412th Test Wing

War-Winning Capabilities … On Time, On Cost

Title: Considering Consequences in Managing Decision Risks in Flight Test

Date: 12 May 2017

Author: Robert Poulson, PhD

Author Office: 412 TW

Author Phone: 661-277-4130

U.S. AIR FORCE

Integrity - Service - Excellence

Overview

- Introduction: Presentation is based on *"Decision Risk management: Trade Space and Statistical Risks in Flight Tests"* Poulson, Poindexter (2014)
- Background: Statistical Precision Metrics
- Innovative Solutions: Consequential Thinking and the Consequence Ratio
- Consequence Ratio and Leadership
- Example: Threat Identification Simulation
- Conclusion

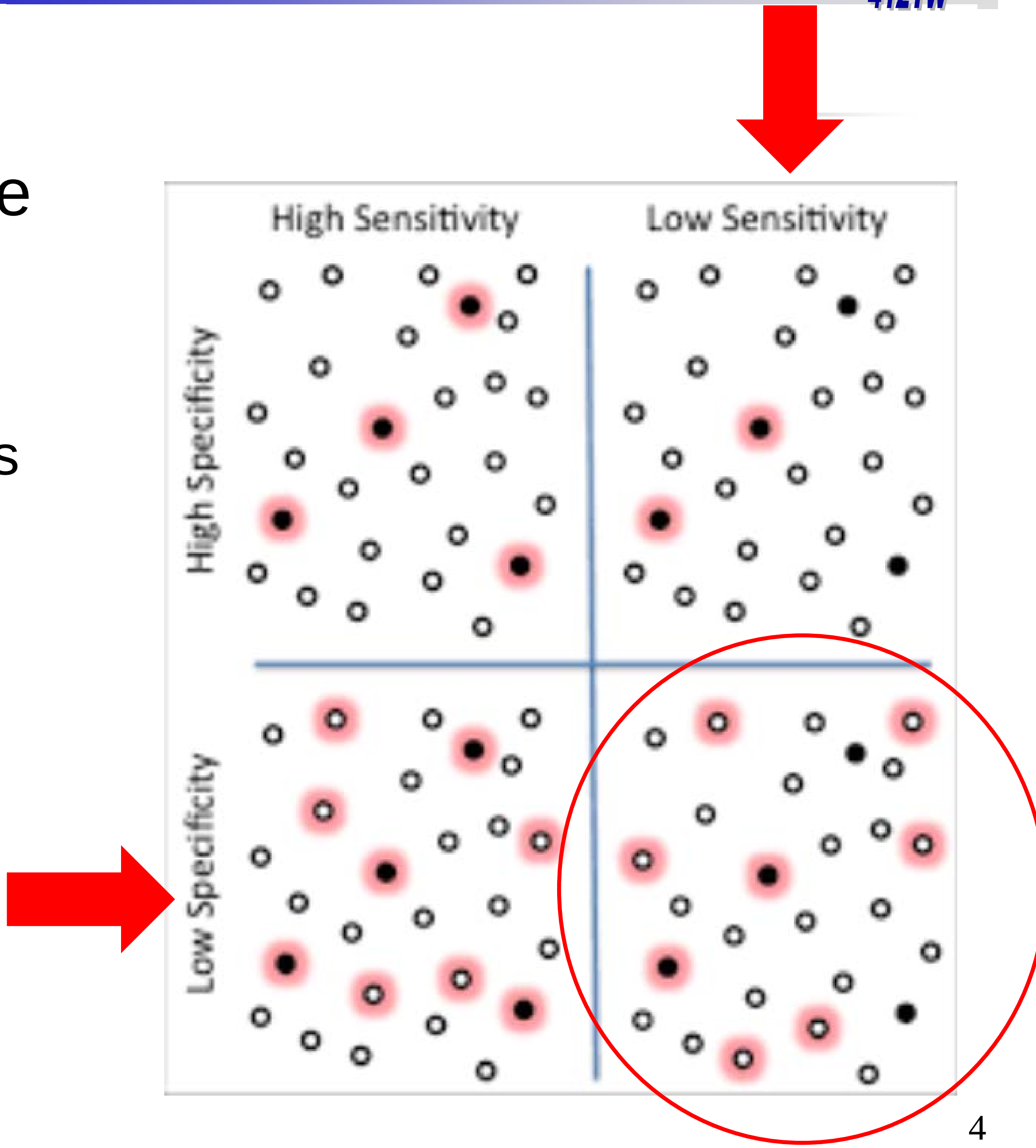
Example: Threat Identification

- Identification Threshold: Electronic Warfare suite requires that a threat system be ID'd within 5 seconds
  - Observed identification times range from to 0-35
  - Draw a conclusion to pas or fail the system based on
    1. Observed times
    2. System specification
- This presentation is about mistakes. The risk of either
  - Concluding a system should <u>pass</u> when it should have <u>failed</u>
  - Or concluding a system <u>failed</u> when it should have <u>passed</u>

I think it's important for scientists to be ... a bit more humble, recognizing we are capable of making mistakes ... - which is increasingly serious in a society where our work may have unpredictable consequences. – *Robert Winston, MD (2011)*

Statistical Precision Metrics

- Risk of Error vs. Power and Confidence
- Example: Drug Testing
  - Goal: identify solid dots
- High Risk Areas:
  - Low Sensitivity
  - Low Specificity
- Sensitivity is Power
- Specificity is Confidence.



Statistical Precision Metrics

- Decision Matrix
- Example: Biological / Pharmaceutical
  - False positives result from low specificity
  - False negatives result from low sensitivity.
- Example: Flight Testing
  - Pass / Fail
  - Satisfactory / Unsatisfactory

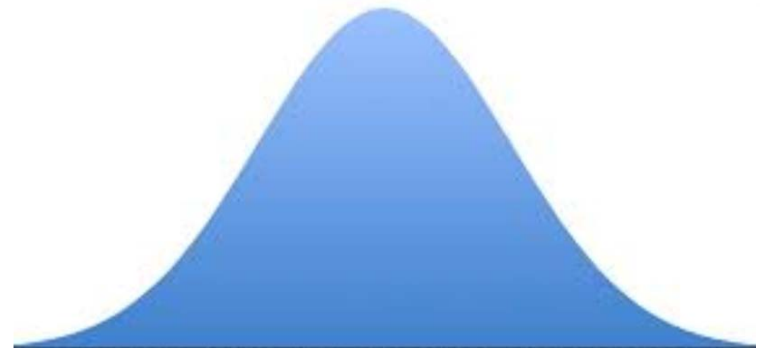
		True Condition	
		Overweight	Not Overweight
Answer from Test	Overweight? - Yes	**True Positive** *(Sensitivity)*	**False Positive** *(Type I Error)*
	Overweight? - No	**False Negative** *(Type II Error)*	**True Negative** *(Specificity)*

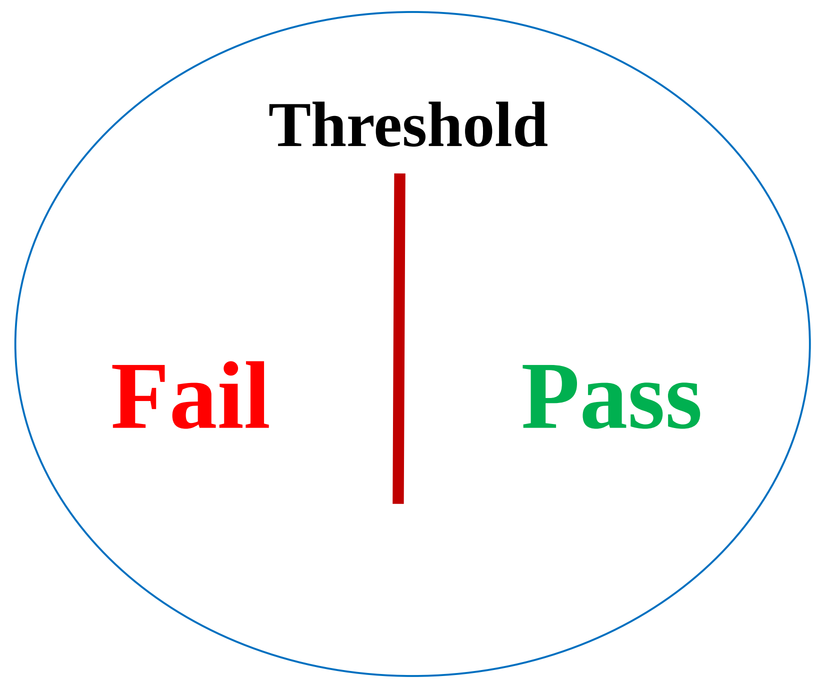
		Truth (The system is actually...)	
		H_0: Satisfactory	H_A: Unsatisfactory
Decision (The system is rated...)	Unsatisfactory	Type I Error α	Power $(1-\beta)$
	Satisfactory	Confidence $(1-\alpha)$	Type II Error β

Statistical Precision Metrics

- Source of Risks
  - What is being observed is a Random Outcome (A.K.A. sampling risk)
  - Distribution of outcomes is unknown
  - Pass/Fail scenario
- Sampling Risk cannot be overcome by discipline expertise.
  - Aside: discipline experts play an extensive role in reducing noise.

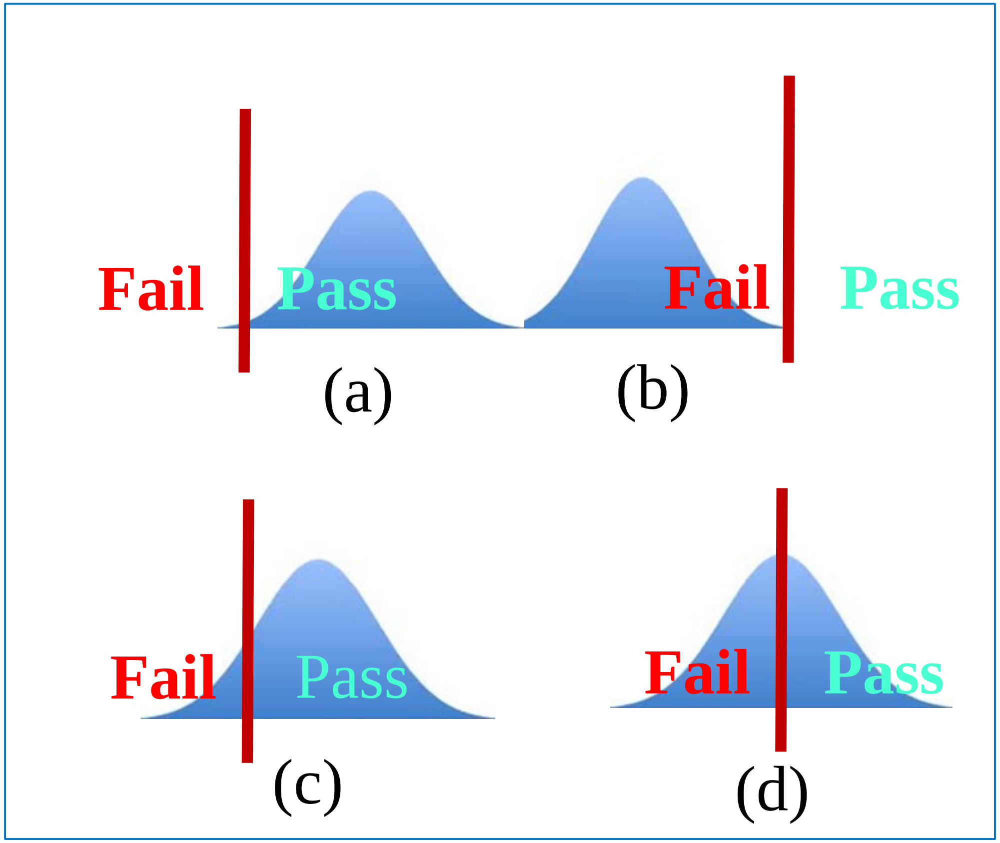
Distribution of Possible Outcomes





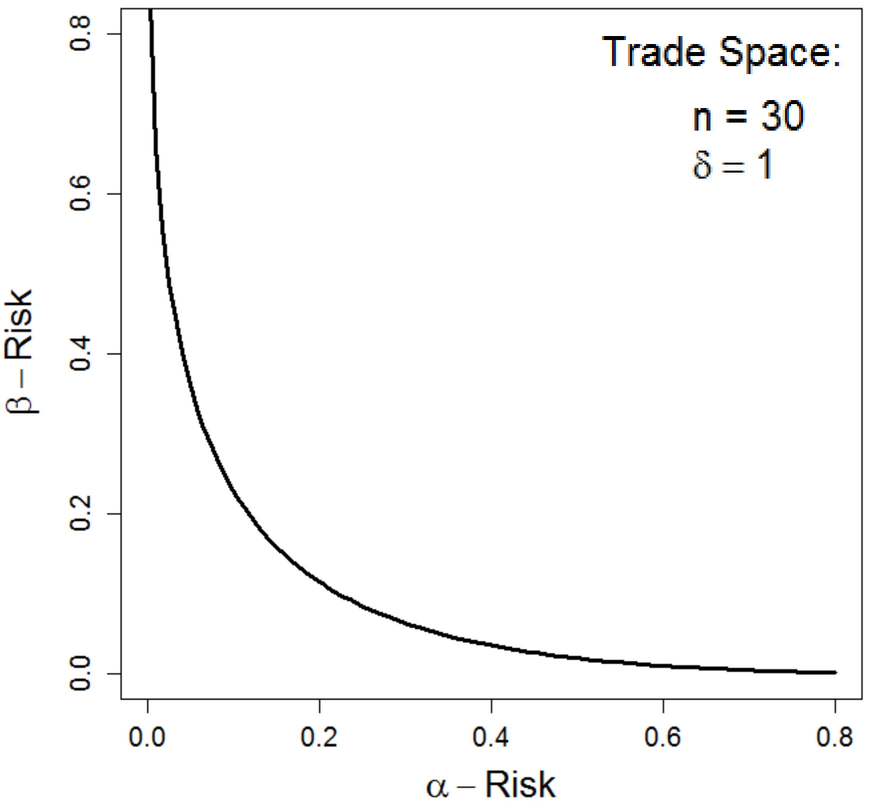
Statistical Precision Metrics

- Consider Pass / Fail risks for the four cases (large is good)
- If you are lucky to get (a) or (b), then no management is necessary.
- However, (c) and (d) require management (i.e. where are you going to ‘draw the line’.
- Unfortunately the distribution is unknown!



Statistical Precision Metrics

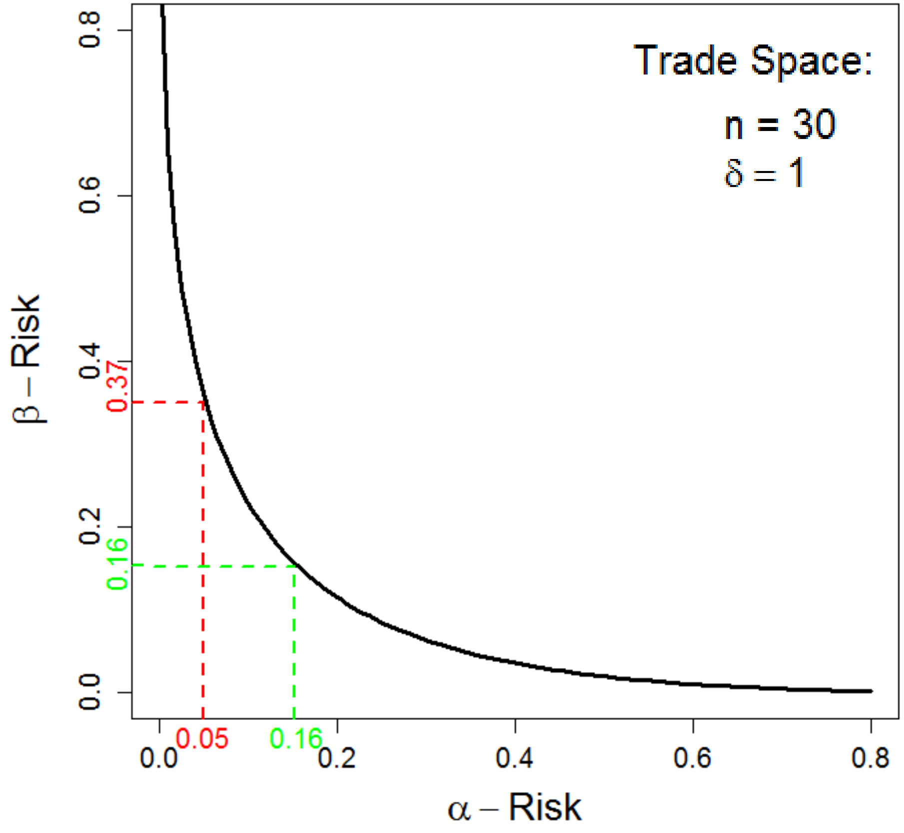
- *How Critical Thinking Shapes the Military Decision Making Process*, Usry, Floyd, 2004
  - Intuition vs. Analytical
  - Approximately 90% of all decisions are intuitive.
  - Analytical decisions typically require structure (e.g. reduce dimension of problem).
- Exists a 1:1 functional relationship between the α-risk and the β-risk.



Low Sensitivity vs. Low Specificity
False Negative vs. False Positive
Consumer Risk vs. Producer Risk

Statistical Precision Metrics

- The α-β Relationship:
  - Any change in the α-risk necessitates a change in the β-risk
  - Example1: Set $\alpha = 0.05$ then $\beta = 0.37$
  - Example2: Increase $\alpha = 0.16$ then β decreases to $\beta = 0.16$
- Decision Risk Trade Space:
  - $\alpha, \beta, n,$ and δ.
  - $n = 30, \delta = 1$ second fixed.



Low Sensitivity vs. Low Specificity
False Negative vs. False Positive
Consumer Risk vs. Producer Risk

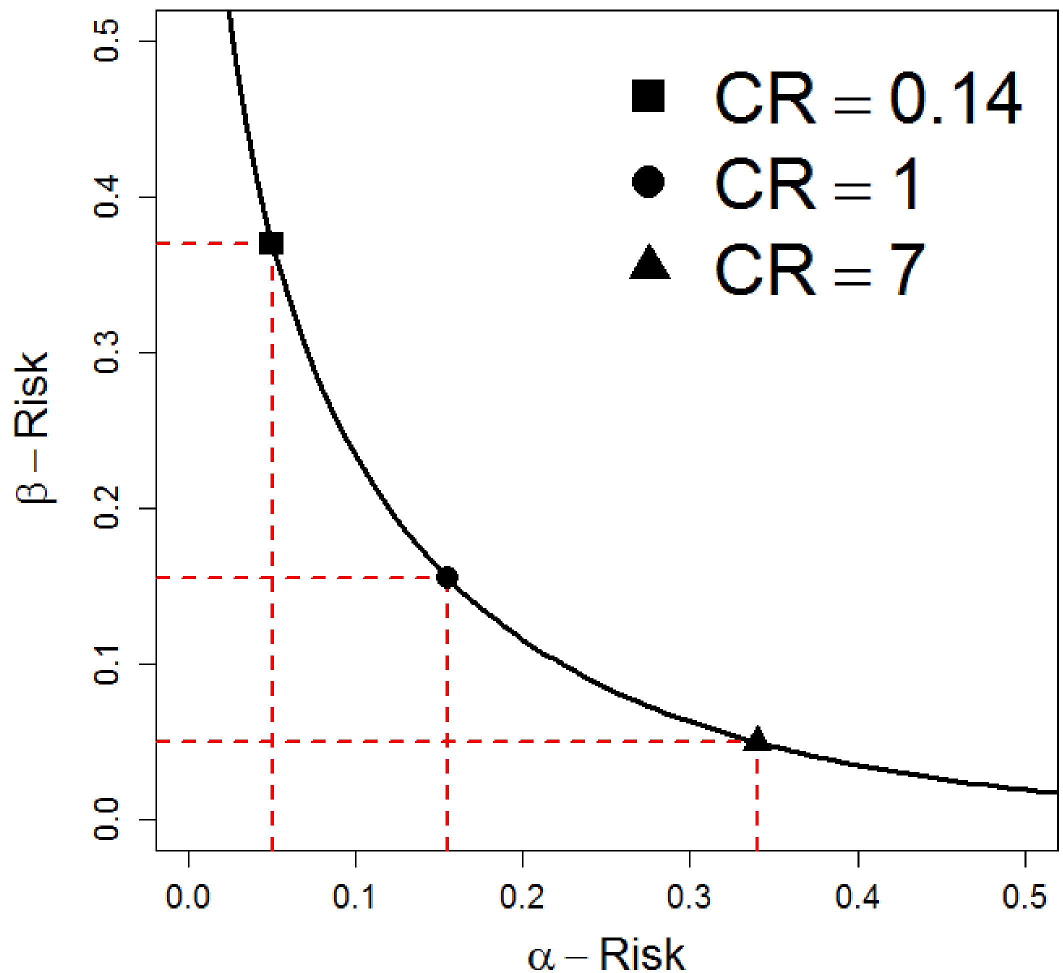
Consequential Thinking

- What are the consequences to making the wrong decision?
  - Example: "industry standards in drug testing have hampered the ability of drug manufacturers to get efficacious drugs to needy patients." Intriligator, 1996

TEST	RISK	CONSEQUENCE
Drug Testing	Low Sensitivity	Accuse a non-user of taking drugs
	Low Specificity	Drug user cleared for competition
Marketplace	Consumer Risk	Own a product that does not work / use time and resources to return product
	Producer Risk	Sales lower than expected or should be
Test & Evaluation	Acquisition Risk	Own a system that does not meet warfighter specifications / Retrofit system
	Developer Risk	Time and Cost to retest / Re-engineer system

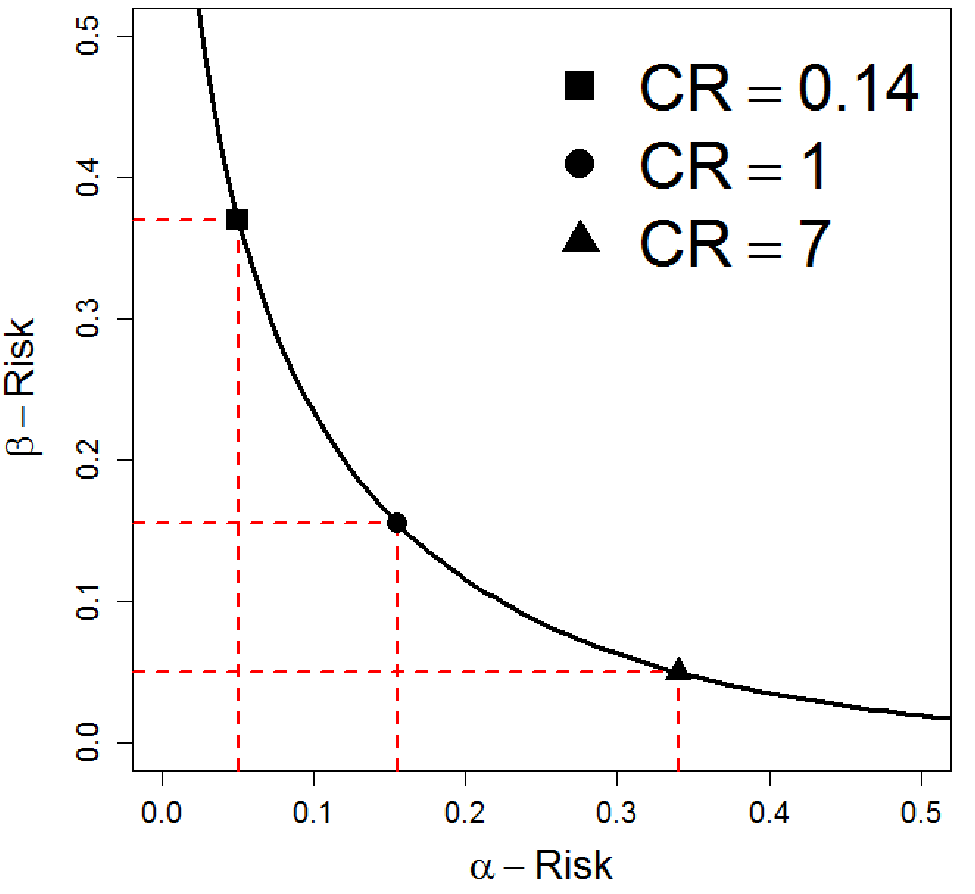
Consequence Ratio

- Consequence Ratio:
  - $CR = \frac{\alpha\text{−}risk}{\beta\text{−}risk}$
  - Should be small if α-risk is more serious than the β-risk
  - Should be large if β-risk is more serious than α-risk
  - Closer to 1 if both α and β risks are equally serious



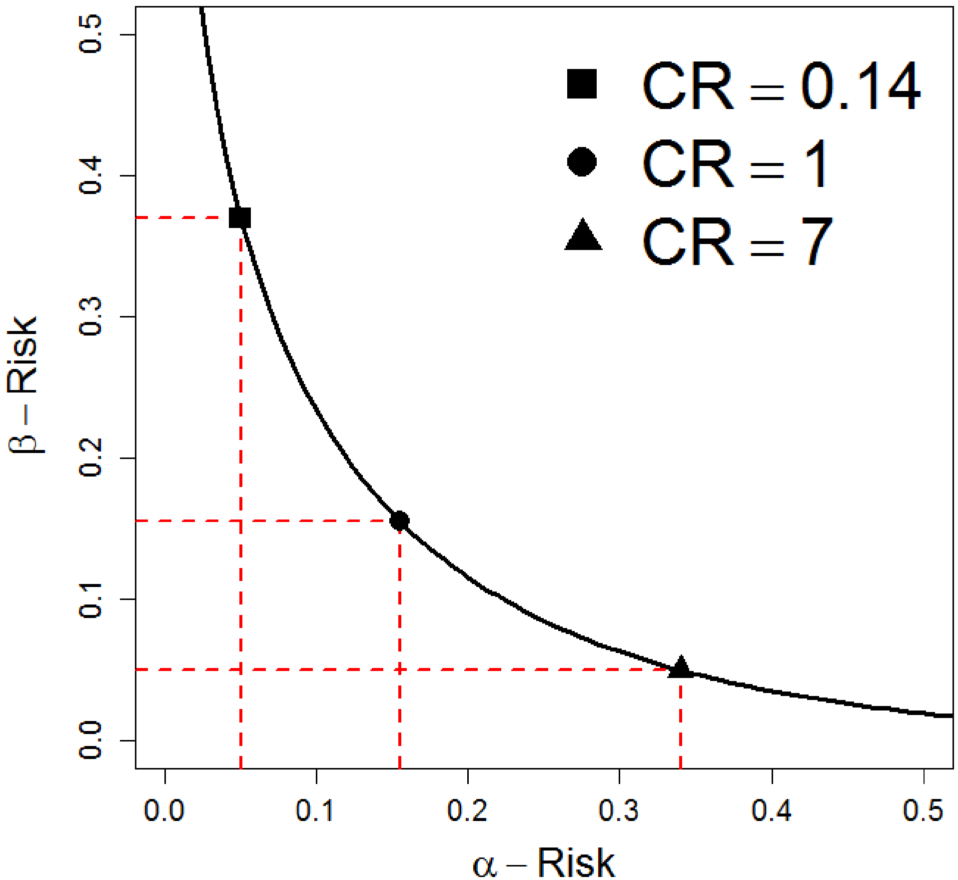
Consequence Ratio & Leadership

- What role does leadership play in managing α and β (acquisition and developer) risks?
  - The practice of holding alpha to 0.05 has led some to the faulty belief that type I errors are more serious than type II errors. –Lipsey and Hurley (2008)
  - "A more balanced set of FDA drug approval standards…could result in better outcomes." –Intriligator (1996)



Consequence Ratio & Leadership

- Science and Discipline Experts (DAU-ACQ 101):
  - Timeliness: "provide information to decision-makers and help them identify and manage risks."
  - Process: "the goal is not to 'eliminate risks', but to manage it."
- Leadership Use of CR:
  - Boundaries / Limits on risks for all tests.
  - Determine risks on a test by test basis (i.e. situational)

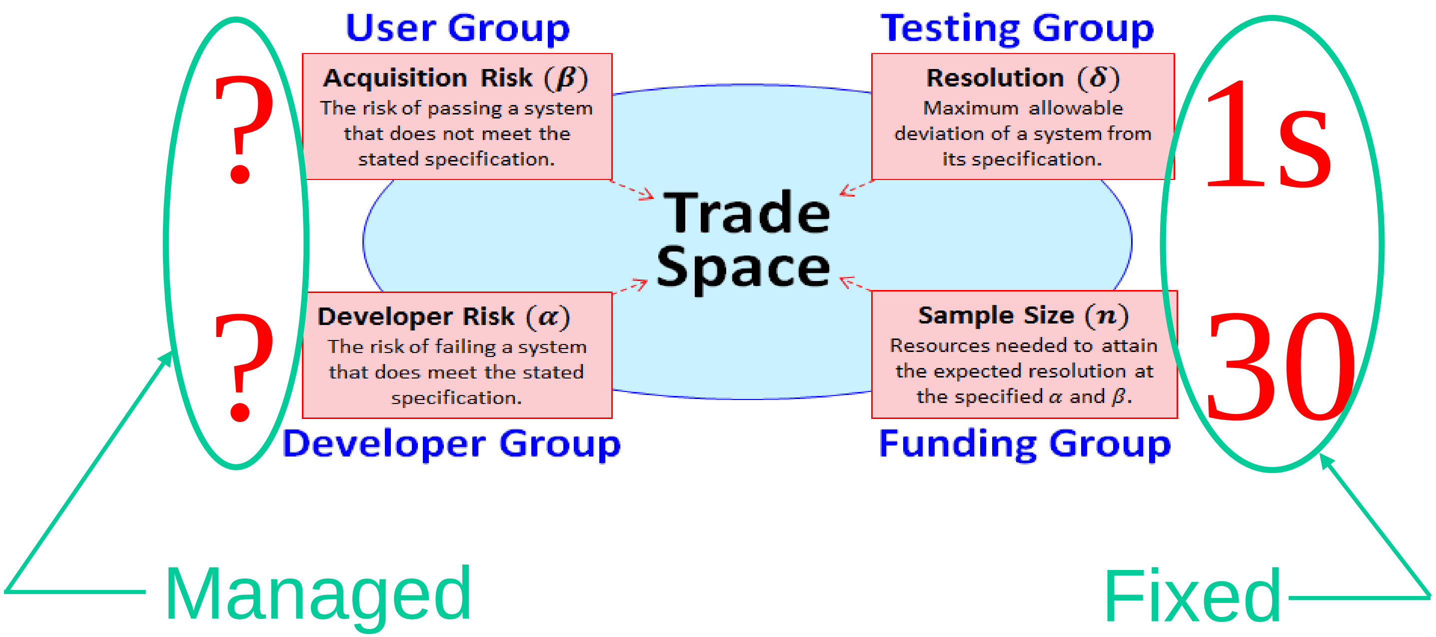


Example: Threat Identification

- Identification Threshold: Electronic Warfare requires that a threat be ID'd within 5 seconds
  - Observed identification times range from to 0-35 sec.
- Threshold: Mean identification within 5 sec.
  - Note: The mean ($\boldsymbol{\mu_{IT}}$) is a characteristic of the distribution of outcomes
  - $\boldsymbol{\mu_{Identification\,Time\,(IT)} = 5\ seconds}$
- Test
  - $\boldsymbol{H_0: \mu_{IT} \leq 5s\ (Pass)\ vs.\ H_A: \mu_{IT} > 5s\ (Fail)}$
  - What is the decision rule (evaluation criteria) to determine whether μ_{IT} has crossed the threshold? Otherwise, μ_{IT} is considered to be within specification.

Example: Threat Identification

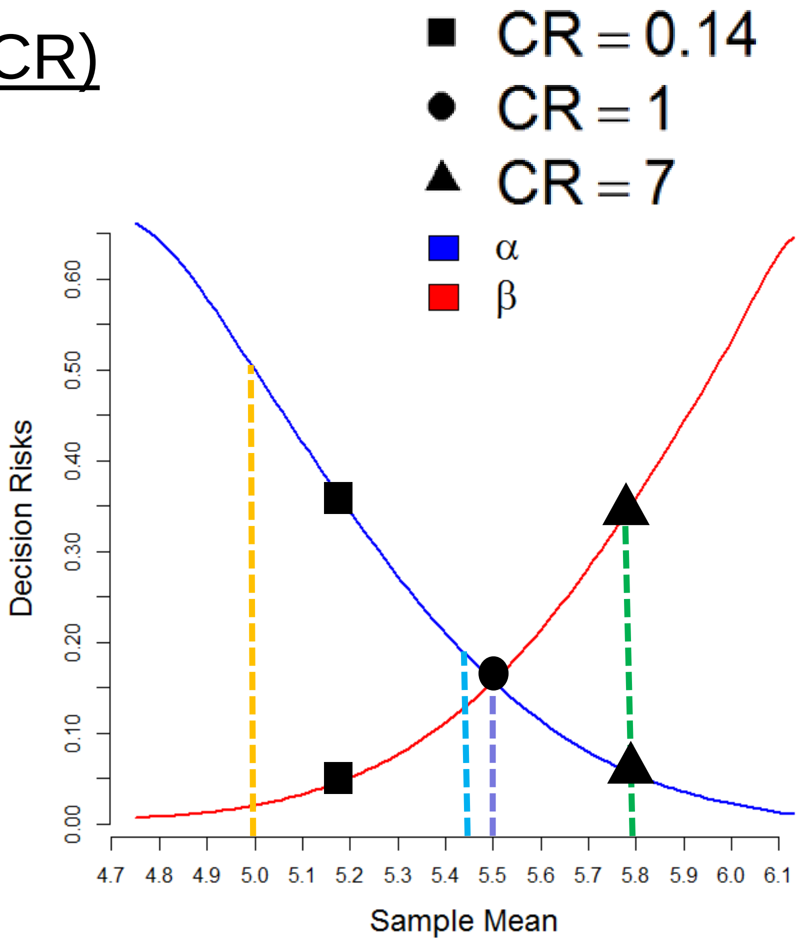
- Decision Risk Management
  - Statistical risk from being forced to sample from a distribution of possible outcomes is managed through appropriate evaluation criteria.
  - For this example the decision rule will be based on the observed sample average.



Example: Threat Identification

- Consequence Ratio: Manage Raw Risks
  - Based on industry or convention
  - Based on consequence ratio (CR)
  - Based on cumulative risk minimization ($\alpha + \beta$)

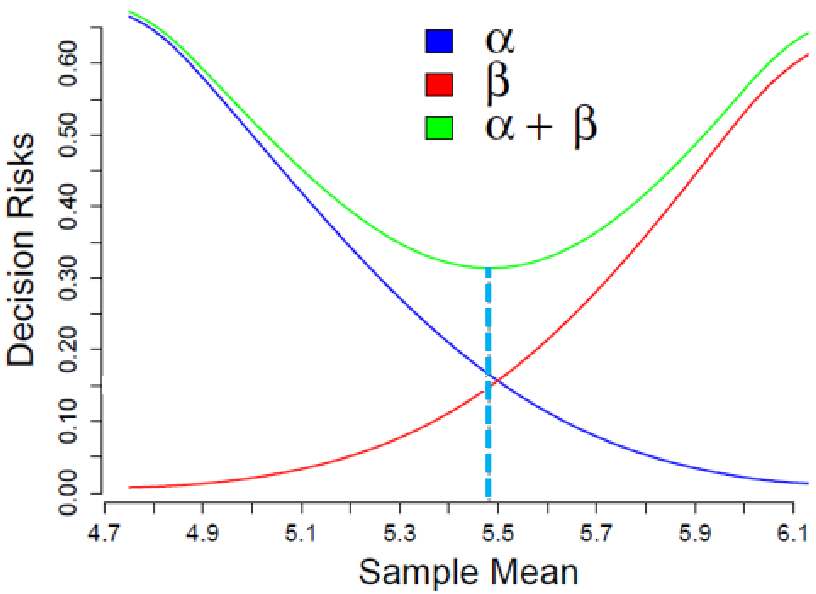
Evaluation Criteria (Sample Mean Decision Rule)	α	β	CR	$\alpha + \beta$
5 sec	50%	2%	25	52%
5.8 sec	5%	37%	0.14	42%
5.5 sec	16%	16%	1	32%
5.48 sec	16.3%	14.7%	1.12	31%



Example: Threat Identification

- Consequence Ratio: Manage Costs
  - $242 cost if either Acquisition or Developer risk is realized (i.e. if a mistake is made either way).
  - Expected loss by convention ($\alpha = 0.05$) = **$101.64**
  - Expected Loss for CR = 1: **$77.44**
  - Expected Loss by cumulative risk min. = 1.12: **$75.02**

Evaluation Criteria (Sample Mean Decision Rule)	α	β	CR	$\alpha + \beta$
5 sec	50%	2%	25	52%
5.8 sec	5%	37%	0.14	42%
5.5 sec	16%	16%	1	32%
5.48 sec	16.3%	14.7%	1.12	31%



Conclusion

- A statistical framework provides the structure to make Analytic decisions when sampling risk is unavoidable
- Consequential thinking and the Consequence Ratio are tools to help bridge the gap between the science / discipline experts and leadership who are managing test programs.
- The Consequence Ratio can be expanded into include additional ways of driving innovative solutions to complicated problems (e.g. cumulative risk minimization, expected loss minimization)

References

- Poulson, Poindexter (2014). *"Decision Risk management: Trade Space and Statistical Risks in Flight Tests"* Internal white paper, 412 TW
- Floyd (2004). *"How Critical Thinking Shapes the Military Decision Making Process."* Naval War College.
- Intriligator, M. D., (1996). *"Drug Evaluations: Type I vs. Type II Errors,"* UCLA Research Program in Pharmaceutical Economics and Policy UC Los Angeles.