

NORTHROP GRUMMAN

DEFINING THE FUTURE

Making Risk Management Tools More Credible: *Calibrating the Risk Cube*

SCEA 2006
Washington, DC



Richard L. Coleman, Jessica R. Summerville,
Megan E. Dameron
Northrop Grumman Corporation

Outline

- **The General Risk Cube Method**
 - How it is viewed by cost estimators and engineers
 - How it is used
- **The limitations and biases in the Risk Cube**
- **Calibrating the Risk Cube to maximize its usefulness**

Proposed Metrics for Likelihood and Consequence for the “Risk Cube” Method of Risk Management

- **Objective: To provide metrics and cutoff scores for Risk Management metrics**
 - Metrics to be incorporated into the most common method of Risk Management (RM), the so-called Likelihood and Consequence, or “Risk Cube” method
- **Users are expected to be DoD Acquisition Program Risk Managers, agencies that oversee risk, and contractor program offices working for them**
 - The consequence score cutoffs are calibrated specifically for such programs based on analysis of DoD Selected Acquisition Reports (SARs)
- **The reports^{1,2} upon which they are based were presented at DoDCAS and SCEA**

[1] *NAVAIR Cost Growth Study, ISPA/SCEA 2001, 34th DoDCAS and ISPA/SCEA 2001*, R.L. Coleman, M.E. Dameron, C.L. Pullen, J.R. Summerville, D.M. Snead

[2] *The Relationship Between Cost Growth and Schedule Growth, 35th DoDCAS and SCEA 2002*, Richard L. Coleman, Jessica R. Summerville, Megan E. Dameron

Background

- **This presentation provides general guidelines for applying the likelihood and consequence method of risk assessment**
- **The purpose of this briefing is to ensure that:**
 - Each issue that might affect the success of the program (technical, schedule, and cost) is identified and assessed as to likelihood and consequence of occurrence
 - A standard format is used for evaluation and reporting of program Risk Assessment findings to facilitate common understanding of program risks at all levels of the organization

Why the Risk Cube Method?

Engineers' and Cost Analysts' Views of Risk

Engineers

- Work in structure & physical materials, with
 - Physics-based *causal* responses
 - Physical connections
- Typically examine or discuss a *specific, discrete* outcome set
 - Point designs
 - Specific system parameters such as weight, size
- Typically seek to know:
 - Given this solution, what will go wrong?
 - Are *design* margins enough?
- Usually prefer Risk Management methods
- Believe that spending money to avoid bad outcomes is the *avoidance of technical risk*
- Engineers thus prefer a discrete system that addresses specific possibilities

Cost Analysts

- Work in dollars & parameters, with
 - Statistical relationships
 - Correlation
- Typically examine or discuss a *general, continuous* outcome set
 - Probability distributions
 - Statistical parameters such as mean and standard deviation
- Typically seek to know:
 - Given this relationship, what is the range of possibilities?
 - Are *cost* margins enough?
- Usually prefer Cost Risk methods
- Believe that spending money to avoid bad outcomes is the *manifestation of cost risk*
- Cost estimators thus prefer a continuous system that shows the range of possibilities

Both views are valid. The Risk Cube is, however, tailored to the engineer.

The Risk Cube Method

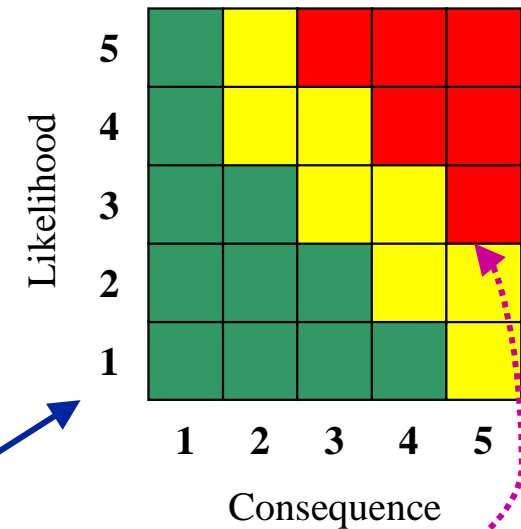
Likelihood

Level	Likelihood
1	Not Likely
2	Low Likelihood
3	Likely
4	Highly Likely
5	Near Certainty

Consequence

Level	Technical	Schedule	Cost
1	Minimal or no impact	Minimal or no impact	Minimal or no impact
2	Minor technical shortfall	Slip < * month(s)	< (1% of Budget)
3	Moderate technical shortfall	Slip < * month(s) of critical path. Sub-system slip > * month(s).	< (5% of Budget)
4	Unacceptable, workarounds available	Slip < * months	< (10% of Budget)
5	Unacceptable, no alternative exist	Cannot achieve key program milestones	> (10% of Budget)

Note: Generic Risk Cube

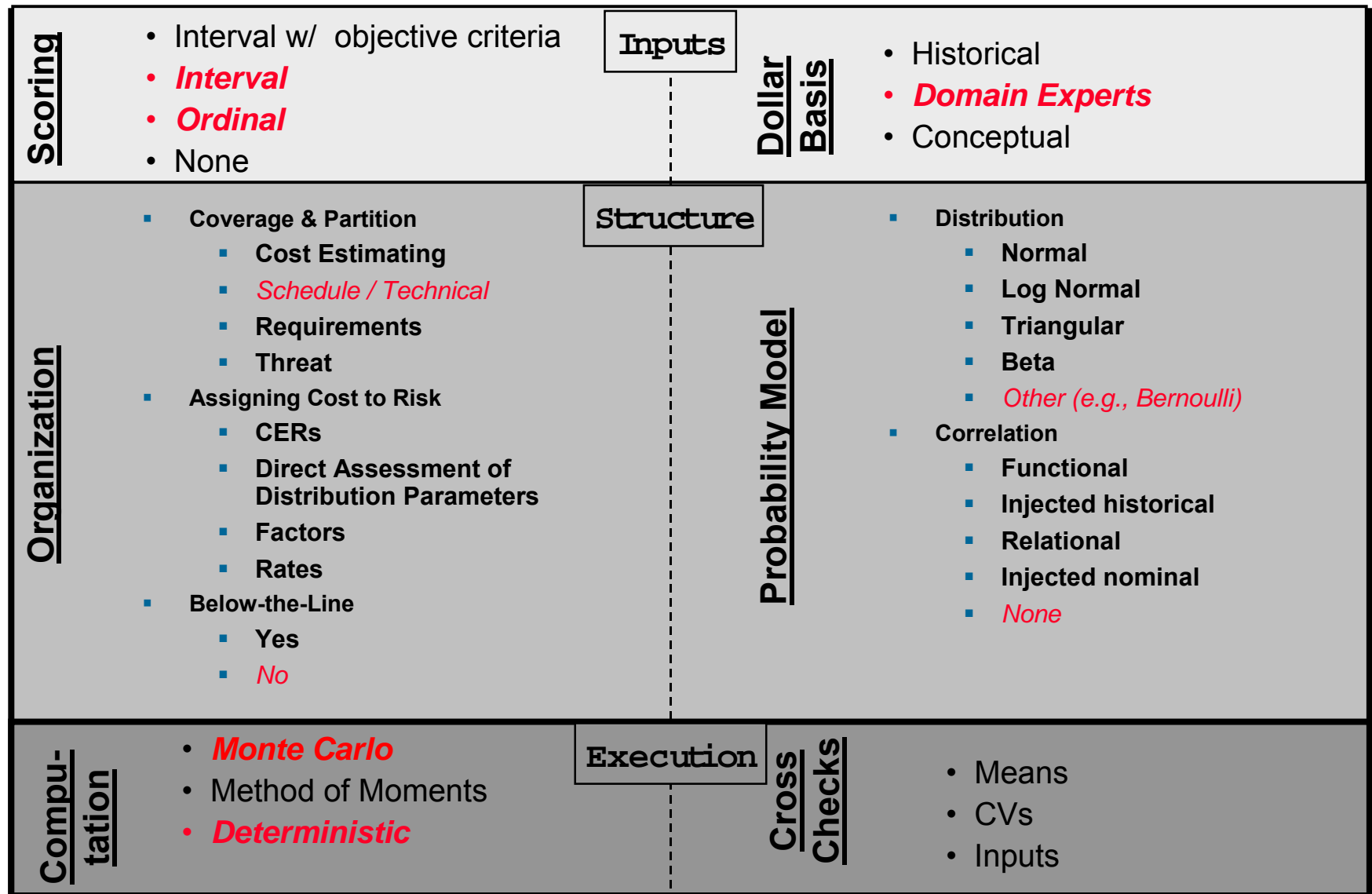


Risk Item Assessments:

Category	Level	Likelihood	Consequence
Statement			
Cause			
Mitigation			

General Model Architecture

The Risk Cube Approach to Risk Management



The Risk Cube Method

- Outcome oriented - begins with analysis of all factors that can cause designs to fail or be wrong, by Subject Matter Experts (engineers), who:
 - Identify each factor (risk item)
 - For each item, estimate
 - The probability of occurrence (Pf) and
 - The cost impact if it occurs (Cf)
- Can be represented by Bernoulli Random Variables
 - The expected cost overrun is the sum of cost impacts multiplied by their respective probabilities

$$\text{Cost Risk} = \sum Pf * Cf$$

$$\text{Mean} = Pf * Cf$$

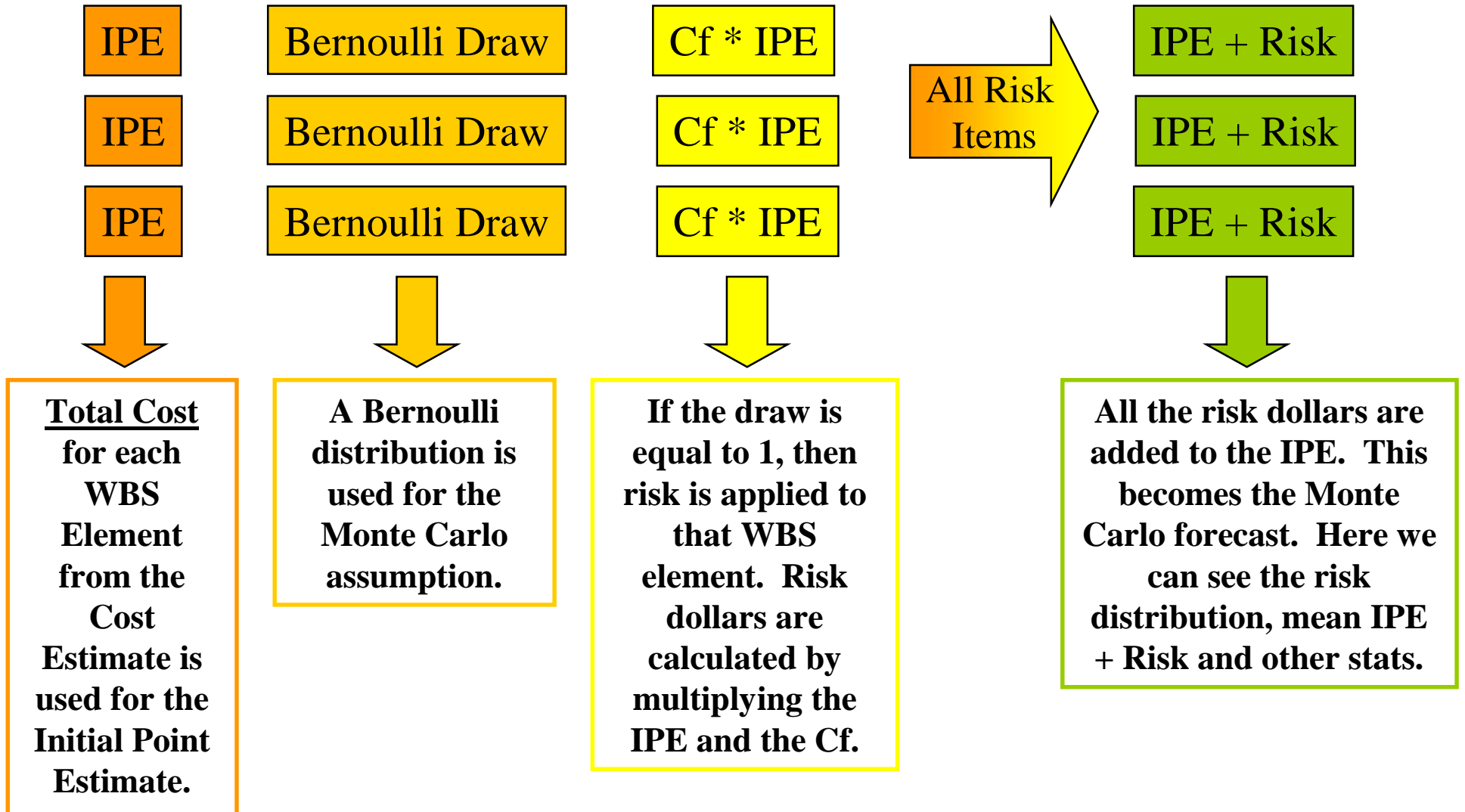
$$\text{Std Dev} = \text{SQRT}(Pf * (1 - Pf) * Cf) = \text{SQRT}(Pf * Qf * Cf)$$

- The minimum, mean and maximum of the risk list are easily computed
 - These values are deterministic (above)
 - Percentiles , including the much-sought-after 50th percentile, must be determined by Monte Carlo

Note: $Qf = 1 - Pf$

Setting up the Monte Carlo Model

Per Risk Item



Risk Cube Assessment Process

Steps:

1. Convert risk scoring to Probabilities and Consequences
2. Map risk items to CWBS
3. Setup Monte Carlo Simulation (using Bernoulli distributions) combining CWBS cost estimate with risk impacts
4. Run model and assess results (i.e., determine biggest hitters, look for potential errors, etc)
5. Crosscheck results with historical data (based on program size)

Level of Effort Needed:

- A few days for preparation and familiarization of the team
- A day or two for mapping of risk items to the WBS
- Completion approximately one week after risk items are mapped to the WBS

Pros and Cons of the Risk Cube

- The Risk Cube relies on:
 - Complete lists of what could happen
 - Accurate Pf's and Cf's
 - Mapping of risk items to the WBS
- Pros:
 - Intuitive and Engineer/Designer-oriented outcome
 - Amenable to mitigation - specific
- Cons:
 - Almost always understates risk because:
 - Identification and scoring by SMEs makes bias (low!) likely
 - Enumeration of “all the things that can go wrong” will inevitably leave out unknown unknowns
 - The aggregate values of trivial, unlisted risks is often bigger than even the largest listed risk

The Risk Cube vs. Historical

- Risk Cube methods can be adjusted to produce results that are comparable to historical cost growth
- However, the Risk Cube method cannot substitute for a historically based risk estimate
 - SMEs tend to be biased or lack adequate familiarity with the program
 - Unknown unknowns are not included
 - Small risks get omitted

Risk Cube vs. Historical

- Risk Cubes *do* add value
 - They are intuitive to engineers
 - Connect with risk management processes
- But, we *expect* Risk Cube results to be somewhat lower than historical
 - If only somewhat lower, the difference may be accounted for by unknown unknowns, small risks, and SME optimism
 - **If much lower, be skeptical of the risk register**
 - **If higher, be alarmed ... experts are rarely pessimistic ... there may be much more risk than anticipated**

Calibrating the Risk Cube

- **We've seen what the risk cube is, and how it is used**
- **We will now proceed to the subject of this paper ... the calibration of the risk cube to make it most useful and "accurate"**
 - By accurate, we do not mean the avoidance of bias, unknown unknowns and omission of a myriad of trivial risks, we mean only making it more likely to be accurate *as far as it goes*

The Likelihood of Failure (of Risks)

- The below categories of likelihood are to be used to indicate the probability of the risk occurring
 - Usually denoted Pf = Probability of Failure
- Equal bin sizes for probability (0.2 each) are desirable, as these are intuitive
- The definition terms were chosen to be easily recognizable
 - Some systems use terms such as “unlikely”, “possible”, and “likely,” but these are subjective as to the implied probability, and should be avoided

Level	Probability (Pf)	Definition
1	$0.0 < Pf \leq 0.2$	Low likelihood
2	$0.2 < Pf \leq 0.4$	Low-to-medium likelihood
3	$0.4 < Pf \leq 0.6$	Medium
4	$0.6 < Pf \leq 0.8$	Medium-to-high likelihood
5	$0.8 < Pf \leq 1.0$	High likelihood

Calibration for Risk Register

- The goal for this methodology is to construct a Cf table with cost impacts that are reasonably well-aligned with cost growth experienced on similar historical programs
- The table will be used to score risks identified for a program, and will result in a risk register with a total expected value that compares well with trends supported by historical data
- To start, we need to research a historically-based cost growth factor to benchmark where the total expected value for cost risk should be centered
 - One such study¹ considers two important attributes in determining this value, for a generic DoD program. This generic example is shown on the following slide, and considers:
 - Commodity
 - Phase of Acquisition cycle
- We will look at 2 examples:
 - Developing RDT&E consequence values for a generic DoD R&D program
 - Developing Procurement consequence values for a ship program that is at the start of its Engineering Manufacturing Development (EMD) phase²

[1] *Proposed Metrics for Likelihood and Consequence, for the "Risk Cube" method of Risk Management*, white paper by T. L. Eng, M. E. Dameron, J. R. Summerville, R. L. Coleman, 28 May 2002

[2] *Ship Program Risk Register Cf Scoring Table Methodology*, white paper by Jessica R. Summerville, Noelle A. Shaw, Peter J. Braxton, Richard L. Coleman, 27 May 2005

Example 1: RDT&E Calibration for a Generic DoD R&D Program

- **Limits were set up to have equal “bin width” (10% each for schedule, and 15% each for cost) this method is very intuitive**
 - There are not equal numbers of programs in each “bin”, in the historical data, but this is not felt to be a problem
- **The bins were set up so that the historical average was near the middle (Level 2 or Level 3)**
 - This differs by commodity, so will not work exactly in every case, but uniformity across DoD might be better than calibrating bins for each commodity or service
- **It is desirable to cover most of programs in the first four levels, so that the unboundedness of the 5th level will not cause too many programs to fall in that bin, rendering the scale meaningless**
- **Scoring applies by WBS element**
 - Pro: Easy application to cost estimate
 - Con: Not as helpful for direct risk register application

Level	Technical (T)	Schedule (S)	Cost (C)
1	Minimal or no impact	Schedule slip to the scored area of $S \leq 10\%$	Cost increases to the scored area of $0\% < C \leq 15\%$
2	Minor technical shortfall, no impact to high level technical requirements	Schedule slip to the scored area of $20\% < S \leq 30\%$	Cost increases to the scored area of $15\% < C \leq 30\%$
3	Moderate technical shortfall but workaround available which will eliminate impact to high level technical requirements	Schedule slip to the scored area of $30\% < S \leq 40\%$	Cost increases to the scored area of $30\% < C \leq 45\%$
4	Unacceptable, workarounds available which will eliminate impact to high level technical requirement	Schedule slip to the scored area of $30\% < S \leq 40\%$	Cost increases to the scored area of $45\% < C \leq 60\%$
5	Unacceptable, no alternative exist	Schedule slip to the scored area of $40\% < S$	Cost increases to the scored area of $60\% < C$

Example 2: Procurement Calibration for a Ship Program

- **To calibrate the expected cost risk to be incurred by the program, the calculation for the expected value of the risk register first needs to be understood**
 - Total Expected Value = $(\sum (P_{fj} * C_{fj})) / \text{Total NRE Estimate}$ (j=1,2....N risks)
 - If the risks are all of an average score (0.4-0.5), then the expected value should match the mean value taken from the historical study
 - If the risks are lower or higher than average, then the expected value should reflect a corresponding impact relative to the mean
- **We can use this formulation with our historical factors to derive a table that will produce the results discussed above**
- **We will use what we know to algebraically derive the Average Cf value for a Risk Register item, then create a scale around that value that reflects the historical range. The details of this methodology are as follows:**

Calibration for a Ship Risk Register

- The first task is to derive the Average Cf.
 - We want to first simplify the formula to reflect a Risk Register where all risks are scored at an average level. This results in the following:

$$\text{Total Program Level Expected Value} = \text{Average Risk Cf} * \text{Average Risk Pf} * \text{Total Number of Risk Items}$$

- Now, we replace the average Risk Pf with an average Pf of 0.5

$$\text{Total Program Level Expected Value} = \text{Average Risk Cf} * 0.5 * \text{Total Number of Risk Items}$$

- Next, we calibrate to historical data by substituting the historical cost growth for the Total Expected Value. Suppose ship procurement estimates typically grow ~4%. Then in the case of ship Procurement, the equation is as follows:

$$4\% = \text{Average Risk Cf} (\%) * 0.5 * \text{Total Number of Risk Items}$$

- And finally, we need to know about how many risk items are expected to comprise the risk register.

- In similar programs, there were ~100 items, therefore we will assume there will be about that many for this one (the count does not need to be exact, just an approximate value that will get us in the right ballpark)

$$4\% = \text{Average Risk Cf} (\%) * 0.5 * 100$$

- We can now solve for the Average Risk Cf in terms of a percentage

$$\text{Average Risk Cf} (\%) = 4\% / (0.5 * 100) = 0.08\%$$

- We now have a value that represents the Cf for the average risk score. To convert this value to a dollar amount, simply multiply by the total estimate.

$$\text{Average Risk Item} (\$) = \text{Proc Estimate} * \text{Average Risk Cf} (\%) = 800\$K$$

NORTHROP GRUMMAN

Calibration for a Ship Risk Register

- **Avg. Growth per Risk Item** = $\text{Total Avg. Proc Growth} / \# \text{ Risk Items} * P_f \text{ Factor}$
- **“Max” Consequence** = $\text{Prod Threshold} * 10\%$

Pf Factor	2
# of Proc Risk Items	91
Total Avg Proc Growth	4%
Avg Proc Growth per Risk Item	0.08%
Proc Cost Estimate (T1) (\$K)	\$1,000,000
Max Consequence (\$K)	\$100,000
Avg Consequence (\$K)	\$800

If each of the 91 risk items are scored at 0.08% growth, the total growth will be 4%

Bins need to go high enough to cover the supposed Max Consequence (10% of Threshold)

$\text{Avg. Consequence} = \text{Avg. Proc Growth per Risk Item} * \text{Proc Threshold}$

Put the Average Consequence at the middle of the C_f scale

Calibration for a Ship Risk Register

- **The next task is to populate the scoring table by figuring out an appropriate range of Cfs for each score in the table**
- **We want the following characteristics:**
 - Average Risk Cf to connect somewhere in the average (0.4-0.5) scoring range
 - Maximum Risk Cf to connect to the highest possible score.
 - In addition, we'd like the Cfs in the table to span something on the order of +/- one standard deviation of cost growth observed in historical programs and be close to symmetric around the average.
 - Based these considerations, and results from the study previously referenced, we have determined that the value of 10% is a good proxy for the Maximum Risk Cf.

$$\text{Max Risk Cf (\%)} = 10\%$$

$$\text{Max Risk Cf (\$)} = \text{Proc Estimate} * 10\% = \$100,000 (\$K)$$

- **Now the bin sizes must be adjusted in order to accommodate the constraints of values we've determined so far This is done by setting a starting bin size for the bin of score 0.1, ((\$0-\$50) or (\$0-\$100) are usually good starting points), and then applying an increase factor to the bin size of each progressive score (0.2, 0.3, etc).**
 - Roughly doubling the bin size for each progressive score (i.e., using an increase factor in the 2.0-2.5 range) will generally create a table that connects our Average and Maximum Cf values to the appropriate scores

$$\text{Bin Size}_i = \text{Bin Size}_{i-1} * \text{Bin Increase Factor} \quad (i = 0.2, \dots, 0.9)$$

Calibration for a Ship Risk Register

Proc Score	Dollar Range (TY02\$K)		Percent Range (% of Total Prod Cost)	
	Min	Max	Min	Max
0.0	\$0	\$0	0.00%	0.00%
0.1	\$0	\$100	0.00%	0.01%
0.2	\$100	\$200	0.01%	0.01%
0.3	\$200	\$500	0.01%	0.03%
0.4	\$500	\$1,400	0.03%	0.08%
0.5	\$1,400	\$4,000	0.08%	0.22%
0.6	\$4,000	\$10,000	0.22%	0.55%
0.7	\$10,000	\$25,000	0.55%	1.38%
0.8	\$25,000	\$65,000	1.38%	3.59%
0.9	\$65,000	\$170,000	3.59%	9.39%
1.0	\$170,000	Over \$170,000	9.39%	Over 9.39%

Cost C_f 's assess the *remaining* risk ... assumes that historical levels of risk were incurred during Program PDRR phase

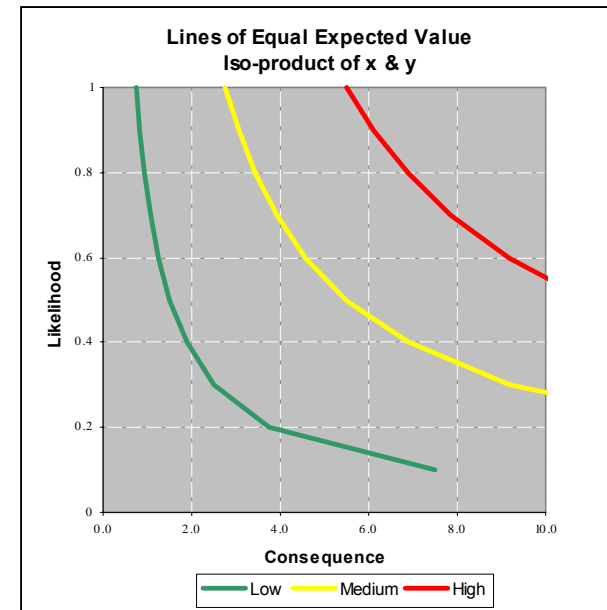
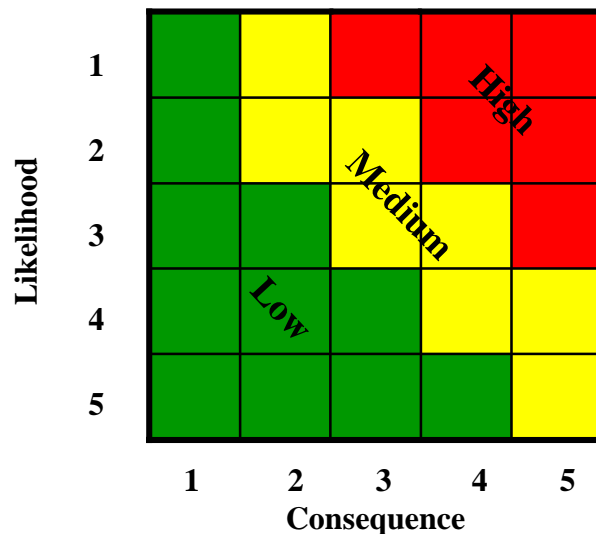
Estimated Average lies at the middle bin

The last bin starts near the assumed Max Consequence ... fulfills need to avoid having too many items fall in this unbounded bin

**Score Proc Cost
at the *Total* Level**

The Shape of the Bands of Expected Value

- Risks are evaluated based on their “expected” outcome
 - Computed as the product of probability and consequence
- Risks are usually considered equivalent if they have the same expected value
- Accordingly, the “cube” is color coded based on equal product, which results in the color scheme on the left, and the graph on the right
 - The lower right box is normally coded yellow and is reserved for *very unlikely* events with *very high* consequences



Conclusion

- **We have discussed the Risk Cube**
 - Why it is preferred by engineers
 - How it works
 - That it is prone to understate risk
- **We have noted that historically based methods are preferred by cost estimators**
- **We have shown a calibration that makes the Risk Cube method more accurate**
 - Engineers will never forsake it, so we need to make it as good as it can be made
 - It will continue to be biased low, but at least it will be as accurate as it can be made