NASA System Safety Framework and Concepts for Implementation

Presented at the System Safety/Risk Management Virtual Workshop

December 12, 2012

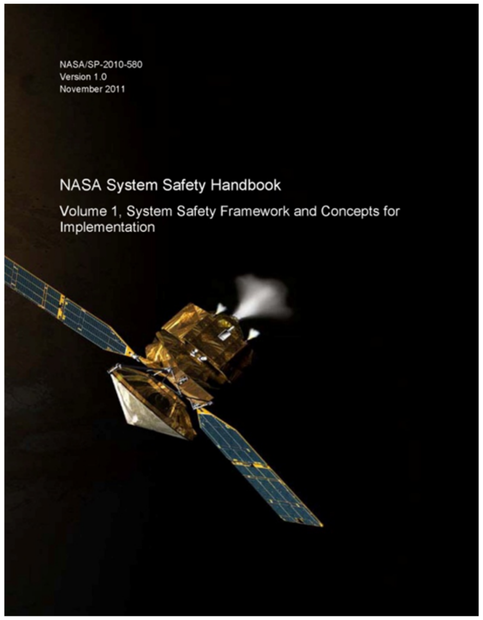
Homayoon Dezfuli, Ph.D.
NASA Technical Fellow for System Safety

Acknowledgements

- This presentation is derived from
  - NASA/SP-2010-580: Volume 1 of NASA System Safety Handbook
  - The work conducted by OSMA in conjunction with the development of the NASA System Safety Standard and Volume 2 of NASA System Safety Handbook

Introduction

- In 2011, we developed a System Safety Framework under which system safety activities are conducted and communicated
- The core elements of the framework are:
  - Safety objectives
  - System safety activities
  - Risk-Informed Safety Case (RISC)
  - Evaluation of RISC



NASA/SP-2010-580
Version 1.0
November 2011

NASA System Safety Handbook

Volume 1, System Safety Framework and Concepts for Implementation

- Explanation of the framework and its core elements are contained in:

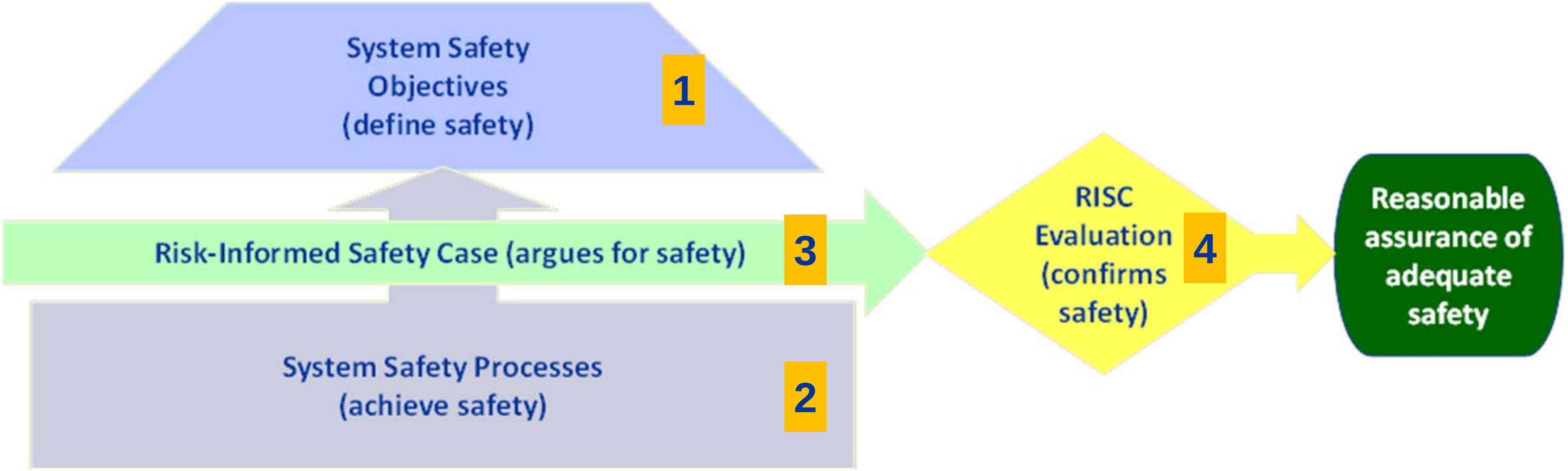
  NASA/SP-2010-580: NASA System Safety Handbook Volume 1: System Safety Framework and Concepts for Implementation

- NASA System Safety Standard and Volume 2 of the handbook are currently in development

Motivation

- Development of the System Safety Framework is motivated by a desire to:
  - Foster objectives-driven analysis and execution of system safety activities (as opposed to techniques- or product-driven approaches)
  - Foster integrated, holistic view of safety to address system-level considerations (recognition that safety is an emergent property)
  - Establish a process for defining “adequate safety”
  - Codify expected safety performance, so that later monitoring and precursor analysis activities have a baseline
  - Establish a means for presenting a coherent case for the safety of the system to decision makers
  - Establish a process that is compatible with the growing trend toward relying on commercial providers for transportation of crew and cargo

The NASA System Safety Framework


System Safety Objectives (define safety)
1
Risk-Informed Safety Case (argues for safety)
3
System Safety Processes (achieve safety)
2
RISC Evaluation (confirms safety)
4
Reasonable assurance of adequate safety

1. System Safety Objectives (Defined by Acquirer)
   – Development of safety objectives that collectively define adequate safety for a system
2. System safety processes (Achieved by Provider)
   – Conduct of safety activities that are implemented to achieve defined safety objectives
3. RISC Development (Argued by Provider)
   – Substantiation of claims that the safety objectives of the system have been met, or are on track to being met, within the specific decision context
4. RISC Evaluation (Confirmed by or on behalf of Acquirer)
   – Evaluation of RISC in order to accept the current or forecasted safety performance of the system as well as the risk that its realized safety performance might be less than its claimed performance

Safety Objectives

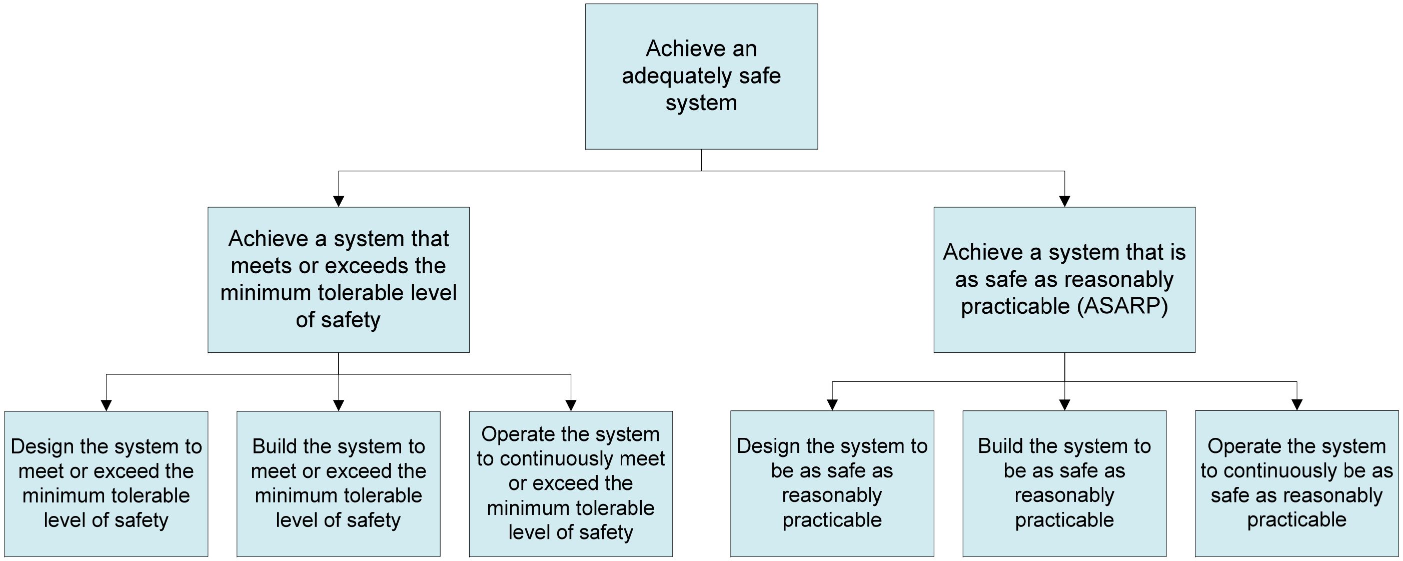
What is Safety?

> “Safety is freedom from those conditions that can cause death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment”
>
> NPR 8715.3

- **The specific scope of safety is application-specific, and must be clearly defined by the stakeholders in terms of the entities to which it applies and the consequences against which it is assessed**
- **The degree of safety that is considered acceptable is also application-specific**
  - **We strive to attain a degree of safety that fulfills obligations to the at-risk communities and addresses agency priorities**
  - **We do not expect to attain absolute safety (nor consider it possible to do so)**

Adequate Safety

- Achieving an adequately safe system requires adherence to the following fundamental safety principles:
  - The system meets or exceeds a minimum tolerable level of safety. Below this level the system is considered unsafe
  - The system is as safe as reasonably practicable (ASARP)


Achieve an adequately safe system
Achieve a system that meets or exceeds the minimum tolerable level of safety
Achieve a system that is as safe as reasonably practicable (ASARP)
Design the system to meet or exceed the minimum tolerable level of safety
Build the system to meet or exceed the minimum tolerable level of safety
Operate the system to continuously meet or exceed the minimum tolerable level of safety
Design the system to be as safe as reasonably practicable
Build the system to be as safe as reasonably practicable
Operate the system to continuously be as safe as reasonably practicable

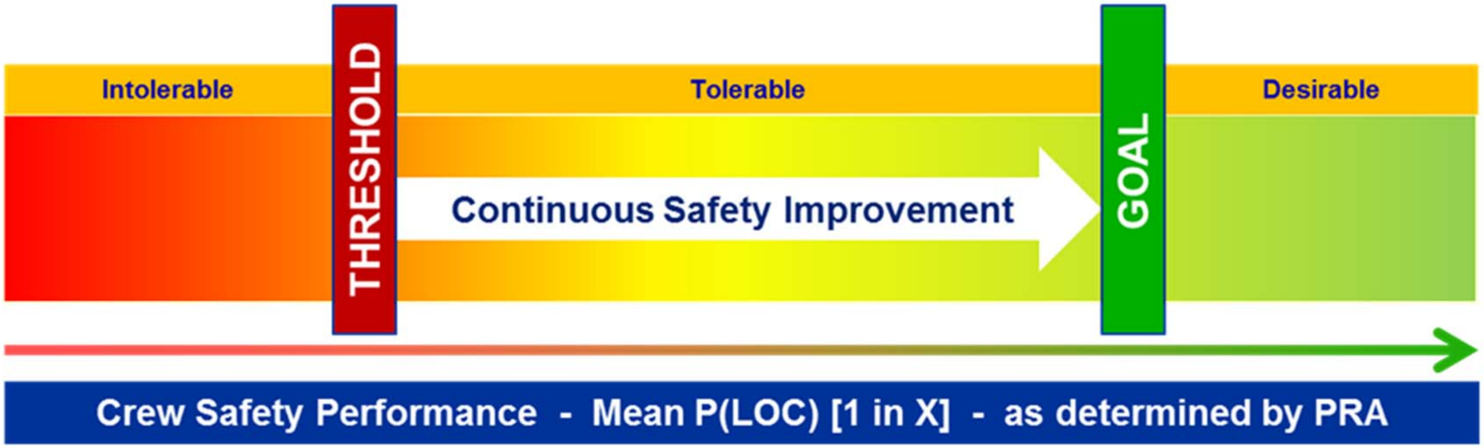
- The minimum tolerable level of safety is not necessarily static, and may evolve over the course of the system life cycle
- The principles of adequate safety must be maintained throughout all phases of the system life cycle

Safety Thresholds & Goals

- NASA's minimum level of tolerable safety for human spaceflight missions is articulated in NASA's agency-level safety goals and thresholds for crew transportation system missions to the ISS
- They reflect a tolerance for an initial safety performance that is acceptable initially but below long-term expectations

Standard of "Minimally Safe Initially"
Less than this would be "intolerable"

Standard of "Minimally Safe for Long Term Operation"
Less than this is tolerable, conditional on continuous safety improvement



Intolerable

THRESHOLD

Tolerable

Continuous Safety Improvement

GOAL

Desirable

Crew Safety Performance - Mean P(LOC) [1 in X] - as determined by PRA

- Don't proceed with the acquisition
- Fix design or operation to meet the threshold
- Termination review

- Actively pursue safety improvements via risk tradeoff studies
- Actively uncover hazards via testing
- Actively identify unaccounted-for hazards via precursor analysis

- Keep alert for safety improvements, but focus more on maintaining the good safety level that has been achieved

Being As Safe As Reasonably Practicable

> "ASARP entails weighing the safety performance of a system against the sacrifice needed to further improve it. A system is ASARP if an incremental improvement in safety would require a disproportionate deterioration of system performance in other areas."
>
> From SS Handbook

- **The ASARP concept is closely related to the "as low as reasonably achievable" (ALARA) and "as low as reasonably practicable" (ALARP) concepts that are found in U.S. nuclear applications and U.K. Health and Safety law**
- **ASARP implies that:**
  - **The performance of each alternative has been analyzed to determine the relative gains and losses in performance (technical, safety, cost, and schedule) that would result from selecting one alternative over another**
  - **Safety performance is given priority in the selection of an alternative, insofar as the selection is within tolerable limits of cost/schedule/technical performance**

Being As Safe As Reasonably Practicable (cont.)

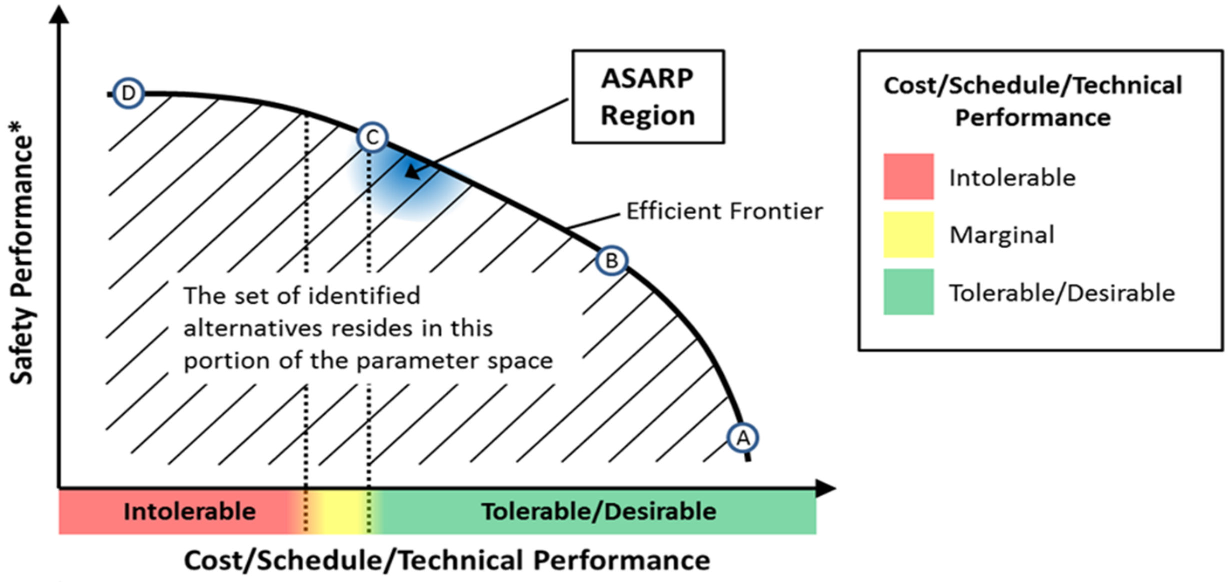
- ASARP reflects a mindset of continuous safety improvement regardless of the current level of safety

(A) Large increases in safety can be achieved by addressing low hanging fruit. Little cost/schedule/technical impact for doing so.

(B) Low hanging fruit has been addressed. but significant increases in safety can still be "bought" without failing to meet cost/schedule/technical performance requirements

(C) Limit of ASARP regime has been reached. Increased safety cannot be "bought" without exceeding tolerable limits of cost/schedule/technical performance

(D) Limit of achievable safety has been reached. Increased safety cannot be "bought" at any cost.



•This figure does not address the "tolerability" of safety performance, but rather the "practicability" of safety performance.

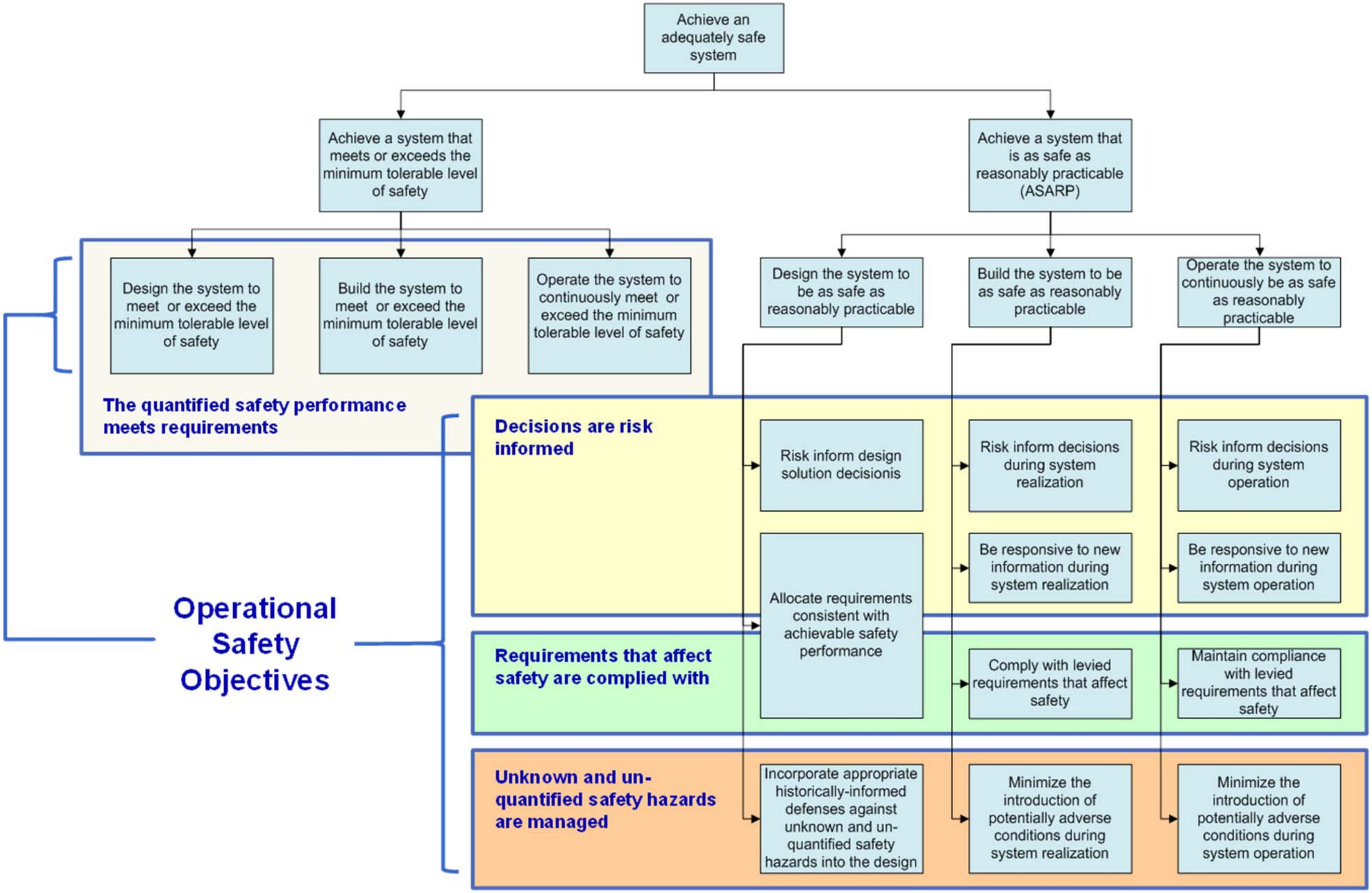
Deriving Operational Safety Objectives

- **The fundamental safety principles set the stage for the further development of safety objectives, negotiated on an application-specific basis**

- **Safety objectives are developed using an objectives hierarchy down to a level where they can be clearly addressed by systems safety activities, thereby creating a link that:**
  - **Assures that system safety activities are directed towards accomplishing defined safety objectives**
  - **Enables the system safety activities to be assessed in terms of the degree to which their target safety objectives have been met**

- **The safety objectives at the bottom level of the objectives hierarchy represent the *operational* definition of safety for the system under consideration, and are referred to as operational safety objectives**

Deriving Operational Safety Objectives



Achieve an adequately safe system

Achieve a system that meets or exceeds the minimum tolerable level of safety

Achieve a system that is as safe as reasonably practicable (ASARP)

Design the system to meet or exceed the minimum tolerable level of safety

Build the system to meet or exceed the minimum tolerable level of safety

Operate the system to continuously meet or exceed the minimum tolerable level of safety

Design the system to be as safe as reasonably practicable

Build the system to be as safe as reasonably practicable

Operate the system to continuously be as safe as reasonably practicable

The quantified safety performance meets requirements

Decisions are risk informed

Risk inform design solution decisionis

Risk inform decisions during system realization

Risk inform decisions during system operation

Be responsive to new information during system realization

Be responsive to new information during system operation

Allocate requirements consistent with achievable safety performance

Requirements that affect safety are complied with

Comply with levied requirements that affect safety

Maintain compliance with levied requirements that affect safety

Unknown and un-quantified safety hazards are managed

Incorporate appropriate historically-informed defenses against unknown and un-quantified safety hazards into the design

Minimize the introduction of potentially adverse conditions during system realization

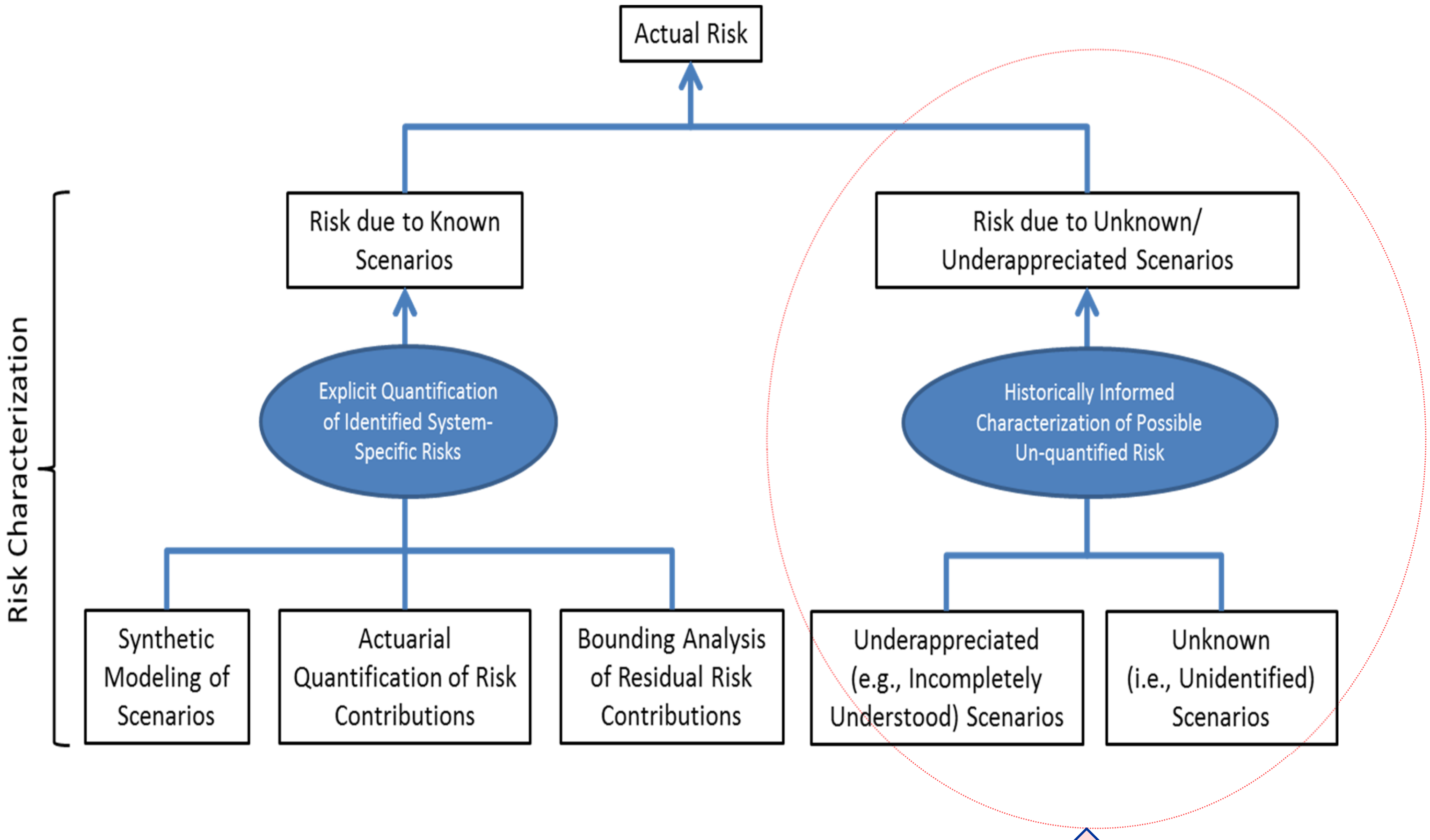
Minimize the introduction of potentially adverse conditions during system operation

Operational Safety Objectives

Characterizing the Actual Risk of a System

- Safety goals and thresholds represent expectations about actual risk, including both known and unknown/underappreciated (UU) sources
  - Known sources of risk are amenable to explicit quantification via synthetic, scenario-based methods of analysis (e.g., PRA), and actuarial methods (when sufficient data are available)
  - UU sources of risk are not amenable to synthetic analysis or direct actuarial characterization, yet are historically recognized as significant contributors to risk
    - They tend to remain latent in the system until revealed by operational failures, precursor analysis, etc.
    - They tend to be most significant early in the system life cycle
    - They disproportionally reflect complex intra-system and environmental interactions

Characterizing the Actual Risk of a System



Actual Risk

Risk due to Known Scenarios

Risk due to Unknown/ Underappreciated Scenarios

Explicit Quantification of Identified System-Specific Risks

Historically Informed Characterization of Possible Un-quantified Risk

Synthetic Modeling of Scenarios

Actuarial Quantification of Risk Contributions

Bounding Analysis of Residual Risk Contributions

Underappreciated (e.g., Incompletely Understood) Scenarios

Unknown (i.e., Unidentified) Scenarios

Risk Characterization

How should we characterize the contribution of UU scenarios to actual risk?

How to Characterize the Contribution of UU Scenarios to Actual Risk

- **One possible approach is imbedded in the concept of *safety performance margin* (referred to in the System Safety Handbook as *safety risk reserve*).**
- **In this approach, the actual risk of a system is understood to be the sum of the risk from known scenarios, as explicitly quantified using traditional risk analysis methods, plus the risk from UU scenarios, as characterized by the safety performance margin.**
- **The limit on the allowable explicitly quantified risk can be derived by subtracting (in risk terms) the safety performance margin from the minimum tolerable level of safety.**
- **If the explicitly quantified risk is within this limit, then by implication there is reasonable assurance that the actual risk is within the minimum tolerable level of safety**
- **Methods for establishing an initial safety performance margin and a margin draw-down profile based on historical data for similar systems are currently being investigated by OSMA**

Conceptual View of Managing Safety Performance Consistent with the Minimum Tolerable Level of Safety

A: The design concept and operating environment have been established and a preliminary design has been developed.

A1: A preliminary safety performance margin is derived to account for unknown and underappreciated scenarios.
A2: The risk from known scenarios is explicitly quantified using PRA or alternative synthetic + actuarial methods.

B: Improvements are made to bring the concept of operations within the safety threshold.

B1: The safety performance margin is reduced based on provisions made to reduce the UU risks.
B2: Mitigations are introduced to reduce the top known risks.
B3: An initial safety requirement and a safety requirement profile are derived.

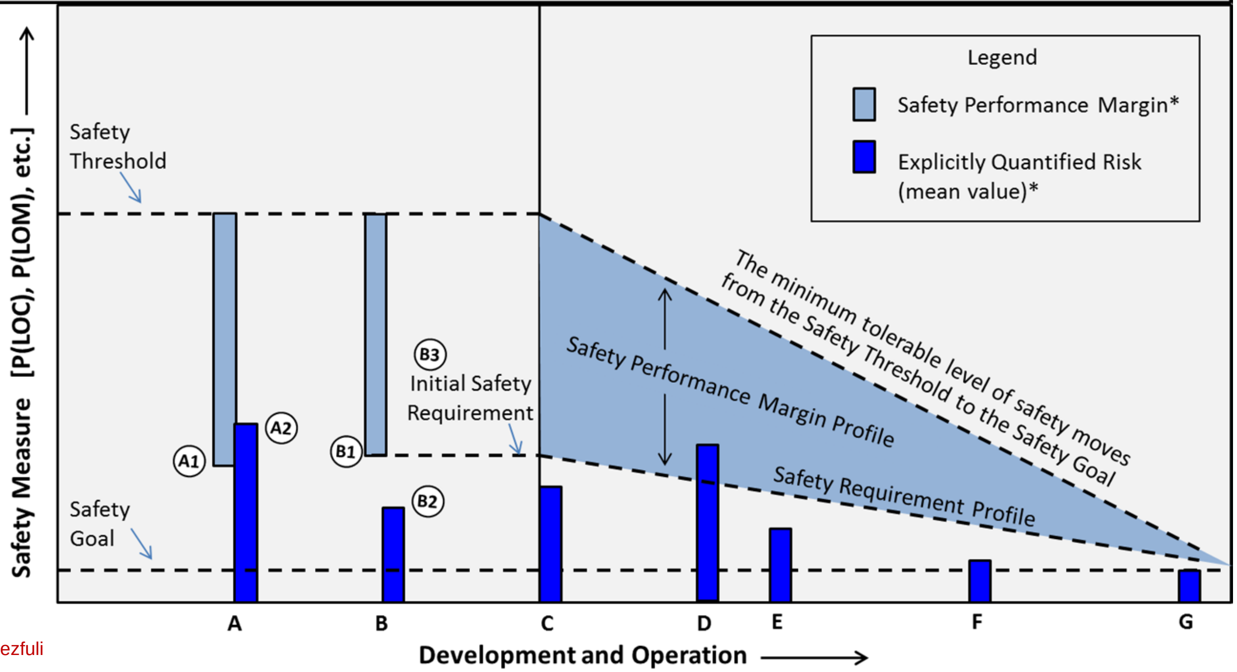
C: The known risk of the as-built system is within the safety requirement.

D: Newly discovered scenarios bring the known risk beyond the safety requirement, necessitating a mitigation (or a re-baselining of the safety requirement).

E: Mitigations are introduced to reduce the risk to within the safety requirement.

F: Proactive safety upgrade and improvement programs reduce the risk in increments, consistent with ASARP.

G: System performance meets the long-term safety goal.



System Safety Activities

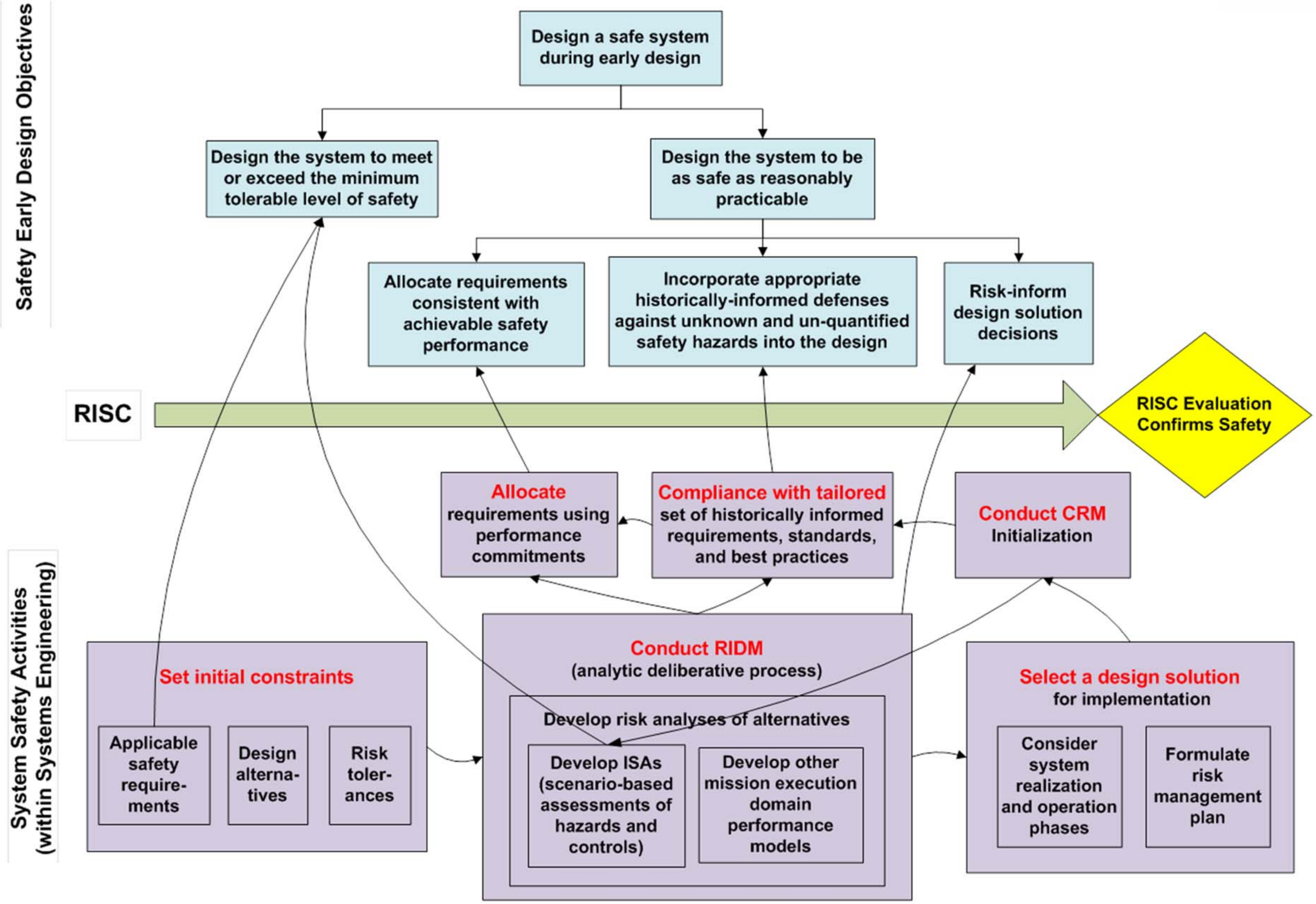
System Safety Activities

- The handbook takes the approach that system safety informs systems engineering and risk management, but does not directly engineer the system or manage risk
- System Safety Activities
  - Are conducted as part of overall systems engineering / risk management
  - Are focused on achieving a system that meets the operational safety objectives
- The system safety framework recognizes the need for flexibility in the nature and composition of SS activities, so long as objectives are met
- Integrated safety analysis (ISA) is the central system safety activity

Integrated Safety Analysis (ISA)

- There are system safety activities that support the ISA
  - Uncertainty reduction (e.g., testing, performance monitoring, Accident Precursor Analysis)
- There are system safety activities that are supported by the ISA
  - System design, requirements development, requirements verification, performance monitoring, program control and commitments
- A graded approach to ISA
  - Comprehensive identification of “hazard scenarios” (i.e., deviations from the envelope of normal operation) using qualitative techniques
  - Detailed analysis of “accident scenarios” (i.e., the subset of the hazard scenarios that produce accident conditions) using quantitative techniques
  - Identification and management of high-priority controls based on their credited effectiveness in the ISA (capability, reliability, availability)

System Safety Activities (Early Design)



Safety Early Design Objectives
Design a safe system during early design
Design the system to meet or exceed the minimum tolerable level of safety
Design the system to be as safe as reasonably practicable
Allocate requirements consistent with achievable safety performance
Incorporate appropriate historically-informed defenses against unknown and un-quantified safety hazards into the design
Risk-inform design solution decisions
RISC
RISC Evaluation Confirms Safety
System Safety Activities (within Systems Engineering)
Allocate requirements using performance commitments
Compliance with tailored set of historically informed requirements, standards, and best practices
Conduct CRM Initialization
Set initial constraints
Applicable safety require-ments
Design alterna-tives
Risk toler-ances
Conduct RIDM (analytic deliberative process)
Develop risk analyses of alternatives
Develop ISAs (scenario-based assessments of hazards and controls)
Develop other mission execution domain performance models
Select a design solution for implementation
Consider system realization and operation phases
Formulate risk management plan

Presented by Homayoon Dezfuli

Risk-Informed Safety Case (RISC)

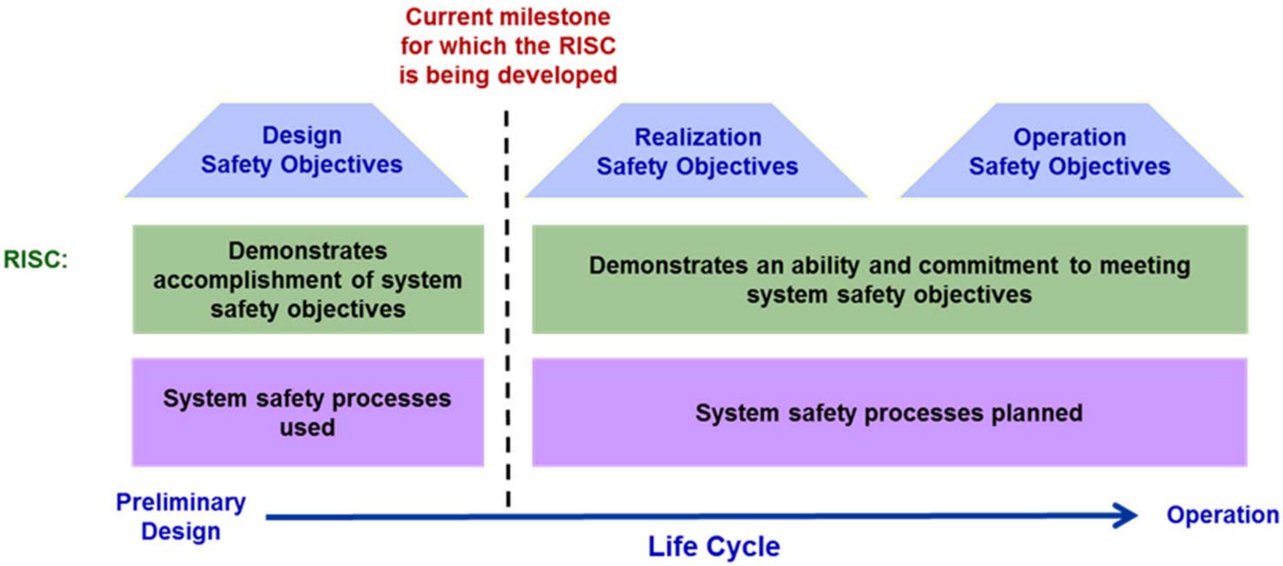
Risk-Informed Safety Case (RISC)

> “A risk-informed safety case (RISC) is a structured argument, supported by a body of evidence, that provides a compelling, comprehensible and valid case that a system is or will be adequately safe for a given application in a given environment. This is accomplished by addressing each of the operational safety objectives that have been negotiated for the system, including articulation of the roadmap for the achievement of safety objectives that are applicable to later phases of the system life cycle.”
>
> From SS Handbook

- **The term ‘risk-informed’ is used to emphasize that adequate safety is the result of a deliberative decision making process that involves an assessment of risks, and strives for a proper balance between safety performance and performance in other mission execution domains**

RISC Life Cycle Considerations

- The RISC (produced by the system provider) addresses the full system life cycle, regardless of the particular point in the life cycle at which the RISC is developed. This results in two types of safety claims:
  - Claims related to the safety objectives of the current or previous phases argue that the objectives have been met
  - Claims related to the safety objectives of future phases argue that necessary planning and preparation have been conducted, and that commitments are in place to satisfy the objectives at the appropriate time


Current milestone for which the RISC is being developed
Design Safety Objectives
Realization Safety Objectives
Operation Safety Objectives
RISC:
Demonstrates accomplishment of system safety objectives
Demonstrates an ability and commitment to meeting system safety objectives
System safety processes used
System safety processes planned
Preliminary Design
Operation
Life Cycle

Evaluation of RISC

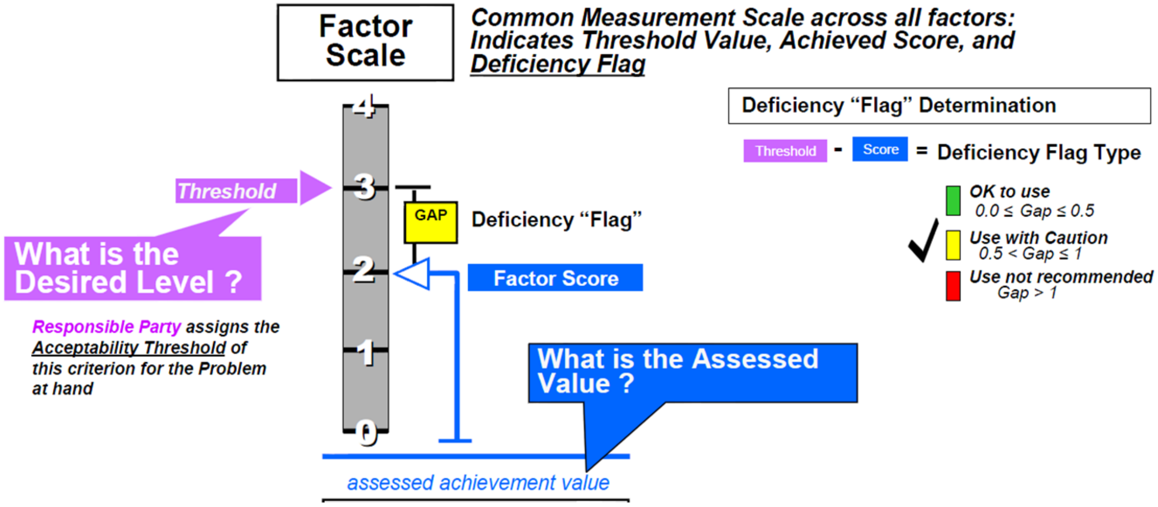
Evaluation of the RISC

- **Acquisition of a system that needs to satisfy safety objectives needs to entail demonstration of the *satisfaction* of those safety objectives and associated requirements**
- **The mechanism for this demonstration is the RISC that is delivered by the system provider**
- **The acquirer needs to address the risk that the level of safety realized in the system is worse than claimed in the safety case. The acquirer must either accept this risk, or reject the system.**
- **In principle, then, the acquirer must *evaluate* the RISC in sufficient depth to co-own the uncertainty in safety performance**
- **This does not mean that the acquirer must *redo* the safety analysis.**
- **It *does* mean that the acquirer's decision needs to be based on**
  - **the acquirer's assessment of the analysis processes, and**
  - **the acquirer's sense of the strength of the evidence presented, including prior information such as historical experience with the technology**

Distinguishing "Safety Performance Risk" From "Programmatic Risk" via an Example

- A program requires a launch vehicle having P(loss on launch) ≤ 1E-3. This implies that the top-level program authority has decided that P(loss on launch) ≤ 1E-3 can be accepted (assuming other objectives are met). This value is furnished as a threshold to the provider by the acquiring organization
  - Provider-assessed safety performance: 7E-4
  - Programmatic risk for the acceptor: Probability that the actual P(loss) is higher than 7E-4. (In particular, there is a risk that it is higher than the 1E-3 threshold)
- The acquiring organization must decide whether to accept the system, as well as the programmatic risk, in part by analyzing the probability that P(loss) > 1E-3
  - This includes an evaluation of the provider's analysis
  - This includes an evaluation of the provider as well (design philosophy, organizational factors, programmatic factors, etc.)
- An evaluation protocol is needed for assessing the programmatic risk that P(loss on launch) > 1E-3
- The evaluation can yield one of three possible results:
  - sufficiently confident that P(loss on launch) ≤ 1E-3
  - insufficiently confident that P(loss on launch) ≤ 1E-3
  - confident that P(loss on launch) > 1E-3

Potential Applicability of NASA STD-7009



Factor Scale

Common Measurement Scale across all factors: Indicates Threshold Value, Achieved Score, and <u>Deficiency Flag</u>

Deficiency "Flag" Determination

Threshold - Score = Deficiency Flag Type

🟩 *OK to use* 0.0 ≤ Gap ≤ 0.5

✓ 🟨 Use with Caution 0.5 < Gap ≤ 1

🟥 Use not recommended Gap > 1

Threshold

What is the Desired Level ?

Responsible Party assigns the Acceptability Threshold of this criterion for the Problem at hand

4

3

2

1

0

GAP

Deficiency "Flag"

Factor Score

What is the Assessed Value ?

assessed achievement value

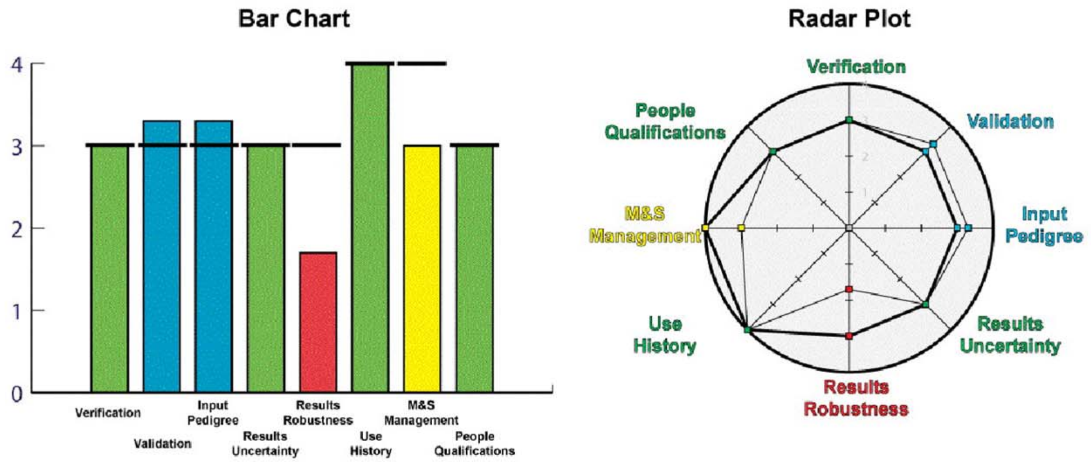


Figure 7—Sufficiency Thresholds and Color Coding on Bar Chart and Radar Plot for Factor Scores

- NASA STD-7009 contains an appendix describing a "Credibility Assessment Scale," or CAS, intended to guide prospective users of modeling and simulation results
- The CAS consists of 8 factors that are used to determine whether the credibility of the analysis used to support decision making is at least equal to the expectations of the DM

Summing it All Up

Summary

- An adequately safe system should adhere to two fundamental principles
  - Meeting or exceeding the minimum tolerable level of safety
  - Being ASARP
- Completeness issues associated with synthetic risk analysis methods should be formally addressed
  - Safety risk margin is proposed as one way to account for this
- Integrated safety analysis is the central system safety activity
  - System safety uses the ISA to inform systems engineering and risk management decisions, but does not directly engineer the system or manage risk
- The RISC makes the coherent case that the system is safe for the intended application
  - The RISC serves as a comprehensive proxy for the safety of the system
  - The safety of the system relative to goals/thresholds is one element of the RISC
  - The RISC is not a radically new idea. It is a formalization and integration of processes and ideas that are already in place or being incorporated to support a certification process
- Evaluation protocols are needed for evaluating RISCs
  - The acquiring organization must accept (or not) the programmatic risk that the system's actual safety performance is not as good as its characterization in the RISC