successful outcomes would include

- fewer and less severe program cost overruns
- faster reconciliation between the program, service, and CAPE estimates
- faster decisions when program change events occur later in the lifecycle

Part of closing the gap in different estimates depends on experts making similar judgments about "size" factors in their analogies, and variability in the potential range of input and efforts.

Within our method, expert judgment plays a vital role at several points, including

- in the identification of significant program change drivers
- in the consideration of various states and the probability of their occurrence within each program change driver
- in the estimation of the probability of one program change driver influencing the state or magnitude of another program change driver
- in providing estimates of the joint probabilities of a change in state of a program change driver state resulting from the joint change of other drivers

We conducted research on methods of improving the accuracy of expert judgment so that it reflects the level of knowledge of the expert. We refer to this concept (a judgment accurately reflecting expert knowledge) as the degree of "calibration" of the expert. Expert calibration is discussed further in Section 4.2.

Our research into enhanced expert judgment via calibration is distinguished on two dimensions: 1) DoD domain specificity, and 2) transparency.

Few research efforts venture beyond generic knowledge into specific domains, and we have found no evidence of calibration techniques applied to the DoD acquisition process. We believe that the most effective calibration of expert judgment may occur when we introduce DoD domain-specific cost estimation materials to help "anchor"[2] expert judgment as described further in Section 4.4.

---

[2] By "anchor" we refer in this report to pertinent factual information on which experts base their judgments. Well calibrated individuals commonly consider several such anchors before making their best judgments. The term "anchor" is sometimes used elsewhere to refer to people's tendency to rush to judgment based on limited information, where they fail to adjust their initial judgments when faced with other information [Meyer 2001].

From a transparency standpoint, we believe that such research and training will dramatically increase the transparency of the basis of early DoD cost estimates.

3.4 Program Change Drivers

Much of the uncertainty in estimating MDAP costs prior to Milestone A arises from the limited information used to construct a cost estimate. We worked with DoD contractors, domain experts, and former DoD program managers in workshops to address how potential program change drivers can affect program costs. Our approach seeks to identify and quantify such drivers so that probable scenarios can be constructed that result in the calculation of probability distributions to be incorporated in modeling the program cost estimate. The identification of program change drivers is best accomplished by the experts who provide programs with the information to consider for cost estimation. Instead of limiting their consideration to the direct inputs needed for any given cost estimate, we ask them to consider aspects of a program that might change (and affect cost) during the program's lifecycle—particularly given the new information developed during the Technology Development Phase in preparation of Milestone B. To initiate the workshop discussion, we chose to use the Probability of Program Success (POPS) factors currently in use by the Navy and Air Force. The POPS criteria are used to evaluate program readiness to proceed and interpose review *gates* on the DoD acquisition process, as represented by the circles in Figure 5.

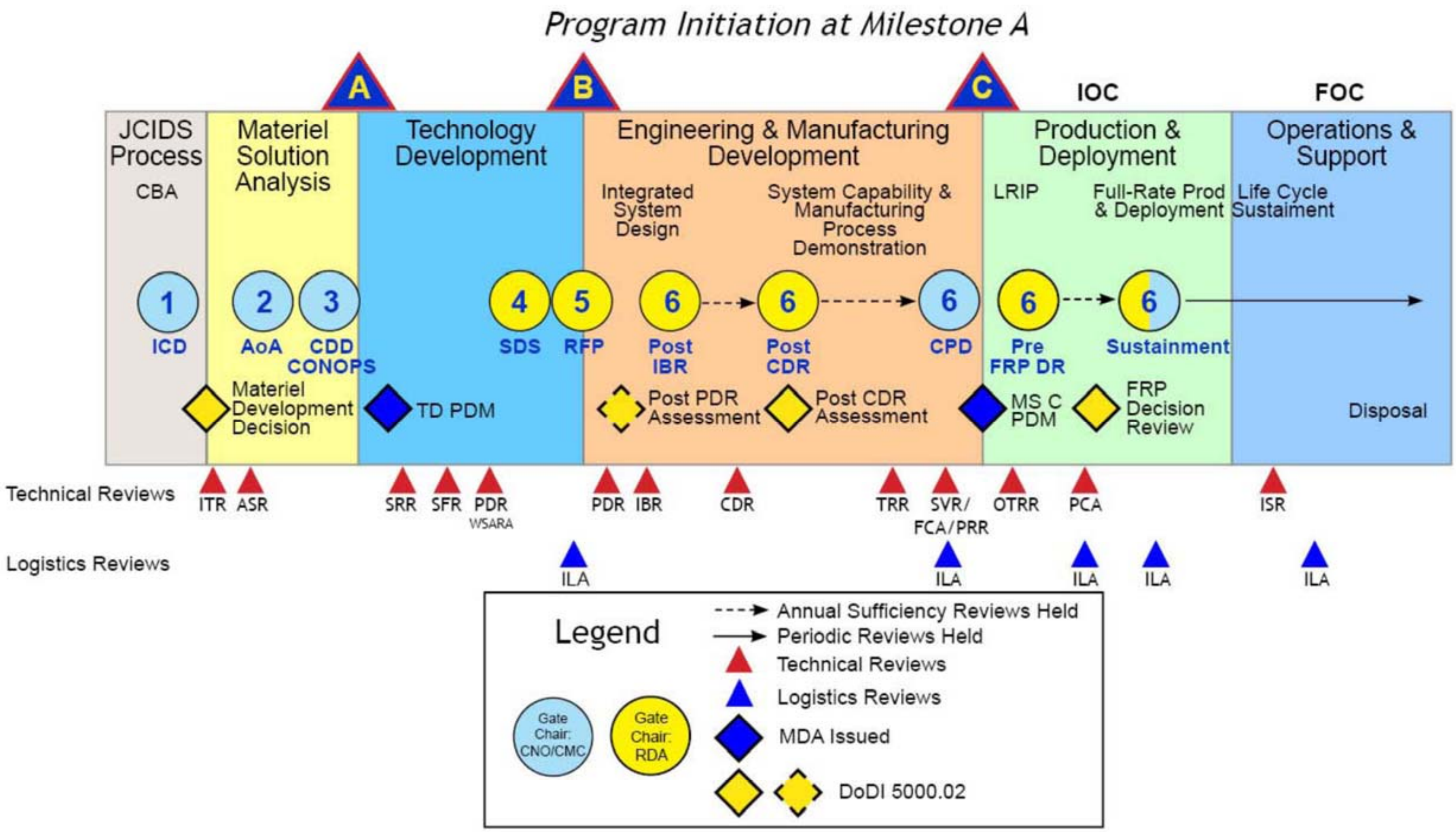


Figure 5 Naval POPs Gate Reviews in the Acquisition Lifecycle

As shown in Figure 5, there are three POPS review gates that take place during the Materiel Solution Analysis phase and before the Milestone A review. Information generated in the MSA is evaluated during the gate reviews according to specified criteria and metrics that are grouped in the following categories:

Program Requirements

- Parameter Status
- Budget and Planning
- CONOPS

Program Resources

- Scope Evolution
- Manning

Program Planning/Execution

- Acquisition Management
- Industry/Company Assessment
- Total Ownership Cost Estimating
- Test and Evaluation
- Technical maturity
- Sustainment
- Software
- Contract Planning and Execution
- Government Program Office Performance
- Technology Production

External Influencers

- Fit in Vision
- Program Advocacy
- Interdependencies

Each gate review has specific criteria which must be met by the program to gain Service approval to proceed, in addition to the DoD Acquisition requirements. In particular, gates 1, 2, and 3 focus on the conceptual requirements. Gate 1 includes the Service review of the Initial Capabilities Document (ICD) and the Analysis of Alternatives (AoA) guidance. Approval is issued to proceed into the MSA phase. Gate 2 concentrates on evaluating all the information generated for the AoA, including lifecycle cost forecasts for all options. Milestone A documentation and a preliminary Technology Readiness Level (TRL) assessment are also reviewed. Gate 3 is the final Service approval required to apply for Milestone A approval. The program manager’s cost estimate is compared to the initial Independent Cost Estimate (ICE). The draft Capability Development Document (CDD) and the Concept of operations (CONOPS) are approved, along with the System Design Specification (SDS) development plan. Similar reviews and documentation for MDAPs occur in all the Services.

The wealth of information required for MDAPs often depends on the contributions of domain experts. However, much of the information generated and required by the pre-Milestone A analyses is not used in the cost estimation process, even though it could potentially enlighten and improve the process, and increase the accuracy of the estimate. In our approach, we used the above groupings at our workshops, as a starter set of concepts to generate ideas by the experts regarding potential program changes that might alter the expected program development and cost. As the workshop proceeds, other program change drivers invariably are identified and added to the list. We used these program change drivers to build a Dependency Matrix, as shown in Figure 6.

The experts are also asked to brainstorm ideas about the status of each program change driver. The specific, assumed state as proposed by the Materiel Solution is labeled as the *nominal state*. We ask the experts to identify possible changes that might occur to the nominal state, and use their best judgment on the probability that the nominal state will change as shown in Table 3.

Cause \ Effect	Scope Responsibility	Scope Definition	Mission / CONOPS	Capability Definition	Funding Schedule	Prog Mgt Structure	Manning at program office	Systems Design	Standards/Certifications	Acquisition Management	Program Mgt - Contractor Relations	Project Social / Dev Env	Supply Chain Vulnerabilities	Information sharing	PO Process Performance	Sustainment Issues	Contract Award	Contractor Performance	Production Quantity	Data Ownership	Change in Strategic Vision	Advocacy Change	Industry Company Assessment	Cost Estimate	Test & Evaluation	Closing Technical Gaps (CBA)	Building Technical Capability & Capacity	Functional Measures	Functional Solution Criteria (measure)	Interdependency	Interoperability
Scope Responsibility																															
Scope Definition																															
Mission / CONOPS																															
Capability Definition																															
Funding Schedule																															
Prog Mgt Structure																															
Manning at program office																															
Systems Design																															
Standards/Certifications																															
Acquisition Management																															
Program Mgt - Contractor Relations																															
Project Social / Dev Env																															
Supply Chain Vulnerabilities																															
Information sharing																															
PO Process Performance																															
Sustainment Issues																															
Contract Award																															
Contractor Performance																															
Production Quantity																															
Data Ownership																															
Change in Strategic Vision																															
Advocacy Change																															
Industry Company Assessment																															
Cost Estimate																															
Test & Evaluation																															
Closing Technical Gaps (CBA)																															
Building Technical Capability & Capacity																															
Functional Measures																															
Functional Solution Criteria (measure)																															
Interdependency																															

Each cell gets a value (0, 1, 2, or 3) to reflect the perceived cause-effect relationship of the row heading to the column heading)

Note: The sum of a column represents a dependency score for the column header. The sum of a row is the value of driving force of the row header

Figure 6 Program Change Driver Dependency Matrix

Driver	Nominal	State 1	State 2	State 3	State 4	State 5
Scope Definition	Stable	Users added	Additional (foreign) customer	Additional deliverable (e.g. training & manuals)	Production downsized	Scope Reduction (funding reduction)
Mission / CONOPS	Defined	New condition	New mission	New echelon	Program becomes Joint	
Capability Definition	Stable	Addition	Subtraction	Variance	Trade-offs [performance vs. affordability, etc.]	
Funding Schedule	Established	Funding delays tie up resources [e.g. operational test]	FFRDC ceiling issue	Funding change for end of year	Funding spread out	Obligated vs. allocated funds shifted
Advocacy Change	Stable	Joint service program loses participant	Senator did not get re-elected	Change in senior Pentagon staff	Advocate requires change in mission scope	Service owner different than CONOPS users
Closing Technical Gaps (CBA)	Selected trade studies are sufficient	Technology does not achieve satisfactory performance	Technology is too expensive	Selected solution cannot achieve desired outcome	Technology not performing as expected	New technology not testing well

Table 3 Example Program Change Drivers and Potential States During Acquisition Lifecycle

The matrix provides the relationship between nominal and dependent states, and contains the conditional probability that one will affect the other—not the impact of the change. The very large number of program change drivers and states identified for an MDAP can be reduced to an efficient set of drivers that capture the impact on cost, using DSM methods[3] as described below.

3.5 The Design Structure Matrix Technique

In order to reduce the number of possible combinations and obtain the set of drivers with the greatest potential impact on cost, we initially create a square matrix using the names of the drivers as row and column labels (same order in both directions), as shown in Figure 6. The row is the program change driver and the column is the effect. For example, if the cell is designated (Advocacy, Funding), then the cell will contain the conditional probability that an Advocacy change will cause a Funding change. The diagonal will be blank.

We then populate the cells with rating values {blank, 1, 2, 3} denoting the probability that a change in driver A will cause or precede a change in driver B, the values defined as follows:

- Blank: no relationship
- 1: low probability of causing a change (<30%).
- 2: moderate probability of causing a change (30%< change <70%)
- 3: high probability of causing a change (>70%)

Figure 7 shows an example of such a matrix of cause and effect ratings that were formed by domain experts from the SEI Acquisition Support Program (ASP) who participated in a pilot workshop (see Section 5.2).

The next step is to form an upper triangular matrix, which means that it is a directed graph with no cycles (iterated loops). This form is required for the construction of the BBN. The upper triangular matrix in Figure 8 will be the basis for drawing the graph for a BBN that has no cycles.

A transformation is the movement of a row-column pair (to preserve the blank diagonal) and is carried out by hand.[4] If you have followed the steps correctly, the diagonal will again have all the blacked out cells. This is called a “unitary transformation” in matrix algebra.

If a perfectly triangular upper matrix cannot be created, the implication is that the directed graph will contain some number of cycles (A causes B causes C causes A). Cycles cannot be allowed in constructing the BBN [Ben-Gal 2008]. Three strategies can be used to reduce the matrix to upper triangular.

---

3 www.dsmweb.org

4 The following procedure shows how to do this in Excel:

1. Right-click on the row you want to move and select “Cut” from the popup menu.
2. Right-click on the target row below where you would like to move the cut cells and select “Insert the cut cells.”
3. Right-click on the column of the same name and select “Cut.”
4. Right click on the column to the right of where you want to move the cut cells and select “Insert cut cells” from the popup menu.

The first strategy accommodates activities that cannot be separated into component steps. In the workshop matrix below (Figure 8), the drivers Interdependency, Interoperability, and Systems Engineering cannot be ordered into a triangular upper matrix. These three problems must be solved at the same time. They can be treated as a single entity for the estimation process. The situation is not surprising, so we treat them jointly as one thing. Later in the lifecycle we might not make the same decision because the design effects of a change are uniquely identifiable.

The second way to simplify the matrix is to ignore some interactions that have the value “1” (hence low conditional probability) and appear below (and left) of the diagonal. If the value is “1” it is less likely to be selected in a scenario anyway.

Finally, the third method is to add an additional program change driver into the model. The iteration problem, where A causes B causes A, can be turned into separate steps that remove the cyclic behavior. In this case A causes B causes A′, introduces a new step into the development in order to remove the iteration.

Only the salmon-colored cells shown in Figure 6 represent cycles and would be treated as a single driver. All other entries below the diagonal are 1s and will be ignored.

On the basis of the upper triangular matrix, we construct and populate the BBN network with drivers. A list of drivers and their definitions included in our demonstration analysis is found in Appendix B, along with drivers that were eliminated from the analysis.

Change Drivers - Cause & Effects Matrix

Effects / Causes	Scope Responsibility	Scope Definition	Mission / CONOPS	Capability Definition	Funding Schedule	Prog Mgt Structure	Manning at program office	Systems Design	Standards/Certifications	Acquisition Management	Program Mgt - Contractor Relations	Project Social / Dev Env	Supply Chain Vulnerabilities	Information sharing	PO Process Performance	Sustainment Issues	Contract Award	Contractor Performance	Production Quantity	Data Ownership	Change in Strategic Vision	Advocacy Change	Industry Company Assessment	Cost Estimate	Test & Evaluation	Closing Technical Gaps (CBA)	Building Technical Capability & Capacity	Functional Measures	Functional Solution Criteria (measure)	Interdependency	Interoperability	Size	Project Challenge	Product Challenge	Total
Scope Responsibility		2				1	1			1	1	1			1																				8
Scope Definition	3					1	1																												5
Mission / CONOPS	0	3		3																															6
Capability Definition					0	0	0	3	2	2	1	1	2	2	0	1	0	0	2	0															16
Funding Schedule							1			1					2			1																	5
Prog Mgt Structure							2					1		1	2																				6
Manning at program office						2								1	2																				5
Systems Design	1												1	1		1				1			2	2	3	2	1	2	2	2	2				23
Standards/Certifications													1			1	1	1					1		3		1		1						10
Acquisition Management											2	3	1	1	2		2	2		1			1	1				1	1	1				1	20
Program Mgt - Contractor Relations												2		1	1	1		2							1		1	1	1	1	1			2	15
Project Social / Dev Env										1	1			1	2		2	2		1					1		1	1	1	1	1	1	1	1	19
Supply Chain Vulnerabilities					1			1	1	1								1	2																7
Information sharing								1							1	1		1		1						1	1								7
PO Process Performance																		2																2	4
Sustainment Issues																																			0
Contract Award																																			0
Contractor Performance																																		2	2
Production Quantity																																		2	2
Data Ownership																																		2	2
Change in Strategic Vision				3	2									2		2	2		3			3	2	3	2		3			2					29
Advocacy Change	1	2				1	1			1																									6
Industry Company Assessment																																			0
Cost Estimate																																			0
Test & Evaluation																																			0
Closing Technical Gaps (CBA)	1							3	1	1	2	2	2	1	0	2	2	2	1	1	1	2	2	2	1		2	2	2	1	1				37
Building Technical Capability & Capacity (CBA)		1				1				1	2	2	2	3		2	2	1	1	2	0	1	2	1	1	0		1	2	1					29
Functional Measures									1		2	2	1	1	1		1	1		1		1	1		2				2						17
Functional Solution Criteria (measure)									1		2	2						1					1		2					1					10
Interdependency	1	1			1	1	1	2	1		1	1	2	1	2	2		1		1	1	1	1	1	1				1		2	2	2	3	34
Interoperability	1	1						2	1	1	1	1	2	1	1	2		1		1				1	3		1		1	1		2	2	2	29
Size																																			0
Project Challenge																																			0
Product Challenge																																			0
Totals	8	10	0	6	4	7	7	12	8	10	15	18	14	17	17	15	12	19	9	10	2	8	13	11	20	3	11	8	14	11	7	5	5	17	

Figure 7 Dependency Matrix Before Transformation

Change Drivers - Cause & Effects Matrix

Effects / Causes	Mission / CONOPS	Change in Strategic Vision	Capability Definition	Advocacy Change	Closing Technical Gaps (CBA)	Building Technical Capability & Capacity (CBA)	Interoperability	Systems Design	Interdependency	Functional Measures	Scope Definition	Functional Solution Criteria (measure)	Funding Schedule	Acquisition Management	Program Mgt - Contractor Relations	Project Social / Dev Env	Prog Mgt Structure	Manning at program office	Scope Responsibility	Standards/Certifications	Supply Chain Vulnerabilities	Information sharing	PO Process Performance	Sustainment Issues	Contract Award	Production Quantity	Data Ownership	Industry Company Assessment	Cost Estimate	Test & Evaluation	Contractor Performance	Size	Project Challenge	Product Challenge	Total	Number right of diagonal
Mission / CONOPS			3								3								0																6	0
Change in Strategic Vision			3	3		3			2				2									2		2	2	3		2	3	2					29	0
Capability Definition								3					0	2	1	1	0	0		2	2	2	0	1	0	2	0				0				16	0
Advocacy Change											2			1			1	1	1																6	0
Closing Technical Gaps (CBA)						2	1	3	1	2		2		1	2	2			1	1	2	1	0	2	2	1	1	2	2	1	2				34	0
Building Technical Capability & Capacity (CBA)									1	1		2		1	2	2	1				2	3		2	2	1	2	2	1	1	1				27	0
Interoperability						1		2	1		1	1		1	1	1			1	1	2	1	1	2			1		1	3	1	2	2	2	29	1
Systems Design						1	2		2	2		2							1		1	1		1			1	2	2	3					21	3
Interdependency				1			2	2			1	1	1		1	1	1	1	1	1	2	1	2	2			1	1	1	1	1	2	2	3	33	5
Functional Measures												2			2	2				1	1	1	1		1		1	1		2	1				16	0
Scope Definition																	1	1	3																5	0
Functional Solution Criteria (measure)									1						2	2				1								1		2	1				10	1
Funding Schedule														1				1					2								1				5	0
Acquisition Management										1		1			2	3					1	1	2		2		1	1	1		2			1	19	2
Program Mgt - Contractor Relations										1		1				2						1	1	1						1	2			2	12	2
Project Social / Dev Env														1	1							1	2		2		1			1	2	1	1	1	14	2
Prog Mgt Structure																1		2				1	2												6	1
Manning at program office																	2					1	2												5	2
Scope Responsibility														1	1	1	1	1					1												6	5
Standards/Certifications						1						1									1			1	1			1		3	1				10	2
Supply Chain Vulnerabilities								1					1	1						1						2					1				7	4
Information sharing					1	1		1															1	1			1				1				7	3
PO Process Performance																															2			2	4	0
Sustainment Issues																																			0	0
Contract Award																																			0	0
Production Quantity																																		2	2	0
Data Ownership																																		2	2	0
Industry Company Assessment																																			0	0
Cost Estimate																																			0	0
Test & Evaluation																																			0	0
Contractor Performance																																		2	2	0
Size																																			0	0
Project Challenge																																			0	0
Product Challenge																																			0	0
Totals	0	0	6	4	1	9	5	12	8	7	7	13	4	10	15	18	7	7	8	8	14	17	17	15	12	9	10	13	11	20	19	5	5	17		
Below diagonal	0	0	0	1	1	4	4	4	1	2	0	3	1	3	2	2	3	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0		

Figure 8 Dependency Matrix After DSM Transformation

3.6 Bayesian Belief Network (BBN) Modeling

We selected BBNs as a method of probabilistic modeling that offers a basis for quantifying the conditional likelihood of occurrence and relationships among program change drivers. Figure 9 depicts an example fragment of the BBN model for a subset of program change drivers showing the relationships between drivers and three important outcome factors: 1) project challenge, 2) product challenge, and 3) size growth. These outcome factors are used as inputs to a Monte Carlo analysis, which provides probabilistic distributions of input factors to traditional cost estimation tools such as COCOMO or SEER.

In Figure 9, a truncated view of a BBN, program change drivers are represented by circled nodes and are connected to each other and to outcome factors by arrows representing "cause and effect" relationships or "leading indicator" relationships. For example, Figure 9 illustrates that the Interoperability and Interdependency drivers together influence the state of the Project Challenge outcome factor. The Size Growth outcome factor is also forecast by the same two program change drivers. Lastly, the Product Challenge outcome factor is influenced by seven different drivers, four of which are shown in the diagram: Interoperability, Interdependency, Program Management Contractor Relations, and PO Process Performance.

More specifically, each of the three outcome factors are measured on a scale of 1=Very Low, 3=Nominal, 5=Very High (five distinct values). In this report, program change drivers are modeled as binary factors with two possible states: nominal or not. This approach permits BBN modeling of drivers changing state and provides information on the net effect on other program change drivers and the outcome factors.

For example, the Manning at Program Office driver may switch from the nominal state and cause the PO Process Performance driver to change from a nominal state, and thus negatively impact the Product Challenge outcome factor by increasing it from a (hypothetical) value of 2 to a value of 5.

Each of the program change drivers have distributions rather than single point values and are assigned by the domain experts. In our example derived from the workshops, each driver thus possesses a probability of being in the nominal or not nominal state. The BBN calculates outcome factor distributions based on the probability distributions of the program change drivers. For each outcome factor, there are probabilities associated with each of the 1-to-5 scale values that sum to 100%. Consequently, the Product Challenge outcome factor may have a most likely value of 5 and a lower probability of having a value of 4 or 3, to reflect the uncertainty of the actual value of Product Challenge.

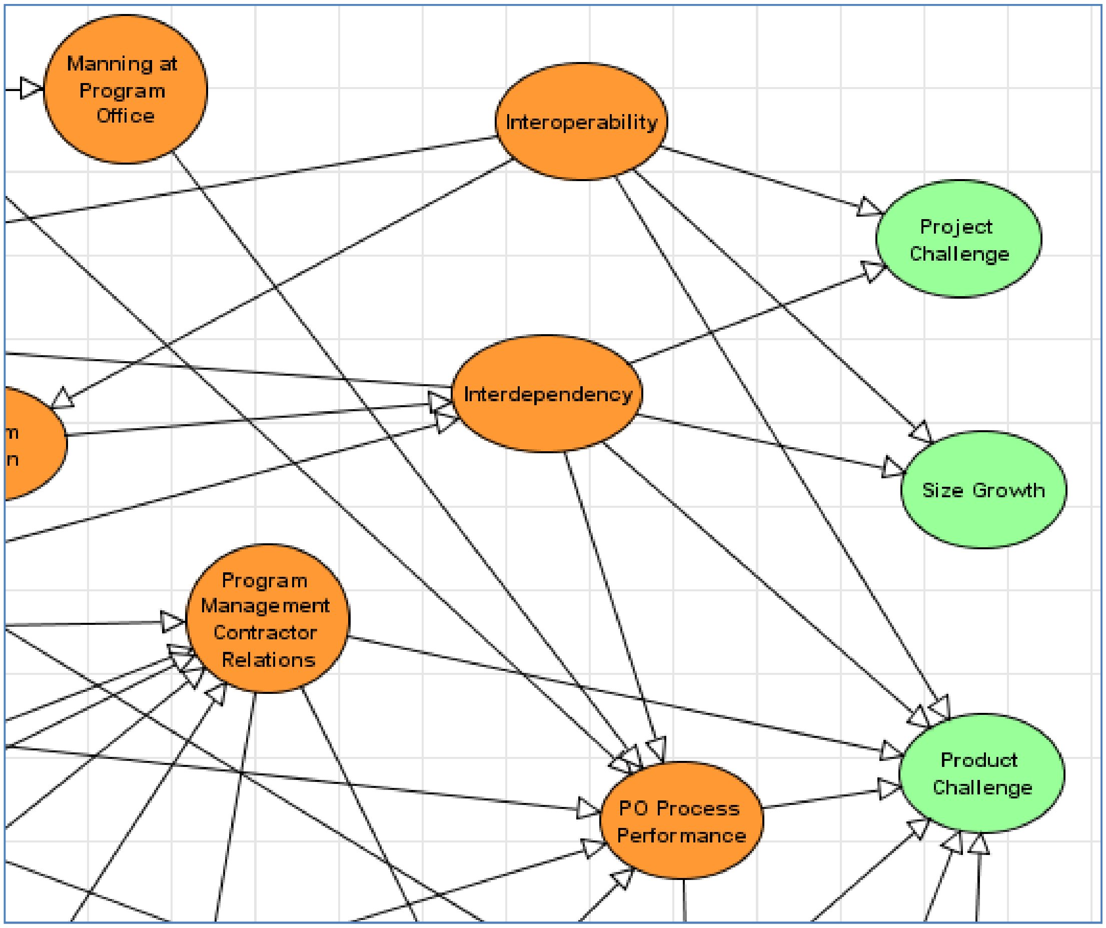


Figure 9 Fragment of a BBN Model

The information provided by the BBN may also shed light on retrospective activities. For example (referring to Figure 9), if there is a high Product Challenge outcome value, the BBN update algorithms will inform the reader as to what degree the PO Process Performance program change driver is responsible as compared to the Program Management Contractor Relations program change driver. If the PO Process Performance driver is more responsible, the BBN update algorithm will also provide information on how much of this is due to changes in drivers arising earlier in the process (e.g., the Manning at Program Office driver or the Interdependency driver).

As noted earlier, BBNs may be populated with objective historical data and informed subjective expert judgment. In our research we are looking at additional statistical methods that can be used to populate a BBN with program change driver information, including correlation studies and predictive modeling techniques (hypothesis testing, statistical and logistic regression analysis, and simulation modeling).

Our research and analysis may demonstrate that some of the program change driver relationships assumed to exist may in fact not exist, while other relationships may be newly ascertained. For example, the relationships shown in Figure 9, derived from subjective expert opinion, may be overturned by empirical analysis that shows different statistical or probability-based relationships of drivers to outcome factors. In this case, non-significant relationships could be dropped from the model.

3.6.1 Populating Relationships Within a Bayesian Belief Network

As previously noted, the DSM matrix produced in the scenario planning workshop provides the significant cause-effect relationships (depicted in the BBN in Figure 9). By modeling only the significant driver cause-effect relationships (e.g., strength of 2 or 3) in the DSM matrix, an overwhelming complexity of driver relationships may be represented in a simplified, manageable BBN.

Figure 10 depicts the resulting BBN for the previously defined DSM matrix, and also shows the state information for each driver node in addition to the three green outcome nodes. Three types of driver nodes exist in this BBN: 1) top-level initiating driver nodes that have no “parent” driver nodes but which have one or more “children” driver nodes, 2) interim driver nodes that have both “parent” and “children” driver nodes, and 3) outcome driver nodes that have no “children” nodes.

For example, in Figure 10, Mission CONOPS is a top-level initiating driver node, Capability Definition is an interim driver node, and Project Challenge is one of three outcome nodes. The outcome nodes are the primary focus of the BBN model in that the model seeks to predict the distributions of the outcome nodes using historical and recent observations of the driver nodes. Once the BBN model produces a prediction of the outcome nodes, the outcome nodes can be used to estimate one or more of the input factors of the CER functions within cost estimation models.

The probabilities for each driver node in the model are also shown in Figure 10 and represent the historical probabilities of the states within each driver node. For driver nodes with no “parent” nodes, the probabilities are directly assessed. For all other nodes, the probabilities are computed as a function of the “parent” node state probabilities and the conditional probabilities assessed by domain experts for the “child” node. For example, the Capability Definition driver node shows a 79% chance of being in a non-nominal state. This probability is calculated across all scenarios of the joint “parent” driver node states of Mission CONOPS and Strategic Vision. Consequently, the probabilities of the nominal versus non-nominal states for all the driver nodes and outcome nodes reflect the probabilities in context of all possible states of all driver nodes (e.g., the probabilities of the nodes in context of all possible scenarios of driver states). In view of this, the distributions shown for the three outcome nodes—Project Challenge, Product Challenge and Size Growth—would be used directly in the next step (the Monte Carlo analysis) to determine the distributions of the input factors to the cost estimation model.

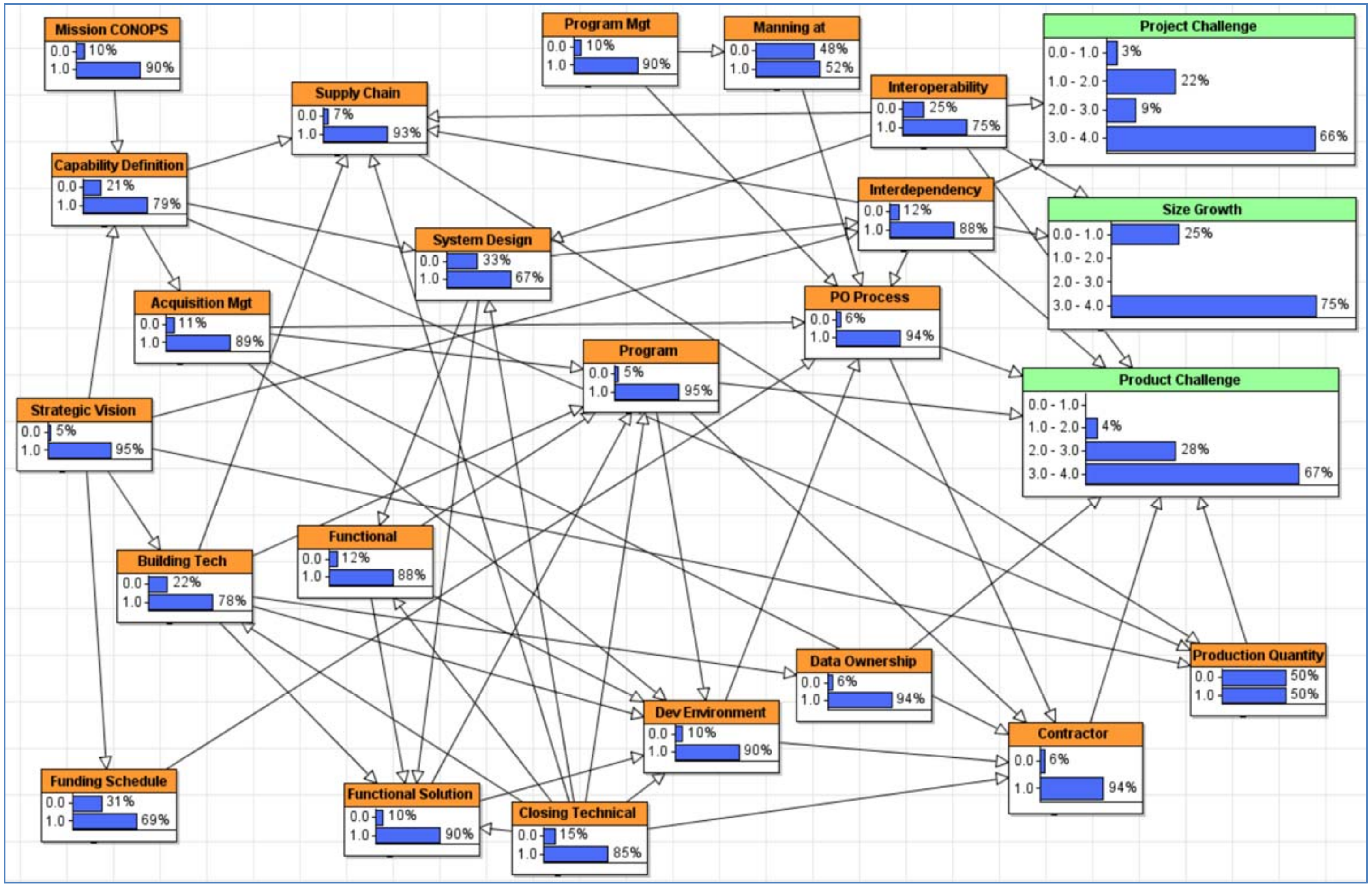


Figure 10 Fully Populated BBN

Figure 11 shows the state probability table for a top-level driver node, Mission CONOPS. Notice that top-level driver nodes merely have a table showing the historical probabilities of each of the possible states of the node. Thus, according to Figure 11, Mission CONOPS historically has a 10% probability of being in a nominal state (0.0) and a 90% probability of being in a non-nominal state (1.0). In our demonstration BBN model, each node can be in one of two states, nominal or not, but more complicated and realistic BBN models can be constructed. A driver node could have n states to match the different n unique states identified by the participants in a scenario planning workshop. In other words, drivers can have more than two states, and different drivers can have varying numbers of states.

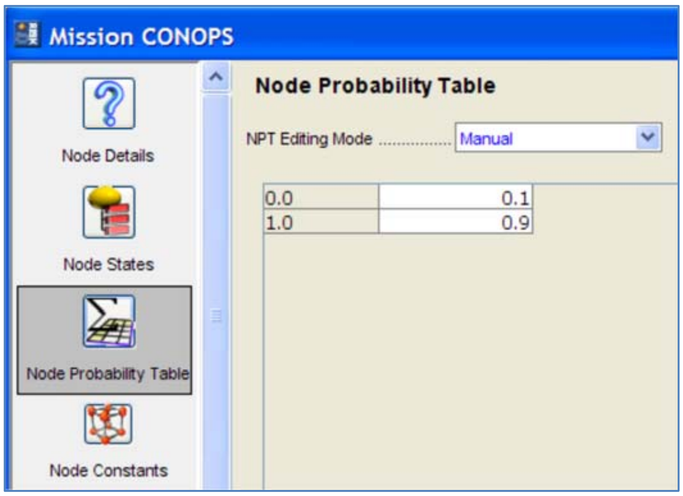


Figure 11 State Probability Table—Top Level Driver Node

Figure 12 shows the state probability table for an interim driver node, Capability Definition. An interim driver node state probability table is more complicated than top level driver node state probability table; it is actually a joint conditional probability table.

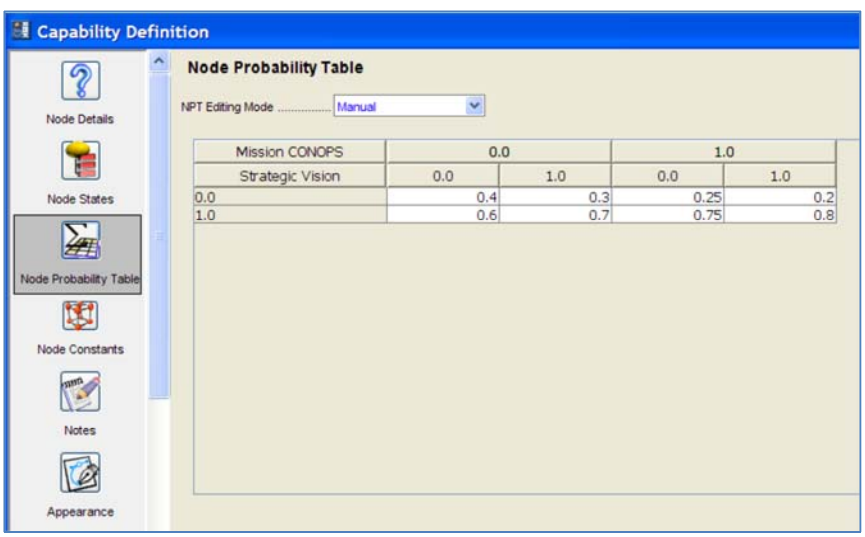


Figure 12 State Probability Table—Interim Driver Node

For example, the Capability Definition driver node is conditional on the joint states of the two "parent" driver nodes: Mission CONOPS and Strategic Vision. Thus, reading the table, when Mission CONOPS and Strategic Vision are both in the nominal state, there is a 40% probability that the Capability Definition driver node will be in a nominal state and a 60% probability that the Capability Definition driver node will be in a non-nominal state. Similarly, when both Mission CONOPS and Strategic Vision driver nodes are in non-nominal states, there is an 80% probability that the Capability Definition driver node will be in a non-nominal state.

The DSM matrix can provide the information needed to create a BBN with significant cause-effect relationships, shown as arrows between the driver nodes. Subsequently, domain experts must provide the joint conditional probabilities for the driver node state probability tables. In this step, the importance of calibrated expert judgment becomes clear. Domain experts must provide reliable and accurate assessments of the joint conditional probabilities to enable a credible and believable BBN model. As such, the BBN model provides a mechanism for the structured use of calibrated expert judgment to predict the outcome nodes needed to estimate the input factors of cost estimation models.

Figure 13 demonstrates an alternative method of populating a joint conditional probability state table for a driver node, Functional Solution Criteria. Notice that Functional Solution Criteria has four parent driver nodes: Closing Technical Gaps, Building Technical Capability and Capacity, System Design, and Functional Measures. With each parent driver node having two possible states, there are now 16 combinations of parent states for which a specification of nominal versus non-nominal must be made for Functional Solution Criteria. Instead of manually populating these 16 joint parent states with probabilities, a mathematical expression can be substituted for the state table. In this case, the state of Functional Solution Criteria is determined by an arithmetic sum giving 40% weight to Functional Measures, 30% weight to System Design, 20% weight to Closing Technical Gaps, and 10% weight to Building Technical Capability and Capacity. Once this sum is calculated, values less than 0.5 are deemed a nominal state for Functional Solution Criteria and values greater than or equal to 0.5 are deemed a non-nominal state.

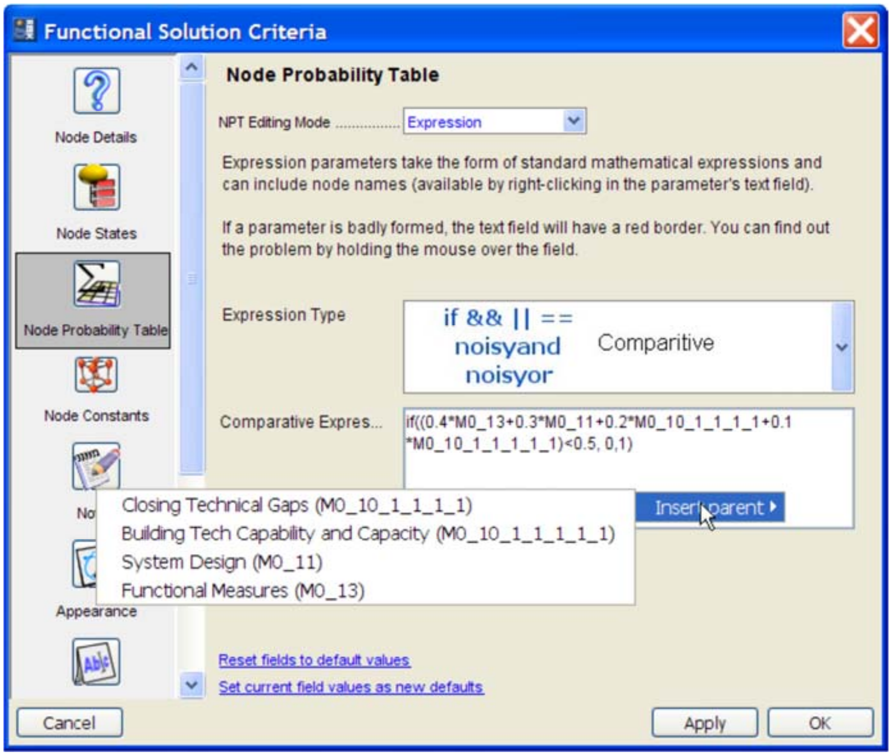


Figure 13 Alternative Method of Populating a Joint Conditional Probability State Table

The three outcome nodes (Project Challenge, Product Challenge, and Size Growth) in this demonstration BBN each have state tables (not shown) produced using similar arithmetic expressions of the "parent" nodes.

3.6.2 Depicting Scenarios Within a BBN

A nominal scenario may therefore be cast as all drivers set to their nominal states. A separate scenario may be cast as a small subset of the drivers, each set to an alternate state.

The Tektronix workshop provided us with the first opportunity to discuss potential future program execution scenarios represented as sets of interrelated program change drivers updated with probabilities for each scenario. Participants worked in large groups to develop the cause-effect matrix for analysis of which drivers influenced other drivers. The group appeared comfortable using a measurement scale of 0=no influence, 1=low probability of influence, 2=moderate probability of influence, and 3=high probability of influence, to describe the probability that a change in one program change driver would cause a change in another driver.

As previously discussed, the use of the graduated scale of probability of influence enabled us to control the explosive growth of the number of scenarios (e.g., combinatorics of associated drivers) by only modeling the scenarios with the strongest influence relationships. Once the exercise to populate the cause-effect matrix was complete, a sanity check of the dominant scenarios proved quite acceptable to workshop participants. Although the Tektronix workshop did not proceed to the step of computing the probabilities of the top ten most likely scenarios, workshop participants did recognize that this remaining step would be straightforward. The final group exercise regarding scenarios involved sharing a documentation template for describing scenarios so that needed context would accompany each scenario, partly to allow sanity testing and agreement among all of the workshop participants. The template shown in

Figure 14 provided a documented thought process of how each scenario may be considered as originating from a nominal context situation, followed by the introduction of a stimulus (external or internal event), and resulting in a response (a change in one or more program change drivers from their nominal states) with a defined outcome (a description of the severity of the effect of the change in the program change drivers).

Figure 14 Template for Scenario Development

During the Tektronix workshop, each subgroup developed one to two unique scenarios from the cause-effect matrix and described each scenario using the template. The structure of the template enabled a focused and concise discussion and rapid agreement among the participants of each scenario. The workshop participants chose to stop at the point of documenting a few scenarios and use the remaining time in the workshop for calibration training and testing of expert judgment. This specific part of the workshop prompted the greatest enthusiasm and participation primarily due to the participants' immediate recognition of the need to calibrate expert judgment in their current project estimation activity.

The BBN model provides a practical and easy method to update cost estimates based on different scenarios that arise later in the acquisition process. A scenario may be thought of as a departure from the baseline BBN, in which the baseline represents all known historical information regarding the program change drivers, their interrelationships, and the subsequent outcome factors. A scenario may be represented as a departure from the baseline BBN in one of two ways:

1. New information, which we will call "hard evidence," may let us conclusively declare one or more program change drivers to be 100% in a nominal or non-nominal state. In this case, the BBN is updated for a given driver to show 100% for a single state and 0% for the other state.
2. New information, which we will call "soft evidence," reflects our latest subjective assessment of the probabilities of the states within a given driver, such that the probabilities are altered from what was originally defined using historical data. For example, the baseline BBN

might show that the driver, such as Program Office Process Performance, has a 50% chance of remaining in a nominal state rather than the 6% chance shown in the baseline BBN. This "soft evidence" would be entered into the Program Office Process Performance BBN node as an observation, which in turn would cause an update throughout the entire BBN of all "un-observed" program change drivers.

With the ability of the BBN to be updated with scenarios based on new observations and knowledge for a given MDAP program execution, the BBN effectively provides a continuing, real-time mechanism to update cost estimates in a transparent manner. Figure 15 illustrates an example of this. In this example, referred to as "Scenario 1," two driver nodes are held in the nominal state. With new "evidence" the BBN shows updated predictions for the three outcome nodes as follows: 1) the Project Challenge is less likely to be at higher values (e.g. the probability of being at level of 4 dropped from 66% to 47%), 2) the Size Growth outcome is reduced (e.g. the probability of a value of 3-4 dropped from 75% to 62%), and 3) the Product Challenge dropped (e.g. the probability of a value of 2 or higher dropped from 95% to 68%). A new cost estimate can be obtained using these latest distributions of the BBN outcome nodes.

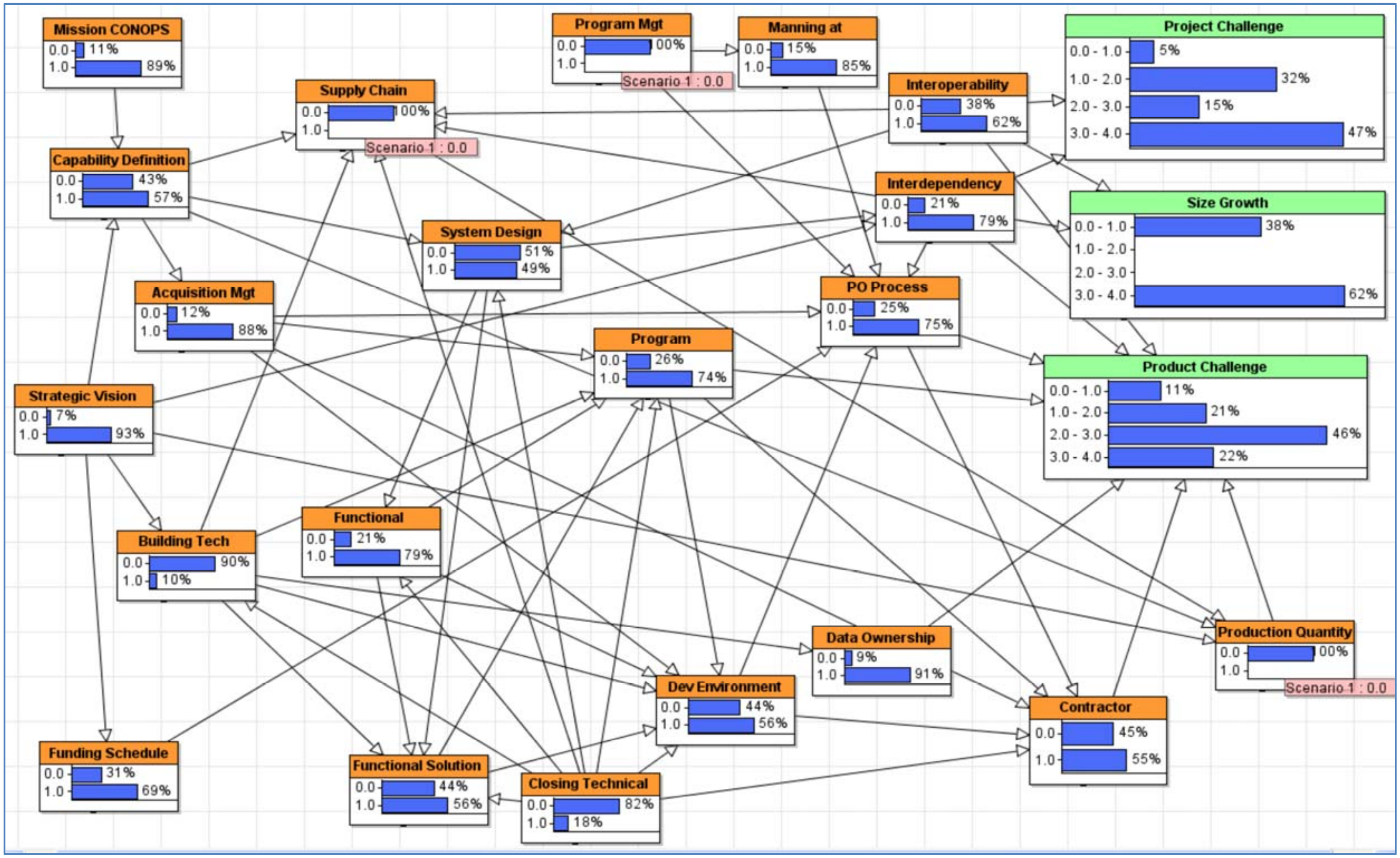


Figure 15 Scenario Of MDAP Actions With Two Driver Nodes In A Nominal State

As new information about drivers becomes available, it can be entered into the BBN model, producing real-time updated predictions of the three BBN outcome nodes. Different scenarios of changes in drivers may be easily modeled and produce updated cost estimates.

The BBN model can be used prior to and after the pre-Milestone A cost estimation to enable stakeholders to ask “what if” types of questions. For example, Figure 16 demonstrates a second scenario analysis (“Scenario 2”) that seeks to understand the cost implications of forcing the following six drivers to remain in their nominal state: Acquisition Management, Program Management Structure, Program Management Contractor Relations, Manning at Program Office, PO Process Performance, and Contractor Performance.

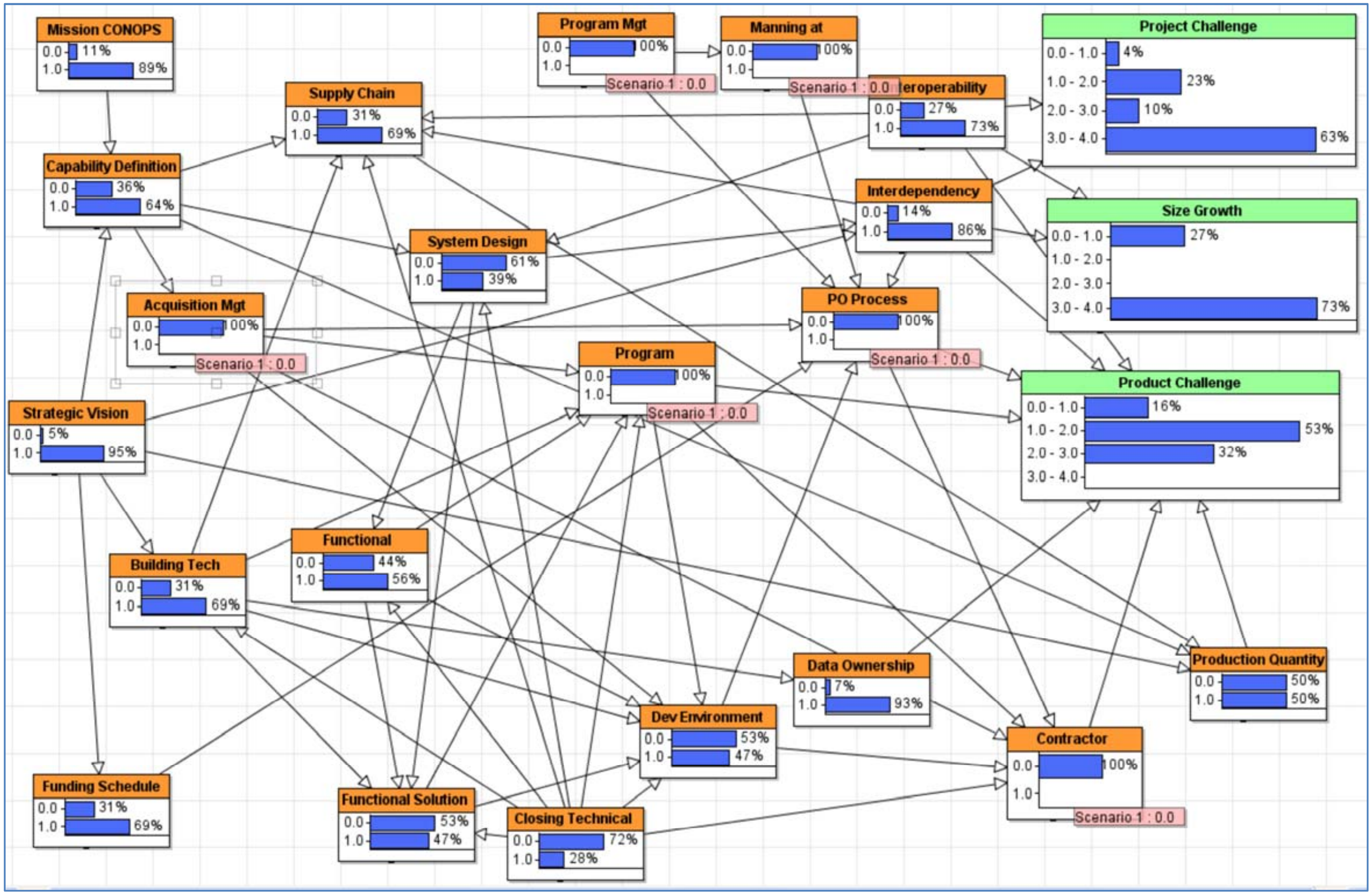


Figure 16 Scenario of MDAP Actions With Six Driver Nodes in a Nominal State

Figure 16 shows that the Project Challenge and Size Growth outcomes remain almost unchanged in distribution, while there is a dramatic change in the Product Challenge outcome node (e.g. for Product Challenge values of 3-4, there is a corresponding drop in probability from 67% to almost 0%). Scenario analyses permit stakeholders to conduct "should-cost" analyses using the BBN model, and plan Program Management Office actions to reduce risk and cost.

Additional analysis made possible by the BBN model includes sensitivity charts that rank drivers in order of most to least influential on a particular outcome node. Figure 17, Figure 18, and Figure 19 are examples of these. In Figure 17, of a ranked list of most influential program change drivers on Project Challenge, the top four drivers are, in order, Interoperability, Interdependency, PO Process Performance, and Supply Chain. Armed with this knowledge, stakeholders and analysts may test the model with their intuition about drivers, and can target specific drivers with the scarce resources of the Program Management Office.

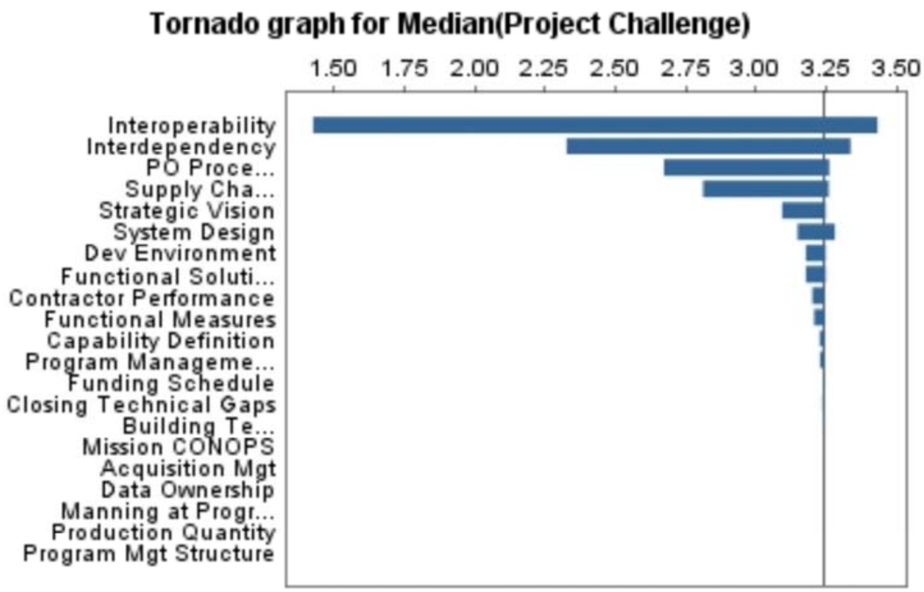


Figure 17 Ranked List of Most Influential Program Change Drivers on Project Challenge

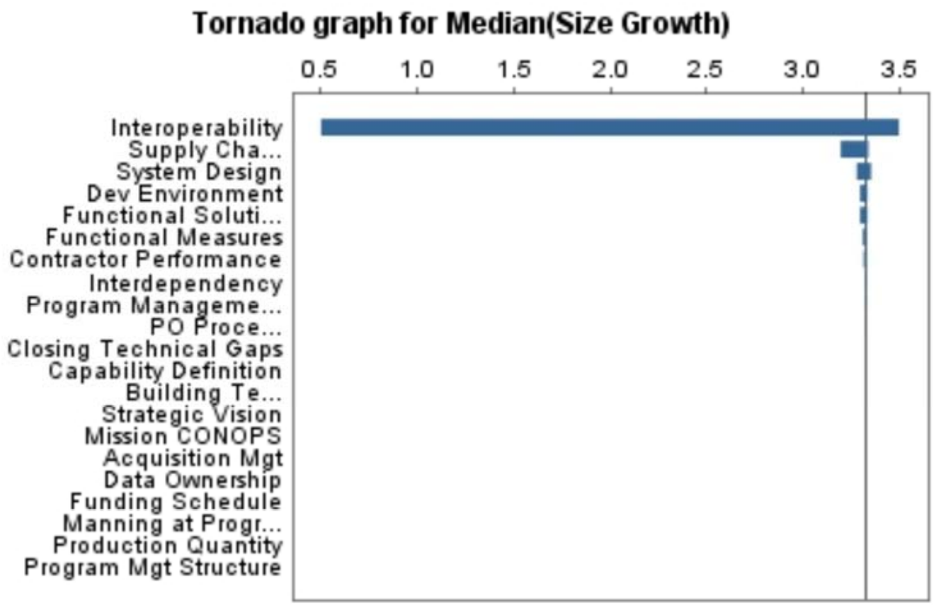


Figure 18 Ranking Drivers for Size Growth

Figure 18 and Figure 19 show the most influential drivers for Size Growth and Product Challenge. The rankings thus provide a way to narrow the number of drivers to study for influence on outputs.

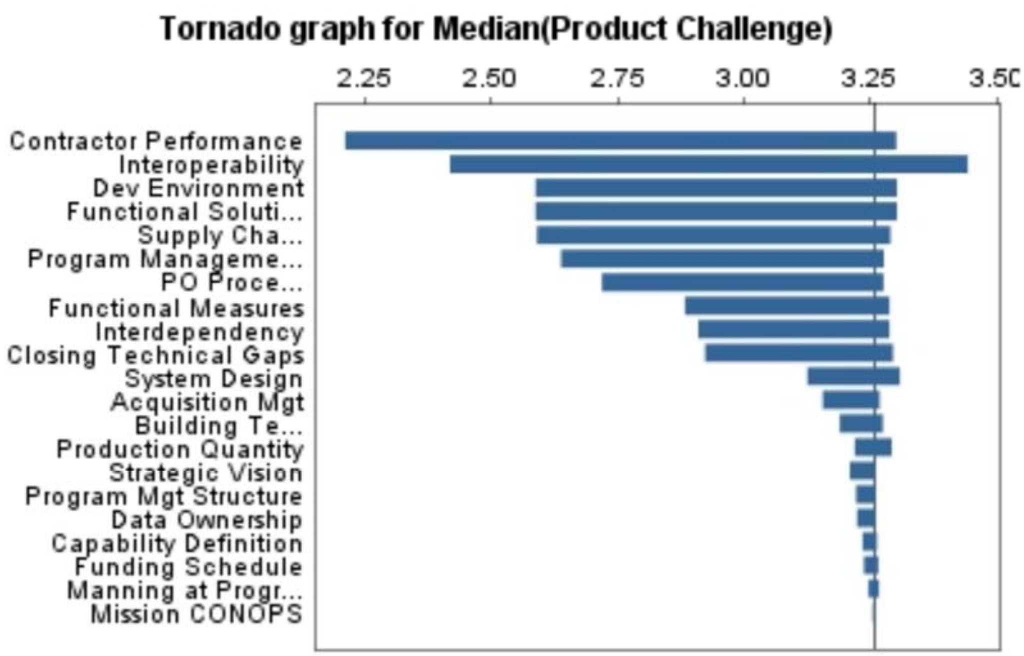


Figure 19 Ranking Drivers for Product Growth

The same analysis can be performed on any interim program change driver node to obtain a picture of what most influences this node. As the BBN is used to model future program execution scenarios in real-time, the sensitivity analysis can explain unmeasured predecessor drivers affect-ing a “downstream” driver. The BBN can also be used as a diagnostic tool to investigate observed changes in driver states.

3.7 Linking the BBN to Existing Cost Estimation Models

Parametric cost estimation models for software use a mathematical equation to calculate effort and schedule from estimates of size and a number of parameters. COCOMO II is a well-known estimation tool and is open source.

COCOMO II uses 22 separate parameters in addition to size. Many of these parameters depend on the development team and its performance, which is unknown at the time the estimate for Milestone A is required. Therefore, our team used the following factors as inputs to the estimation tool. Only the initial “size” is not calculated via the BBN directly.

1. software size
2. size growth/shrinkage range factor
3. project challenge nominal and range (2)
4. product challenge nominal and range (2)

We believe these six factors provide satisfactory coverage and accuracy for the estimate. Still, these factors do not easily match to the 22 COCOMO II factors. COCOMO uses the terms “Effort Multiplier” and “Scale Factor.”

3.8 Mapping BBN Outputs to the COCOMO Inputs

Cost estimation tools capture important cost estimation relationships and have been calibrated on an extensive amount of historical data. For instance, Capers Jones reports in *Applied Software Measurement (3rd edition)* that he has access to data from 13,000+ projects [Jones 2008] .

COCOMO II is a well-known, open source estimation tool.[5] We use it here to demonstrate how we can connect the results from our BBN to an estimation tool. Not every estimation tool is suitable for this approach, but several others can be used in the same manner as demonstrated here.

COCOMO II uses some 13 parameters for pre-architecture cost estimation. These parameters are Size, measured in lines of code, plus five “Scale Factors (SFs)” and seven “Effort Multipliers (EMs).” The main equation of interest here is the effort equation. The base effort equation appears below:

$$PM_{NS} = A \times Size^{E} \times \prod_{i=1}^{n} EM_i$$

$$where\ E = B + 0.01 \times \prod_{j=1}^{5} SF_j$$

5 http://sunset.usc.edu/csse/research/COCOMOII/cocomo_main.html

Table 4 lists the factors in the COCOMO equations.

Table 4 COCOMO Equation Parameters

Symbol	Description
PM	Effort measured in Person-Months
A	A is constant and set at 2.91. It is a productivity factor that can be calibrated to the organization's use and experience.
Size	Source code measured in KSLOC (1000's of source lines of code)
E	The calculation for **E** is shown in the second equation
B	The value of **B** is set at 0.91 based on the current historical data.
EM_i	The $\mathbf{EM_i}$ are the "effort multipliers" from Table xx
SF_j	The $\mathbf{SF_j}$ scale factors are also in Table xx.

The Size parameter must be obtained separately. Frequently it is estimated by using an analogy to some past project along with whatever expert judgment may be available. We use the BBN results from the Project Challenge and Product Challenge to compute the various values for the Effort Multipliers and Scale Factors in the model. In this manner, all of the parameters in the model are established in order to produce an effort estimate.

Detailed explanations of all COCOMO II factors are described in the COCOMO II 2000.0 Model Manual [COCOMO II 2000]. Following are brief definitions of the factors we use in our method along with a statement indicating the relationship to the three primary outputs of the BBN.

The following **Scale Factor** parameters are associated with pre-architecture estimates II:

- **TEAM** describes stakeholder-team cohesion. A program exhibiting discord about requirements and performance criteria will also have low team cohesion, and the program will suffer. A highly cohesive team will be more efficient and will therefore have a lower cost.
- **PMAT** describes the process maturity of the development organization. This is usually measured in CMMI levels. DoD contractors are CMMI Level 2 or above. Low values mean the team is less productive and therefore will increase cost.
- **PREC** describes whether the system is truly new and unprecedented. DoD development projects are usually unprecedented. Low values mean the work is less familiar and hence will increase cost.
- **FLEX** considers whether or not the project has flexibility on requirements. Such things as "safety of flight" tend to limit flexibility in development projects. Low values mean work is more constrained (difficult) and will increase cost.
- **RESL** evaluates whether the project has processes capabilities and schedule to address software architecture and systems engineering concerns. Low values mean risk reduction is less effective and therefore cost will be higher.

Effort Multipliers associated to pre-architecture development are

- **PERS** addresses the capability and continuity of the analysts and developers. Low scores for this factor drive costs up.

- **RCPX** considers the reliability, complexity, data, and documentation challenges of the system. Low scores on this factor reduce cost.
- **PDIF** represents constraints on computer resources utilized, such as performance budgets for CPUs and platform volatility. Low values on this factor indicate less severe or challenging constraints and lower risk of platform volatility. Low values contribute to lower cost.
- **PREX** considers likely personnel experience with tools, language, and platform to be used on the system. Low values indicate less experience and therefore will contribute to higher costs.
- **FCIL** considers the use of tools and multi-site development. Low values indicate less use of automated tools and multi-site development thereby creating a greater challenge to the project and a likely higher cost.
- **RUSE** represents whether the product is required to be reusable. Higher values indicate the additional requirement for the product to be reusable and therefore increase cost.
- **SCED** characterizes the amount of schedule compression demanded of the project. Values on this scale are rated in comparison to a nominal schedule. Hence a low value is demanding schedule compression that will likely require additional effort beyond that which would be expended under a nominal schedule. As a result, cost will increase.

Table 5 shows the values for early project estimation and our mapping of the COCOMO factors to our Product and Project Challenge factors [COCOMO II 2000]. The column headings range from Extremely Low (XL) to Extremely High (XH). The table is used to look up the needed COCOMO values based on the output from the BBN. For example, if the Product Challenge of the BBN output is computed to be “low,” then the values corresponding to low cost impact for the COCOMO factors mapped to the Product Challenge factor would be used. Because some COCOMO factors are scaled in a reverse manner (e.g., high may mean lower cost impact), we had to design our lookup algorithm to take this into account. Factors with this reverse relationship are indicated by ‘**<X>**.’

Table 5 Mapping BBN Outputs to COCOMO Inputs

Drivers	XL	VL	L	N	H	VH	XH	Product	Project
Scale Factors									
PREC		6.20	4.96	3.72	2.48	1.24	0.00	<X>	
FLEX		5.07	4.05	3.04	2.03	1.01	0.00	<X>	
RESL		7.07	5.65	4.24	2.83	1.41	0.00	<X>	
TEAM		5.48	4.38	3.29	2.19	1.10	0.00		<X>
PMAT		7.80	6.24	4.68	3.12	1.56	0.00		<X>
Effort Multipliers									
RCPX	0.49	0.60	0.83	1.00	1.33	1.91	2.72	X	
RUSE			0.95	1.00	1.07	1.15	1.24	X	
PDIF			0.87	1.00	1.29	1.81	2.61	X	
PERS	2.12	1.62	1.26	1.00	0.83	0.63	0.50	<X>	
PREX	1.59	1.33	1.12	1.00	0.87	0.74	0.62		<X>
FCIL	1.43	1.30	1.10	1.00	0.87	0.73	0.62		<X>
SCED		1.43	1.14	1.00	1.00	1.00			<X>

As an example, if the BBN predicts a high score on Product Challenge, then we would select values for the corresponding COCOMO factors that will have a high impact on cost. For precedentedness (PREC) this would mean the product is really new and different, it is unprecedented; therefore in COCOMO terms, it gets a VL and a scaling factor value of 6.2.

Finally, some COCOMO values were treated as single or two-point distributions. Only a single value was allowed for risk resolution (RESL). Generally, risk practices at most contractors and most DoD program offices are adequate, so a nominal value was selected. SCED was also allowed only a single value. TEAM was considered to be a two-point distribution. Either the team of stakeholders is reasonably cohesive or it is not. This was selected as a sample in the versatility of making the selection.

We assume that a large DoD project such as an MDAP will have some minimal level of complexity in both project and product structures. Additionally, there is always significant schedule pressure applied to DoD projects, and therefore the minimum SCED value allowed is the one for “Nominal.” The PERS factor for analyst and developer capability is difficult to determine. We know neither what development team is selected nor their actual capability for the product development work. Also, arguably, this factor can be applied as both a project challenge and a product challenge. Here it is used only as a product challenge factor, implying that finding a team that can do the work will be difficult and we expect a significant training effort. We may have to change the range and use of this parameter with further study.

PREX and FACL were given bi-modal distributions as these factors are determined by the selection of the contractor. For early estimation purposes, these factors are used in simulation only to expand the bounds of the estimates. Again, further study will help to determine whether this expansion is needed and by how much.

Alternative scenarios have the effect of moving the central tendency to the right or left and reducing the spread of the distribution at the same time. This observation is important in demonstrating that reducing uncertainty is very important early in the lifecycle.

We believe that combining the probability distributions from multiple scenarios will not help to improve the decision process. It is better to make each of the potential risks and effects as visible as possible in order to support risk-adjusted decisions. Therefore, a separate effort distribution should be presented to decision makers for each of the simulated scenarios.

Making the connection from the BBN output to the estimation tool input is not yet properly instrumented. As a result, the current probability distributions may be skewed in either range or central tendency. The model seems to behave correctly for the trivial cases and moves in the right direction when we apply scenarios. Therefore we believe the model can be used to provide realistic and useful results.

3.9 Monte Carlo Simulation in the Cost Estimation Process

Monte Carlo simulation is an uncertainty modeling method that has risen in popularity in the past 15 years with the advent of commercially available software tools such as @Risk by Palisade and Crystal Ball by Oracle.[6] Used in the cost estimation process, this method provides the estimator with the ability to produce cost estimates with uncertainty distributions. In this way, decision makers gain insight into both the upside and downside risk of a given cost estimate.

As a step in our cost estimation method, we model the uncertainty of the example COCOMO cost estimation spreadsheet to arrive at an uncertainty distribution for cost. To do this, we will revisit the input section of the example COCOMO spreadsheet in Figure 20 and observe the two green input factors to the COCOMO calculation. In this case, the two factors are synonymous with the two output factors of the BBN, namely, Product Challenge and Project Challenge. To keep this example simple for demonstration purposes, the input of Estimated Size (KSLOC) is set at 50 based on separate knowledge from an analysis of an analogy.

	A	B	C	D
1	Effect			
2	Product Challenge	5		
3	Project Challenge	4		
4	Estimated Size (KSLOC)	50		
5	Product Challenge factors		5	
6	COCOMO Parameter		XL	VL
7	Scale Factors	PREC		5
8		Val		6.2
9		FLEX		5

6 Screenshots are used with permission under the SEI license agreement with Oracle Corporation. Our license requires that each screenshot be labeled "not for commercial use."

Figure 20 Segment of COCOMO Spreadsheet Showing Inputs

Looking at a segment of the output section of the COCOMO spreadsheet in Figure 21, the output factor of COCOMO effort in Person-Months can be seen. This output factor is shaded in blue by Crystal Ball to indicate it is an outcome factor in which the values from each simulation trial calculation should be saved off for subsequent analysis. The value of 1971.566 Person-Months happens to be the current calculated value of Person-Months based on the latest values selected for Product Challenge and Project Challenge. This value will change as the simulation selects random values for the two Challenge factors. The real focus will be on the resulting uncertainty distribution for the Person-Months factor.

	A	B	C	D	E
1	Calculate Effort Multiplier and Scale Factor				
2	EM	is the product of		PERS	RCPX
3	8.98			1.00	2.496
4				rand	LF
5					
6	SF	is the sum of		PREC	FLEX
7	19.264			6.2	4.86
8				LF	LF
9					
10					
11	COCOMO Effort				
12	Person-Months		Size	E=B+0.01*SUM(SF)	
13	1971.566		50	1.10264	
14			74.70545		
15					
16					
17					

Figure 21 Segment of COCOMO Spreadsheet Showing Effort Output

Figure 22 and Figure 23 depict the information entered for the uncertain distributions of the two Challenge factors, taken directly from the BBN outputs. Essentially, Figure 22 and Figure 23 represent the Challenge factors' distributions from the corresponding BBN outcome factors once the BBN updates calculations with all program change drivers set at their default historical distributions. Remember that the BBN models each driver with a probability of being in a nominal versus non-nominal state. Additionally, note that Figure 22 and Figure 23 represent discrete distributions with a probability of occurrence for the different discrete value ranges of the Challenge factor.

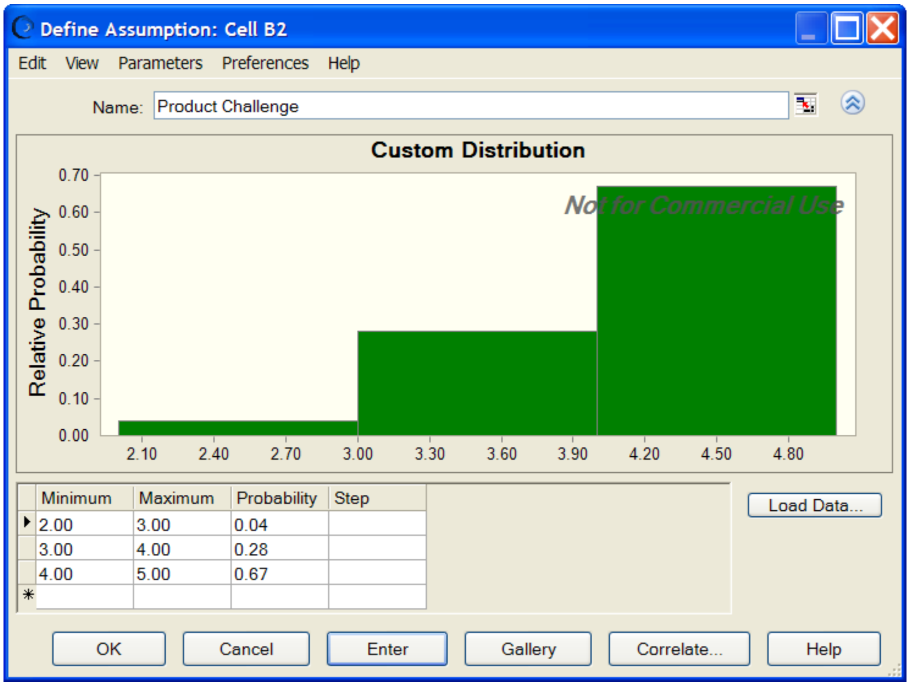


Figure 22 Probability Distribution for Product Challenge Factor

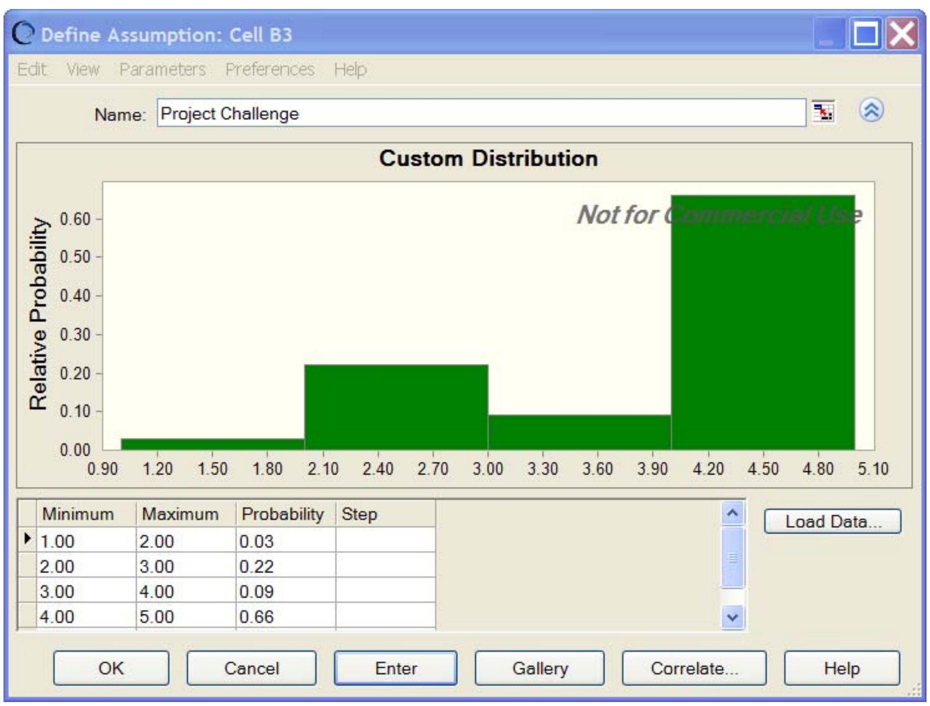


Figure 23 Probability Distribution for Project Challenge Factor

Having updated the two Challenge factors' distributions within the COCOMO spreadsheet we run a Monte Carlo simulation model using the output distributions from the BBN model. The Monte Carlo simulation runs for a set number of trials in which all the values of the outcome, Person-Months, are saved to a log. The resulting distribution for Person-Months, including the upper 90% confidence level of approximately 1,854 Person-Months, is shown in Figure 24.

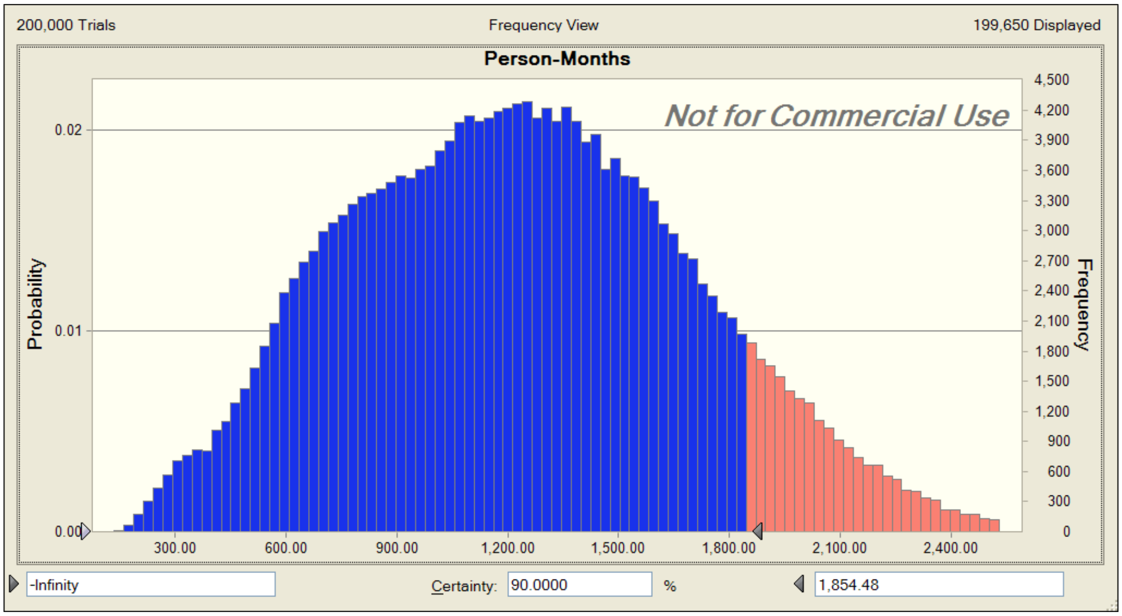


Figure 24 Probability Distribution for Person-Months Output Factor

An alternative way to view the same result is shown in Figure 25 using what is called a reverse cumulative graph. The reverse cumulative graph enables the decision-maker to readily identify upper limits of Person-Months for different confidence levels by starting with the desired confidence level on the y-axis and then tracing horizontally to the edge of the graph and locating the corresponding value of 1,854 Person-Months.

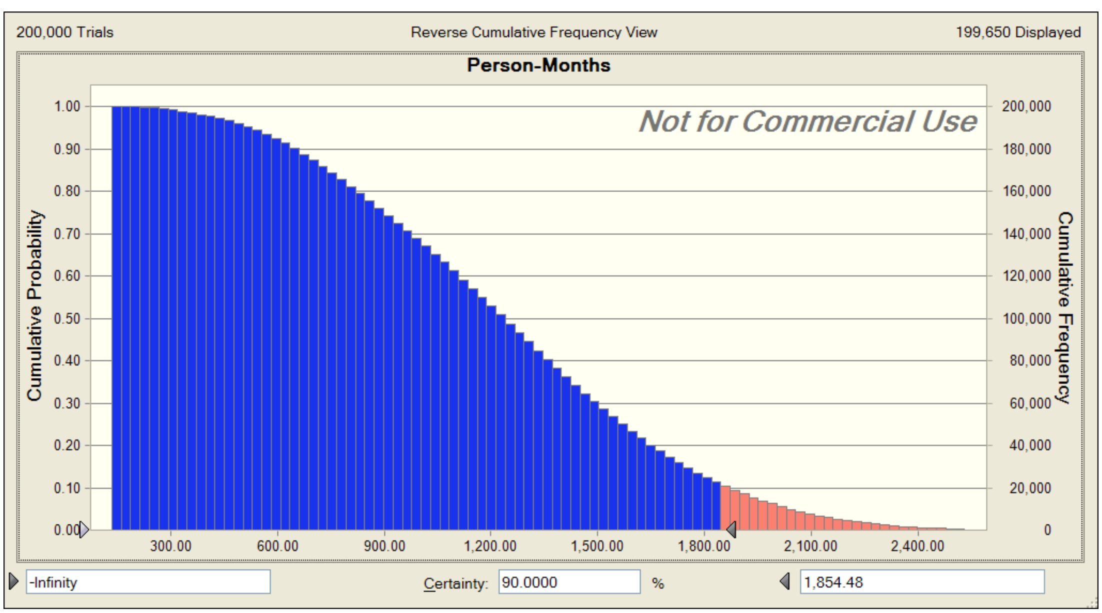


Figure 25 Cumulative Probability Distribution for Person-Months Output Factor

Monte Carlo simulation output includes a statistical table for the outcome factors. Figure 26 and Figure 27 depict such results for the Person-Months factor from the previous simulation and notably show a range of 2,664 Person-Months (2,800 minus 136).

Forecast: Person-Months

Edit View Forecast Preferences Help

200,000 Trials — Statistics View — Sheet2!A13

Statistic	Forecast values
Trials	200,000
Base Case	1,610.61
Mean	1,235.54
Median	1,223.79
Mode	---
Standard Deviation	462.03
Variance	213,475.26
Skewness	0.1958
Kurtosis	2.56
Coeff. of Variability	0.3740
Minimum	135.56
Maximum	2,800.10
Mean Std. Error	1.03

Figure 26 Statistics from Person-Months Simulation Results

Percentiles:	Forecast values
0%	142.64
10%	635.70
20%	811.10
30%	964.55
40%	1,101.78
50%	1,227.09
60%	1,352.45
70%	1,486.25
80%	1,639.48
90%	1,858.65
100%	2,819.27

Figure 27 Percentiles from Person-Months Simulation Results

In summary, the simulation result of an upper 90% confidence limit of 1,854 Person-Months relates to the scenario of all program change drivers left to their default historical distributions of nominal versus non-nominal. However, the next examples will show how specific scenarios of program change drivers drive different results for Person-Months using the Monte Carlo simulation of the COCOMO spreadsheet.

Figure 28 and Figure 29 now show different resulting distributions for the two Challenge factors based on the Scenario 1, where two program change drivers (Supply Chain Vulnerabilities and Program Management Structure), are now set at only their nominal state. The resulting BBN outcome distributions are then entered into the Monte Carlo simulation as depicted in Figure 28 and Figure 29. Note that each of these Challenge distributions reflects lower probabilities for the higher values of the Challenge factor, which is reasonable considering that there is greater confidence that most factors will remain in their nominal state.

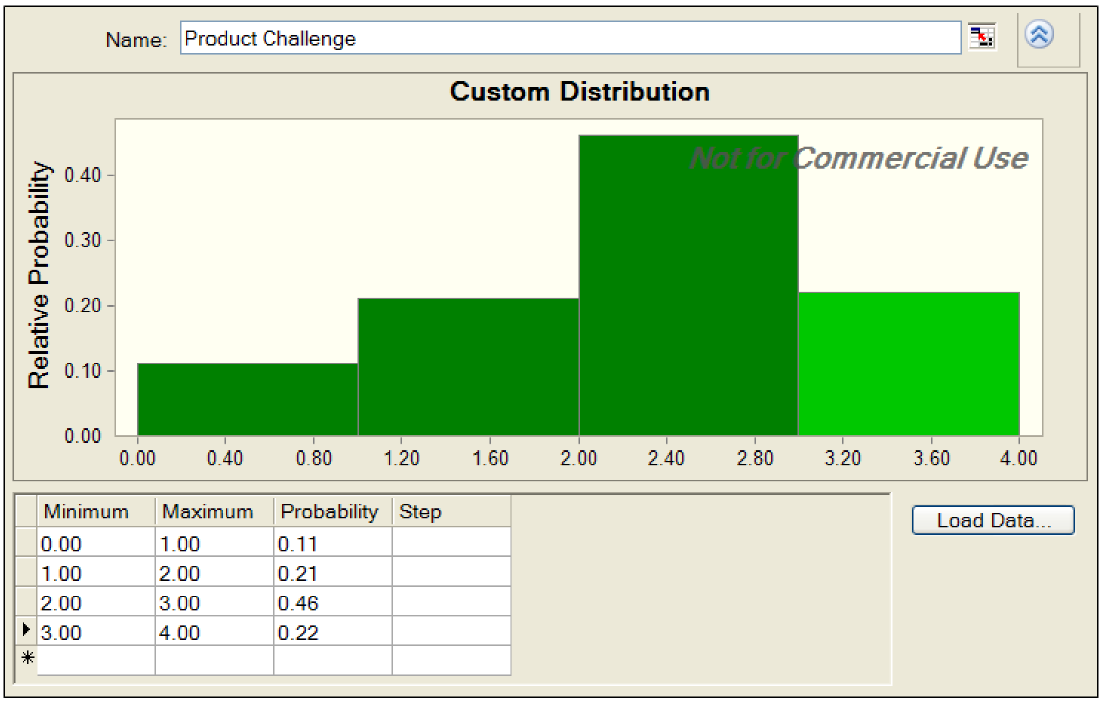
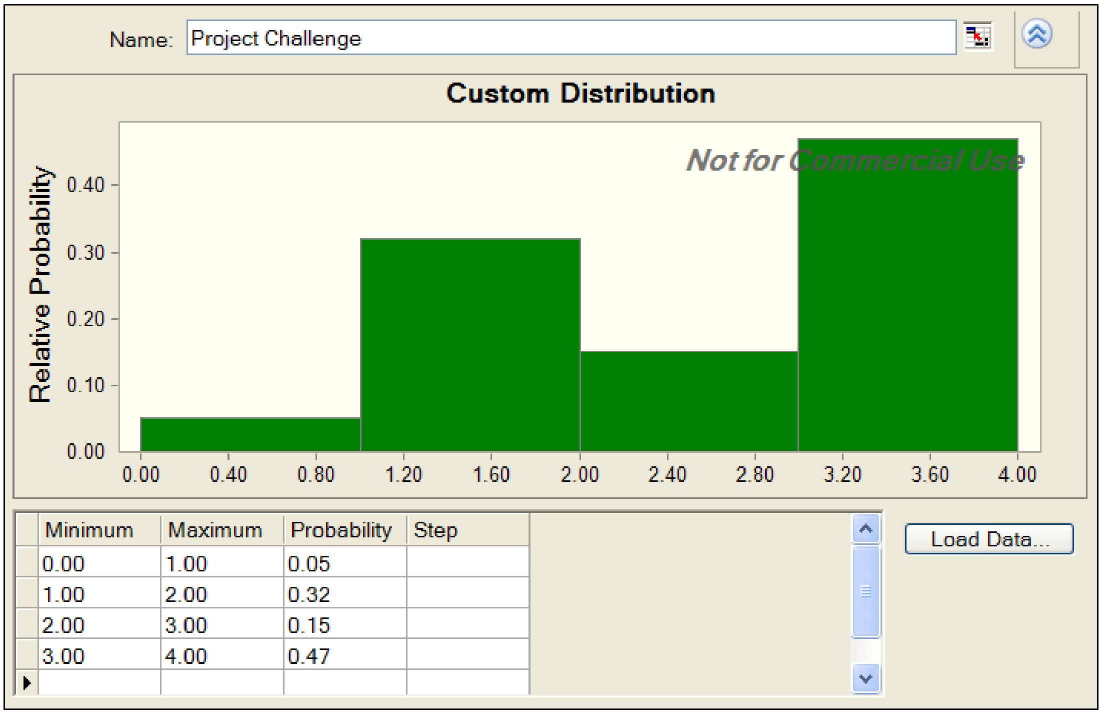


Figure 28 Probability Distribution for Product Challenge Factor (Scenario 1)



Minimum	Maximum	Probability	Step
0.00	1.00	0.05	
1.00	2.00	0.32	
2.00	3.00	0.15	
3.00	4.00	0.47	

Figure 29 Probability Distribution for Project Challenge Input Factor (Scenario 1)

Figure 30 shows the resulting new uncertainty distribution for Person-Months with a 90% confident upper limit of 674.24 Person-Months. This is a significant drop from a previous 90% confident upper limit of 1,854 Person-Months and shows that controlling just three of the BBN program change drivers to nominal state enabled a savings of 1,180 Person-Months of effort (1,854 Person-Months minus 674 Person-Months).

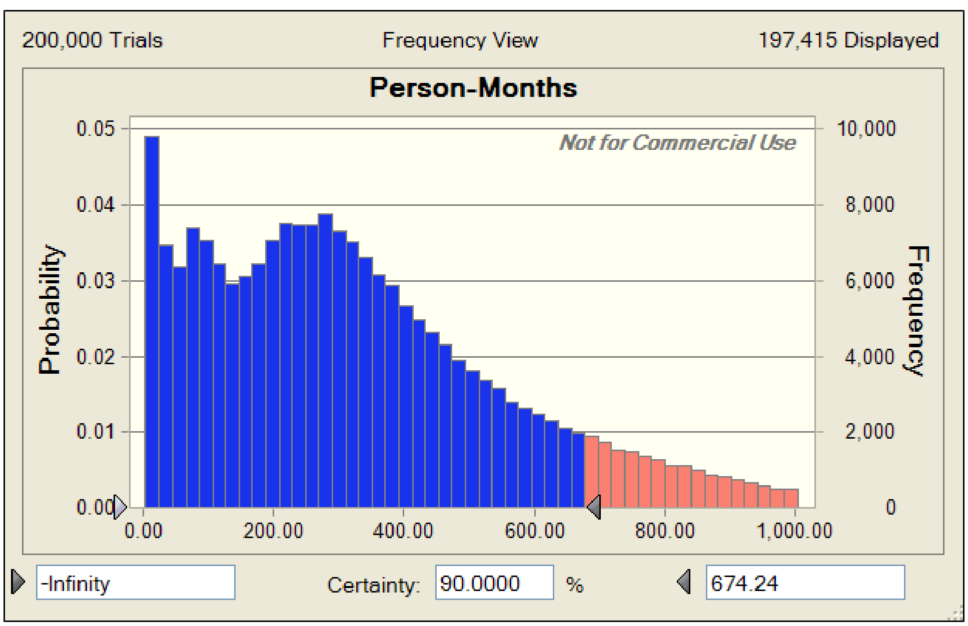
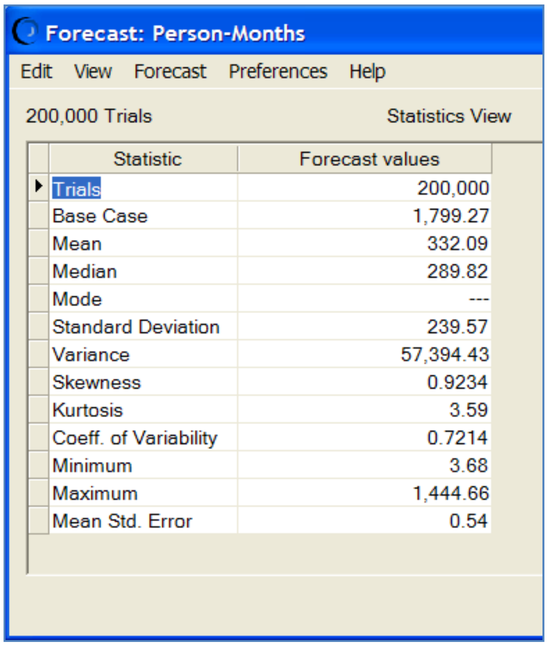


Figure 30 Probability Distribution for Person-Months Output Factor (Scenario 1)

Additionally for this scenario, Figure 31 shows that the median number of Person-Months is 290 but with a 90% confident upper limit of 674 Person-Months. Notably, the range is now 1,441 Person-Months, a significant reduction from the previous range of 2,664 Person-Months. Consequently, controlling some of the program change drivers to remain in a nominal state reduced both the absolute value and the range of expected Person-Months, thereby providing tighter distributions of Person-Months.



Statistic	Forecast values
Trials	200,000
Base Case	1,799.27
Mean	332.09
Median	289.82
Mode	---
Standard Deviation	239.57
Variance	57,394.43
Skewness	0.9234
Kurtosis	3.59
Coeff. of Variability	0.7214
Minimum	3.68
Maximum	1,444.66
Mean Std. Error	0.54

Figure 31 Statistics from Person-Months Simulation Results (Scenario 1)

In another example, Figure 32 and Figure 33 depict the distributions for the two Challenge factors related to the program change driver scenario depicted in Section 3.6.2, in which six drivers (Acquisition Management, Program Management Structure, Manning at Program Office, Program Management Contractor Relations, Program Office Performance, and Contractor Performance) are set at only their nominal states. Note that each Challenge distribution shows a continuing shift of probability to lower values of the Challenge factor. This remains congruent with the fact that even more program change drivers will now be controlled to remain in their nominal states.

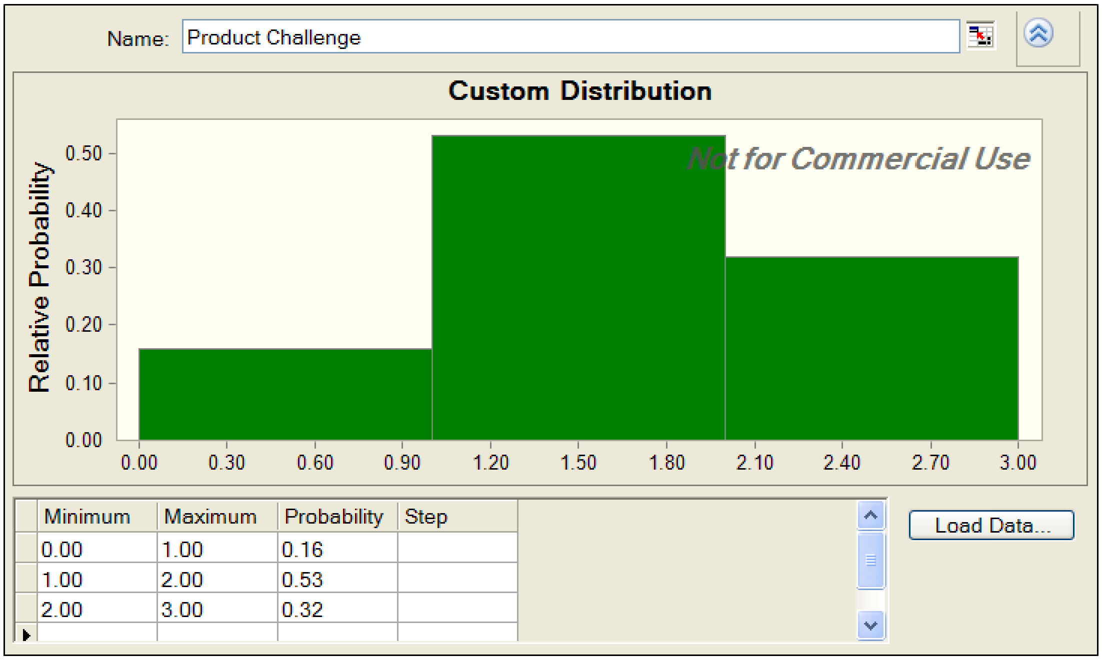


Figure 32 Probability Distribution for Product Challenge Input Factor (Scenario 2)

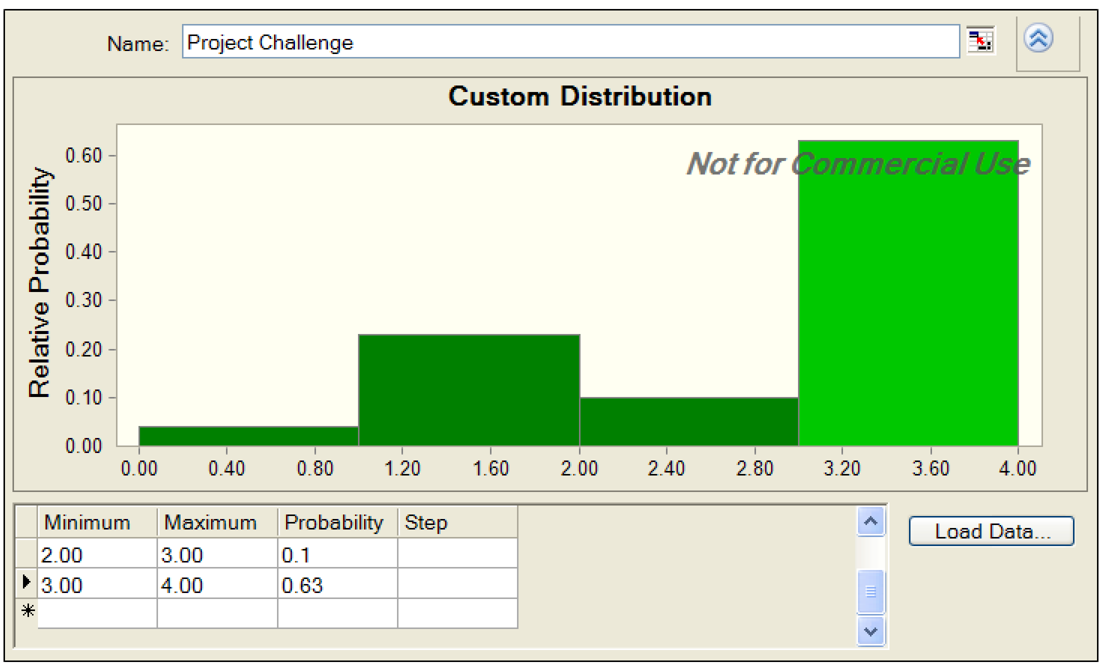


Figure 33 Probability Distribution for Project Challenge Input Factor (Scenario 2)

For this scenario, Figure 34 depicts the resulting uncertainty distribution for Person-Months with a 90% confident upper limit of 389 Person-Months.

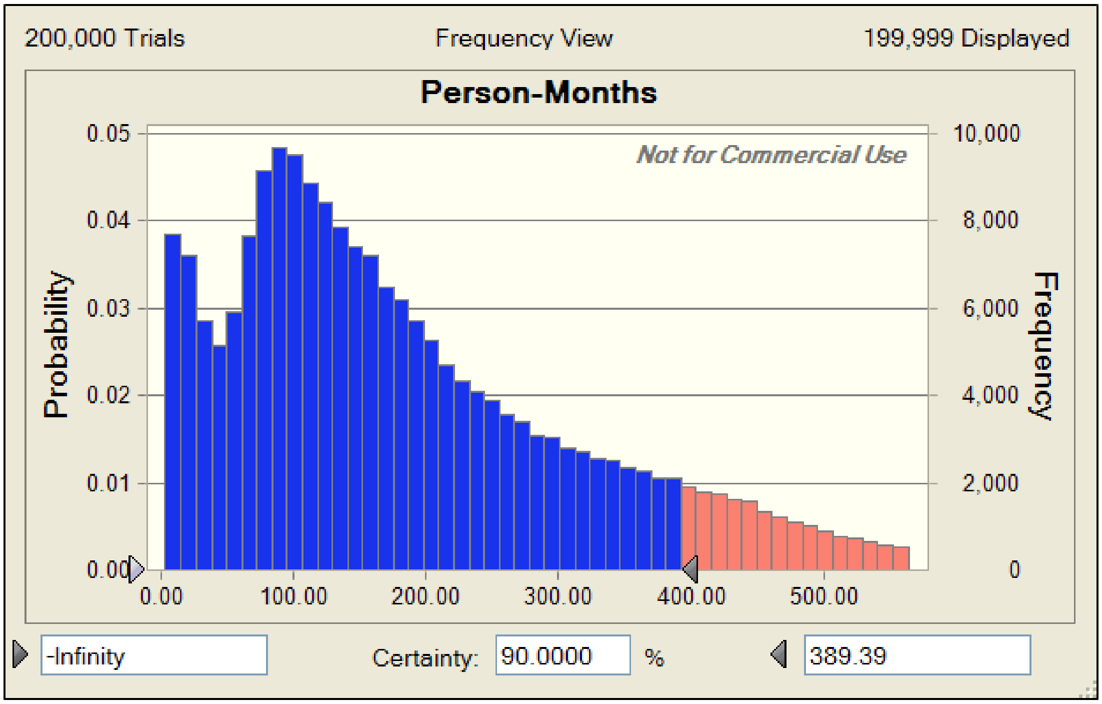
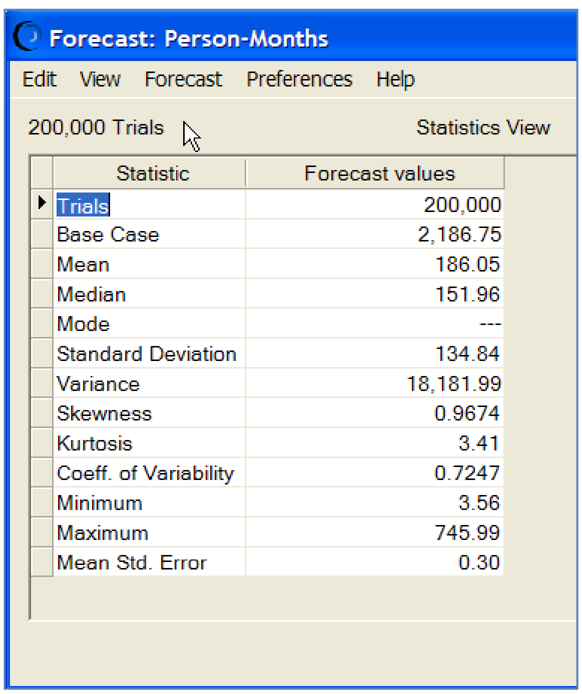


Figure 34 Probability Distribution for Person-Months Output Factor (Scenario 2)



Statistic	Forecast values
Trials	200,000
Base Case	2,186.75
Mean	186.05
Median	151.96
Mode	---
Standard Deviation	134.84
Variance	18,181.99
Skewness	0.9674
Kurtosis	3.41
Coeff. of Variability	0.7247
Minimum	3.56
Maximum	745.99
Mean Std. Error	0.30

Figure 35 Statistics from Person-Months Simulation Results (Scenario 2)

Again, in looking at the simulation statistics for Person-Months in Figure 35, the consequence of controlling six program change drivers to their nominal state produces an even lower number of Person-Months with an even tighter range. The range in this scenario is 742 Person-Months as compared to the previous ranges of 2,664 and 1,441 respectively.

Figure 36 depicts the Person-Months simulation result for each scenario after resizing the graphs so that they have comparable x axis scales. After this resizing, it is visually apparent that the combined BBN and Monte Carlo simulation enable the decision-maker to see the reduction in both the absolute number and range (e.g., uncertainty) of Person-Months based on controlling successively more of the program change drivers in the model.

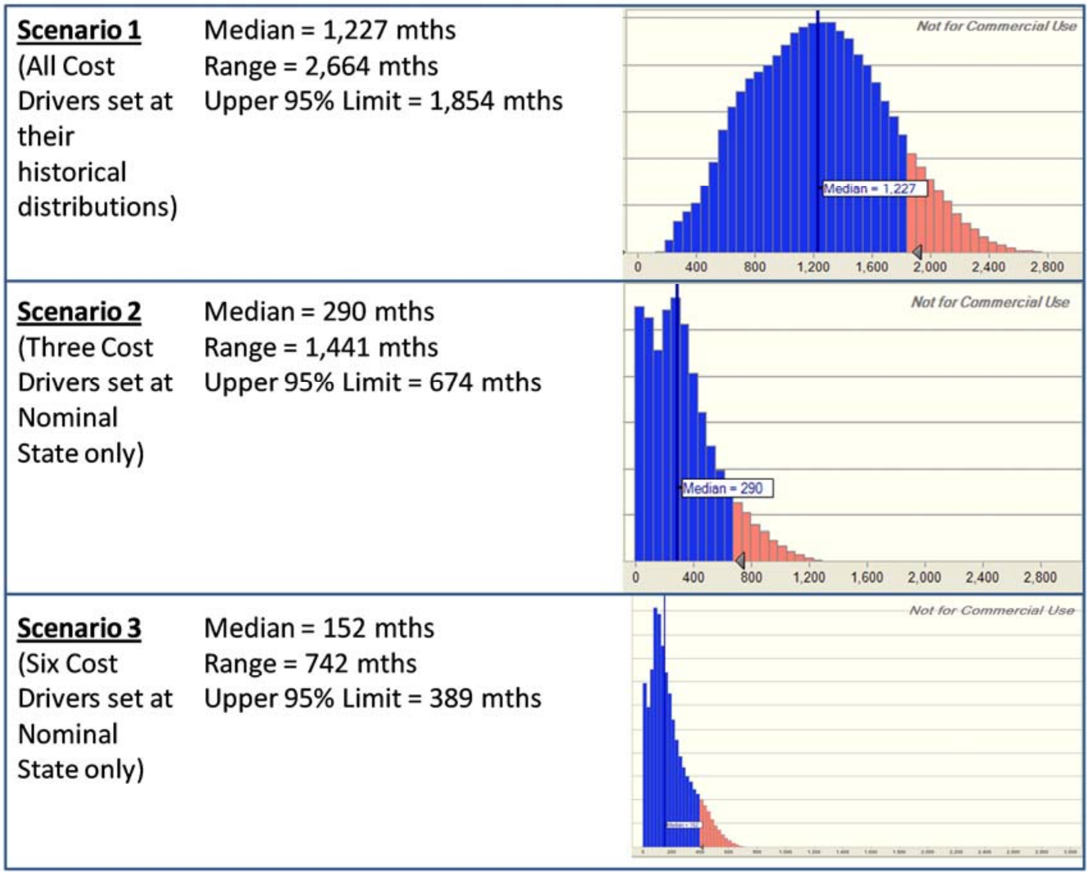


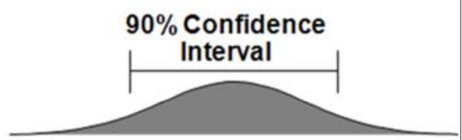
Figure 36 Person-Months Simulation Result for Each Scenario

4 Expert Judgment in Cost Estimation

While formal cost estimation models that rely on quantitative historical data have existed for many years, the most commonly used cost estimation methods continue to be based largely or even entirely on expert judgment. Empirical evidence on estimation accuracy when based on expert judgment remains relatively sparse and often inconsistent, but research in fields including defense estimation shows that experts are often overconfident in the accuracy of their judgments [Gino 2011, Francesca 2007, Hubbard 2010]. Little change is apparent since 2007 in the published research literature that is documented in the BESTweb system and maintained by the Simula Research Laboratory in Norway.[7]

4.1 The Problem with Expert Judgment

Experts are often overly optimistic about the expected costs of a program. Such over-optimism is by no means limited to purposeful underestimates of cost for political purposes or to schedule-driven constraints imposed by management. Experts often overstate how much is possible to complete in a limited amount of time, and some appear to be more prone to doing so than others. Figure 37 shows the amount by which educated professionals often overestimate the correctness of their responses to various categories of questions. At best, they gave correct answers only 50% of the time, even when the questions were specific to their industry, yet they stated that they were 90% certain that they had given a correct answer. As seen in the bottom two rows of the figure, however, overconfidence can be reduced considerably through calibration training.



Group	Subject	% Correct (target 90%)
Harvard MBAs	General Trivia	40%
Chemical Co. Employees	General Industry	50%
Chemical Co. Employees	Company-Specific	48%
Computer Co. Managers	General Business	17%
Computer Co. Managers	Company-Specific	36%
AIE Seminar (before training)	General Trivia & IT	35%-50%
AIE Seminar (after training)	General Trivia & IT	~90%

(AIE = Hubbard Generic Calibration Training)

Used with permission from Douglas Hubbard
Copyright HDR 2008 dwhubbard@hubbardresearch.com

Figure 37 Accuracy Within Subjectively Stated 90% Confidence Intervals

7 www.simula.no/BESTweb

4.2 Calibrating the Experts

Fortunately, research shows that judgment can be markedly improved through training that emphasizes bounding their best estimates within realistic limits of uncertainty (i.e., "I am 90% confident that my answer lies between A and B"). In particular, accuracy can be improved by considering interdependencies among cost factors to properly anchor and guide judgments. Individuals who regularly base their judgments on multiple anchor points are said to be "calibrated" such that wider intervals around their best estimates consistently indicate more uncertainty and narrow intervals reflect more thorough knowledge [Hubbard 2010].

4.3 Existing Calibration Training

Many people (perhaps experts in particular) seem to think that they are expected to "correctly" provide narrow ranges around their best judgments even if wider ranges more realistically represent their uncertainty. A key element of existing methods of calibration training involves getting people to recognize and reduce their uncertainty before rushing to judgment. The training consists of asking trainees a series of general factual questions. Each test is followed by guidance on factors that may affect the trainees' judgments. The guidance aims to increase accuracy and reduce uncertainty by emphasizing that answers to difficult questions depend on related circumstances and encouraging the trainees to consider related factors that may affect the basis of their judgments.

The importance of realistically representing one's uncertainty is emphasized by asking the trainees to think about the consequences of being wrong. They are encouraged to start with very wide ranges for bounding their answers and then narrow the ranges based on what they know about related information (i.e., by being explicit about the bases of their judgments). Similarly, checklists encourage the trainees to consider how and why their initial answers may be wrong as well as right. What else need they think about before rushing to judgment? How and when should they adjust their initial answers?

Notable improvements in the test results have been demonstrated for even the most experienced practitioners using generic question sets. However several test/guidance iterations are usually necessary to achieve those improvements.

Figure 38 summarizes the combined results of 11 studies of how well people subjectively assess the likelihood of being correct, with and without calibration training. The accuracy of the answers of trained individuals tends to be quite consistent with their stated confidence in their answers. Those who have not been trained are much less likely to answer the questions correctly, and the accuracy of their answers actually varies more as their stated confidence increases.

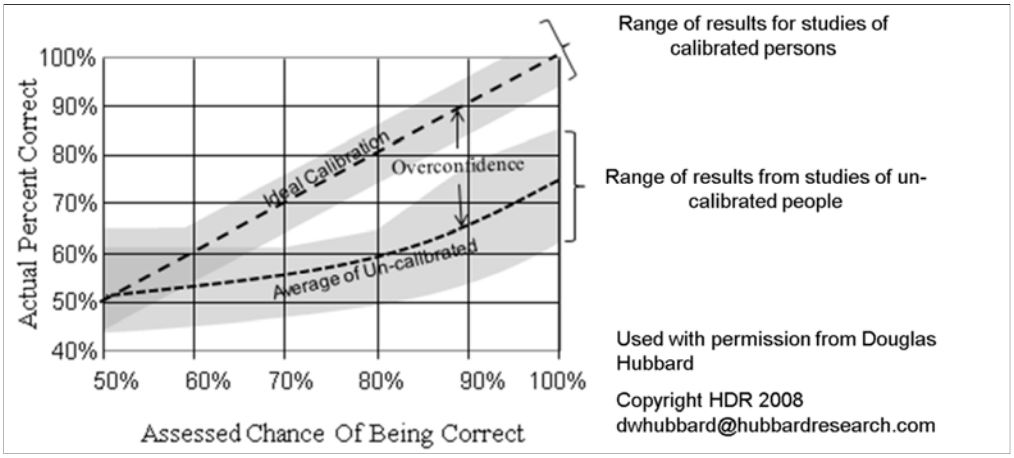


Figure 38 Subjective Assessment of the likelihood of Being Correct, With and Without Calibration Training

4.4 Domain-Specific Calibration

Techniques for calibrating experts by using anchors to guide their judgments can be efficiently adapted for the DoD environment. We propose to develop domain-specific anchors for calibrating DoD estimators to improve their performance in specification of uncertain cost-related inputs. Calibration training can make domain experts and other DoD cost estimation personnel more adept at explicitly identifying and describing likely program change drivers and the interdependencies among them.

The need for domain-specific anchors was reinforced in our conversations with experts at our workshops at Tektronix and the SEI. We hope that the cost estimation anchors developed for this research will be augmented over time with others adapted from program change drivers identified in estimates using our overall method for modeling program uncertainties. Such a library of DoD-specific cost estimation anchors may provide useful guidance for future program estimators as well as better training for them and other DoD personnel.

4.5 Results of Early Workshops

We conducted calibration-training exercises in our workshops at Tektronix and ASP using general questions from Hubbard rather than industry-specific questions. Some of the test questions asked the participants to provide quantitatively stated upper and lower bounds within which they were 90% certain that the correct answer lies. Other questions were true/false questions, where the trainees were asked to quantitatively express their confidence that each answer was correct. These too were used with permission from Hubbard.

We asked the participants in both workshops for their feedback on the exercises. The replies of the 14 participants in the Tektronix calibration training are shown in Figure 39. When asked if they benefitted from the training, most participants said they found it very beneficial (first bar). We also asked about the extent to which “honest communication of uncertainty was welcomed or desired” in their organization in the “past several years” and “would be welcomed and embraced” in future discussions about forecasting and estimation.

Responses to these questions are shown in the second and third bars, and are quite different when comparing the past with the future. Over half of the participants chose answers indicating that communication about uncertainty was historically not welcomed in their organization. However, all but one of them chose answers on the upper half of the continuum when asked if calibration training was likely to add to open discussion of uncertainty in the future. The differences in their replies is statistically significant ($p<.003$) according to the Mann-Whitney U-test, even with such a small number of cases. Finally, all of the workshop participants chose answers 5 or 6 (“high” or nearly “high”) when asked about the degree to which “you believe that supervisors, managers and senior leaders should go through calibration training.”

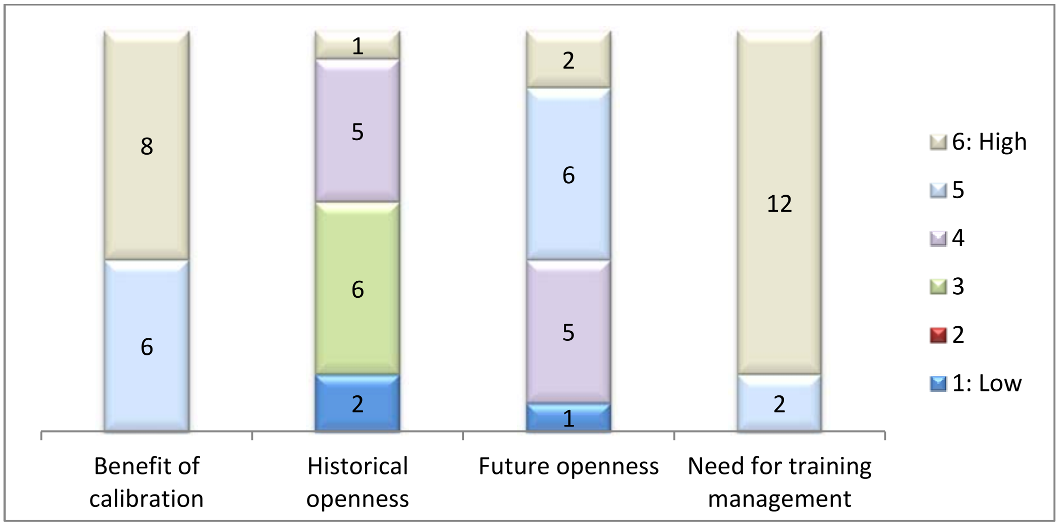


Figure 39 Feedback from Tektronix Workshop

We did not ask specific questions about calibration in the ASP workshop feedback form, and the small number of participants does not justify statistical analysis. However, the discussion during and after the calibration exercises was extremely positive. The participants all agreed that calibration training could be very worthwhile for practical use under real-world DoD circumstances.

Figure 40, Figure 41, and Figure 42 summarize the performance of the participants in our initial abbreviated training sessions at Tektronix and the ASP workshop. The goal for a well-calibrated individual is to be correct 90% of the time in the tests that are administered during the calibration training. As seen in the figures, the participants improved quite noticeably in the accuracy of their judgments over the course of the training. For simplicity’s sake we show the aggregate percentage of correct answers for each group. There were a few notable individual differences within the groups, but the initial differences were lessened considerably over the course of the training exercises.

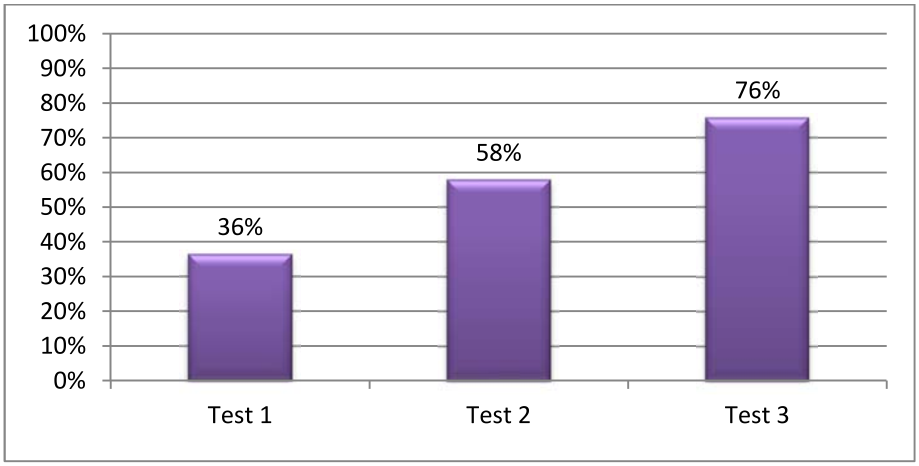


Figure 40 Results of Calibration Training at Tektronix Workshop

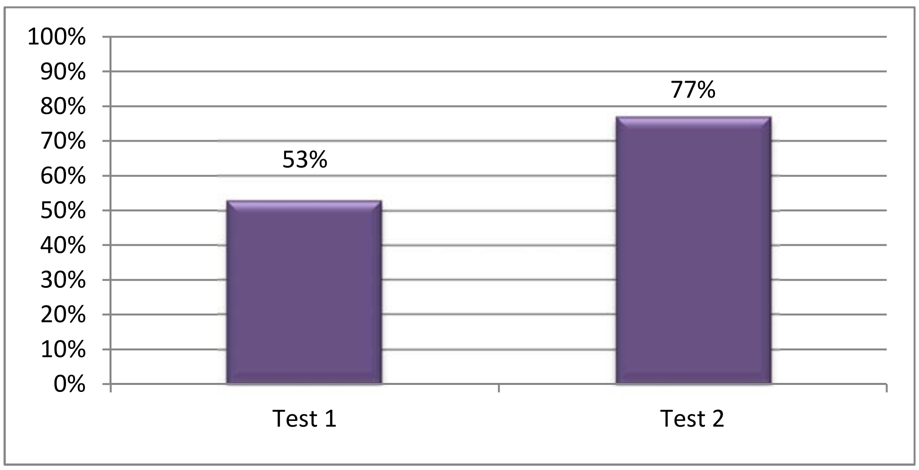


Figure 41 Results of Calibration Training with Tektronix Architects

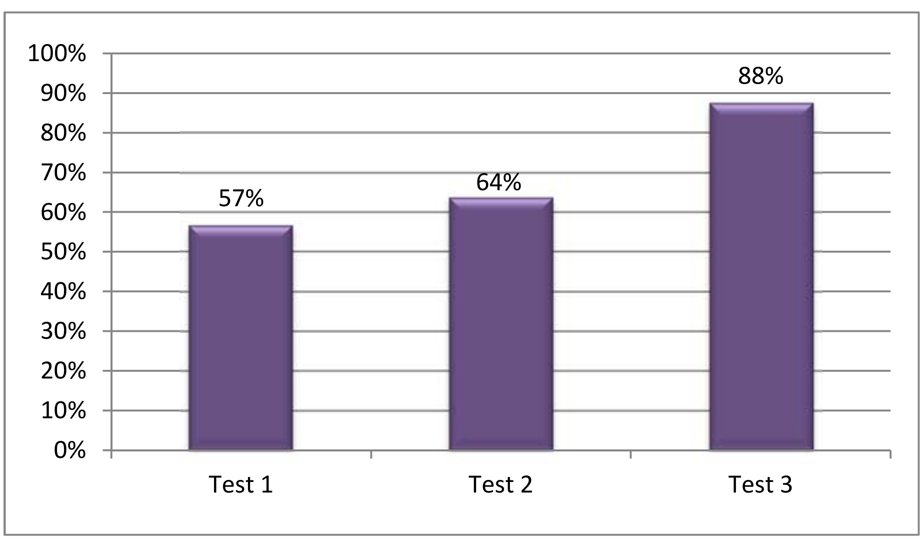


Figure 42 Results of Calibration Training at ASP Workshop

The participants in our workshop at Tektronix were enthusiastic about the calibration work, particularly after they saw that there still was ample room for improvement for most of them after the training exercise was completed. They quickly recognized how calibrated thinking could lead to more effective identification of otherwise unconsidered program change drivers and the cascading dependencies among them. In fact, they immediately launched a project to prepare an inventory of company-specific calibration anchors. The two participants (one engineer and one manager) whose performance improved the most during the training were chosen to lead the project.

The training was also well-received by the SEI's Acquisition Support Program (ASP). They too were impressed with their improvements in correctly answering test questions within narrower bounded intervals. The calibration training exercises followed immediately after further review of the program change driver dependencies developed during and after the first ASP workshop. This sequence led to insights about the conceptual similarities between related program change drivers and calibration anchors. There was a lively discussion about the need for DoD-specific anchors to reduce the endemic over-confidence and cost overruns in existing DoD program estimates.

4.6 Calibrating Teams

Our experience thus far suggests that explicit identification of program change drivers, dependencies among them, and judgments about conditional probabilities minimizes the need for formal methods of reconciling differences among experts. Informal discussion among the teams with whom we have worked has sufficed thus far. However, this approach may not be as well received elsewhere as our approach to early cost estimation becomes more widely adopted.

There is a growing body of research on methods that may help reconcile differences in expert judgment in cost estimation [Valerdi 2011, Gino 2011, Francesca 2007, Jørgensen 2005, Jørgensen 2004, Hora 2004, Shepperd 2001,Miranda 2001]. In one study, groups of experts submitted less-optimistic estimates than did individuals when queried alone. The group estimates were closer to actual effort expended, and "the group discussions led to better estimates than a mechanical averaging of the individual estimates" [Moløkken-Østvold 2004]. However, algorithmic methods of reconciliation of individual differences may prove to be better or faster and are deserving of further research [Hora 2004, Shepperd 2001, Miranda 2001, Ariely 2000, Wallsren 1997, Hihn 2004, Winkler 2004, Unal 2004].

We currently are designing a series of experiments on methods to reconcile differences in judgment among individuals working in small groups. Reconciliation of individual differences in these early experiments will be done using a group consensus method, most likely a variant of wide-band Delphi methods. Differences in individual judgments also will be reconciled algorithmically, and the two methods will be compared with respect to differences in their results and the time necessary to complete them.

These experiments are being preceded by and will be based on a panel study that tracks patterns of improvement during training to calibrate individual judgment capabilities and degradation in those skills over time. In so doing we will begin investigating the need for refresher training. Of course we also will confirm the need for reconciliation of individual differences among the panel study participants. However the work by Hubbard, other research, and our own experience suggest that such a need will very likely exist.

Both the panel studies and our planned classroom experiments are being conducted with software engineering graduate students at Carnegie Mellon University, all of whom have at least some prior work experience. None of the students are experts in cost estimation. The tradeoff, of course, is better experimental control at the expense of being able to generalize the results to the cost estimation domain.

We are starting this work with existing calibration training using general questions and anchors. Later in the year we hope to replicate these studies using anchors developed for software and software intensive systems in collaboration with faculty colleagues at Carnegie Mellon and elsewhere.

5 Workshop Results

5.1 Tektronix Workshop

Throughout March – June 2011, the research team held several pilot workshops to obtain feedback on the concepts of program change drivers, states within each program change driver, and associated probabilities of occurrence for each driver state. The first workshop was conducted at Tektronix Communications in Plano, Texas. The workshop consisted of one day of method introduction and discussion of program change drivers, a half day of scenario derivation using the program change driver cause-effect matrix, and a full day of calibration training. Survey results of the Tektronix workshop participants (Figure 43) indicate significant perceived value in the steps of the method used in the workshop.

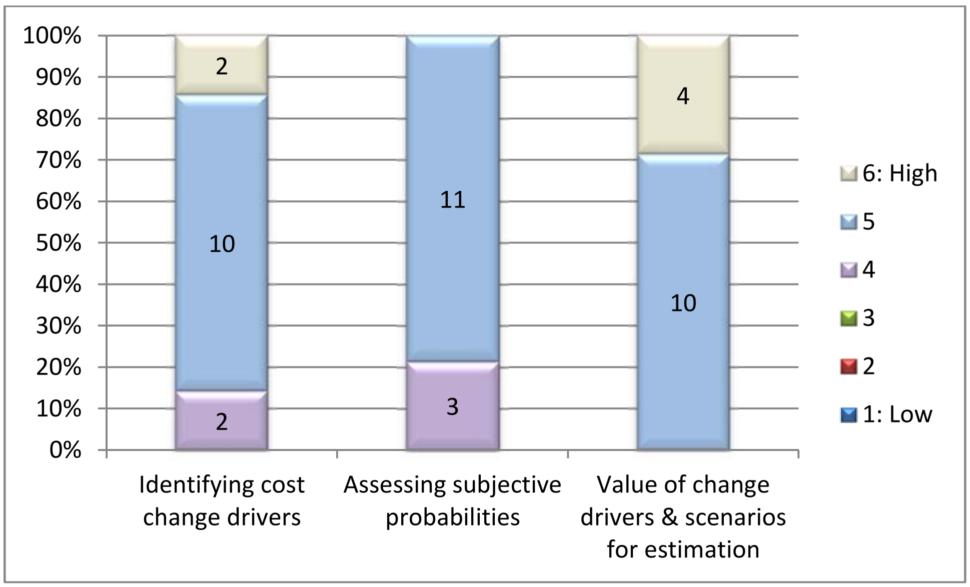


Figure 43 Perceived Value of Workshop at Tektronix

Four primary lessons arose from the Tektronix workshop discussion of program change drivers and their states:

1. The composition of workshop attendees significantly influences the bias of the program change drivers explored and defined. The attendees were employees in engineering, program management, quality and technical marketing. A reasonable balance was achieved, but it was clear that the background of the individual influenced the attention paid to particular drivers. For example, engineers focused on product and technology-related program change drivers, while program managers honed in on programmatic program change drivers. This bias was much stronger than we expected, and we intend to establish minimum participation standards by role within an organization in future workshops.

2. The scenario planning workshop must include sufficient time for discussion, to secure agreement among the participants regarding the time horizon, situation map and boundary of the cost estimation scenario planning. Although we covered these topics in the preparato-

ry slide presentation, we later discovered that insufficient time had been devoted to this topic, as it generated detailed discussion among participants, who had different perspectives due to differences in product line, time horizon, and situation map with competitors. We concluded that a more proactive discussion of these scenario planning workshop elements is warranted including testing for knowledge through role playing or feedback to the group of different hypothetical situations.

3. Participants were grouped in 3-4 person breakout groups during the scenario planning workshop, and each group was assigned one or more categories of program change drivers to brainstorm and discuss. These groups worked very well and enabled a good alignment of individual background and knowledge with the program change driver categories. Without this approach, the program change driver activity alone would have required at least several days. The one negative we observed is that several individuals indicated they had unique information that would have enlightened other groups working on different program change drivers. We still need to identify a solution that is practical, efficient and complete in coverage of program change drivers and their states.

4. The breakout groups had difficulty thinking of mutually exclusive states for drivers. Their natural inclination was to identify possible future conditions that could occur within a program change driver, but not necessarily in a mutually exclusive fashion. As a result, we concluded that the method needs to accommodate domain experts identifying future conditions of program change drivers, whether mutually exclusive or not. In the future it may be useful to include a step to determine a reduced set of mutually exclusive states for program change driver conditions.

5.2 ASP Workshop

We learned some additional lessons in workshops with representatives from the SEI ASP organization:

1. The initial list of program change drivers derived from the POPS categories served well in prompting discussion about the historical sources of surprise in cost.
2. Participants advised that a number of the initial drivers could be grouped together due to their similarity.
3. A number of additional program change drivers were identified to account for events that SEI ASP staff had witnessed in previous program interventions.
4. The SEI ASP staff added rich descriptions of the likely states of many of the drivers reflecting their field experience.
5. Participants quickly embraced the need to calibrate expert judgment and advised that this part of the solution should not only support cost estimation but also more mature risk management and Program Management Office (PMO) operational decision-making.

6 Summary and Conclusions

6.1 Summary

Cost estimation of DoD MDAPs during concept refinement (MSA phase) comprises a set of focused documentation that essentially presents an argument for the Defense Department to spend large portions of budget to build new systems to achieve new capabilities. Proposals assume a steady state of progression in the research, development, production, and sustainment of the solution even though the system will take years or decades to fully accomplish. However, this same documentation contains information that can be utilized to identify elements of uncertainty that are not currently addressed by cost estimation methods. Pre-Milestone A estimates rely heavily on subjective expert judgment to select analogies from which to extrapolate or adjust costs. We believe that by engaging the appropriate domain experts to form judgments of uncertain factors we can produce a more accurate and realistic view of program execution to inform the cost estimation process and lead to improved decision making.

As described throughout this report, our overall method for modeling uncertainties aims to provide credible and accurate program cost estimates within clearly defined, statistically valid confidence intervals. By making visible the potential changes that may occur during program execution, our approach also supports the quick revision of program estimates to better mitigate risk and respond more quickly to the program changes that often arise over a program's lifecycle.

Equally important the same flexibility enables the early consideration of the likely impact of different possible future scenarios on the estimates . Intuitive visual representations of the data explicitly model influential relationships and interdependencies among the program change drivers on which the ultimate estimates depend. The assumptions and constraints underlying the estimates are well documented and available for evaluation and further use. This contributes to better management of cost, schedule, and adjustments to program scope as more is learned and conditions change.

Our method synthesizes scenario building, BBN modeling, and Monte Carlo simulation into an estimation method that quantifies uncertainties, allows subjective inputs, visually depicts influential relationships and outputs, and assists with the explicit description and documentation underlying an estimate. As described more fully in Section 3.6.2, we use scenario analysis and design structure matrix (DSM) techniques to limit the combinatorial effects of multiple interacting program change drivers. Scenarios of combined program change drivers are represented in the BBNs. The BBNs and Monte Carlo simulation are then used to predict variability of what become the inputs to existing, commercially available cost estimation methods and tools. As a result, interim and final cost estimates are embedded within clearly defined confidence intervals. An overview of these methods is provided in Appendix A.

6.1.1 Our Approach to Meeting DoD Needs

Paul Kaminski—chairman of the Defense Science Board and a former DoD official in charge of Acquisition—presented a rigorous checklist to the Senate Armed Services Committee in 2009 for programs to implement in the pre-Milestone A phase [Kaminski 2009]. Cost risk is addressed by the following items:

- Are the major known cost and schedule drivers and risks explicitly identified, and is there a plan to track and reduce uncertainty?
- Has the cost confidence level been accepted by the stakeholders?

Our experience with DoD acquisition of software intensive systems led us to investigate several methods to address how to characterize and quantify the effects of using uncertain information to forecast program execution for an MDAP's lifecycle, particularly as embodied by the pre-Milestone A cost estimate. As detailed in this report, we start with the people who conceptualize the program details necessary to obtain approval to proceed into the Technology Development Phase. We work with these experts to form explicit judgments about the likelihood of specific change factors which could impact program performance and cost. This visibility of uncertainty distinguishes our approach from current practices. Combinations of these quantified factors are used to build scenarios with the experts, where conditional probabilities allow for the calculation of variability with defined statistical confidence levels. These models are then used as input factors to existing cost estimation tools and methods. Rather than use the output of a cost model to adjust or extrapolate the range of potential cost, we use the identified potential variability of the model's inputs to derive a range for the cost estimate. We believe this explicit use of uncertainty in the inputs result in more robust estimates.

Three recent MDAP programs illustrate the effects of uncertainty on the DoD budget. The Army's Comanche helicopter, the Navy's DDG-1000 destroyer, and the Air Force's Transformational Satellite Communications System (TSAT) cost upwards of $35 billion before cancellation. This investment represents a loss of expected capability. Program interventions and mitigation address problems once they have occurred but do not prevent problems from occurring. Better information on risk earlier in the lifecycle enables better decisions.

6.1.2 Review of the TSAT Reports for Program Change Drivers

We recently obtained access to several of the TSAT briefings prepared when the program sought approval at the Pre-Acquisition Key Decision Point B (KDP-B) in 2003. This phase in space acquisition programs is no longer used but is similar to the current pre-milestone A phase. The documents provide a rich set of program change drivers identified by the program and the AoA technical comparisons.

In addition to the starter set of POPS drivers, some of the potential program change driver categories considered included:

- Coverage/Capacity/Configuration
- Survivability
- Operational Management System
- Interoperability/Operations

- Non-Communications Functions

Within each category several technological concerns are identified as issues that could impact mission, schedule, and cost. The program's Satellite Operations Center software alone listed 16 major functions, many of which required unique software. Indeed, 27 of the 37 thresholds established by TSAT's Capability Development Document (CDD) were then considered at risk of impacting cost and feasibility of the program.

By 2006, the program itself reported significant impacts to costs and schedule due to budget instability from Congressional budget cuts in FY03, FY04, FY05, and FY06. The instability of TSAT funding resulted in program execution inefficiencies and increased program lifecycle cost [DAES 2006]. We also know, for example, that emergent cryptographic requirements to TSAT drove up costs and delayed schedule. These types of program change drivers are exactly what we expect to elicit from experts as we engage them in the previously mentioned program change driver workshops and Hubbard calibration exercises.

At the time of Milestone B certification, the program was under heavy criticism for cost growth, schedule delays, and performance shortfalls. By 2009, the Air Force cancelled TSAT after spending $3.5 billion and no usable outcome.

The question of whether decision makers would have made different choices earlier in a program's lifecycle if better estimates of total cost and schedule were available will always be subject to conjecture. In the case of TSAT, GAO recognized by 2005 that TSAT was an ambitious new military communications program that would enable laser crosslinks capable of 20 GB/sec, compared to the already developed Advance Extremely High Frequency (AEHF) satellite with radio frequency links of 60 MB/sec. Even so, AEHF was over budget and behind schedule even though it was considered a mature technology [Cancian 2011]. For the money invested in TSAT, the Air Force could have constructed and deployed seven modernized AEHF satellites.

6.1.3 The Results So Far

QUELCE as an approach to early cost estimation is unprecedented in many ways. We spent much of the past year developing and refining our analytical methods. We have begun to establish sufficient proof of concept about the value of the work. We started by trying out the earlier steps in the overall method in small-scale workshops with Tektronix participants and with senior technical staff from the SEI's Acquisition Support Program who have wide experience in DoD program development and cost estimation.

Feedback about the value of our approach from the participants in both workshops was quite positive. Each workshop started with an overview presentation of all aspects of our approach to early cost estimation. The workshop at Tektronix included hands-on exercises on identifying program change drivers, populating the design structure matrix, and some early experimentation with our scenario methods (Section 3.6.2). The ASP workshops included hands-on exercises and definition of the dependency matrix, which led to rich discussion of our scenario methods. Subsequently we followed up by prototyping the entire approach including a working Bayesian Belief Network along with use of the outputs from instantiated BBNs populated with probabilities for example scenarios, and we used Monte Carlo simulation to feed the calculation of cost estimates within probabilistic bounds using COCOMO II (see Section 3.9). Both workshops included calibration training exercises (see Section 4).

The work that we have started with our colleagues at Aerospace on the Air Force's former TSAT program is an example of how we think one can use existing artifacts and expertise for retrospective studies to "postdict" the results of using our methods. From a major program that was cancelled, the TSAT documents we have reviewed thus far confirm our belief in the efficacy of QUELCE to identify the uncertainties of important change drivers and the consequent impact to costs. Following-up with the subsequent steps in our overall method with the participation of key personnel who worked on TSAT promises to be a very valuable exercise. Resources permitting, we plan to do additional retrospective and prospective studies in collaboration with our colleagues at Aerospace.

In addition we have received very positive feedback from leaders in estimation research and DoD estimation experts at the most recent (2011) Department of Defense Cost Analysis Symposium, as well as in ongoing discussions and presentation to colleagues, contacts in program offices, the service cost centers, and other DoD agencies. We are continuing discussions about conducting empirical trials with some of them. Others have said that they would like to be included in proposals for further work in this area. Everyone in the DoD cost estimation and program management community affirms that this is an important problem area in need of a solution.

6.2 Further Research

Stated broadly, we intend to evaluate the extent to which the probabilistic methods that we propose improve the accuracy and precision of cost estimates for the DoD programs with which we work, as compared to their previous approaches to cost estimation at pre-Milestone A. We will use the results of our evaluative research to refine our approach to early cost estimation as well as demonstrate the added value we have observed thus far.

Our initial results are based largely on trials of the earlier steps of the method in workshops and *post hoc* review of previous estimation artifacts. However additional empirical research activities are in the pipeline. Our focus will be on the feasibility of the implementation of our approach in the Program Office, which is where all the work generated in the Materiel Solution Analysis comes together. We will continue doing studies that focus on the individual steps of our overall approach for continual refinement and the cataloging of quantified program change drivers. Over time, knowledge of the program change drivers' impact on program performance can be used across analogous systems and components, much like the use of CERs. The smaller scale studies will be followed by more comprehensive studies covering the overall early estimation method.

In addition to the actual cost estimates, we will track estimation effort, elapsed time, and total cost during these trials. Our quantitative measures will include time and effort expended on training as well as model implementation and interpretation of the estimation results. As such we will track effort and elapsed time for each step in the overall method. We will supplement these data with any available existing records of time and effort previously spent on comparable program estimates using other methods. As stated earlier, we expect that time and effort required for the rework of estimates will be greatly reduced.

We will elicit additional feedback through structured interviews of the personnel doing the estimation tasks. Participants will be asked for their judgments about clarity of the method definitions and ease of task performance, especially as more is learned and requirements or available technical solutions change over time. We also will ask them directly about the value of the insights provided by performing the tasks and their confidence in the resulting estimates as compared to their prior experience with other estimation methods. In addition to asking about the realism and likely accuracy of the estimates we will ask the participants about the realism of the proposed scope of work. We will continue soliciting similar information about the refined methods as they are used more widely over time.

Similar questions can be posed to the management for whom the estimations are done. If possible, we will query other key DoD stakeholders, particularly those responsible for management decision authority (MDA), source selection, preliminary design review (PDR), and other key decisions about continuance of the proposed programs into the post-Milestone A Technology Development Phase.

An early focus on retrospective analysis: Our initial field studies of actual program estimates will focus on proposed, existing, and canceled programs that are experiencing or recently have experienced difficulties early in the program lifecycle. This will allow us to make direct comparisons with estimates using other methods as well as our own. Such retrospective studies can be particularly valuable in providing timely comparisons of estimates or re-estimates. Under the right circumstances they may allow comparisons of the estimates with actual expenditures early in the lifecycle.

Mechanisms are being established to ensure that retrospective re-estimates exclude information that was previously unknown to the participants. We will ask the participants to consider only facts they knew at the time of the original estimates, which may or may not have been considered explicitly.

As noted in the TSAT discussion above, we have already begun analysis of pre-Milestone A and early Milestone B documentation that was used in early estimation for the former TSAT program. Such real-world examples are crucial to provide compelling evidence in support of further research in this important area. Of course it often takes many years before initial estimates can be compared the actual costs expended. However comparisons also can be made early on with independent cost estimates (ICE).

Prospective studies: We currently are exploring possibilities for participation in this research with other DoD programs and military service offices. We would particularly like to work with proposed programs whose estimates have recently failed to be certified for Milestone A. Proposed or existing programs that are not confident in their preliminary estimates could provide even better testbeds for our methods. In either case estimates may already exist to compare with the QUELCE approach. However it can be difficult at best to get people to do new things when they are under time pressure, perhaps especially for teams working on certification to become MDAPS. For that reason we also will consider doing early trials with proposed or existing major Acquisition Category II (ACAT II) programs. Programs for large-scale software intensive systems outside of the defense industry also may be able to provide more rapid feedback on the accuracy or perceived realism of their estimates.

Small-scale field studies: Small studies that focus on selected steps of our overall method can be conducted in several venues. These can be done in workshop settings similar to our initial trials with Tektronix, Inc. Such studies can be useful in building interest to participate in more comprehensive coverage of our entire suite of method steps.

Graduate seminar projects for master's and doctoral students are another likely venue, especially where the students already have practical experience in the field. Among other things, such graduate student practicum projects can start by using existing methods such as SEER or the COCOMO suite of parametric models, followed by studies of our method's individual steps, and leading to exercise of the full method in year-long project courses. Advising senior graduate students or teams of students who are working on their longer term thesis projects would be even better. We currently are discussing such opportunities with colleagues at Carnegie Mellon and the University of Arizona. We intend to initiate similar discussions with faculty at DoD educational institutions such as the Defense Acquisition University (DAU), the Naval Post Graduate School, and the Air Force Institute of Technology,

Classroom experiments: We are currently designing classroom experiments with graduate students in Carnegie Mellon's Master of Software Engineering degree program. Initial discussions are underway for similar studies with systems engineering graduate students at the University of Arizona. We also are considering following up soon with our faculty colleagues at DAU; working with experienced DoD personnel taking continuing education/in-service refresher training courses would be especially useful in achieving valid, generalizable experimental results.

Like other large-scale interventions, our overall approach to early estimation clearly is too unwieldy for controlled experimentation. However designed experimental methods can be applied to the individual steps of our overall method. Likely experiments in the long run would use real program histories to compare student solutions with actual program results for selected steps of our overall method (e.g., in identifying program change drivers).

A series of panel studies are underway at Carnegie Mellon, where we are tracking the results of calibration training aimed at improving individuals' capabilities to make accurate judgments under uncertain conditions. These will be followed by a series of experiments on the effectiveness of group consensus and algorithmic methods to reconcile differences in judgment among individuals working in small groups.

6.3 Conclusion

Extensive cost overruns have been endemic in defense programs for many years. These overruns have often been associated with optimistic expectations about achievable program scope that can be delivered on schedule and within budget. The problem has been exacerbated by the fact that a great deal of uncertainty typically exists about large-scale, unprecedented systems that take years to develop and deploy. Needed capabilities and yet-to-be-developed technical solutions are not yet well understood, both before and after Milestone A certification. And the costs are compounded when very large programs are cancelled after millions or even billions have already been spent on systems that are never delivered.

Cost estimates for unprecedented systems must rely heavily on expert judgments made under uncertain conditions. QUELCE aims to reduce the adverse effects of that uncertainty by making it explicit. Important program change drivers and the dependencies among them that may not otherwise be considered are made explicit to improve the realism and likely accuracy of the estimates. The basis of an estimate is documented explicitly, which facilitates updating the estimate during program execution and helps others to make informed judgments about their accuracy. We explicitly consider variations in the range of possible states of the program change drivers that may occur under different likely scenarios, as specified by the involved domain experts. Hence our use of probabilistic methods combining Bayesian Belief Systems and Monte Carlo simulation places the cost estimates within what may prove to be a more defensible range of uncertainty than heretofore has been possible.

To formulate QUELCE, we have leveraged our team's considerable experience with DoD acquisition programs and expertise with analytical techniques. Our experience and the results that we have achieved thus far suggest that our approach to early cost estimation has considerable merit. We look forward to refining it over time based on the results of our continuing research. More importantly, we hope to collaborate with DoD MDAPs in applying QUELCE to the cost estimation process. And we certainly welcome your ideas and participation as we move forward.

Appendix A: Rationale for Analytical Elements of the QUELCE Method

The QUELCE method integrates three proven analytical tools—the design structure matrix, Bayesian Belief Networks, and Monte Carlo simulation—in a novel way to model the uncertain program change drivers early in the acquisition lifecycle. This appendix describes the tools and provides a brief rationale for their use in the QUELCE solution.

Design Structure Matrix Technique

The Design Structure Matrix (DSM) and associated analytical techniques are frequently used to represent and analyze components of processes and products. DSM provides techniques for analyzing and restructuring a system to highlight the dependencies. The components (e.g. program change drivers) are displayed in a square matrix , rows and columns having the same names, and the values in the cells of the matrix represent the relationships among the components. The relationships may be coded to reflect simple cause and effect dependency (e.g., yes or no) or by the strength of the relationship (e.g., scale of 0 to 3). Through restructuring the matrix we can identify those components having the greatest impacts on the overall set of program change drivers. DSM thus enables us to reduce the complexity of the problem by identifying the set of program change drivers with less influence. Those drivers which exhibit little impact can be dropped from further modeling.

Another use of DSM is to help identify loops in systems. Identifying loops is important to the work here because the BBN requires an acyclic, directed graph. DSM techniques quickly identify loops within a system. While it may not be possible to eliminate all loops via DSM, it provides a means to preserve as much as possible of the original model as possible and identifying the relationships that must be eliminated via expert judgment, consolidation, or other means.

A useful resource for Design Structure Matrix for learning more about this technique can be found at http://www.dsmweb.org.

Bayesian Belief Network (BBN) Models

Probabilistic modeling is a viable alternative to statistical regression modeling as an analytical modeling approach for multiple independent and dependent variables. Members of the research team saw the opportunity to capitalize on the strengths of BBN models once the reality of the nature of data and factors during pre-Milestone A became apparent. Although BBN models are not the only method that could have been implemented, they offer a number of benefits that directly map to the challenges of the early lifecycle cost estimation. They can be used to

1. gain some analytical freedom from many of the statistical assumptions behind classical statistical methods (e.g., they are not limited to statistical regression to explain relationships between program change drivers in the BBN model)
2. create a holistic quantitative model of all program change drivers and their inter-relationships
3. predict future costs and explain or diagnose problems in prior estimates

4. analyze and model both objective and subjective data, thus capitalizing on expert judgment when historical data is not available
5. operate and make predictions with incomplete data (e.g., we may not know the status of several of the drivers such as the Program Management Contractor Relations and the Interdependency program change drivers, but we can use accepted conditional probabilistic algorithms within BBNs to update all of the unobserved program change drivers and still compute the resulting value of the BBN outcome factors)
6. make predictions for program change driver situations or scenarios that have not been experienced before
7. incorporate a learning mechanism similar to artificial intelligence in which experience with the program change driver conditions and resulting outcome factor values enable an update to the BBN relationships

In summary, the research team concluded that the use of BBNs would enable early lifecycle program change driver modeling in light of the uncertainties of data completeness and accessibility for the array of program change drivers.

Monte Carlo Simulation

Monte Carlo simulation is an uncertainty modeling method that has risen in popularity in the past 15 years with the advent of commercially available software tools such as @Risk by Palisade and Crystal Ball by Oracle. Used in the cost estimation process, this method provides the estimator with the ability to produce effort estimates with uncertainty distributions. This allows decision makers to gain insight into both the upside and downside risk of a given effort estimate. This section provides an explanation of Monte Carlo simulation and explains the application of Monte Carlo in the cost estimation process.

Figure 44 depicts a simple calculation: net income as a function of income minus expenses. Traditionally, decision makers might enter the best- and worst-case values for income and expense to get a range of values for net income. For example, the best-case values of income and expense may be $200,000 and $45,000, respectively, resulting in a net income of $155,000. Conversely, the worst-case values of income and expense may be $90,000 and $100,000, respectively, resulting in a worst-case net loss of $10,000. However, these two extremes are very unlikely to occur, and decision makers may find this large range of net income to be impractical and unsupportable. With Monte Carlo simulation, a decision-maker may now see the likely occurrence of each factor varying in context of the other for a more realistic analysis of the net income outcome.

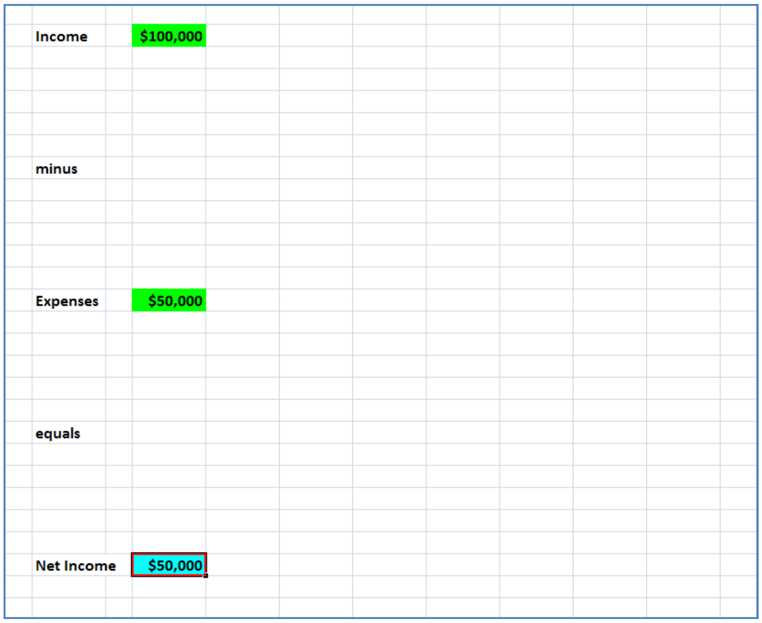


Figure 44 Example Point Calculation—Net Income

In Figure 45 the uncertainty of income is captured as an unbalanced uncertainty represented by the Gamma distribution, with a range as previously discussed of $90,000 to $200,000. Likewise in Figure 45 , notice that the uncertainty of expense is now captured as an unbalanced uncertainty represented by the Gamma distribution with a range as previously discussed of $45,000 to $100,000. Without a method such as Monte Carlo simulation, a decision maker would find great difficulty in concluding the uncertainty distribution of net income. This difficulty would be magnified in the presence of an uncertainty model with dozens of model factors predicting net income.

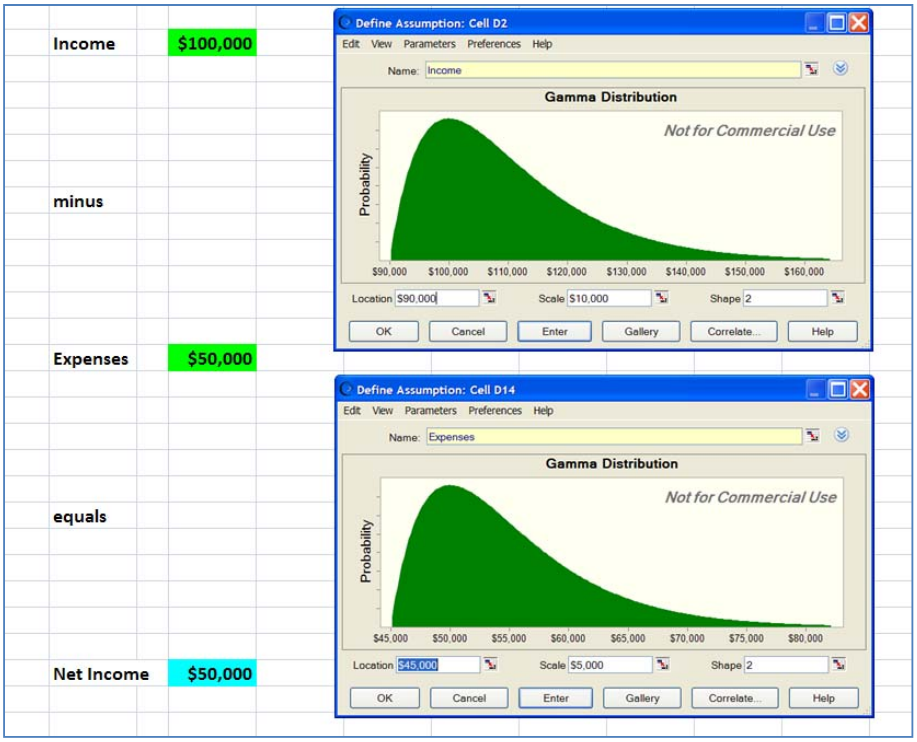


Figure 45 Example Distribution—Net Income

After identifying uncertain distributions for the income and expense factors, a Monte Carlo simulation can be conducted to identify the resulting uncertain distribution for net income. To conduct the Monte Carlo simulation, a tool such as @Risk or Crystal Ball (each are add-ons to Microsoft Excel) will use a random number generator to randomly select a value for income and a value for expense from their corresponding distributions. The resulting value for net income is calculated and saved to a log. This is considered a single trial in the Monte Carlo simulation.

The simulation can be set to run for thousands or hundreds of thousands of times, resulting in a set of values for net income that can be visualized as a distribution, as shown in Figure 46. Notice that the uncertain distribution for net income is not balanced but realistically pictures different upside and downside ranges of possible value. Also notice in Figure 46 the graph can be used to identify the lower 90% confidence limit for net income, in this case $37,764. Thus, using this value, there would only be a 10% chance of observing net income values lower than $37,764. Additionally, Figure 47 displays the statistical results of the simulation for net income and reflects a mean of $55,044 and a median of $52,798. This type of result enables the decision maker to be more informed of the uncertain behavior of the cost estimate knowing that the most likely cost estimate is $52,798 but acknowledging with 90% confidence that net income will not drop below $37,764. This is much more useful information than the traditional analysis results of worst case (-$10,000) and best case ($155,000).

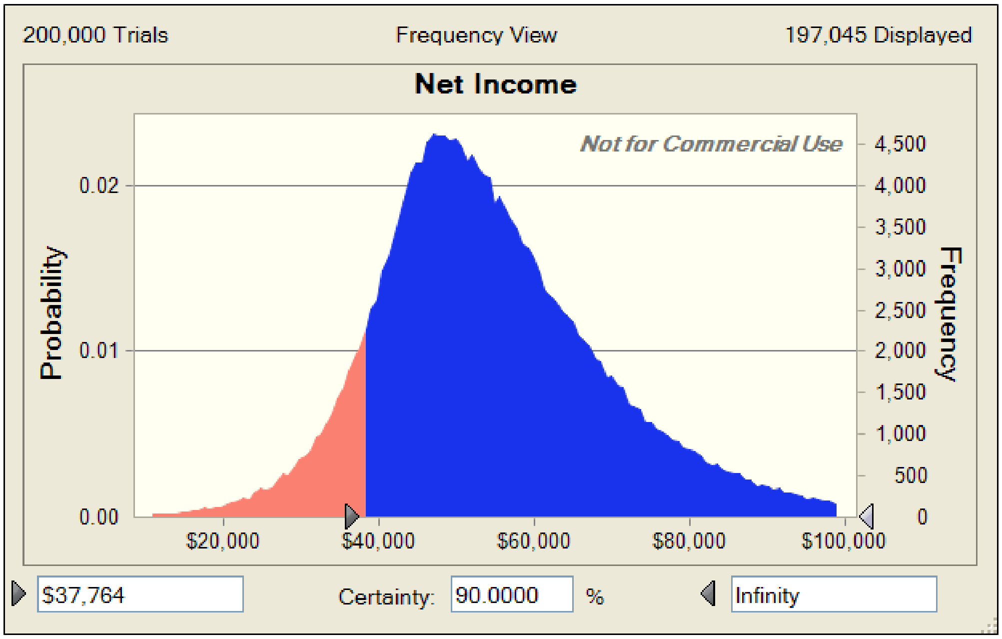


Figure 46 Net Income as a Distribution

Statistic	Forecast values
Trials	200,000
Base Case	$50,000
Mean	$55,044
Median	$52,798
Mode	---
Standard Deviation	$15,828
Variance	$250,527,448
Skewness	0.9028
Kurtosis	5.16
Coeff. of Variability	0.2876
Minimum	-$29,908
Maximum	$185,765
Mean Std. Error	$35

Figure 47 Net Income Simulation Statistical Results

Although not shown, Monte Carlo simulation enables the modeling of uncertain factors that tend to be correlated with each other. For example, the previous simulation could be re-run with the income and expense factors highly correlated so that if a random, high value was selected for income in a simulation trial then a random, high value would be selected for the expense factor, reflecting that income and expense are positively correlated factors. Modeling correlation among the model factors tends to give more accurate and credible simulation results and better reflects reality. This increased accuracy would generally result in tighter distributions for the outcome factor, net income.

As shown in the QUELCE method, Monte Carlo simulation serves a vital role in accepting distributions of three output factors of the BBN, and using that information to populate the input factors of the cost estimation model and produces a probability distribution for effort (Person-Months). Consequently, Monte Carlo simulation enables the analysis of uncertainty throughout the analytical process rather than processing single point values for factors and outcomes.

Appendix B: Program Change Drivers

This appendix contains a table of program change drivers that were included in our prototype implementation of QUELCE and one of those that were not.

Table 6 Program Change Drivers Included in the BBN

Acquisition Management	Changes in program management staff or emphasis on different aspects of program can affect performance. During TDP the actual acquisition strategy may change; in addition to other considerations, such changes might include deciding whether to add contractors or change fee structure.
Mission and CONOPS	When there is a mission or concept of operations change, the effect on the program is all-encompassing. An advocacy change can be the stimulus, as can the prospect of a conflict in a new geo-political environment. CONOPS for fighting in Kosovo were different from concerns in Afghanistan, and Iraq presented its own new challenges. Fortunately, changes to Mission and CONOPs are rare.
Capability Definition	Capability Definition (CD) is defined as "the ability to execute a specified course of action." The CD is effectively the requirements piece of Capability Based Analysis. The CD does not include the intent.
Change in Strategic Vision	Strategic vision tends to change slowly. The horizon is usually 5-20 years. Since that horizon is similar to the deployment schedule for some munitions, these changes do affect product technology and design. The potential for a change is fairly high prior to Milestone A and decreases significantly after Milestone B.
Closing Technical Gaps	Identification and closing of technical gaps is a significant source of change whenever a technology is considered for solution but is not yet ready for manufacturing and deployment. Estimators must determine how much study and experimentation will be needed to determine the technical fit and cost for a new technology.
Building Technical Capability and Capacity	While identifying a technology fit is important, it is essential that the product designers build sufficient resources to utilize and support the use of the technology. Skills, suppliers, testers, and logisticians are affected as well as the designers. These factors are all subject to change.
Interdependency	Program interdependency suggests that two or more programs are cooperating to optimize schedule or resources. If Program A is waiting on Program B and Program B is late, then Program A will also be late. Other forms of interdependency are also possible.
Interoperability	Often a system is required to interoperate with another system developed under an independent program. Any of several deficiencies or changes can affect the development effort. The greater the number of interoperable systems, the more frequently the current program will be affected. Interdependency, interoperability, and systems design tend to interact strongly. During the research project, we represented these as a single program change driver. In the future, they will probably be separated.
Functional Measures	For purposes of the initial research, we joined together the Key Performance Parameters and Technical Performance Measures. These are not identical and may not have the same change effects.

Functional Solution Criteria	In the Functional Solutions Analysis guidance [CJCSM 2007], the program office must provide estimates of performance parameters and technology readiness. During TDP, Key Performance Parameters and Technology Performance Parameters must be finalized. The estimate must consider how many experiments, trade studies and prototypes will be evaluated during this work.
Funding Schedule	Since DoD Acquisitions may take 5-10 years or more, progress payments are an essential part of the DoD acquisition process. The funding schedule itself can be a factor in contractor decisions and performance. Changes in the funding schedule often have a dramatic effect.
Program Management and Contractor Relations	The relationship is expected to be professional, efficient and effective. If the relationship deteriorates, then work is slowed by more communications, more meetings and more data calls. This pattern of cost growth was documented by Aerospace [Eslinger 2004].
Program Social Structure and Development Environment	These factors have been identified as program change drivers by benchmarking studies [Jones 2008]. They were combined during this study.
Program Management Structure and Manning at Program Office	These were combined during the current study. As program change drivers, they appear to have the same effects on program execution.
Supply Chain Vulnerabilities	Parts reaching end-of-life, sole-source relationships and many other factors can make it difficult to maintain timely access to critical parts and supplies.
Systems Design	System design develops the rules for mapping functional requirements onto the component pieces of the product to achieve acceptable performance. Changes in function definition or external performance criteria affect system design.
Program Office Process Performance	Program offices make commitments based on schedules but often do not have effective processes. In these cases, the schedule may be achieved but the quality of the result is questionable. The quality may affect a work product under review (CDRL). Also action items may not be addressed promptly between program office and other government organizations.
Production Quantity	Changes in production quantity expose the program to many consequential changes. A quantity reduction may even expose the program to a breach when a program will then commit many of its scarce resources to responding to Congressional requests for information.
Data Ownership	Critical data may belong to a contractor. If the program has not anticipated this need it may be difficult to obtain the data. Even a contractual change may be needed.
Contractor Performance	Pre-Milestone A, Contractor Performance must be assumed on the basis of past programs and industry benchmarks. During the TDP, the actual performance may be significantly different. This is one of the reasons to use ranges during the estimation process.

Table 7   Program Change Drivers Not Included in the BBN

Advocacy Change	Advocacy changes occur when the potential value of the program diminishes or increases relative to the advocate's constituency. A senior officer may drop his advocacy when he believes other sponsors will not address concerns of his service or because he believes his concerns are already addressed by the program and the funding is needed elsewhere. A member of Congress may drop sponsorship because the program provides little apparent benefit to either the service or his voters. Similarly, sponsors can be added when they are convinced that some benefit accrues to their constituency.
Scope Definition	Participants in the workshop felt that changes in scope definition were very unlikely and could not cite strong connections as a result.
Scope Responsibility	This change involves re-assigning some scope of work to a different party. Those present at the workshop knew of no such examples.
Standards/Certifications	Changes in standards and requirements for certification do occur but affect hardware more often than software. The group had no experience with effects on software.
Information Sharing	Failure to share information between the program office and contractor or among contractors is fairly common and can create significant problems. Participants agreed to include this as a program change driver but were unable to suggest a scenario that involved connection to other program change drivers.
Sustainment Issues	Sustainment is primarily a hardware and logistics concern. No strong connections were suggested in the workshop.
Contract Award	Contract award can be a program change driver in the situation where there is a protest whether successful or not. Workshop participants did not identify a related example.
Industry Company Assessment	This type of assessment is usually done when qualifying bidders on a proposal. Bidders at Milestone A usually have the necessary qualifications, so this was not identified as a program change driver.
Cost Estimate	The cost estimate becomes a program change driver if the initial estimate was too optimistic. Changing this estimate may require very high-level approval (possibly even at the Congressional level). Hence it can be a significant program change driver. In the workshop, however, it was not an important factor.
Test & Evaluation	At Milestone A it is difficult to see how test and evaluation will be a program change driver. As a program change driver, this may need additional definition. In the workshop this driver was omitted by participants.

Bibliography

[Air Force 2006]
Air Force Policy Directive 10-6, May 31, 2006. “Capabilities-Based Planning & Requirements Development.”
http://www.survivalebooks.com/free%20manuals/2006%20US%20Air%20Force%20CAPABILITIES-BASED%20PLANNING%20&%20REQUIREMENTS%20DEVELOPMENT%203p.pdf

[Air Force 2007]
The Air Force Cost Risk and Uncertainty Analysis Handbook, April 2007.

[Angelis 2000]
Angelis, L. & Stamelos, I. “A Simulation Tool for Efficient Analogy Based Cost Estimation.” *J Empirical Software Engineering 5*, 1(March 2000).

[ArgenaRisk 2011]
ArgenaRisk. “Argena: Bayesian Network and Simulation Software for Risk Analysis and Decision Support.” Argena, 2011. http://www.agenarisk.com/

[Ariely 2000]
Ariely, Dan, Bender, Randall H., Dietz, Christiane B., Gu, Hongbin, Wallsten, Thomas S., Tung Au, Wing, Budescu, David V., & Zauberman, Gal. “The Effects of Averaging Subjective Probability Estimates Between and Within Judges.” *Journal of Experimental Psychology: Applied 6*, 2 (2000) 130-147.

[Bajaj 2006]
Bajaj, Nonika, Tyagi, Alok, & Agarwal, Rakesh. “Software Estimation: a Fuzzy Approach,” *Software Engineering Notes 31*, 3 (May 2006): 1-5.

[Baretta 2009]
Baretta, James. “Gate Review Process.” *Presentation at DODCAS DON Session*, 2009.

[Baretta 2010]
Baretta, James. “Department of the Navy Probability of Program Success (PoPS).” *Presentation at DODCAS DON Service Breakout*, 2010.

[Ben-Gal 2008]
Ben-Gal, Irad. "Bayesian Networks" in *Encyclopedia of Statistics in Quality and Reliability*, John Wiley & Sons, 2008.

[Bibi 2010]
Bibi, S., Stamelos, J., Gerolimos, G. & Kollias, V. “BBN based approach for improving the software development process of an SME – a case study.” *Journal of Software Maintenance and Evolution: Research and Practice 22*, 2 (March 2010).
http://onlinelibrary.wiley.com/doi/10.1002/smr.451/abstract

[Bliss 2011]
Bliss, Gary. “Root Cause Findings After One Year,” *PARCA, 44th DoD Cost Analysis Symposium*, Feb 15-18, 2011.

[Boehm 1981]
Boehm, Barry W. *Software Engineering Economics*. Prentice-Hall, 1981.

[Boehm 2009]
Boehm, B., Abts, C., Brown, A. W., Chulani, S., Clark, B. K., Horowitz, K., Madachy, R., Reifer, D. & Steece, B. *Software Cost Estimation with COCOMO II*. Prentice Hall, 2000 (ISBN 0130266922).

[Bolger 1994]
Bolger, F. & Wright, G. “Assessing the Quality of Expert Judgment: Issues and Analysis,” *Decision Support Systems 11*, 1 (January 1994): 1-24.

[Bolten 2008]
Bolten, Joseph G., Leonard, Robert S., Arena, Mark V., Younossi, Obaid & Sollinger, Jerry M. *Sources of Weapon System Cost Growth: Analysis of 35 Major Defense Acquisition Programs.* Santa Monica, CA: RAND Corporation, 2008. http://www.rand.org/pubs/monographs/MG670.

[Book 2001]
Book, S. A. “Estimating Probable System Cost.” *Crosslink 2*, 1,Winter 2000/2001.

[Cancian 2010]
Cancian, Mark F. “Cost Growth: Perception and Reality.” *Defense AR Journal,7,* 3 (2010): 391.

[CJCSM 2007]
Chairman of the Joint Chiefs of Staff. Manual for Operation of the Joint Capabilities Integration and Development System (CJCSM 3170.01C), 2007.

[CAPE 2010]
Director, Cost Assessment and Program Evaluation, Department of Defense. *Annual Report on Cost Assessment Activities FY 2009*. February 2010

[COCOMO II 2000]
http://sunset.usc.edu/csse/research/COCOMOII/cocomo_main.html

[DAES 2006]
Defense Acquisition Executive Summary: Transformational Satellite Communications System (TSAT) , RCS: DD-A&T(Q&A)823-382, December 2006, p.5.

[Dahmann 2009]
Dahmann, J. S. & Kelley, M. *Systems Engineering During the Materiel Solution Analysis and Technology Development Phases*. Office of the Director, Defense Research and Systems Engineering, September 2009.

[Darwiche 2009]
Darwiche, Adnan. *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press, 2009 (ISBN: 978-0521884389).

[Darwiche 2010]
Darwiche, Adnan. "Bayesian Networks." *Communications of the ACM 53*, 12 (December 2010): 80-90. http://cacm.acm.org/magazines/2010/12/102122-bayesian-networks/comments

[DAU 2011]
Defense Acquisition University, DOD 5000 Defense Acquisition System. https://dap.dau.mil/aphome/das/Pages/Default.aspx

[Dixon 2007]
Dixon, Jr., Richard J. "The Case For A Knowledge Based DoD Software Enterprise: An Exploratory Study Using System Dynamics." Naval Postgraduate School Thesis, Monterey, CA, September 2007.

[DoD 2010]
Department of Defense & Department of the Navy. *Acquisition Instructions and DON Cost Estimating Standard*, 2010.

[Don 2007a]
Don, A. M. "Not So Above Average After All: When People Believe They Are Worse than Average and Its Implications for Theories of Bias in Social Comparison." *Organizational Behavior and Human Decision Processes* 102, 1(2007): 42.

[Don 2007b]
Don, A. M. & Deborah, A. S. "Error and Bias in Comparative Judgment: On Being Both Better and Worse Than We Think We Are." *Journal of Personality and Social Psychology* 92, 6(2007): 972.

[Dysert 2006]
Dysert, Larry R. "Is 'Estimate Accuracy' an Oxymoron?" *AACE International Transactions* (2006).

[Engelman 2009]
Engelman, J. B., Capra C. M., Noussair C., & Berns G. S. "Expert Financial Advice: Neurobiologically 'Offloads' Financial Decision-Making Under Risk." *Public Library of Science* ONE, 4(3) (2009): e4957.

[Erwin 2011]
Erwin, Sandra I. "Defense Watch." *National Defense Magazine,* Sept 2011.

[Eslinger 2004]
Eslinger S. and Abelson, L. A., "Recognizing Common Patterns of Software Acquisition Failures," *17th Annual Systems and Software Technology Conference.* Salt Lake City, UT, Apr. 18–21, 2005.

[Feickert 2011]
Feickert, Andrew. *The Army's Ground Combat Vehicle (GCV) and Early Infantry Brigade Combat Team (E-IBCT) Programs*. Congressional Research Service Report for Congress, July 8, 2011.

[Fenton 2009]
Fenton, N., Marquez, D., and Neil, M. "Predicting Project Velocity in XP Using a Learning Dynamic Bayesian Network Model," *IEEE Transactions on Software Engineering, 35*,1 (2009):124-137.

[Flynn 2011]
Flynn, Brian & Garvey, Paul. "Weapon System Acquisition Reform Act (WSARA) and the Enhanced Scenario-Based Method for Cost Risk Analysis," *44th Annual Department of Defense Cost Analysis Symposium*. 2011.

[Francesca 2007]
Francesca, G. & Don, A. M. "Effects of Task Difficulty on Use of Advice." *Journal of Behavioral Decision Making 20*, 1 (2007): 21.

[GAO 2008a]
General Accounting Office. *Defense Acquisitions: Results of Annual Assessment of DoD Weapon Programs, Statement of Michael J. Sullivan*, Director Acquisition and Sourcing Management, Testimony Before the Committee on Oversight and Government Reform and the Subcommittee on National Security and Foreign Affairs, House of Representatives, Tuesday, April 29, 2008, GAO-08-674T.

[GAO 2008b]
General Accounting Office. *Defense Acquisitions: A Knowledge-Based Funding Approach Could Improve Major Weapon System Program Outcomes*. Report to the Committee on Armed Services, U.S. Senate, July 2008, GAO-08-619.

[GAO 2009]
Government Accountability Office. *Cost Estimating and Assessment Guide: Best Practices for Developing and Managing Capital Program Costs.* March 2009, GAO-09-3SP.

[GAO 2010a]
Government Accountability Office. *Defense Acquisitions: Assessments of Selected Weapon Programs*. Report to Congressional Committees, March 2010, GAO-10-388SP.

[GAO 2010b]
Government Accountability Office. *Defense Acquisitions: Observations on Weapon Program Performance and Acquisition Reforms*. Testimony Before the Subcommittee on National Securi-

ty and Foreign Affairs, Committee on Oversight and Government Reform, House of Representatives United States, Wednesday, May 19, 2010, GAO-10-706T.

[GAO 2011]
Government Accountability Office. *DoD Cost Overruns: Trends in Nunn-McCurdy Breaches and Tools to Manage Weapon Systems Acquisition Costs, Statement of Michael J. Sullivan*, Director Acquisition and Sourcing Management, Testimony before the Committee on Homeland Security and Governmental Affairs, Subcommittee on Federal Financial Management, Government Information, Federal Services and International Security, United States Senate, March 29, 2011, GAO-11-499T.

[Garvey 2011]
Garvey, P. R. & Flynn, B. J. "Enhanced Scenario-Based Method (eSBM) for Cost Risk Analysis." *DoD Cost Analysis Symposium*. February 2011.

[Gino 2011]
Gino, F. & Sharek Z., et al. "Keeping the Illusion of Control Under Control: Ceilings, Floors, and Imperfect Calibration." *Organizational Behavior and Human Decision Processes 114*, 2 (2011): 104.

[Gounatidis 2006]
Gounatidis, N. "How Does the Political Nature of the Defense Acquisition Process Affect Cost Growth". Master's Thesis, Department of the Air Force, Air University, Air Force Institute of Technology (September 2006).

[Hearty 2009]
Hearty, P., Fenton, N., Marquez, D., & Neil, M. "Predicting Project Velocity in XP Using a Learning Dynamic Bayesian Network Model," *IEEE Transactions on Software Engineering 35*, 1 (Jan-Feb 2009). http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4624275

[Hihn 2004]
Hihn, Jairus & Lum, Karen T. "Improving Software Size Estimates by Using Probabilistic Pairwise Comparison Matrices," 140-150. *Proceedings of the 10th International Symposium on Software Metrics (METRICS'04)*. Chicago, IL, September 2004.

[Hill 2005]
Hill, Thomas. *Statistics: Methods and Applications.* StatSoft, Inc., 2005 (ISBN: 978-1884233593). http://www.statsoft.com/textbook/

[Hofbauer 2011]
Hofbauer, Joachim, Sanders, Gregory, Ellman, Jesse & Morrow, David. *Cost and Time Overruns for Major Defense Acquisition Programs.* Defense-Industrial Initiatives Group, Center for Strategic & International Studies, April 2011.

[Hora 2004]
Hora, Stephen C. "*Probability Judgments for Continuous Quantities: Linear Combinations and Calibration Management Science 50*, 5 (May 2004): 597-604.

[Hubbard 2010]
Hubbard, D.W. *How to Measure Anything: Finding the Value of Intangibles in Business*. Hoboken, New Jersey: John Wiley & Sons, Inc, 2nd edition, 2010.

[Hull 2009]
Hull, David. *Methods and Challenges in Early Cost Estimating.* Office of the Deputy Assistant Secretary of the Army for Cost and Economics (ODASA-CE), March, 2009.

[ISPA 2008]
International Society of Parametric Analysts and the Society of Cost Estimating and Analysis. Parametric Estimating Handbook, Fourth Edition, 2008.

[Johnson 2000]
Johnson, P. M., Moore, C.A., Dane, J. A. & Brewer, R. S. "Empirically Guided Software Effort Guesstimation," *IEEE Software 17*, 6 (January/February 2000): 51-56.

[Jones 2008]
Jones, Capers. *Applied Software Measurement: Global Analysis of Productivity and Quality*. McGraw-Hill Osborne Media, April, 2008.

[Jørgensen 2004a]
Jørgensen, Magne. "Realism in Assessment of Effort Estimation Uncertainty: It Matters How You Ask," *IEEE Transactions on Software Engineering,30,* 4 (April 2004):209-217.

[Jørgensen 2004b]
Jørgensen, Magne, Teigan, Karl Halvor, & Moløkken, Kjetil. "Better Sure than Safe? Overconfidence in Judgement Based Software Development Effort Prediction Intervals," *Journal of Systems and Software,70,* 1-2 (February 2004):79-93.

[Jørgensen 2005]
Jørgensen, M. "Practical Guidelines for Expert-Judgment-Based Software Effort Estimation." *IEEE Software 22*, *3*(May-June 2005): 57-63.

[Jørgensen 2007]
Jørgensen, M. & Shepperd, M. "Systematic Review of Software Development Cost Est Studies," *IEEE Transactions On Software Engineering 33*, 1 (January 2007): 33-53.

[Hamdan 2009]
Hamdan, K., Bibi, S., Angelis, L., & Stamelos, I. "A Bayesian Belief Network Cost Estimation Model that Incorporates Cultural and Project leadership Factors." *IEEE Symposium on Industrial Electronics and Applications*. Kuala Lumpur, Malaysia, October 2009.

[Kaminski 2009]
Kaminski, Dr. Paul G. Statement before the Committee on Armed Services, U.S. Senate, March 3, 2009.

[Leotta 2010]
Leotta, J. *Software Cost Estimating Relationships*. Navy Engineering Logistics Office (2010)

[Lindemann 2010]
Lindemann, Udo, Maurer, Maik & Braun, Thomas. *Structural Complexity Management: An Approach for the Field of Product Design.* Springer, 2010.

[Lindgren 2009]
Lindgren, M. and Bandhold, H. *Scenario Planning—Revised and Updated Edition: The Link Between Future and Strategy.* Macmillan, 2009.

[Long 2007]
Long, E. Andrew, Forbes, James, Hees, Jing & Stouffer, Virginia. *Empirical Relationships between Reliability Investments And Life-Cycle Support Costs* (Report Sa701t1). LMI Government Consulting, 2007.

[Meyer 2001]
Meyer, Mary A. & Booker, Jane M..*Eliciting and Analyzing Expert Judgment: A practical Guide.* Society for Industrial and Applied Mathematics, 2001. ISBN 0-89871-474-5.

[Miranda 2001]
Miranda, Eduardo. "Improving Subjective Estimations Using Paired Comparisons" *IEEE Software Magazine 18,* 1 (January 2001).

[Mishra 2011]
Mishra, H., Mishra A., & Shiv, B. "In Praise of Vagueness: Malleability of Vague Information as a Performance Booster", *Psychological Science 22*, 6 (June 2011): 733-738.

[Moløkken-Østvold 2004]
Moløkken-Østvold, Kjetil & Jørgensen, Magne. "Group Processes in Software Effort Estimation." *Empirical Software Engineering 9* (2004): 315–334.

[Moløkken-Østvold 2005]
Moløkken-Østvold, K. & Jørgensen, M. "Expert Estimation of the Effort of Web-Development Projects: Why Are Software Professionals in Technical Roles More Optimistic Than Those in Non-technical Roles," *Empirical Software Engineering*, *10*, 1 (June 2005): 7-30.

[Monarch 2009]
Monarch, I., Goldenson, D., & Stoddard, R. "Mapping the Dynamics of Software Engineering: The Case of Software Estimation. *Proceedings of the 2009 COCOMO Forum*. November 2009.

[Mooney 1997]
Mooney, Christopher Z. *Monte Carlo Simulation (Quantitative Applications in the Social Sciences)*. Sage Publications, 1997 (ISBN: 978-0803959439).

[NCCA 2008]
Naval Center for Cost Analysis and Air Force Cost Analysis Agency, Software Technology Support Center. *Software Development Cost Estimating Handbook Volume I,* September 2008.

[NCCA 2009]
Naval Center for Cost Analysis and Air Force Cost Analysis Agency, Software Technology Support Center, Cost Analysis Group. *Software Development Cost Estimating Guidebook*, 2009.

[NRC 2008]
National Research Council, Committee on Pre-Milestone A Systems Engineering. *A Retrospective Review and Benefits for Future Air Force Systems Acquisition*, 2008 (ISBN: 0-309-11476-4).

[NRC 2011]
Committee on Evaluation of U.S. Air Force Preacquisition Technology Development, National Research Council. *Evaluation of U.S. Air Force PreacquisitionTechnology Development*, 2011. (ISBN: 0-309-16276-9).

[Oracle 2011]
Oracle. "Oracle Crystal Ball." Oracle, 2011.
http://www.oracle.com/us/products/applications/crystalball

[OSD 2007]
Office Of The Secretary Of Defense, Cost Analysis Improvement Group. *Operating And Support Cost-Estimating Guide*, October 2007.

[Pendharkar 2005]
Pendharkar, P. C., Subramanian, G. H. & Rodger, J. A. "A Probabilistic Model for Predicting Software Development Effort." *IEEE Transactions on Software Engineering 31*, 7 (July 2005).
http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1492375&tag=1

[Pfleeger 2005]
Pfleeger, Shari Lawrence, Wu, Felicia & Lewis, Rosalind. *Software Cost Estimation and Sizing Methods: Issues and Guidelines*. Santa Monica, CA: RAND Corporation, 2005.
http://www.rand.org/pubs/monographs/MG269.

[POPs 2010]
Naval PoPs Guidebook Version 2.2., Guidance for the Implementation of Naval PoPS.
A Program Health Assessment Methodology For Navy and Marine Corps Acquisition Programs, 2010.

[Roper 2007]
Roper, Martha. *Pre-Milestone-A Decision-Making: Can We Cost Capabilities?* Office of the Deputy Assistant Secretary of the Army for Cost and Economics, Defense Acquisition University, 2007.

[Roper 2010]
Roper, Martha A. *Pre-Milestone A Cost Analysis: Progress, Challenges, and Change*. Defense Acquisition University, 2010. https://acc.dau.mil/CommunityBrowser.aspx?id=428598

[RTO 2007]
Methods and Models for Life Cycle Costing (*Méthodes et modèles d'évaluation*

du coût de possession). RTO Technical Report TR-SAS-054, Final Report of Task Group SAS-054, June 2007.

[Saaty 1980]
Saaty, Thomas L. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation.* McGraw-Hill, 1980.

[Shepperd 2001]
Shepperd M. & Cartwright, M. “Predicting with Sparse Data.” *IEEE Transactions on Software Engineering 27*, 11 (November 2001): 987-998.

[Stephen 2004]
Stephen, C. H. "Probability Judgments for Continuous Quantities: Linear Combinations and Calibration." *Management Science 50,* 5(2004): 597.

[Unal 2004]
Unal, Resit, Keating, Charles, Conway, Bruce & Chytka, Trina. *Development of an* Expert *Judgement Elicitation Methodology Using Calibration and Aggregation for Risk Analysis in Conceptual Vehicle Design* (ODU PROJECT NO: 130012). Old Dominion University, 2004.

[Valerdi 2011]
Valerdi, Ricardo, “Convergence of Expert Opinion via the Wideband Delphi Method: An Application in Cost Estimation Models.” *21st Annual INCOSE International Symposium*. Denver, CO, June 2011.

[Wallsren 1997]
Wallsten, Thomas S., Budescu, David V., Erev, Ido & Diederich, Adele. “Evaluating and Combining Subjective Probability Estimates.” *Journal of Behavioral Decision Making 10,* 3 (September 1997): 243-268.

[Winkler 2004]
Winkler, Robert L. & Clemen, Robert T. “Multiple Experts vs. Multiple Methods: Combining Correlation Assessments.” *Decision Analysis 1*, 3 (September 2004): 167-176.

[Younossi 2007]
Younossi, Obaid, Mark V. Arena, Robert S. Leonard, Charles Robert Roll, Arvind Jain and Jerry M. Sollinger. *Is Weapon System Cost Growth Increasing? A Quantitative Assessment of Completed and Ongoing Programs*. Santa Monica, CA: RAND Corporation, 2007. http://www.rand.org/pubs/monographs/MG588.

[Younossi 2008]
Younossi, Obaid , Lorell, Mark A., Brancato, Kevin, Cook, Cynthia R., Eisman, Mel, Fox, Bernard, Graser, John C., Kim, Yool, Leonard, Robert S., Pfleeger, Shari Lawrence & Sollinger, Jerry M. *Past Lessons and Future Recommendations.* Santa Monica, CA: RAND Corporation, 2008. http://www.rand.org/pubs/monographs/MG690.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. **AGENCY USE ONLY**

2. **REPORT DATE**
December 2011

3. **REPORT TYPE AND DATES COVERED**
Final

4. **TITLE AND SUBTITLE**
Quantifying Uncertainty in Early Lifecycle Cost Estimation (QUELCE)

5. **FUNDING NUMBERS**
FA8721-05-C-0003

6. **AUTHOR(S)**
Robert Ferguson, Dennis Goldenson, James McCurley, Robert Stoddard, David Zubrow, Debra Anderson

7. **PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

8. **PERFORMING ORGANIZATION REPORT NUMBER**
CMU/SEI-2011-TR-026

9. **SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
HQ ESC/XPK
5 Eglin Street
Hanscom AFB, MA 01731-2116

10. **SPONSORING/MONITORING AGENCY REPORT NUMBER**
ESC-TR-2011-026

11. **SUPPLEMENTARY NOTES**

12A **DISTRIBUTION/AVAILABILITY STATEMENT**
Unclassified/Unlimited, DTIC, NTIS

12B **DISTRIBUTION CODE**

13. **ABSTRACT (MAXIMUM 200 WORDS)**

Difficulties with estimating the costs of developing new systems have been well documented, and are compounded by the fact that estimates are now prepared much earlier in the acquisition lifecycle, before there is concrete technical information available on the particular program to be developed. This report describes an innovative synthesis of analytical techniques into a cost estimation method that models and quantifies the uncertainties associated with early lifecycle cost estimation.

The method described in this report synthesizes scenario building, Bayesian Belief Network (BBN) modeling and Monte Carlo simulation into an estimation method that quantifies uncertainties, allows subjective inputs, visually depicts influential relationships among program change drivers and outputs, and assists with the explicit description and documentation underlying an estimate. It uses scenario analysis and design structure matrix (DSM) techniques to limit the combinatorial effects of multiple interacting program change drivers to make modeling and analysis more tractable. Representing scenarios as BBNs enables sensitivity analysis, exploration of scenarios, and quantification of uncertainty. The methods link to existing cost estimation methods and tools to leverage their cost estimation relationships and calibration. As a result, cost estimates are embedded within clearly defined confidence intervals and explicitly associated with specific program scenarios or alternate futures.

14. **SUBJECT TERMS**
Cost estimation, pre-milestone A, scenario building, design structure matrix, bayesian belief network, monte carlo simulation, calibration of expert judgment, probabilistic modeling, uncertainty

15. **NUMBER OF PAGES**
101

16. **PRICE CODE**

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL