DevTreks –social budgeting that improves lives and livelihoods

Conservation Technology Assessment (CTA) 1 Introduction

Last Updated: September 13, 2018; First Released: January 09, 2015

Author: Kevin Boyle, President, DevTreks

Version: DevTreks 2.1.6

A. Introduction

This reference introduces techniques for analyzing the risk and uncertainty associated with Resource Stock and Monitoring and Evaluation (M&E) Indicators and communicating these results in terms of likelihoods and probabilities. These techniques will be used to analyze the risk and uncertainty associated with the Resource Stock and M&E data explained in the *Resource Stock Calculation and Analysis* and the *Monitoring and Evaluation* references. The Stock references define CTA as follows:

> *Conservation Technology Assessment is the analysis of resource stock flows and balances, and conservation technologies that are designed to prevent or correct imbalances in the stocks.*

| Section	Page
Introduction to Risk and Uncertainty	2
Statistical and Probabilistic Risk Analysis	3
Cost, Benefit, and Performance Uncertainty	4
CTA Analysis	5
Online Algorithms	6
Mathematical and Statistical Libraries	6
Communicating CTA Results	6
CTA Algorithms	7
Sustainable CTAs and the RCA Value Framework	8
Summary and Conclusions	9
Appendix A. Supplemental Analyses	15

B. Introduction to Risk and Uncertainty

When NASA launches a space explorer to Mars, they carefully assess the likelihood that the space ship will accomplish its mission (see US, NASA 2011). When a country's defense agency builds a plane or ship, they carefully assess the riskiness of the plane crashing, or ship sinking (see US GAO, 2009). When many countries assess the effectiveness of a new health care technology, they carry out a Health Technology Assessment that carefully assesses the probability of the medical treatment working (see WHO, 2011). What do all these "high priced" assessments have in common? They use formal risk analysis to quantify probable outcomes. The risk analysis plays a central decision support role that determines how large sums of money get allocated, and lives saved, based on the quantified results.

When a local extension agent travels to a local dairy to assess a problem with milk production, they make recommendations for improvements based on the dairies they have worked with in the past and on recent professional dairy publications they have read. When a local low income housing agency builds a new low income housing project, they assess the likelihood of success based on previous experience building similar types of projects. When a local family doctor makes a recommendation for a prescription, they base the prescription on the guidance offered by their professional associations and on what they have found to work in the past. What do all of these "low priced" services have in common? Much less reliance on quantitative decision support tools, and much more reliance on tacit knowledge gained from experience and study (see Kahneman, 2011, for further explanation about these types of decisions).

What do "low and high priced" risk analysis share in common? Most of the data used to support decisions has been lost or can't be accessed in a way that easily supports similar decisions that will be faced by future decision makers. This reference introduces a set of risk analysis tools that can help "high or low priced" services and projects quantify the likelihood of work outcomes.
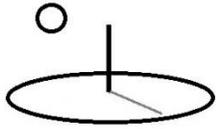
That structured data can be added to both online, and tacit, knowledge banks. Clubs that desire quantified tools will have concrete, online, tools and knowledge banks available to help guide their decisions. Clubs that rely on their experience will have easier access to additional background data that goes into their tacit knowledge banks. Future decision makers will have easier access to historical data that can assist them with current problems.

C. Statistical and Probabilistic Risk Analysis

This tutorial relies on the references for fuller explanation of the mathematics and analytic techniques used in statistical and probabilistic risk analysis (PRA). The references do a professional job defining key risk assessment concepts (Probability), formulas (Lognormal Distributions), methodologies (Monte-Carlo), and reporting techniques (Probability Density Functions). **Appendix A** summarizes their major points. Most of the references recommend that PRAs should be reported using confidence intervals for uncertain numbers, along with probability density functions. Sibling references include examples demonstrating how these recommendations can be carried out.

PRAs can generate completely wrong numbers when models used to analyze data are defined badly, the data used in the models is spurious, or wrong assumptions are made about the distribution of indicators. The likelihood of conducting a bad PRA can be reduced by closely following the recommendations found in the major references (i.e. WHO, IPCC, NASA).

In additional to PRAs, a variety of statistical risk analysis techniques can be used to determine the uncertainty of resource stock data. The following list describes some of these assessment techniques (US OMB 2006, World Bank 2012). Many of these techniques overlap. References and algorithms that demonstrate these techniques will be included in future releases.

Personal, Social, and Managerial Decision Support: Data associated with personal preferences, societal biases, and subjective probabilities, is collected and analyzed to determine how to improve personal, social, and managerial decisions. For example, researchers have found
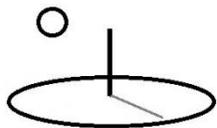
that investors weigh losses twice as high as equivalent gains and tend to overestimate low probability events while underestimating high probability events.

Damage Assessments and Disaster Risk Reductions: Historical and simulated data is used to assess the probable damages associated with various intensity levels of natural disasters. For example, flood damage assessments show the probable damages associated with various depths of flooding. The Social Performance Analysis tutorial has examples of algorithms that demonstrate Disaster Risk Reduction techniques.

Actuarial Analysis of real-world data: Historical data is analyzed using standard statistical methods. For example, historic data on car crashes has been analyzed to determine the likelihood of safety features working. The Social Performance Analysis tutorial has examples of algorithms that demonstrate formal Impact Evaluations that use these techniques.

Dose-Response Analysis using experimental data: Experiments are conducted on humans or animals to determine how they react to various doses of substances. For example, experiments on rodents have been conducted to determine the likelihood that threshold amounts of substances lead to the development of cancer tumors.

Infectious Disease and Epidemic Modeling: The occurrence and severity of epidemics are simulated using computer algorithms. For example, the spread of the HIV virus has been modeled to understand transmission pathways among select groups of people. The Social Performance Analysis tutorial includes examples that use introductory population algorithms to examine the "epidemiology" of social performance data.

Failure Analysis of Physical Structures: The failure of physical structures is commonly analyzed using low risk and high consequence statistical techniques. For example, concrete floors are commonly analyzed to determine whether they can withstand earthquakes.

Appendix A presents a framework for conducting risk analysis. **Appendix B** contains a list of additional analytic techniques, such as Sensitivity Analysis, that can be used to support, or replace, PRAs.
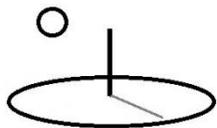
D. Cost, Benefit, and Performance Uncertainty (1*)

A key requirement of CTA is to tie base element economic cost and benefit amounts to indicator amounts. That allows common CTA Performance Analysis techniques, such as cost effectiveness analysis, to be supported. The NIST (1988) and World Bank (2013) reference introduces traditional approaches to account for risk when conducting economic evaluations. Those references discuss how risk preferences influence decision making. The World Bank and United Nations (2010) reference explains how to tie physical science indicators associated with climate change, such as projected flood depths, to economic damages. The NASA (2008) and GAO (2009) references cover cost risk analysis at greater depth.

The Resource Stock Calculation reference demonstrates how to use indicators to calculate uncertain costs and benefits. Several examples in this reference also demonstrate how to calculate uncertain costs and benefits. As uncertainty is incorporated into more calculators, additional techniques for analyzing uncertain costs and benefits will be explained. For example, *The Monitoring and Evaluation* tutorials demonstrate how to use Indicators to measure risk and uncertainty for each, separate, base element.

E. CTA Analysis (2*)

All CTA algorithms use Resource Stock Calculators and M&E Calculators and therefore their calculations can be analyzed further using Resource Stock Analyzers and M&E Analyzers. When the data being analyzed is observational data stored in Data URL datasets, these analyzers produce automated metadata analysis (analysis of analyses). Metadata analysis of randomized control trial data is the primary technique employed in Health Technology Assessments (HTAs) (see Example 1m below, DevPacks Stock Analysis, in the *Resource Stock Analysis* reference for examples).

Keep in mind the conventions used by those analyzers. For example, the CTA-01 reference includes examples demonstrating that Operations and Operating Budgets use Input.OCAmount as a multiplier for Input *Stock and M&E* calculations, while Components and Capital Budgets

use Input.CAPAmount as the multiplier. Even though most algorithms are run asynchronously and simultaneously, keep in mind that this technology has limits on what data can be realistically processed (and full tests have not been run to find out yet). This version does not support cancelling an analysis and does not show numeric progress during analyses.

F. Online Algorithms

DevTreks recommends limiting the actual analyses that can be accessed from online production servers to proven algorithms. We recommend using development servers, whether online or offline, for testing algorithms prior to putting them "into production". When possible, use validated algorithms, containing known indicator and score analysis properties, including vetted Math Expression and scripts. For example, Algorithm 1 includes a basic regression analysis algorithm that can be investigated using standard statistical packages. Algorithms 2 and 3 use mathematical and statistical libraries that can be tested on a variety of platforms before being made available through DevTreks. After the analyses have been satisfactorily tested on the development server, see if the final algorithms can be run, or at least displayed, on the production server. If not, make recommendations for improvements to the algorithms and online tools.

G. Mathematical and Statistical Libraries (3*)

Algorithms that use statistical libraries, such as the R Project and Python, began appearing in version 1.8.2. Virtual machine-based libraries showed up in version 2.0.2. The current version runs most of these libraries as command line executables or through web apis (Application Programming Interfaces) and web services. That introduces some latency, or slowness, as executables load associated libraries and urls get routed to and from Internet destinations. The related CTA references demonstrate that some of the statistical packages, such as R Server, are starting to address scalability.

H. Communicating CTA Results
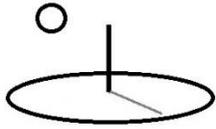
Appendix A demonstrates that the analytic results of CTA can be explained using techniques that vary from simple tables to complicated graphics. URLs referenced with R Project and Python algorithms demonstrate the power of using graphics to explain statistical results. The risk assessment framework introduced in that Appendix explains the importance of basing the CTA analysis on the objectives of stakeholders and then communicating the results of the CTA in a manner that meets stakeholder needs. Most of the climate change reference demonstrate the value of analyzing and communicating some types of risks using GIS techniques. Future releases will improve the tools available for communicating the results of CTA.

I.   CTA Algorithms (4*)

All CTAs are completed using automated, quantitative, algorithms. The algorithms are defined using Indicator or Score's Math Type (algorithm1) and Sub Math Type (subalgorithm1) properties. The Math Type property must select an option identifying the statistical library to use, using a generic name such as "algorithm1" or "algorithm2". The Sub Math Type property must identify the name of the subalgorithm within the library to use, using a generic name such as "subalgorithm1" or "subalgorithm2". The Social Performance Analysis tutorial introduces Machine Learning algorithms which use an underscore in their name, such as "subalgorithm_01"

Not every algorithm supports running the algorithm for both indicators and scores. For example, algorithms that run regression analysis require TEXT datasets, but Scores for Resource Stock calculations do not currently use datasets and therefore can't use that algorithm. The Resource Stock Calculation reference explains properties of CTA-related calculators. The data standards defined in that reference, particularly data column names and Math Expression terms, must be followed exactly for these algorithms to work. In addition, most algorithms rely on text data stored in URLs. Those URLs must contain the data expected by the algorithm. The convention of many of those TEXT files is to use the last 3 rows of data as scoring datasets (i.e. used to generate a sensitivity analysis of confidence intervals) and the rest of the rows as the training dataset.
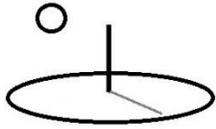
Datasets must be cleaned up prior to using them with any CTA algorithm. Unless an algorithm shows a different example, all data matrixes must have column vectors that have the same number of rows (positive definite) –if not, zeros will be inserted to make them that way. The manipulation of very large matrixes presents technical software challenges that will continue being addressed (i.e. by using algorithms that don't rely on matrix manipulation or that use multi-threaded mathematical libraries, such as algorithm2's Intel Math Kernel library).

The following algorithms are currently available.

algorithm1 (MathNet and System.Math): The sibling reference, CTA Algorithm 1, gives examples of subalgorithms used with this algorithm. The examples use probabilistic statistics that employ the open source MathNet and System.Math mathematical libraries.

algorithm2 (R): The sibling reference, CTA Algorithm 2, gives examples of subalgorithms used with this algorithm. The examples use probabilistic statistics that employ the open source R statistical library.

algorithm3 (Python): The sibling reference, CTA Algorithm 3, gives examples of subalgorithms used with this algorithm. The examples use probabilistic statistics that employ the open source Python statistical library.

algorithm4 (AML): The sibling reference, CTA Algorithm 4, gives examples of subalgorithms used with this algorithm. The examples use probabilistic statistics that employ the Azure Machine Learning statistical library.

algorithm5 (**Display**): The sibling reference, CTA Algorithm 5, gives examples of subalgorithms used with this algorithm. The examples display probabilistic statistics generated by any statistical library. Metadata analysis of those results can be carried out using Resource Stock and M&E Analyzers.
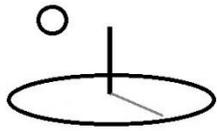
Algorithm6 (**Julia**): Version 2.0.2 stubbed out most of the source code to support this open source statistical library, but it hasn't been tested yet. The main point is to reinforce the importance of economies of scope in this type of IT product. A future release will add a CTA 6 reference to demonstrate this algorithm.

Future releases will include additional algorithms, examples, and statistical libraries.

J. Sustainable CTAs and the RCA Value Framework

The Performance and Social Performance Analysis tutorials use the term, Resource Conservation Accounting Value (RCA Value Framework), to introduce an underlying scientific framework for carrying out CTAs. The framework makes heavy use of "social impact pathways" and "stakeholder value perspectives" to understand cause and effect attribution related to business and social actions. Better understanding of what works and doesn't work helps businesses, stakeholders, and communities, to take more effective mitigation and adaptation actions in support of Sustainable Development Goals and Sendai Disaster Risk Reduction goal accomplishment.

Versions 2.0.4+ began focusing on algorithms and techniques that help producers make sustainable production choices and consumers to make sustainable consumption choices. The techniques introduced in the Social Performance Analysis (SPA) tutorial include Life Cycle Impact Analysis, Product Life Cycle Analysis, Organization Life Cycle Analysis, Social Life Cycle Analysis, Life Cycle Cost and Benefit Budgeting, Incremental Cost Effectiveness Analysis, and custom algorithms related to population impact states, Performance Monitoring, Impact Evaluation, and Machine Learning.

These techniques support international efforts to improve sustainability as codified in the SDG. The following statement (from the SPA tutorial) summarizes the expected outcome from this approach.
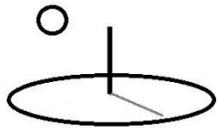
Stakeholders, investors, asset managers, supply chain participants, consumers, employees, and impacted communities, can more easily assess the degree of SDG target accomplishment, or "social soundness", achieved by specific communities, companies, and their executives. Companies and communities can anticipate a large proportion of stakeholders, or more accurately, informed stakeholders, to act accordingly.

The Social Performance and DevPacks tutorials introduce examples demonstrating how to determine the sustainability of products in all parts of supply chains, as well as the actions taken by company executives in support of sustainability. This increased accountability helps consumers and supply chain participants to eliminate products, goods, and services that come from companies who employ unsustainable business practices, or that are associated with company executives who undermine SDG accomplishments.

Although DevTreks is not consumer software, it allows sustainability workers to provide evidence that helps consumers to make informed, accountable, sustainable, product choices. The sooner that consumers adopt the recommendations from these improved accountability systems, the greater is the likelihood of societal improvement through SDG goal accomplishment.

Summary and Conclusions

This reference introduces CTAs. CTAs may help Indian smallholders to adapt effectively to changing monsoons, Somalian health care administrators to deliver more cost effective medical treatments, Philippine city managers to mitigate rising typhoon intensity more efficiently, California vegetable growers to manage water wisely, Egyptian municipalities to manage energy distribution better, Chilean mariners to adopt sustainable fish stock management practices faster, domestic consumers to make sustainable consumption choices more easily, and people to improve their lives and livelihoods in sustainable ways.

Footnotes

1. Version 1.7.6 started using mathematical parsers to calculate mathematical expressions and mathematical libraries to carry out analysis. Version 1.8.0 changed the default

software pattern used in calculators and analyzers to accommodate heavier traffic loads and larger datasets, but this pattern still requires full testing with large datasets. Version 1.8.2 added statistical libraries. Version 2.0.2 added the option of running the libraries from virtual machines. Versions 2.0.4 and 2.0.6 allowed all of the CTA algorithms to be used with all of the Monitoring and Evaluation calculators. Version 2.1.4's Social Performance algorithms upgraded calculator patterns to rely more heavily on TEXT datasets stored in Indicator.URL and Score.URL properties. The full capability of these types of tools extends well beyond the examples in this reference. Future releases will continue to find and use improved software patterns.

2. Many algorithms generate results that are not intuitive. It's hard to be confident of their accuracy without third party, vetted, reference datasets. The point is that algorithms cannot simply be used as black boxes that magically spit out answers. All of the algorithms that are used in this reference will continue to be researched, improved, or replaced, as more knowledge is gained about their workings.

3. We investigated using additional open source, or free, statistical libraries. Two algorithms were built using a Bayesian statistical software library (Minka et al, 2014) prior to developing algorithms based on scripting languages such R and Python. We decided not to release those algorithms for now because the scripting languages are popular. Related open source statistical software libraries, such as OpenBugs and STAN, were investigated and may be supported in future releases.

4. Several tutorials point out that, while it may be the role of a software development company like DevTreks to build and document basic algorithms, that full-time role is the responsibility of social networks and their clubs. Users of this source code are encouraged to build and document their own algorithms targeting the needs of their own networks, clubs, and members.

5. A performance commitment defines the acceptable risk tolerance as defined by the decision makers. NASA (2011) describes this amount as follows: "A performance commitment is a performance measure value set at a particular percentile of the performance measure's pdf, so as to anchor the decision-maker's perspective to that performance measure value as if it would be his/her commitment, were he/she to select
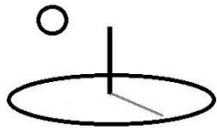
that alternative. For a given performance measure, the performance commitment is set at the same percentile for all decision alternatives, so that the probability of failing to meet the different alternative commitment values is the same across alternatives." The media used with the URLs used in the CTA-01 references includes a graphic of the data's cumulative density function with a potential performance commitment shown on the graphs.

References

Ireland, Health Information and Quality Authority. Guidelines for the Economic Evaluation of Health Technologies in Ireland. 2010 (www.hiqa.ie)

Lurie, Philip, Goldberg Matthew and Robinson, Mitchell. US Institute for Defense Analyses. A Handbook of Cost Risk Analysis Methods. April, 1993

Kahneman, Daniel. Thinking, Fast and Slow. Farrar, Straus, and Giroux. New York, USA. 2011

T. Minka, J. Winn, J. Guiver, S. Webster, Y. Zaykov, B. Yangel, A. Spengler, and J. Bronskill. Infer.NET 2.6, Microsoft Research Cambridge, 2014. http://research.microsoft.com/infernet

U.S. Air Force. Cost Risk and Uncertainty Analysis Handbook. 2007

U.S. Department of Commerce. National Institute of Standards and Technology. Engineering Statistics Handbook, http://www.itl.nist.gov/div898/handbook/index.htm (last accessed on the web in May, 2015)

U.S. Department of Commerce. National Institute of Standards and Technology. Techniques for Treating Uncertainty and Risk in the Economic Evaluation of Building Investments. NIST Special Publication 757. 1988

US Government Accountability Office. Applied Research and Methods. GAO Cost Estimating and Assessment Guide. Best Practices for Developing and Managing Capital Program Costs. March, 2009.
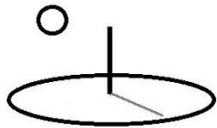
U.S. National Aeronautical and Space Administration. NASA Risk Management Handbook, NASA/SP-2011-3422, Version 1.0, November, 2011.

U.S. National Aeronautical and Space Administration. Probabilistic Risk Assessment Procedures Guide for NASA Managers and Practitioners, NASA/SP-2011-3421, Version 2.0, December, 2011.

U.S. National Aeronautical and Space Administration. 2008 NASA Cost Estimating Handbook, 2008.

U.S. Office of Management and Budget. Proposed Risk Assessment Bulletin. 2006

Westen, Alkema, Damen, Kerle, and Kingma. Multi-hazard risk assessment, Distance Education Course, Guide Book. United Nations University. ITC School on Disaster Geoinformation Management. 2011

World Bank. Risk Analysis Course Manual. Global Facility for Disaster Reduction and Recovery. 2011

World Bank and United Nations. Natural hazards, unnatural disasters: the economics of effective prevention. 2010

World Health Organization. Health Technology Assessment of Medical Devices. 2011

References Note

We try to use references that are open access or that do not charge fees.

Improvements, Errors, and New Features

Please notify DevTreks (devtrekkers@gmail.com) if you find errors in these references. Also please let us know about suggested improvements or recommended new features.

A video tutorial explaining this reference can be found at:
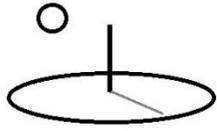
https://www.devtreks.org/commontreks/preview/commons/resourcepack/Technology Assessment 1/1526/none
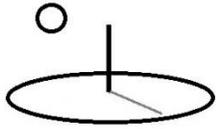
Appendix A. Probabilistic Risk Analysis (PRA) Introduction

The following three images introduce a framework (NASA, 2011) for analyzing risk and uncertainty in a project, program, or technology:


Risk-Informed Decision Making (RIDM)

Part 1 - Identification of Alternatives
Step 1 – Understand Stakeholder Expectations
and Derive Performance Measures
Step 2 – Compile Feasible Alternatives

Part 2 - Risk Analysis of Alternatives
Step 3 – Set the Framework and Choose the
Analysis Methodologies
Step 4 – Conduct the Risk Analysis and
Document the Results

Part 3 - Risk-Informed Alternative Selection
Step 5 – Develop Risk-Normalized Performance
Commitments
Step 6 – Deliberate, Select an Alternative, and
Document the Decision Rationale

To Requirements (Re)Baselining

Figure 13. RIDM Process Steps
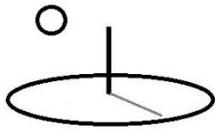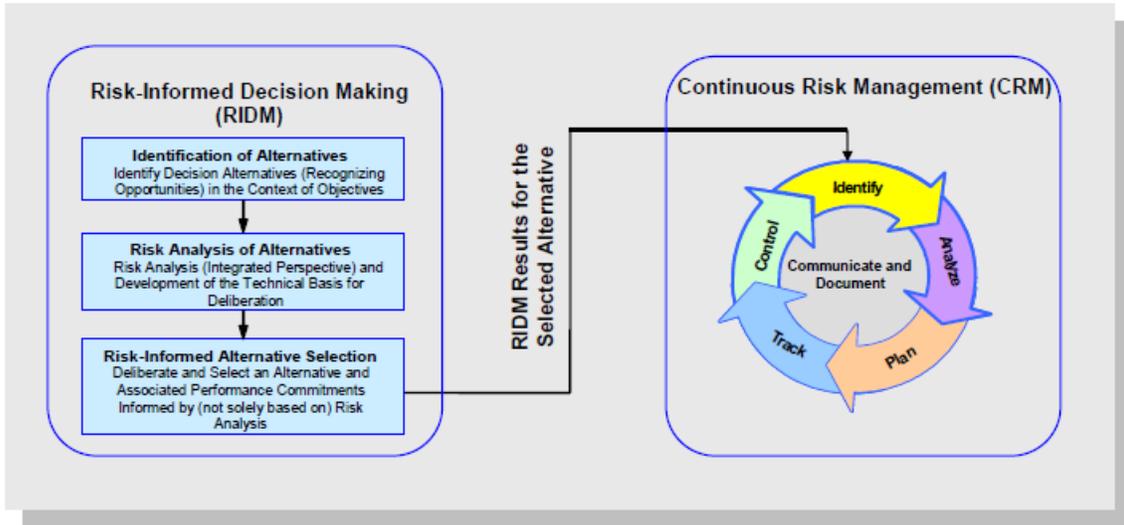
Figure 10. RIDM Input to CRM Initialization
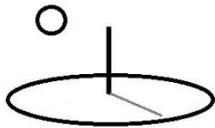
Planetary Science Mission Example: Set the Analysis Framework

The figure below shows the risk analysis framework used to integrate the domain-specific analyses. Each alternative is characterized by its performance parameters, some of which are uncertain (shown in red text) and others of which have definite, known deterministic values (shown in black text). In order to calculate the performance measures previously selected for illustrative purposes, four separate performance models have been developed for radiological contamination, data collection, schedule, and cost. Some performance parameters, such as *spacecraft structure mass*, *launch reliability*, and *science package TRL*, are used in multiple models. Some models (e.g., the data collection model) produce outputs (e.g., *science package mass*) that are inputs to other models (e.g., the schedule model).

Risk Analysis Framework
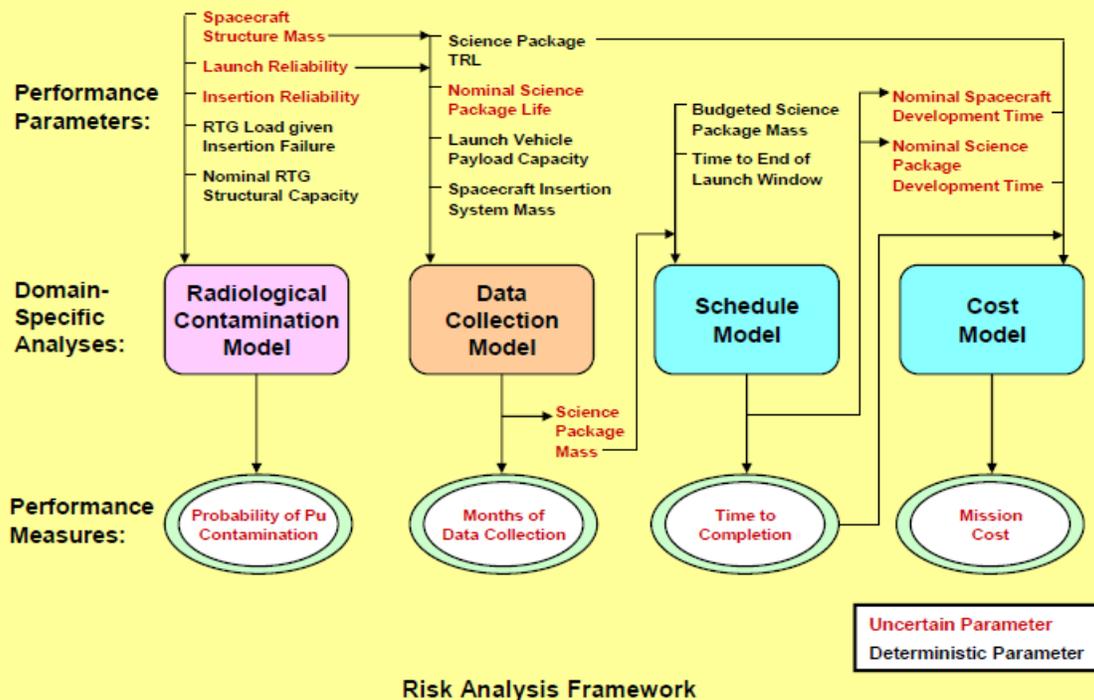
The analysis framework shown above was used for all four alternatives selected for risk analysis. However, in general, each alternative may require its own analysis framework, which may differ substantially from other alternatives' frameworks based on physical or operational differences in the alternatives themselves. When this is the case, care should be taken to assure analytical consistency among alternatives in order to support valid comparisons.

The US, GAO (2009) points out that "Point estimates (i.e. the Performance Parameters displayed in the previous image) alone are insufficient for good decisions. … Quantitative risk and uncertainty (R&U) analysis provide a way to assess the variability in the point estimate". The following images (GAO, 2009) demonstrate that the variability of the alternative point estimates (i.e. Performance Parameters and Measures) can be modeled and summed using probability (or sampling) distributions:
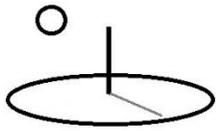
Figure 19: A Point Estimate Probability Distribution Driven by WBS Distributions

Source: NASA.

Note: RPE = reference point estimate.
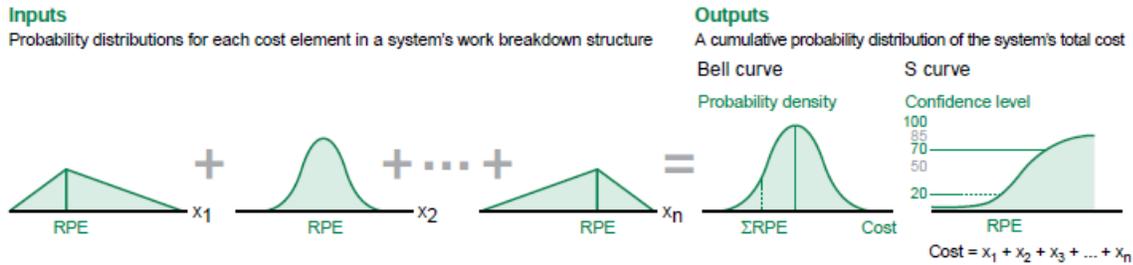
The following image demonstrates that US NASA (2011) uses "sampled" Performance Parameter (i.e. point estimates) to derive distributions of Performance Measures:
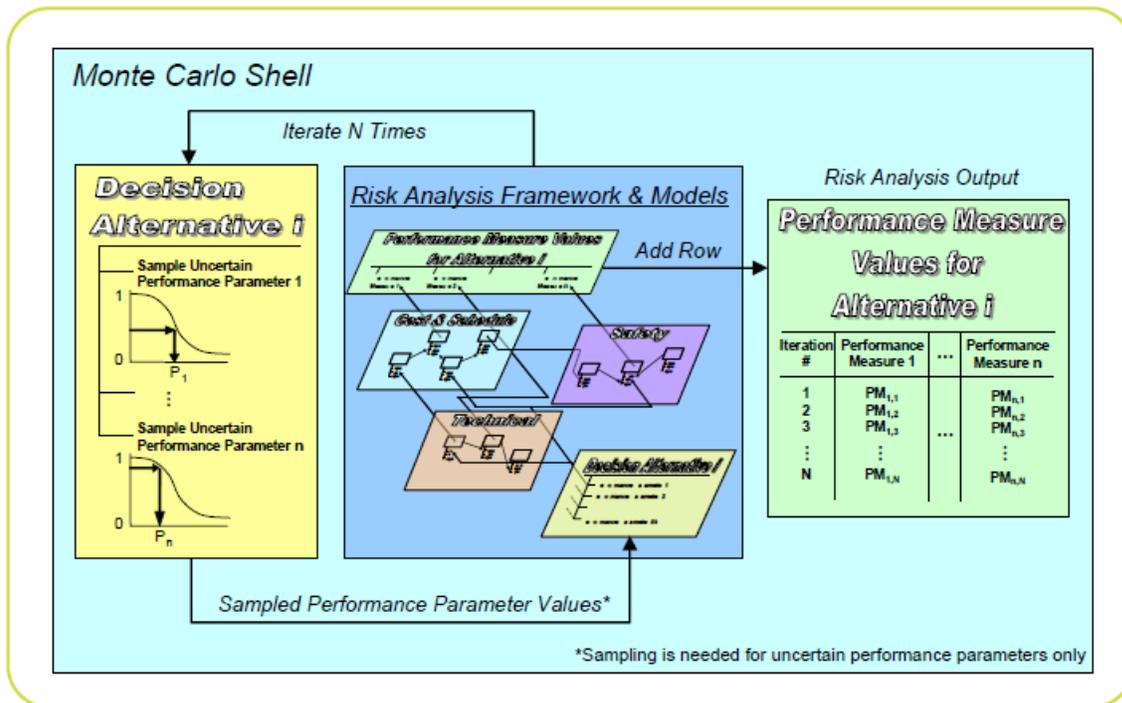

**Figure 23. Risk Analysis Using a Monte Carlo Sampling Procedure**
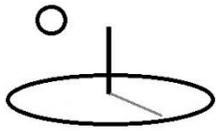
The following image demonstrates that the US GAO (2009) concurs with using simulated point estimates (i.e. in this case, Cost) to analyze risk and uncertainty:

One way to determine whether a program is realistically budgeted is to perform an uncertainty analysis, so that the probability associated with achieving its point estimate can be determined. A cumulative probability distribution, more commonly known as an S curve—usually derived from a simulation such as Monte Carlo—can be particularly useful in portraying the uncertainty implications of various cost estimates. Figure 16 shows an example of a cumulative probability distribution with various cost estimates mapped to a certain probability level.

Figure 16: A Cumulative Probability Distribution, or S Curve
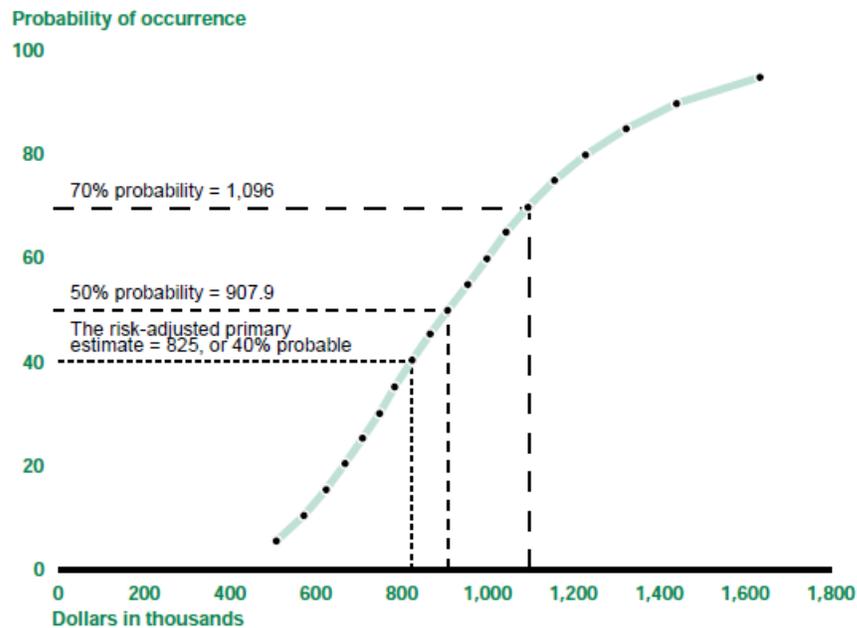
Source: GAO and NASA.

In figure 16, one can readily see that given what is known about program risks and uncertainties, the least this hypothetical program could cost is about $500,000, at about 5 percent probability; the most, $1,700,000 or less, at about 95 percent probability. Using an S curve, decision makers can easily understand what the likelihood of different funding alternatives will imply.[54]

The following image (US GAO 2009) demonstrates that some types of risk analysis can be satisfactorily completed using simple qualitative techniques:
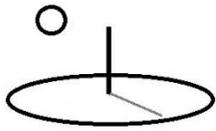
Risk Cube (Probability Impact Matrix) Method

The risk cube method prioritizes uncertainties that could jeopardize program cost, schedule, performance, and quality objectives in terms of probability of occurrence and cost effect. Subject matter experts, typically engineers and others familiar with the program, define the risk factors, probabilities, and cost effect for each identified risk. Using these data, the cost estimator develops the expected cost overrun by multiplying the cost impact by each risk factor's probability of occurrence. A common technique for engaging those knowledgeable about the program is creating a two-dimensional matrix like the one in figure 17.

Figure 17: A Risk Cube Two-Dimensional Matrix

Probability
The likelihood that an objective will not be met if the current plan is used

Consequence
The program penalty incurred if the objective is not obtained
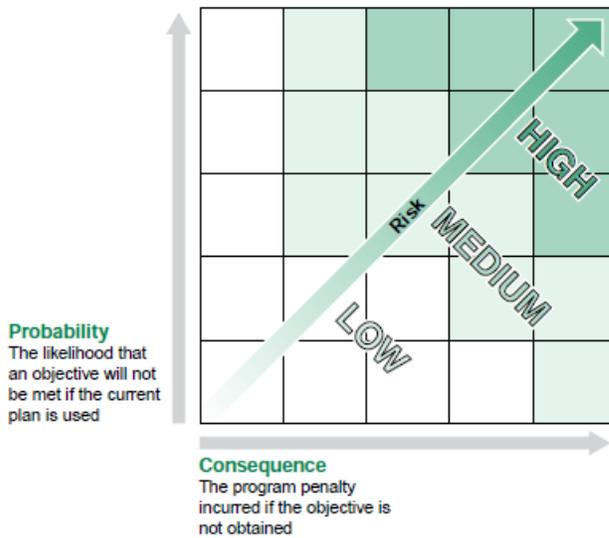
Source: GAO.

In the risk cube (P-I matrix) method, risks are mapped onto the matrix, based on the severity of the consequence—ranging from low risk = 1 to high risk = 5—and the likelihood of their occurring—ranging from low likelihood = 1 to high = 5. Risks that fall in the upper right quadrant are the most likely to occur and have the greatest consequences for the program, compared to risks that fall into the lower left quadrant.

The following image (Ireland, HIQA 2010) demonstrates that some types of risk can be understood better by using more comprehensive analytic techniques (in this image, three health care treatment alternatives are shown using scatter plots with incremental costs on the y-axis and incremental benefits, or QALYs, on the x-axis):
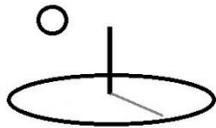
Scatter Plot

For a probabilistic sensitivity analysis (PSA), the analysis is encouraged to present the ICER results using the scatter plot on the cost-effectiveness plane, as depicted in Figure 5.4. Each symbol on the scatter plot represents one simulation of the parameter set. The level of uncertainty in the model is characterised by the spread of the point estimates.

Figure 5.4 Cost-effectiveness of the core scenarios: probabilistic sensitivity analysis

Note: Each symbol represents one simulation of the parameter set.
Source: Health Information and Quality Authority (2009) *Health technology assessment (HTA) of a population-based colorectal screening programme in Ireland.*
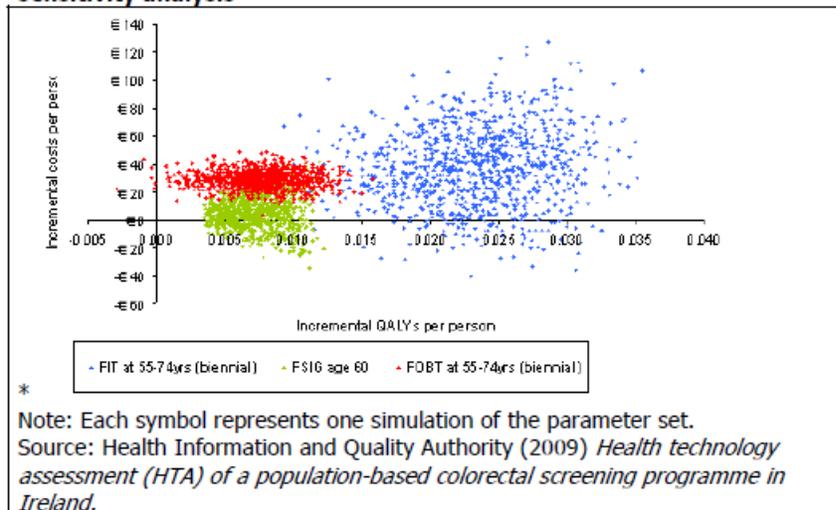
In Figure 5.4, the spread of both the incremental costs and QALYs was wider for the FIT-based screening scenario than for the other options, indicating greater uncertainty for this option. Although considerable uncertainty is evident in the scatter plot, all three scenarios analysed remained cost-effective in all simulations compared to a policy of "no screening". In addition, there were instances where both FSIG and FIT-based screening appear to be cost-saving compared to "no screening". There is a clear distinction in terms of incremental QALYs between FIT screening and screening based on either gFOBT or FSIG, with almost all simulations of FIT-based screening associated with greater gains in QALYs than the other two options. In the majority of simulations, the incremental costs of screening using gFOBT exceeded those for FSIG, without an associated increase in incremental QALYs.

The following image (US NASA, 2011) demonstrates that the results of risk analysis are often reported in terms of imposed constraints such as budget constraints or minimal acceptable levels of safety (**5***):
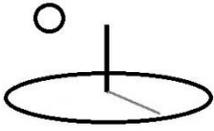
Planetary Science Mission Example: Conduct the Risk Analysis

Each of the modeled performance measures is quantified, using a Monte Carlo shell to sample the uncertain performance parameter pdfs and propagate them through the analysis framework, producing the performance measure results shown below. Each chart presents the assessed performance of the alternatives for a single performance measure, as well as the applicable imposed constraint, which defines the level of performance needed in order to fully meet the top-level objectives.

The risk analysis of alternatives produced the following pdfs:
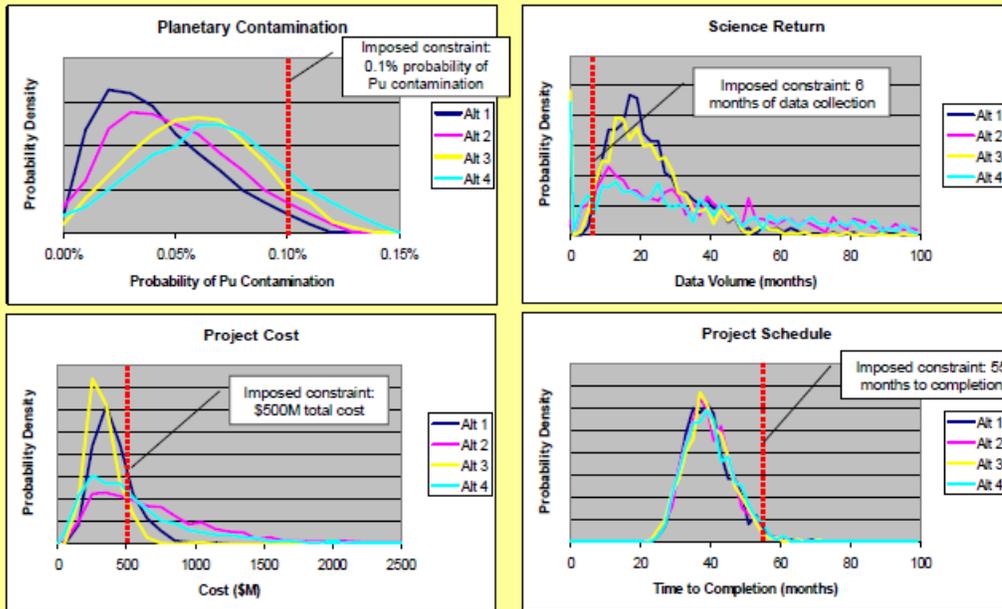
Performance Measure Results for the Planetary Science Mission Example

There is substantial overlap among the pdfs, particularly for *time to completion*. In practice, consideration is given to performing additional analysis to resolve such overlap in cases where doing so is expected to illuminate the decision. However, additional analysis might not help to distinguish among alternatives, especially when the underlying uncertainties are common to them. This is considered to be the case for the notional analysis of the Planetary Science Mission Example.

The following image (NASA, 2011) demonstrates that some types of probabilities can be estimated better by supplementing prior beliefs about model parameters with subsequent observations of real data – a technique known as Bayesian statistics.
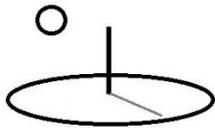
An important tool for developing these probability distributions is Bayes' Theorem, which shows how to update a "prior" distribution over basic event probability to reflect new evidence or information, and thereby obtain a "posterior" distribution. (Refer to Figure 3-11.) Application of Bayes' Theorem is discussed at length in Chapter 5. The general idea is that as more evidence is applied in the updating process, the prior distribution is mapped into a posterior distribution that comports with the new evidence. If there is substantial uncertainty in the prior, corresponding to relatively few data supporting the prior, then new evidence will tend to dominate the characteristics of the posterior distribution.


$$\pi_0(\lambda)$$

$$L(E|\lambda) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

Evidence: K failures in t hours of operation
K=2; t=1000 hours

The prior pdf of λ without knowledge of the evidence

The likelihood of observing conditional on the value λ

$$\pi_1(\lambda|E) = \frac{L(E|\lambda) \cdot \pi_0(\lambda)}{k}$$

The posterior pdf of λ given evidence of E

**Figure 3-11. Application of Bayes' Theorem.**

3-22
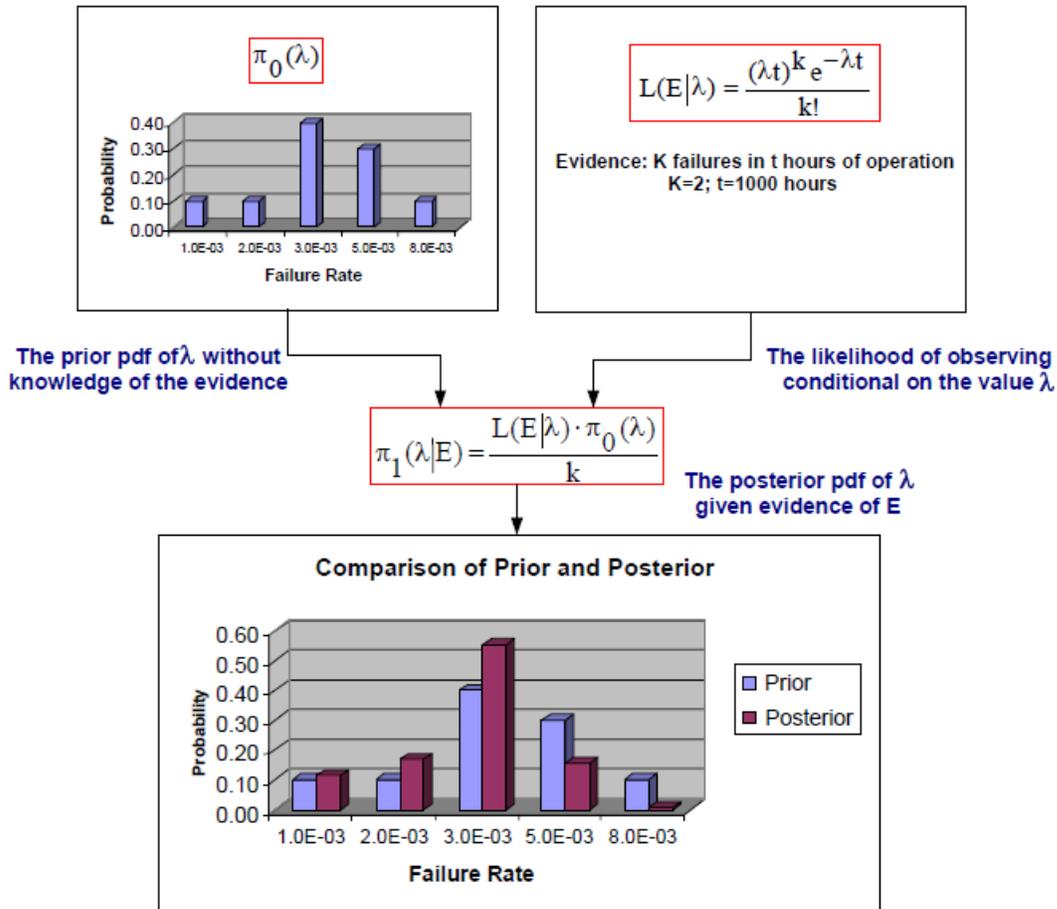
US NASA (2011) identifies the following best practices for carrying out probabilistic risk analysis. Substitute CTA for the NASA space mission project context.
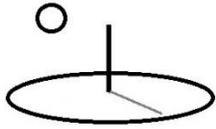
Conduct the analysis in the context of a full Risk Analysis Framework (i.e. NASA's CRM) that continually assesses risk and makes appropriate implementation adjustments that ensure the accomplishment of a project's objectives. Complete the actual risk analysis in a manner that can fully support the decision making (i.e. NASA's RIDM) needed in the encompassing framework. Conduct an actual risk analysis by following these steps:

1. *Use Stakeholder Expectations to Identify Alternatives that meet their Performance Objectives*: Describe the Performance Objectives required by Stakeholder's. Identify Alternatives that can meet those objectives. In terms of CTA, Alternatives will be alternative technologies, practices, or projects. Risk analysts who specialize in health or safety, as contrasted to building and machine risk assessments, point out the importance of incorporating the correct target populations being affected by risk (i.e. Thompson, 2011, HIQA, 2010). These targeted populations are important stakeholders for most social improvement-related technologies, practices, and projects,

2. *Identify Performance Measures and Constraints*: Determine the most important Performance Measures or Risk Factors (i.e. Cost, QALY, Data Collection Labor), and Constraints (i.e. Safety Allowance), for each Performance Objective.

3. *Identify Important Scenarios, Likelihoods, and Consequences.* Evaluate each Performance Measure in terms of potential scenarios that can lead to likely positive and negative consequences. Not every Performance Measure will need to be evaluated in terms of scenarios. Scenarios are equivalent to "comparators" in DevTreks Change and Progress Analyzers.

4. *Identify Performance Parameters*: Identify initial, but uncertain, Performance Parameters that can be used to measure each Performance Measure. The Resource Stock and M&E references demonstrate that Indicators can be used as Performance Parameters.

5. *Specify the Uncertainty of each Performance Parameter*: Use sampling or ranking distributions (i.e. min-max, scenario likelihood, triangular, normal, ranking matrix) to model the uncertainty of the initial Performance Parameters.

24
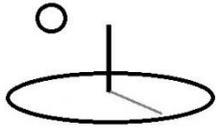
6. *Account for Correlated Performance Parameters*: Consider appropriate correlations for each Performance Parameter.

7. *Generate Performance Measures*: Use a modeling technique (i.e. weighted average, algebraic, Monte Carlo, Latin Cube, Markov, Bayes) to generate sampled point estimates, or Performance Measures, from running calculations that use the uncertain Performance Parameters.

8. *Analyze the Performance Measures*: Display and analyze the simulated point estimates, or Performance Measures (i.e. scatter plots, histograms, probability distribution graphs) with any imposed constraint levels. Eliminate or add Performance Measures.

9. *Use Performance Commitments to Present the Risk Analysis to Decision Makers (5*)*: Use risk tolerance levels, or Performance Commitments, for each Performance Measure. Make sure these levels meet stakeholder expectations. Report the Performance Measures and Performance Commitment for each alternative to the decision makers. If necessary, examine additional alternatives.

10. *Report the Selected Alternatives*: Report the selected Alternatives using a Risk-Informed Selection Report that expresses the final Performance Measure results in terms of their Performance Commitments (i.e. Alternative 1, Performance Measure Cost 1 = $10,000 with probability). As discussed in item 1, social improvement-related risk assessments need to communicate risks in terms of effects on targeted populations.

Not every risk analysis needs every "best practice". For example, some types of risk analysis (i.e. rank matrices) use limited, or no, sampling distributions of point estimates. Some types of cost estimates or benefit estimates do not need to compare alternatives. A simple sensitivity analysis of how different fuel prices may change tillage and planting costs does not need every step.

The following items demonstrate general software features that support carrying out PRAs of CTA data. Substitute CTA for the NASA space mission project context (and no, this is not rocket science).
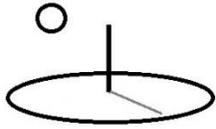
A. Domain-Specific (i.e. Safety and Radiological Contamination) Models: These types of custom models can be built in DevTreks using standard Linked Views (i.e. custom calculator and analyzers that carry out numerical analysis in technical domains such as Agriculture, Food Nutrition, Health Care, Natural Resources, Disaster Prevention/Recovery, or in the NASA example, Safety and Radiological Contamination).

B. Technical Data Collection Models: DevTreks itself is an online data collection model. Multimedia is stored in the Resources application. Stories, calculators, and analyzers, are stored in the Linked Views application. The DevPacks application supports a wide variety of hierarchical content management techniques (i.e. objective hierarchies, trade trees, risk taxonomies).

C. Schedule Models: DevTreks' Monitoring and Evaluation (M&E) calculators and analyzers can replicate, in part, this model. The full replication of this model will require additional types of M&E calculators and analyzers (i.e. Scheduling Risk Analyzers, Least Cost Scheduling Analyzers).

D. Benefit and Cost Models: This model can be replicated using DevTreks' standard economic evaluation calculators and analyzers (i.e. Net Present Value and Life Cycle calculators and analyzers). DevTreks treats benefit estimation just as importantly as cost estimation.

E. Risk Analysis of Models: This reference introduces the first set of CTA tools available for conducting risk analysis. These generic indicators, mathematics, and statistics, support a wide assortment of risk measurement. They can also be used to set the initial costs and benefits found in the Benefit and Cost models.

PRAs of modeled data can still generate completely wrong numbers when models are defined badly, the data used in the models is spurious, or wrong assumptions are made about the distribution of indicators. Most of the tutorials refer to scientific publications (i.e. IPCC) that, if used properly, can improve the models, the assumptions, and the data.

Several references (i.e Lurie, Goldberg, and Robinson, 1993) further clarify the limitations of PRAs as follows: "Although a formal analysis of cost [and benefit] risk will not reduce the risk
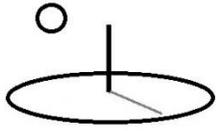
inherent in a program, it will help program managers understand the nature of the risks involved, and quantify and display the uncertainty associated with cost [and benefit] estimates".
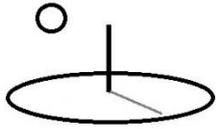
Appendix B. Supplemental Analyses

Numerous related analyses can be conducted in support of a risk analyses. Some of these include:

> *Sensitivity Analysis* examines changes in one risk indicator while keeping all other indicators constant. This procedure is used to identify (and manage) the most important drivers of project and program results. Some sensitivity indicators compare the percentage change in one risk indicator with percent changes in other indicators (i.e. a 1% change in yields results in a 10% change in net profit). The usual steps in a sensitivity analysis include (US GAO 2009):
>
> 1. identify key cost drivers, ground rules, and assumptions for sensitivity testing;
> 2. reestimate the total cost by choosing one of these cost drivers to vary between two set amounts —for example, maximum and minimum or performance thresholds;
> 3. document the results;
> 4. repeat 2 and 3 until all factors identified in step 1 have been tested independently;
> 5. evaluate the results to determine which drivers affect the cost estimate most.

The following image (US GAO, 2009) demonstrates a typical sensitivity analysis of gross project costs. Examples of sensitivity analysis of Benefit Cost Ratios and Cost Effectiveness Ratios can be found in the CTAP reference.
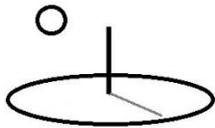
Figure 14: A Sensitivity Analysis That Creates a Range around a Point Estimate

| | Increase estimate: | +$40.0 million (0.4%) | +$50.0 million (0.5%) | +$1,009 million (10.0%) | +$22.0 million (0.2%) | +$1,668 million (15.0%) |
| | Description: | Increase the number of cost penalties in airframe development CER | Double the development testing | Increase airframe weight | Eliminate concurrent production quantities | Increase quality of materials in aircraft |

Life-cycle cost (FY 07 constant dollars)

$13 billion — $12.789

$12 billion

$11.099 $11.121

$11 billion

Increase in life-cycle cost estimate — $10.040 — $10.090

$10 billion (Point estimate)

Decrease in life-cycle cost estimate — $9.940 — $9.89 — $9.79 — $9.75

$9.36

$9 billion

| | Decrease estimate: | -$60.0 million (0.6%) | -$50.0 million (0.5%) | -$100.0 million (1.0%) | -$40.0 million (0.4%) | -$390.0 million (4.0%) |
| | Description: | Use 88% learning curve instead of 91% | Eliminate integration and assembly cost add-on | Reduce airframe weight | Improve aircraft maintainability | Reduce peacetime flying hours |

Source: GAO.

Incremental Analysis is a type of cost effectiveness analysis that measures how an incremental change in one indicator affects another indicator. Some Health Technology Assessments (Ireland HIQA, 2010) use measurements of incremental costs and incremental benefits as the initial point estimates that go into a risk analysis. The following image (US, NIH, HTA Course 101) displays a typical incremental analysis of costs vs. number of medical tests:

x 21
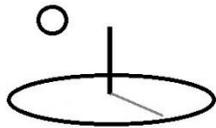
erage Cost Analysis vs. Marginal Cost Analysis

e importance of determining marginal costs is apparent in the analysis by Neuhauser and Lewicki of a proposed protocol of sequential stool guaiac testin scure a steep rise in marginal costs of testing because the high detection rate from the initial tests is averaged over subsequent tests that contribute ps to demonstrate how it is possible to spend steeply increasing health care resources for diminishing returns in health benefits.

Cancer screening and detection costs with sequential guaiac tests

No. tests	No. of cancers detected	Additional cancers detected	Total cost ($) of diagnosis	Additional ($) cost of diagnosis	Average cost ($) per cancer detected	Marginal cost ($) per cancer detected
1	65.9469	65.9469	77,511	77,511	1,175	1,175
2	71.4424	5.4956	107,690	30,179	1,507	5,492
3	71.9004	0.4580	130,199	22,509	1,810	49,150
4	71.9385	0.0382	148,116	17,917	2,059	469,534
5	71.9417	0.0032	163,141	15,024	2,268	4,724,695
6	71.9420	0.0003	176,331	13,190	2,451	47,107,214

s analysis assumed that there were 72 true cancer cases per 10,000 population. The testing protocol provided six stool guaiac tests per person to det sitive, a barium-enema test was performed, which was assumed to yield no falsepositive and no false-negative results. Other assumptions: the true-po: t was 91.667%; the false-positive rate of any single guaiac test was 36.508%; the cost of the first stool guaiac test was $4 and each subsequent gu s $100. The marginal cost per case detected depends on the population screened and the sensitivity of the test used.

urce: Neuhauser 1975.

Weighted Average Analysis relies on simple subjective assessments of risk using weighted risk scores such as 1 for low levels, 2 for medium levels, and 3 for high levels. The following image (US GAO, 2009) displays an example:
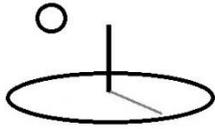
Using an algorithm that assigns relative weights and scales to each risk area and classifies risk as low, medium, or high can help determine which programs would most benefit from surveillance. Table 43 shows how an algorithm can be used to evaluate a candidate program.

Table 43: A Program Surveillance Selection Matrix

Risk factor	Weight	Risk level High = 3	Medium = 2	Low = 1	Risk score
Contract value	0.05	More than 20% of business base	5%–20%	Less than 5%	3
Nature of work	0.05	High-risk, many unknowns		Low-risk content	3
Program office experience	0.05	Inexperienced staff		Very experienced staff	1
Program type	0.05	Development	Production	Operations and maintenance	3
Baseline resets	0.10	Many per year	Once a year	Less than one a year	3
Historic trends	0.10	Worsening		Trends are improving	3
Previous findings	0.10	Many unresolved		Few or easily closed	1

Risk factor	Weight	Risk level High = 3	Medium = 2	Low = 1	Risk score
Variance percent	0.10	Worse than –10%	–5% to –10%	Better than –5%	3
Management interest	0.40	High visibility		Low visibility	3
Total					2.6

Source: © 2004 National Defense Industrial Association (NDIA) Program Management Systems Committee (PMSC), *Program Management Systems Committee Surveillance Guide* (October 2004).

Scenario Analysis (US Air Force, 2007; or ADB, 2002's Simplified Probabilistic Analysis) is a variant of sensitivity analysis that defines a group of indicators (a scenario) that change together and that represent a logical scenario for a project. For example, the Fukushima nuclear reactor disaster may be related to a failure to adequately consider feasible scenarios, such as a large tsunami, in its design. Algebraic formulas are used to generate these types of probability distributions:

31
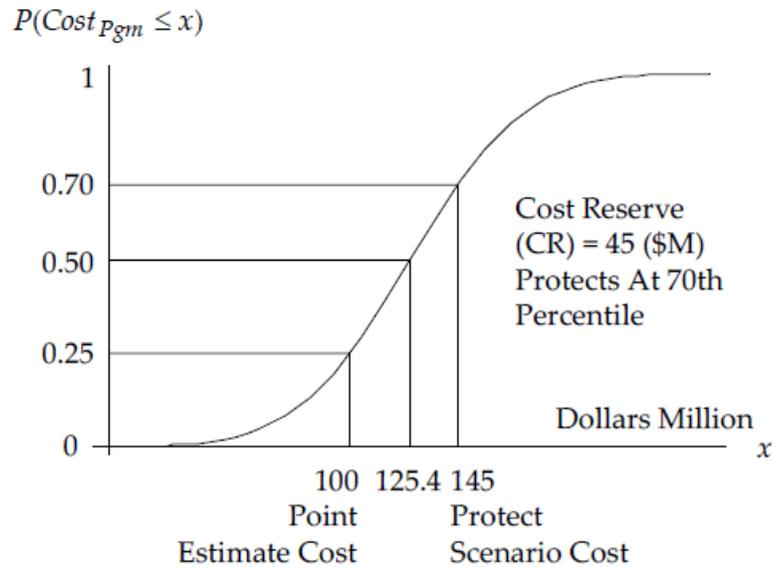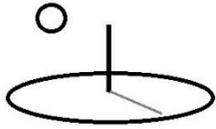
Figure 5. Example 4-1 Illustrated: Assumed *Normal* Distribution

Numerous studies [1] have empirically shown the normal or lognormal to be excellent approximations to the overall distribution function of a program's total cost, even in the presence of correlations among cost element costs. The decision to use one over the other is really a matter of analyst judgment. In practice, it is

Decision Tree Analysis depicts risk using branches to display the probability of alternative actions. The following image (US, NIST, 1988) displays a building construction investment which can use alternative insulation (R1, R3 …). Each alternative can be quantified using a probable (i.e. minimum, most likely, maximum) net investment return ($3 and $2 in the top node):
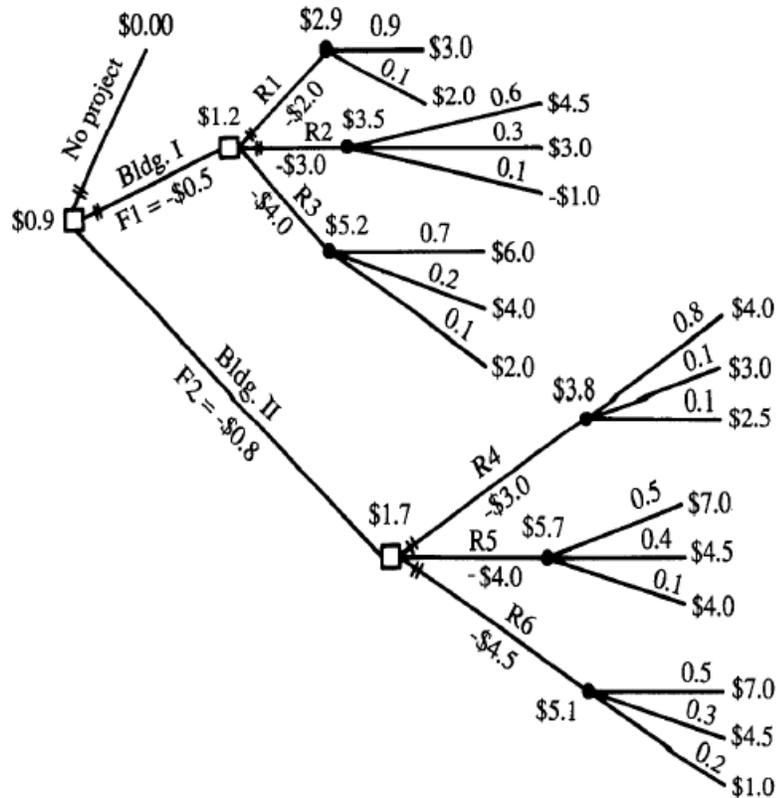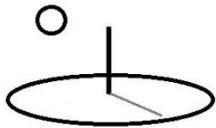
Figure 5-1. Decision tree for conservation investment ($ values are in millions).
□ = Decision Node
● = Chance Node
R = Retrofit Package

Event Tree and Fault Tree Analysis: (NASA 2011) risk is modeled as a series of events that collectively lead to the likelihood of a consequence, such as an accident.

Behavioral and Psychological Decision Analysis (see page 14 in US NASA, 2011 or Kahnemann, 2011) uses techniques developed by psychological and behavioral scientists.

*Constrained Choice, or Optimization, Analysis (*refer to economics texts) measures an optimal level of one or more outcomes (i.e. net returns) given constrained input choices, such as cost, acres of land, or minimal level of safety.
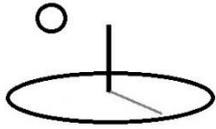
Schedule Analysis (see GAO, 2009) assesses the probability of falling behind schedule and incurring additional project costs.

Markov Analyses (see Ireland HIQA, 2010) measures R&U indicators that change over time, such as when a series of health care treatments change health care states over time.

Parametric Analysis (see NASA, 2011 and GAO, 2009) uses regression analysis techniques to develop probabilistic models that predict a dependent variable based on datasets of observed dependent and independent variable data.

Computer Algorithms (see the McCaffrey reference) use a variety of numeric analyses techniques (i.e. neural network, bayes, matrix decomposition) to assess the probability of outcomes.

Risk Scoring and Risk Cubes: (see the GAO 2009 reference) Factors that have large potential impacts on results are identified and then ranked and scored, using both qualitative and quantitative means.

Machine Learning (ML) Algorithms: The Social Performance Analysis tutorial used the following image to introduce the differences between standard statistical models and Machine Learning algorithms.
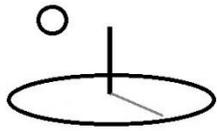
ML versus Statistics

Statistics:

- Hypothesis testing
- Experimental design
- Anova
- Linear regression
- Logistic regression
- GLM
- PCA

Machine Learning:

- Decision trees
- Rule induction
- Neural Networks
- SVMs
- Clustering method
- Association rules
- Feature selection
- Visualization
- Graphical models
- Genetic algorithm

http://statweb.stanford.edu/~jhf/ftp/dm-stat.pdf